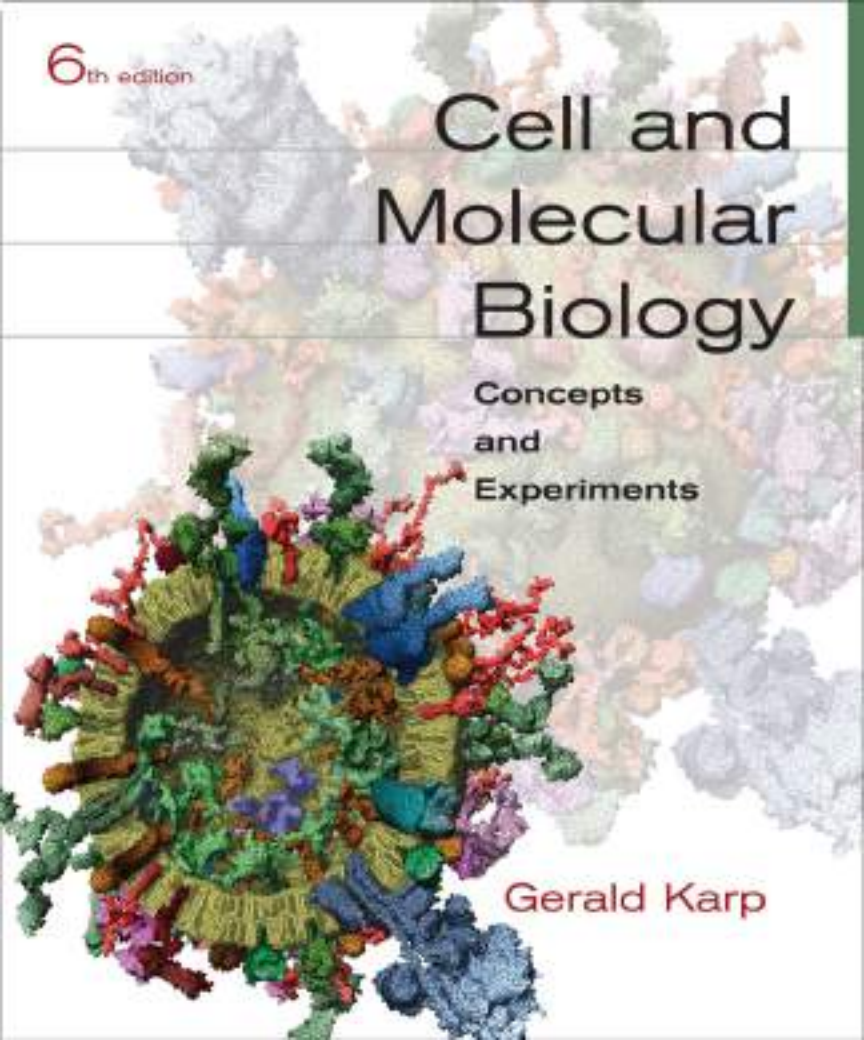# 6th edition

Cell and Molecular Biology

Concepts and Experiments

Gerald Karp

Nobel Prizes Awarded for Research in Cell and Molecular Biology Since 1958

Year	Recipient*	Prize	Area of Research	Pages in Text
2008	Francoise Barré-Sinoussi Luc Montagnier	M & P**	Discovery of HIV	23
	Harald zur Hausen		Role of HPV in cancer	654
	Martin Chalfie Osamu Shimomura Roger Tsien	Chemistry	Discovery and development of GFP	267, 720
2007	Mario R. Capecchi Martin J. Evans Oliver Smithies	M & P	Development of techniques for knockout mice	760
2006	Andrew Z. Fire Craig C. Mello	M & P	RNA Interference	449, 762
	Roger D. Kornberg	Chemistry	Transcription in eukaryotes	427, 481
2004	Richard Axel Linda B. Buck	M & P	Olfactory receptors	622
	Aaron Ciechanover Avram Hershko Irwin Rose	Chemistry	Ubiquitin and proteasomes	529
2003	Peter Agre Roderick MacKinnon	Chemistry	Structure of membrane channels	146, 148
2002	Sydney Brenner John Sulston	M & P	Introduction of *C. elegans* as a model organism	17
	H. Robert Horvitz		Apoptosis in *C. elegans*	643
	John B. Fenn	Chemistry	Electrospray ionization in MS	740
	Koichi Tanaka		MALDI in MS	740
	Kurt Wüthrich		NMR analysis of proteins	56
2001	Leland H. Hartwell Tim Hunt Paul Nurse	M & P	Control of the cell cycle	564, 600
2000	Arvid Carlsson	M & P	Synaptic transmission and signal transduction	163
	Paul Greengard Eric Kandel			605
1999	Günter Blobel	M & P	Protein trafficking	276
1998	Robert Furchgott Louis Ignarro Ferid Murad	M & P	NO as intercellular messenger	641
1997	Jens C. Skou	Chemistry	Na^+/K^+-ATPase	153
	Paul Boyer		Mechanism of ATP synthesis	195
	John Walker			
	Stanley B. Prusiner	M & P	Protein nature of prions	64
1996	Rolf M. Zinkernagel Peter C. Doherty	M & P	Recognition of virus-infected cells by the immune system	709
1995	Edward B. Lewis Christiane Nüsslein-Volhard Eric Wieschaus	M & P	Genetic control of embryonic development	EP12
1994	Alfred Gilman Martin Rodbell	M & P	Structure and function of GTP-binding (G) proteins	610
1993	Kary Mullis	Chemistry	Polymerase chain reaction (PCR)	751
	Michael Smith		Site-directed mutagenesis (SDM)	760
	Richard J. Roberts Phillip A. Sharp	M & P	Intervening sequences	438

Year	Recipient*	Prize	Area of Research	Pages in Text
1992	Edmond Fischer Edwin Krebs	M & P	Alteration of enzyme activity by phosphorylation/dephosphorylation	112, 614
1991	Erwin Neher Bert Sakmann	M & P	Measurement of ion flux by patch-clamp recording	147
1989	J. Michael Bishop Harold Varmus	M & P	Cellular genes capable of causing malignant transformation	677
	Thomas R. Cech Sidney Altman	Chemistry	Ability of RNA to catalyze reactions	469
1988	Johann Deisenhofer Robert Huber Hartmut Michel	Chemistry	Bacterial photosynthetic reaction center	213
1987	Susumu Tonegawa	M & P	DNA rearrangements responsible for antibody diversity	696
1986	Rita Levi-Montalcini Stanley Cohen	M & P	Factors that affect nerve outgrowth	372
1985	Michael S. Brown Joseph L. Goldstein	M & P	Regulation of cholesterol metabolism and endocytosis	312
1984	Georges Köhler Cesar Milstein	M & P	Monoclonal antibodies	763
	Niels K. Jerne		Antibody formation	687
	Bruce Merrifield	Chemistry	Chemical synthesis of peptides	746
1983	Barbara McClintock	M & P	Mobile elements in the genome	402
1982	Aaron Klug	Chemistry	Structure of nucleic acid-protein complexes	76
1980	Paul Berg	Chemistry	Recombinant DNA technology	748
	Walter Gilbert Frederick Sanger		DNA sequencing technology	753
	Baruj Bennacerraf Jean Dausset George D. Snell	M & P	Major histocompatibility complex	699
1978	Werner Arber Daniel Nathans Hamilton O. Smith	M & P	Restriction endonuclease technology	746
	Peter Mitchell	Chemistry	Chemiosmotic mechanism of oxidative phosphorylation	181
1976	D. Carleton Gajdusek	M & P	Prion-based diseases	64
1975	David Baltimore Renato Dulbecco Howard M. Temin	M & P	Reverse transcriptase and tumor virus activity	676
1974	Albert Claude Christian de Duve George E. Palade	M & P	Structure and function of internal components of cells	267
1972	Gerald Edelman Rodney R. Porter	M & P	Immunoglobulin structure	693
	Christian B. Anfinsen	Chemistry	Relationship between primary and tertiary structure of proteins	62
1971	Earl W. Sutherland	M & P	Mechanism of hormone action and cyclic AMP	614
1970	Bernard Katz Ulf von Euler	M & P	Nerve impulse propagation and transmission	160
	Luis F. Leloir	Chemistry	Role of sugar nucleotides in carbohydrate synthesis	280
1969	Max Delbrück Alfred D. Hershey Salvador E. Luria	M & P	Genetic structure of viruses	22, 415

Year	Recipient*	Prize	Area of Research	Pages in Text
1968	H. Gobind Khorana Marshall W. Nirenberg	M & P	Genetic code	456
	Robert W. Holley		Transfer RNA structure	457
1966	Peyton Rous	M & P	Tumor viruses	676
1965	Francois Jacob Andre M. Lwoff Jacques L. Monod	M & P	Bacterial operons and messenger RNA	500, 421
1964	Dorothy C. Hodgkin	Chemistry	X-ray structure of complex biological molecules	740
1963	John C. Eccles Alan L. Hodgkin Andrew F. Huxley	M & P	Ionic basis of nerve membrane potentials	159
1962	Francis H. C. Crick James D. Watson Maurice H. F. Wilkins	M & P	Three-dimensional structure of DNA	386
	John C. Kendrew Max F. Perutz	Chemistry	Three-dimensional structure of globular proteins	56
1961	Melvin Calvin	Chemistry	Biochemistry of CO_2 assimilation during photosynthesis	221
1960	F. MacFarlane Burnet Peter B. Medawar	M & P	Clonal selection theory of antibody formation	687
1959	Arthur Kornberg Severo Ochoa	M & P	Synthesis of DNA and RNA	538, 456
1958	George W. Beadle Joshua Lederberg Edward L. Tatum	M & P	Gene expression	420
	Frederick Sanger	Chemistry	Primary structure of proteins	54

*In a few cases, corecipients whose research was in an area outside of cell and molecular biology have been omitted from this list.

**Medicine and Physiology

ALL THE HELP, RESOURCES, AND PERSONAL SUPPORT YOU AND YOUR STUDENTS NEED!

1st DAY OF CLASS ...AND BEYOND!

2-Minute Tutorials and all of the resources you & your students need to get started
www.wileyplus.com/firstday

WILEY PLUS Student Partner Program

Student support from an experienced student user
Ask your local representative for details!

WILEY FACULTY NETWORK

Collaborate with your colleagues, find a mentor, attend virtual and live events, and view resources
www.WhereFacultyConnect.com

WILEY PLUS QuickStart

Pre-loaded, ready-to-use assignments and presentations
www.wiley.com/college/quickstart

FAQ

Technical Support 24/7 FAQs, online chat, and phone support
www.wileyplus.com/support

Your *WileyPLUS* Account Manager
Training and implementation support
www.wileyplus.com/accountmanager

WILEY PLUS
www.wile**yplus**.com

MAKE IT YOURS!

6th edition

Cell and Molecular Biology

Concepts

and

Experiments

Gerald Karp

About the Author

Gerald C. Karp received a bachelor's degree from UCLA and a Ph.D. from the University of Washington. He conducted postdoctoral research at the University of Colorado Medical Center before joining the faculty at the University of Florida. Gerry is the author of numerous research articles on the cell and molecular biology of early development. His interests have included the synthesis of RNA in early embryos, the movement of mesenchyme cells during gastrulation, and cell determination in slime molds. For 13 years, he taught courses in molecular, cellular, and developmental biology at the University of Florida. During this period, Gerry coauthored a text in developmental biology with N. John Berrill and authored a text in cell and molecular biology. Finding it impossible to carry on life as both full-time professor and author, Gerry gave up his faculty position to concentrate on writing. He hopes to revise this text every three years

About the Cover

A molecular model of the membrane of a synaptic vesicle. Within nerve cells, a synaptic vesicle consists of a cellular membrane surrounding a soluble compartment filled with neurotransmitter molecules. Vesicles of this type are assembled in the vicinity of a nerve cell's nucleus and then transported to the tip of the axon. There the vesicle awaits the arrival of a nerve impulse that will induce it to fuse with the overlying plasma membrane, releasing its contents into the narrow cleft that separates the nerve cell from a neighboring cell. The three-dimensional model of this membrane was constructed using known structures of the various proteins along with information on their relative numbers obtained from the analysis of purified synaptic vesicles. The image on the front cover shows a synaptic vesicle that has been cut in half; the lipid bilayer that forms the core of the vesicle membrane is shown in green. The image on the back cover shows the surface structure of an intact vesicle. Most of the proteins present in this membrane are required for the interaction of the vesicle with the plasma membrane. The large blue protein at the lower right of the vesicle contains a ring of subunits that rotates within the lipid bilayer as the protein pumps hydrogen ions into the vesicle. The elevated concentration of hydrogen ions within the vesicle is subsequently used as an energy source for the uptake of neurotransmitter molecules from the surrounding cytosol. These images provide the most comprehensive model of any cellular membrane yet to be studied and they reveal how much this membrane is dominated by protein—both within the bilayer itself and on both membrane surfaces. (From Shigeo Takamori et al., courtesy of Reinhard Jahn of the Max-Planck Institute for Biophysical Chemistry, *Cell* 127:841, 2006.)

To Patsy and Jenny

Preface to the Sixth Edition

Before I began work on the *first* edition of this text, I drew up a number of basic guidelines regarding the type of book I planned to write.

● I wanted a text suited for an introductory course in cell and molecular biology that ran either a single semester or 1–2 quarters. I set out to draft a text of about 800 pages that would not overwhelm or discourage students at this level.

● I wanted a text that elaborated on fundamental concepts, such as the relationship between molecular structure and function, the dynamic character of cellular organelles, the use of chemical energy in running cellular activities and ensuring accurate macromolecular biosynthesis, the observed unity and diversity at the macromolecular and cellular levels, and the mechanisms that regulate cellular activities.

● I wanted a text that was grounded in the experimental approach. Cell and molecular biology is an experimental science and, like most instructors, I believe students should gain some knowledge of how we know what we know. With this in mind, I decided to approach the experimental nature of the subject in two ways. As I wrote each chapter, I included enough experimental evidence to justify many of the conclusions that were being made. Along the way, I described the salient features of key experimental approaches and research methodologies. Chapters 8 and 9, for example, contain introductory sections on techniques that have proven most important in the analysis of cytomembranes and the cytoskeleton, respectively. I included brief discussions of selected experiments of major importance in the body of the chapters to reinforce the experimental basis of our knowledge. I placed the more detailed aspects of methodologies in a final "techniques chapter" because (1) I did not want to interrupt the flow of discussion of a subject with a large tangential section on technology and (2) I realized that different instructors prefer to discuss a particular technology in connection with different subjects.

For students and instructors who wanted to explore the experimental approach in greater depth, I included an Experimental Pathways at the end of most chapters. Each of these narratives describes some of the key experimental findings that have led to our current understanding of a particular subject that is relevant to the chapter at hand. Because the scope of the narrative is limited, the design of the experiments can be considered in some detail. The figures and tables provided in these sections are often those that appeared in the original research article, which provides the reader an opportunity to examine original data and to realize that its analysis is not beyond their means. The Experimental Pathways also illustrate the stepwise nature of scientific discovery, showing how the result of one study raises questions that provide the basis for subsequent studies.

● I wanted a text that was interesting and readable. To make the text more relevant to undergraduate readers, particularly premedical students, I included The Human Perspective. These sections illustrate that virtually all human disorders can be traced to disruption of activities at the cellular and molecular level. Furthermore, they reveal the importance of basic research as the pathway to understanding and eventually treating most disorders. In Chapter 11, for example, The Human Perspective describes how small synthetic siRNAs may prove to be an important new tool in the treatment of cancer and viral diseases, including AIDS. In this same chapter, the reader will learn how the action of such RNAs were first revealed in studies on plants and nematodes. It becomes evident that one can never predict the practical importance of basic research in cell and molecular biology. I have also tried to include relevant information about human biology and clinical applications throughout the body of the text.

● I wanted a high-quality illustration program that helped students visualize complex cellular and molecular processes. To meet this goal, many of the illustrations have been "stepped-out" so that information can be more easily broken down into manageable parts. Events occurring at each step are described in the figure legend and/or in the corresponding text. I also sought to include a large number of micrographs to enable students to see actual representations of most subjects being discussed. Included among the photographs are many fluorescence micrographs that illustrate either the dynamic properties of cells or provide a means to localize a specific protein or nucleic acid sequence. Wherever possible, I have tried to pair line art drawings with micrographs to help students compare idealized and actual versions of a structure.

The most important changes in the sixth edition can be delineated as follows:

● The references that have always appeared at the end of each chapter in previous editions will now appear as a section at the end of the book.

● The body of information in cell and molecular biology is continually changing, which provides much of the excitement we all feel about our selected field. Even though only three years have passed since the publication of the fifth edition, nearly every discussion in the text has been modified to a greater or lesser degree. This has been done without allowing the chapters to increase significantly in length.

● Every illustration in the fifth edition has been scrutinized and many of those that were reutilized in the sixth edition have been modified to some extent. Many of the drawings from the fifth edition have been deleted to make room for new pieces. Instructors have expressed particular approval for figures that juxtapose line art and micrographs, and this style of illustration has been expanded in the sixth edition. Altogether, the sixth edition contains more than 60 new micrographs and computer-derived images, all of which were provided by the original source.

WileyPLUS

This online teaching and learning environment integrates the **entire digital textbook** with the most effective instructor and student resources to fit every learning style.

With *WileyPLUS:*

- Students achieve concept mastery in a rich, structured environment that's available 24/7

- Instructors personalize and manage their course more effectively with assessment, assignments, grade tracking, and more.
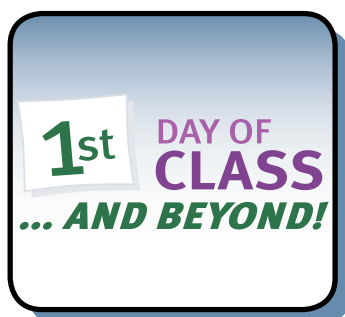
WileyPLUS can complement your current textbook or replace the printed text altogether.

For Students

Personalize the learning experience

Different learning styles, different levels of proficiency, different levels of preparation—each of your students is unique. *WileyPLUS* empowers them to take advantage of their individual strengths:

- Students receive timely access to resources that address their demonstrated needs, and get immediate feedback and remediation when needed.

- Integrated, multi-media resources include

  - Animations of key concepts based on the illustrations of the text.

  - Video library of clips from leading journals which can now be assigned with accompanying questions by Anne Hemsley, Antelope Valley Community College.

- *WileyPLUS* includes many opportunities for self-assessment linked to the relevant portions of the text. Students can take control of their own learning and practice until they master the material.

For Instructors

Personalize the teaching experience

WileyPLUS empowers you with the tools and resources you need to make your teaching even more effective:

- You can customize your classroom presentation with a wealth of resources and functionality from PowerPoint slides to a database of rich visuals. You can even add your own materials to your *WileyPLUS* course.

- With *WileyPLUS* you can identify those students who are falling behind and intervene accordingly, without having to wait for them to come to office hours.

- *WileyPLUS* simplifies and automates such tasks as student performance assessment, making assignments, scoring student work, keeping grades, and more.

  - Pre and Post Lecture Assessment by Jose Vazquez, New York University.

  - Test Bank, Instructor's Manual, and "Clicker" Questions by Joel Piperberg, Millersville University.

  - **NEW** Lecture PowerPoint Presentations by Jose Vazquez, New York University.

NEW *Clinical and Experimental Focus!*

- **NEW** Clinical Case Studies and accompanying questions by Claire Walczak, Indiana University.

- **NEW** Clinical Connections Questions by Sarah VanVickle-Chavez, Washington University in St. Louis.

- Experimental Pathways Questions by Joel Piperberg, Millersville University.

Book Companion Site (www.wiley.com/college/karp)

For the Student

- *Quizzes* for student self-testing.

- *Biology NewsFinder; Flash Cards; and Animations.*

- Answers to the end-of chapter Analytic Questions.

- Additional reading resources provide students with an extensive list of additional useful sources of information.

- Experimental Pathways for Chapters 5, 6, 7, 9, 12, 13, and 15.

For the Instructor

- Biology Visual Library; all images in jpg and PowerPoint formats.

- Instructor's Manual; Test Bank; Clicker Questions; Lecture PowerPoint Presentations.

- Transparencies on Demand. Located on the book companion site, instructors can choose only those transparencies they need, eliminating waste and protecting our environment.

Instructor Resources are password protected.

Acknowledgments

Editor, served as the central nervous system, coordinating the information arriving from compositors, copyeditors, proofreaders, illustrators, photo editors, designers, and dummiers, as well as the constant barrage of text changes ordered by the author. Always calm, organized, and meticulous, she made sure everything was done correctly. Hilary Newman and Anna Melhorn were responsible for the photo and line-art programs respectively. It has been my good fortune to work with Hilary on all six editions of this text. Hilary is skillful and perseverant, and I have utmost confidence in her ability to obtain any image requested. It was also a great pleasure working with Anna for the fourth time. The book has a complex illustration program and Anna did a superb job in coordinating all the many facets required to guide it to completion. The elegant design of the book and cover is due to the efforts of Madelyn Lesure, whose talents are evident. Thanks to Alissa Etrheim who served as editorial assistant for most of the project but moved to the wilds of Alaska just before publication. Thanks also to Claire Walczak for all of her help in revising Chapter 9 and contributing a section on fluorescence imaging techniques and the accompanying artwork. A special thanks is owed Laura Ierardi who skillfully laid out the pages for each chapter.

I am especially thankful to the many biologists who have contributed micrographs for use in this book; more than any other element, these images bring the study of cell biology to life on the printed page. Finally, I would like to apologize in advance for any errors that may occur in the text, and express my heartfelt embarrassment. I am grateful for the constructive criticism and sound advice from the following reviewers:

To the Student

At the time I began college, biology would have been at the bottom of a list of potential majors. I enrolled in a physical anthropology course to fulfill the life science requirement by the easiest possible route. During that course, I learned for the first time about chromosomes, mitosis, and genetic recombination, and I became fascinated by the intricate activities that could take place in such a small volume of cellular space. The next semester, I took Introductory Biology and began to seriously consider becoming a cell biologist. I am burdening you with this personal trivia so you will understand why I wrote this book and to warn you of possible repercussions.

Even though many years have passed, I still find cell biology the most fascinating subject to explore, and I still love spending the day reading about the latest findings by colleagues in the field. Thus, for me, writing a text on cell biology provides a reason and an opportunity to keep abreast with what is going on throughout the field. My primary goal in writing this text is to help generate an appreciation in students for the activities in which the giant molecules and minuscule structures that inhabit the cellular world of life are engaged. Another goal is to provide the reader with an insight into the types of questions that cell and molecular biologists ask and the experimental approaches they use to seek answers. As you read the text, think like a researcher; consider the evidence that is presented, think of alternate explanations, plan experiments that could lead to new hypotheses.

You might begin this approach by looking at one of the many electron micrographs that fill the pages of this text. To take this photograph, you would be sitting in a small, pitch-black room in front of a large metallic instrument whose column rises several meters above your head. You are looking through a pair of binoculars at a vivid, bright green screen. The parts of the cell you are examining appear dark and colorless against the bright green background. They are dark because they've been stained with heavy metal atoms that deflect a fraction of the electrons within a beam that is being focused on the viewing screen by large electromagnetic lenses in the wall of the column. The electrons that strike the screen are accelerated through the evacuated space of the column by a force of tens of thousands of volts. One of your hands may be gripping a knob that controls the magnifying power of the lenses. A simple turn of this knob can switch the image in front of your eyes from that of a whole field of cells to a tiny part of a cell, such as a few ribosomes or a small portion of a single membrane. By turning other knobs, you can watch different parts of the specimen glide across the screen, giving you the sensation that you're driving around inside a cell. Once you have found a structure of interest, you can turn a handle that lifts the screen out of view, allowing the electron beam to strike a piece of film and produce a photographic image of the specimen.

Because the study of cell function requires the use of considerable instrumentation, such as the electron microscope just described, the investigator is physically removed from the subject being studied. To a large degree, cells are like tiny black boxes. We have developed many ways to probe the boxes, but we are always groping in an area that cannot be fully illuminated. A discovery is made or a new technique is developed and a new thin beam of light penetrates the box. With further work, our understanding of the structure or process is broadened, but we are always left with additional questions. We generate more complete and sophisticated constructions, but we can never be sure how closely our views approach reality. In this regard, the study of cell and molecular biology can be compared to the study of an elephant as conducted by six blind men in an old Indian fable. The six travel to a nearby palace to learn about the nature of elephants. When they arrive, each approaches the elephant and begins to touch it. The first blind man touches the side of the elephant and concludes that an elephant is smooth like a wall. The second touches the trunk and decides that an elephant is round like a snake. The other members of the group touch the tusk, leg, ear, and tail of the elephant, and each forms his impression of the animal based on his own limited experiences. Cell biologists are limited in a similar manner as to what they can learn by using a particular technique or experimental approach. Although each new piece of information adds to the preexisting body of knowledge to provide a better concept of the activity being studied, the total picture remains uncertain.

Before closing these introductory comments, let me take the liberty of offering the reader some advice: Don't accept everything you read as being true. There are several reasons for urging such skepticism. Undoubtedly, there are errors in this text that reflect the author's ignorance or misinterpretation of some aspect of the scientific literature. But, more importantly, we should consider the nature of biological research. Biology is an empirical science; nothing is ever proved. We compile data concerning a particular cell organelle, metabolic reaction, intracellular movement, etc., and draw some type of conclusion. Some conclusions rest on more solid evidence than others. Even if there is a consensus of agreement concerning the "facts" regarding a particular phenomenon, there are often several possible interpretations of the data. Hypotheses are put forth and generally stimulate further research, thereby leading to a reevaluation of the original proposal. Most hypotheses that remain valid undergo a sort of evolution and, when presented in the text, should not be considered wholly correct or incorrect.

Cell biology is a rapidly moving field and some of the best hypotheses often generate considerable controversy. Even though this is a textbook where one expects to find material that is well tested, there are many places where new ideas are presented. These ideas are often described as models. I've included such models because they convey the current thinking in the field, even if they are speculative. Moreover, they reinforce the idea that cell biologists operate at the frontier of science, a boundary between the unknown and known (or thought to be known). Remain skeptical.

Brief Contents

Contents

1

Introduction to the Study of Cell and Molecular Biology

C ells, and the structures they comprise, are too small to be directly seen, heard, or touched. In spite of this tremendous handicap, cells are the subject of hundreds of thousands of publications each year, with virtually every aspect of their minuscule structure coming under scrutiny. In many ways, the study of cell and molecular biology stands as a tribute to human curiosity for seeking to discover, and to human creative intelligence for devising the complex instruments and elaborate techniques by which these discoveries can be made. This is not to imply that cell and molecular biologists have a monopoly on these noble traits. At one end of the scientific spectrum, astronomers are searching the outer fringes of the universe for black holes and whirling pulsars whose properties seem unimaginable when compared to those of Earth. At the other end of the spectrum, nuclear physicists are focusing their attention on subatomic particles that have equally inconceivable properties. Clearly, our universe consists of worlds within worlds, all aspects of which make for fascinating study.

As will be apparent throughout this book, cell and molecular biology is *reductionist*; that is, it is based on the view that knowledge of the parts of the whole can explain the character of the whole. When viewed in this way, our feeling for the wonder and

An example of the role of technological innovation in the field of cell biology. This light micrograph shows a cell lying on a microscopic bed of synthetic posts. The flexible posts serve as sensors to measure mechanical forces exerted by the cell. The red-stained elements are bundles of actin filaments within the cell that generate forces during cell locomotion. When the cell moves, it pulls on the attached posts, which report the amount of strain they are experiencing. The cell nucleus is stained green. (FROM J. L. TAN ET AL., PROC. NAT'L. ACAD. SCI. U.S.A., 100 (4), 2003; COURTESY OF CHRISTOPHER S. CHEN, THE JOHNS HOPKINS UNIVERSITY.)

mystery of life may be replaced by the need to explain every-thing in terms of the workings of the "machinery" of the living system. To the degree to which this occurs, it is hoped that this loss can be replaced by an equally strong appreciation for the beauty and complexity of the mechanisms underlying cellular activity. ■

1.1 THE DISCOVERY OF CELLS

Because of their small size, cells can only be observed with the aid of a **microscope**, an instrument that provides a magnified image of a tiny object. We do not know when humans first discovered the remarkable ability of curved-glass surfaces to bend light and form images. Spectacles were first made in Europe in the thirteenth century, and the first compound (double-lens) light microscopes were constructed by the end of the sixteenth century. By the mid-1600s, a handful of pioneering scientists had used their handmade microscopes to uncover a world that would never have been revealed to the naked eye. The discovery of cells (Figure 1.1*a*) is generally credited to Robert Hooke, an English microscopist who, at age 27, was awarded the position of curator of the Royal Society of London, England's foremost scientific academy. One of the many questions Hooke attempted to answer was why stoppers made of cork (part of the bark of trees) were so well suited to holding air in a bottle. As he wrote in 1665: "I took a good clear piece of cork, and with a Pen-knife sharpen'd as keen as a Razor, I cut a piece of it off, and . . . then examining it with a *Microscope*, me thought I could perceive it to appear a little porous . . . much like a Honeycomb." Hooke called the pores *cells* because they reminded him of the cells inhabited by monks living in a monastery. In actual fact, Hooke had observed the empty cell walls of dead plant tissue, walls that had originally been produced by the living cells they surrounded.

Meanwhile, Anton van Leeuwenhoek, a Dutchman who earned a living selling clothes and buttons, was spending his spare time grinding lenses and constructing simple microscopes of remarkable quality (Figure 1.1*b*). For 50 years, Leeuwenhoek sent letters to the Royal Society of London describing his microscopic observations—along with a rambling discourse on his daily habits and the state of his health. Leeuwenhoek was the first to examine a drop of pond water under the microscope and, to his amazement, observe the teeming microscopic "animalcules" that darted back and forth before his eyes. He was also the first to describe various forms of bacteria, which he obtained from water in which pepper had been soaked and from scrapings of his teeth. His initial letters to the Royal Society describing this previously unseen world were met with such skepticism that the society dispatched its curator, Robert Hooke, to confirm the observations. Hooke did just that, and Leeuwenhoek was soon a worldwide celebrity, receiving visits in Holland from Peter the Great of Russia and the queen of England.

It wasn't until the 1830s that the widespread importance of cells was realized. In 1838, Matthias Schleiden, a German lawyer turned botanist, concluded that, despite differences in

(a)

(b)

FIGURE 1.1 **The discovery of cells.** (*a*) One of Robert Hooke's more ornate compound (double-lens) microscopes. (Inset) Hooke's drawing of a thin slice of cork, showing the honeycomb-like network of "cells." (*b*) Single-lens microscope used by Anton van Leeuwenhoek to observe bacteria and other microorganisms. The biconvex lens, which was capable of magnifying an object approximately 270 times and providing a resolution of approximately 1.35 μm, was held between two metal plates. (FROM THE GRANGER COLLECTION; (INSET AND FIGURE 1-1*B*) CORBIS BETTMANN)

the structure of various tissues, plants were made of cells and that the plant embryo arose from a single cell. In 1839, Theodor Schwann, a German zoologist and colleague of Schleiden's, published a comprehensive report on the cellular basis of animal life. Schwann concluded that the cells of plants and animals are similar structures and proposed these two tenets of the **cell theory**:

- All organisms are composed of one or more cells.
- The cell is the structural unit of life.

Schleiden and Schwann's ideas on the *origin* of cells proved to be less insightful; both agreed that cells could arise from noncellular materials. Given the prominence that these two

scientists held in the scientific world, it took a number of years before observations by other biologists were accepted as demonstrating that cells did not arise in this manner any more than organisms arose by spontaneous generation. By 1855, Rudolf Virchow, a German pathologist, had made a convincing case for the third tenet of the cell theory:

■ Cells can arise only by division from a preexisting cell.

1.2 BASIC PROPERTIES OF CELLS

Just as plants and animals are alive, so too are cells. Life, in fact, is the most basic property of cells, and cells are the smallest units to exhibit this property. Unlike the parts of a cell, which simply deteriorate if isolated, whole cells can be removed from a plant or animal and cultured in a laboratory where they will grow and reproduce for extended periods of time. If mistreated, they may die. Death can also be considered one of the most basic properties of life, because only a living entity faces this prospect. Remarkably, cells within the body generally die "by their own hand"—the victims of an internal program that causes cells that are no longer needed or cells that pose a risk of becoming cancerous to eliminate themselves.

The first culture of human cells was begun by George and Martha Gey of Johns Hopkins University in 1951. The cells were obtained from a malignant tumor and named HeLa cells after the donor, Henrietta Lacks. HeLa cells—descended by cell division from this first cell sample—are still being grown
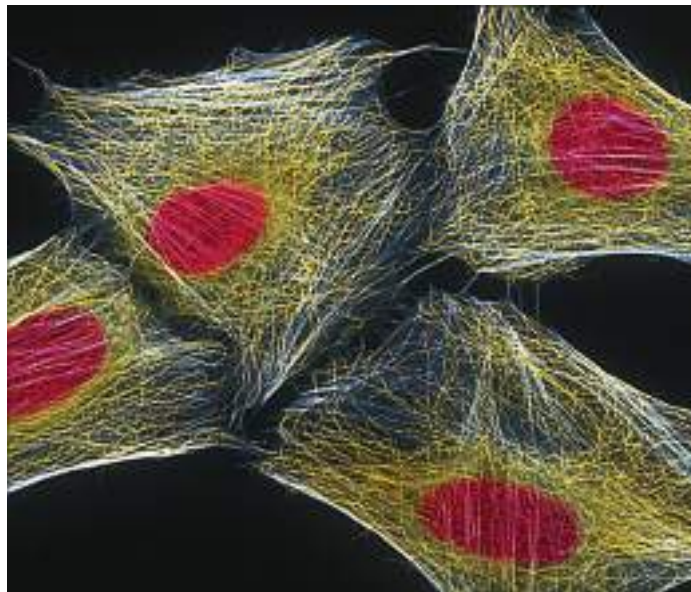
FIGURE 1.2 HeLa cells, such as the ones pictured here, were the first human cells to be kept in culture for long periods of time and are still in use today. Unlike normal cells, which have a finite lifetime in culture, these cancerous HeLa cells can be cultured indefinitely as long as conditions are favorable to support cell growth and division. (TORSTEN WITTMANN/PHOTO RESEARCHERS INC.)

in laboratories around the world today (Figure 1.2). Because they are so much simpler to study than cells situated within the body, cells grown **in vitro** (i.e., in culture, outside the body) have become an essential tool of cell and molecular biologists. In fact, much of the information that will be discussed in this book has been obtained using cells grown in laboratory cultures.

We will begin our exploration of cells by examining a few of their most fundamental properties.

Cells Are Highly Complex and Organized

Complexity is a property that is evident when encountered, but difficult to describe. For the present, we can think of complexity in terms of order and consistency. The more complex a structure, the greater the number of parts that must be in their proper place, the less tolerance of errors in the nature and interactions of the parts, and the more regulation or control that must be exerted to maintain the system. Cellular activities can be remarkably precise. DNA duplication, for example, occurs with an error rate of less than one mistake every ten million nucleotides incorporated—and most of these are quickly corrected by an elaborate repair mechanism that recognizes the defect.

During the course of this book, we will have occasion to consider the complexity of life at several different levels. We will discuss the organization of atoms into small-sized molecules; the organization of these molecules into giant polymers; and the organization of different types of polymeric molecules into complexes, which in turn are organized into subcellular organelles and finally into cells. As will be apparent, there is a great deal of consistency at every level. Each type of cell has a consistent appearance when viewed under a high-powered electron microscope; that is, its organelles have a particular shape and location, from one individual of a species to another. Similarly, each type of organelle has a consistent composition of macromolecules, which are arranged in a predictable pattern. Consider the cells lining your intestine that are responsible for removing nutrients from your digestive tract (Figure 1.3).

The epithelial cells that line the intestine are tightly connected to each other like bricks in a wall. The apical ends of these cells, which face the intestinal channel, have long processes (*microvilli*) that facilitate absorption of nutrients. The microvilli are able to project outward from the apical cell surface because they contain an internal skeleton made of filaments, which in turn are composed of protein (*actin*) monomers polymerized in a characteristic array. At their basal ends, intestinal cells have large numbers of mitochondria that provide the energy required to fuel various membrane transport processes. Each mitochondrion is composed of a defined pattern of internal membranes, which in turn are composed of a consistent array of proteins, including an electrically powered ATP-synthesizing machine that projects from the inner membrane like a ball on a stick. Each of these various levels of organization is illustrated in the insets of Figure 1.3.

Fortunately for cell and molecular biologists, evolution has moved rather slowly at the levels of biological organiza-
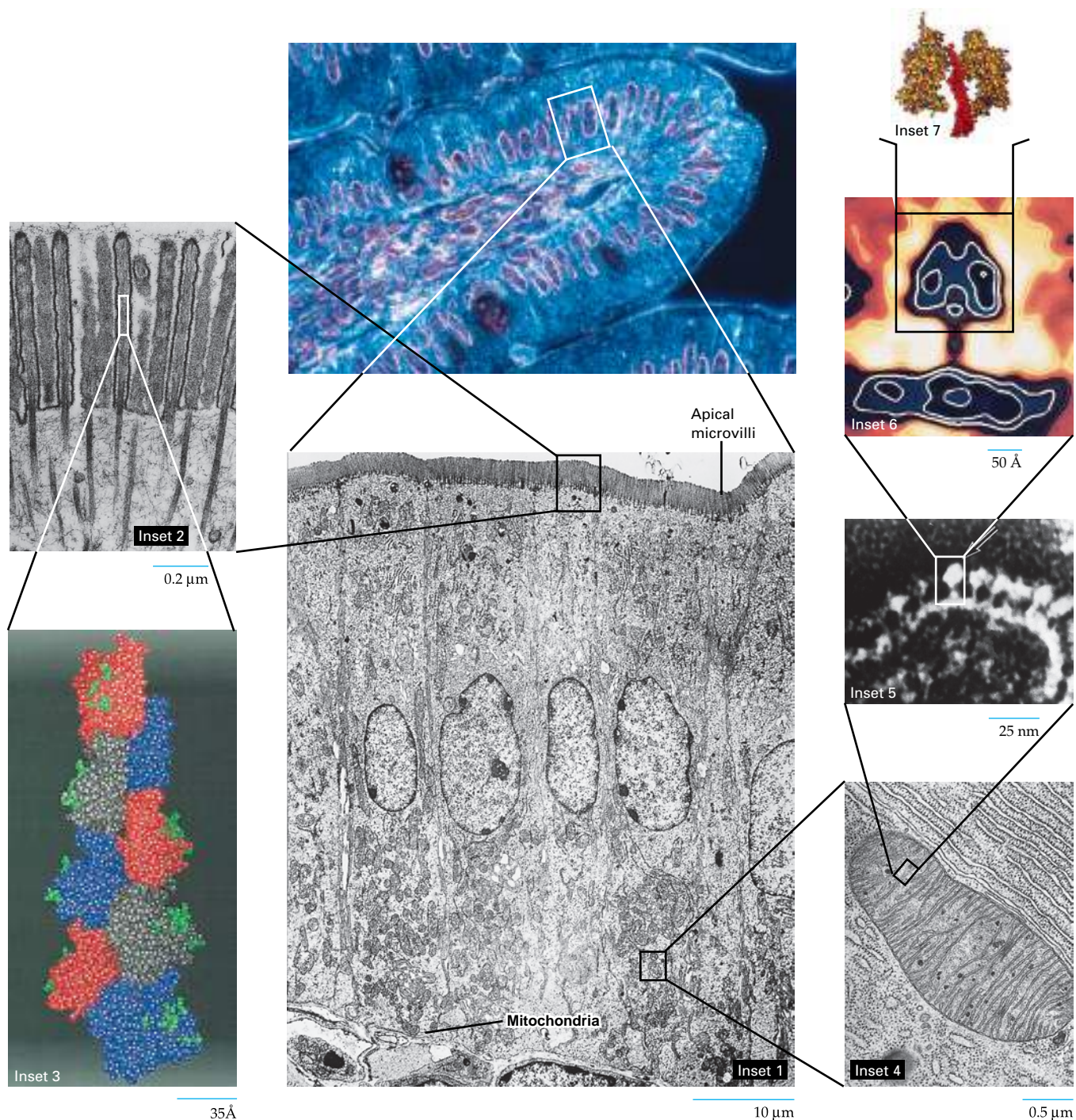
FIGURE 1.3 Levels of cellular and molecular organization. The brightly colored photograph of a stained section shows the microscopic structure of a villus of the wall of the small intestine, as seen through the light microscope. Inset 1 shows an electron micrograph of the epithelial layer of cells that lines the inner intestinal wall. The apical surface of each cell, which faces the channel of the intestine, contains a large number of microvilli involved in nutrient absorption. The basal region of each cell contains large numbers of mitochondria, where energy is made available to the cell. Inset 2 shows the apical microvilli; each microvillus contains a bundle of microfilaments. Inset 3 shows the actin protein subunits that make up each filament. Inset 4 shows an individual mitochondrion similar to those found in the basal region of the epithelial cells. Inset 5 shows a portion of an inner membrane of a mitochondrion including the stalked particles (upper arrow) that project from the membrane and correspond to the sites where ATP is synthesized. Insets 6 and 7 show molecular models of the ATP-synthesizing machinery, which is discussed at length in Chapter 5. (LIGHT MICROGRAPH CECIL FOX/PHOTO RESEARCHERS; INSET 1 COURTESY OF SHAKTI P. KAPUR, GEORGETOWN UNIVERSITY MEDICAL CENTER; INSET 2 COURTESY OF MARK S. MOOSEKER AND LEWIS G. TILNEY, J. CELL BIOL. 67:729, 1975, BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS; INSET 3 COURTESY OF KENNETH C. HOLMES; INSET 4 COURTESY OF KEITH R. PORTER/PHOTO RESEARCHERS; INSET 5 COURTESY OF HUMBERTO FERNANDEZ-MORAN; INSET 6 COURTESY OF RODERICK A. CAPALDI; INSET 7 COURTESY OF WOLFGANG JUNGE, HOLGER LILL, AND SIEGFRIED ENGELBRECHT, UNIVERSITY OF OSNABRÜCK, GERMANY.)

tion with which they are concerned. Whereas a human and a cat, for example, have very different anatomical features, the cells that make up their tissues, and the organelles that make up their cells, are very similar. The actin filament portrayed in Figure 1.3, Inset 3, and the ATP-synthesizing enzyme of Inset 6 are virtually identical to similar structures found in such diverse organisms as humans, snails, yeast, and redwood trees. Information obtained by studying cells from one type of organism often has direct application to other forms of life. Many of the most basic processes, such as the synthesis of proteins, the conservation of chemical energy, or the construction of a membrane, are remarkably similar in all living organisms.

Cells Possess a Genetic Program and the Means to Use It

Organisms are built according to information encoded in a collection of genes. The human genetic program contains enough information, if converted to words, to fill millions of pages of text. Remarkably, this vast amount of information is packaged into a set of chromosomes that occupies the space of a cell nucleus—hundreds of times smaller than the dot on this *i*.

Genes are more than storage lockers for information: they constitute the blueprints for constructing cellular structures, the directions for running cellular activities, and the program for making more of themselves. The molecular structure of genes allows for changes in genetic information (mutations) that lead to variation among individuals, which forms the basis of biological evolution. Discovering the mechanisms by which cells use their genetic information has been one of the greatest achievements of science in recent decades.

Cells Are Capable of Producing More of Themselves

Just as individual organisms are generated by reproduction, so too are individual cells. Cells reproduce by division, a process in which the contents of a "mother" cell are distributed into two "daughter" cells. Prior to division, the genetic material is faithfully duplicated, and each daughter cell receives a complete and equal share of genetic information. In most cases, the two daughter cells have approximately equal volume. In some cases, however, as occurs when a human oocyte undergoes division, one of the cells can retain nearly all of the cytoplasm, even though it receives only half of the genetic material (Figure 1.4).

Cells Acquire and Utilize Energy

Every biological process requires the input of energy. Virtually all of the energy utilized by life on the Earth's surface arrives in the form of electromagnetic radiation from the sun. The energy of light is trapped by light-absorbing pigments present in the membranes of photosynthetic cells (Figure 1.5). Light energy is converted by photosynthesis into chemical energy that is stored in energy-rich carbohydrates, such as sucrose or starch. For most animal cells, energy arrives prepackaged, of-

20 μm

FIGURE 1.4 Cell reproduction. This mammalian oocyte has recently undergone a highly unequal cell division in which most of the cytoplasm has been retained within the large oocyte, which has the potential to be fertilized and develop into an embryo. The other cell is a nonfunctional remnant that consists almost totally of nuclear material (indicated by the blue-staining chromosomes, arrow). (Courtesy of Jonathan van Blerkom.)

ten in the form of the sugar glucose. In humans, glucose is released by the liver into the blood where it circulates through the body delivering chemical energy to all the cells. Once in a cell, the glucose is disassembled in such a way that its energy content can be stored in a readily available form (usually as ATP) that is later put to use in running all of the cell's myriad energy-requiring activities. Cells expend an enormous amount of energy simply breaking down and rebuilding the macromolecules and organelles of which they are made. This continual "turnover," as it is called, maintains the integrity of cell components in the face of inevitable wear and tear and enables the cell to respond rapidly to changing conditions.

Cells Carry Out a Variety of Chemical Reactions

Cells function like miniaturized chemical plants. Even the simplest bacterial cell is capable of hundreds of different chemical transformations, none of which occurs at any sig-

FIGURE 1.5 Acquiring energy. A living cell of the filamentous alga *Spirogyra*. The ribbon-like chloroplast, which is seen to zigzag through the cell, is the site where energy from sunlight is captured and converted to chemical energy during photosynthesis. (M. I. Walker/Photo Researchers.)

nificant rate in the inanimate world. Virtually all chemical changes that take place in cells require *enzymes*—molecules that greatly increase the rate at which a chemical reaction occurs. The sum total of the chemical reactions in a cell represents that cell's **metabolism**.

Cells Engage in Mechanical Activities

Cells are sites of bustling activity. Materials are transported from place to place, structures are assembled and then rapidly disassembled, and, in many cases, the entire cell moves itself from one site to another. These types of activities are based on dynamic, mechanical changes within cells, many of which are initiated by changes in the shape of "motor" proteins. Motor proteins are just one of many types of molecular "machines" employed by cells to carry out mechanical activities.

Cells Are Able to Respond to Stimuli

Some cells respond to stimuli in obvious ways; a single-celled protist, for example, moves away from an object in its path or moves toward a source of nutrients. Cells within a multicellular plant or animal respond to stimuli less obviously. Most cells are covered with *receptors* that interact with substances in the environment in highly specific ways. Cells possess receptors to hormones, growth factors, and extracellular materials, as well as to substances on the surfaces of other cells. A cell's receptors provide pathways through which external agents can evoke specific responses in target cells. Cells may respond to specific stimuli by altering their metabolic activities, moving from one place to another, or even committing suicide.

Cells Are Capable of Self-Regulation

In addition to requiring energy, maintaining a complex, ordered state requires constant regulation. The importance of a cell's regulatory mechanisms becomes most evident when they break down. For example, failure of a cell to correct a mistake when it duplicates its DNA may result in a debilitating mutation, or a breakdown in a cell's growth-control safeguards can transform the cell into a cancer cell with the capability of destroying the entire organism. We are gradually learning how a cell controls its activities, but much more is left to discover.

Consider the following experiment conducted in 1891 by Hans Driesch, a German embryologist. Driesch found that he could completely separate the first two or four cells of a sea urchin embryo and each of the isolated cells would proceed to develop into a normal embryo (Figure 1.6). How can a cell that is normally destined to form only part of an embryo regulate its own activities and form an entire embryo? How does the isolated cell recognize the absence of its neighbors, and how does this recognition redirect the course of the cell's development? How can a part of an embryo have a sense of the whole? We are not able to answer these questions much better today than we were more than a hundred years ago when the experiment was performed.

Throughout this book we will be discussing processes that require a series of ordered steps, much like the assembly-
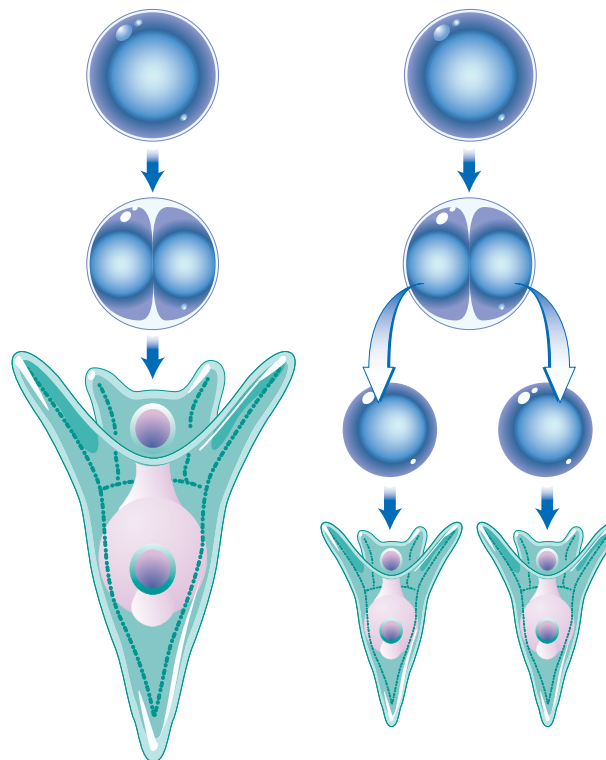
FIGURE 1.6 Self-regulation. The left panel depicts the normal development of a sea urchin in which a fertilized egg gives rise to a single embryo. The right panel depicts an experiment in which the cells of an embryo are separated from one another after the first division, and each cell is allowed to develop in isolation. Rather than developing into half of an embryo, as it would if left undisturbed, each isolated cell recognizes the absence of its neighbor, regulating its development to form a complete (although smaller) embryo.

line construction of an automobile in which workers add, remove, or make specific adjustments as the car moves along. In the cell, the information for product design resides in the nucleic acids, and the construction workers are primarily proteins. It is the presence of these two types of macromolecules that, more than any other factor, sets the chemistry of the cell apart from that of the nonliving world. In the cell, the workers must act without the benefit of conscious direction. Each step of a process must occur spontaneously in such a way that the next step is automatically triggered. In many ways, cells operate in a manner analogous to the orange-squeezing contraption discovered by "The Professor" and shown in Figure 1.7. Each type of cellular activity requires a unique set of highly complex molecular tools and machines—the products of eons of natural selection and biological evolution. A primary goal of biologists is to understand the molecular structure and role of each component involved in a particular activity, the means by which these components interact, and the mechanisms by which these interactions are regulated.

Cells Evolve

How did cells arise? Of all the major questions posed by biologists, this question may be the least likely ever to be an-

Orange Juice Squeezing Machine

FIGURE 1.7 **Cellular activities** are often analogous to this "Rube Goldberg machine" in which one event "automatically" triggers the next event in a reaction sequence. (RUBE GOLDBERG™ AND © OF RUBE GOLDBERG, INC.)

swered. It is presumed that cells evolved from some type of precellular life form, which in turn evolved from nonliving organic materials that were present in the primordial seas. Whereas the origin of cells is shrouded in near-total mystery, the evolution of cells can be studied by examining organisms that are alive today. If you were to observe the features of a bacterial cell living in the human intestinal tract (see Figure 1.18*a*) and a cell that is part of the lining of that tract (Figure 1.3), you would be struck by the differences between the two cells. Yet both have evolved from a common ancestral cell that lived more than three billion years ago. Those structures that are shared by these two distantly related cells, such as their similar plasma membrane and ribosomes, must have been present in the ancestral cell. We will examine some of the events that occurred during the evolution of cells in the Experimental Pathways at the end of the chapter. Keep in mind that evolution is not simply an event of the past, but an ongoing process that continues to modify the properties of cells that will be present in organisms that have yet to appear.

REVIEW ?

1. List the fundamental properties shared by all cells. Describe the importance of each of these properties.

2. Describe the features of cells that suggest that all living organisms are derived from a common ancestor.

3. What is the source of energy that supports life on Earth? How is this energy passed from one organism to the next?

1.3 TWO FUNDAMENTALLY DIFFERENT CLASSES OF CELLS

Once the electron microscope became widely available, biologists were able to examine the internal structure of a wide variety of cells. It became apparent from these studies that there were two basic classes of cells—prokaryotic and eukaryotic—distinguished by their size and the types of internal structures, or **organelles**, they contain (Figure 1.8). The existence of two distinct classes of cells, without any known intermediates, represents one of the most fundamental evolutionary divisions in the biological world. The structurally simpler, **prokaryotic** cells include bacteria, whereas the structurally more complex **eukaryotic** cells include protists, fungi, plants, and animals.[1]

We are not sure when prokaryotic cells first appeared on Earth. Evidence of prokaryotic life has been obtained from rocks approximately 2.7 billion years of age. Not only do these rocks contain fossilized microbes, they contain complex organic molecules that are characteristic of particular types of prokaryotic organisms, including cyanobacteria. It is unlikely that such molecules could have been synthesized abiotically, that is, without the involvement of living cells. Cyanobacteria almost certainly appeared by 2.4 billion years ago, because that is when the atmosphere become infused with molecular oxy-

---

[1]Those interested in examining a proposal to do away with the concept of prokaryotic versus eukaryotic organisms can read a brief essay by N. R. Pace in *Nature* 441:289, 2006.
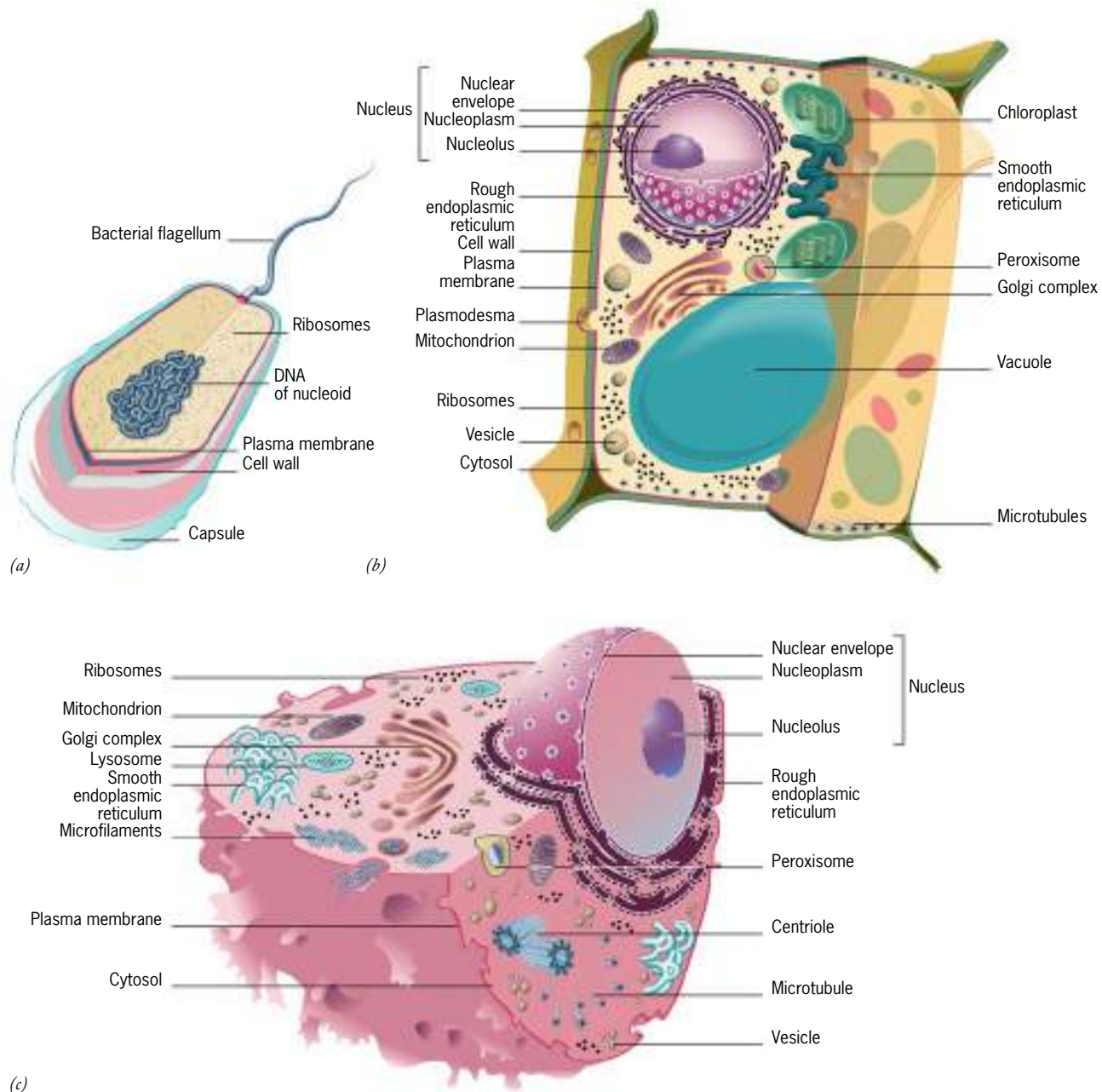
FIGURE 1.8   The structure of cells. Schematic diagrams of a "generalized" bacterial (*a*), plant (*b*), and animal (*c*) cell. Note: Organelles are not drawn to scale.

gen (O_2), which is a by-product of the photosynthetic activity of these prokaryotes. The dawn of the age of eukaryotic cells is also shrouded in uncertainty. Complex multicellular animals appear rather suddenly in the fossil record approximately 600 million years ago, but there is considerable evidence that simpler eukaryotic organisms were present on Earth more than one billion years earlier. The estimated time of appearance on Earth of several major groups of organisms is depicted in Figure 1.9. Even a superficial examination of Figure 1.9 reveals how "quickly" life arose following the formation of Earth and cooling of its surface, and how long it took for the subsequent evolution of complex animals and plants.

Characteristics That Distinguish Prokaryotic and Eukaryotic Cells

The following brief comparison between prokaryotic and eukaryotic cells reveals many basic differences between the two types, as well as many similarities (see Figure 1.8). The similarities and differences between the two types of cells are listed in Table 1.1. The shared properties reflect the fact that eukaryotic cells almost certainly evolved from prokaryotic ancestors. Because of their common ancestry, both types of cells share an identical genetic language, a common set of metabolic pathways, and many common structural features. For example, both types of cells are bounded by plasma membranes
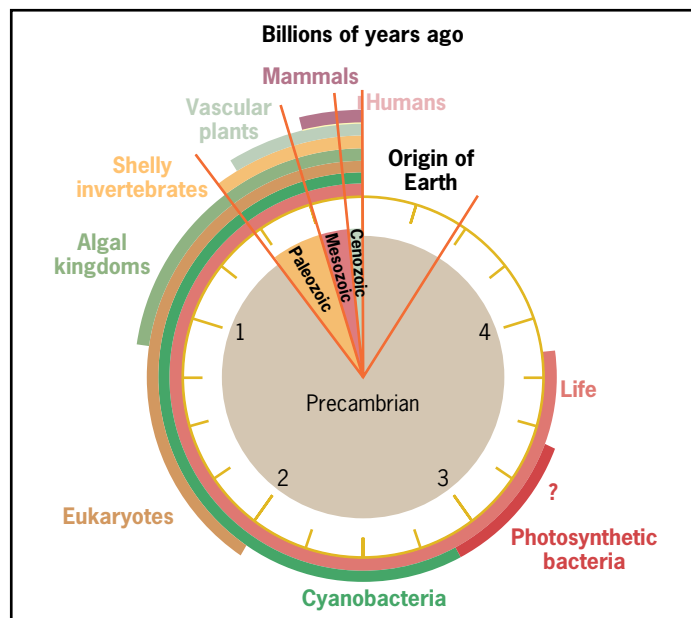
FIGURE 1.9 Earth's biogeologic clock. A portrait of the past five billion years of Earth's history showing a proposed time of appearance of major groups of organisms. Complex animals (shelly invertebrates) and vascular plants are relatively recent arrivals. The time indicated for the origin of life is speculative. In addition, photosynthetic bacteria may have arisen much earlier, hence the question mark. The geologic eras are indicated in the center of the illustration. (REPRINTED WITH PERMISSION FROM D. J. DES MARAIS, SCIENCE 289:1704, 2001. COPYRIGHT © 2000 AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

TABLE 1.1 A Comparison of Prokaryotic and Eukaryotic Cells

Features held in common by the two types of cells:

- Plasma membrane of similar construction
- Genetic information encoded in DNA using identical genetic code
- Similar mechanisms for transcription and translation of genetic information, including similar ribosomes
- Shared metabolic pathways (e.g., glycolysis and TCA cycle)
- Similar apparatus for conservation of chemical energy as ATP (located in the plasma membrane of prokaryotes and the mitochondrial membrane of eukaryotes)
- Similar mechanism of photosynthesis (between cyanobacteria and green plants)
- Similar mechanism for synthesizing and inserting membrane proteins
- Proteasomes (protein digesting structures) of similar construction (between archaebacteria and eukaryotes)

Features of eukaryotic cells not found in prokaryotes:

- Division of cells into nucleus and cytoplasm, separated by a nuclear envelope containing complex pore structures
- Complex chromosomes composed of DNA and associated proteins that are capable of compacting into mitotic structures
- Complex membranous cytoplasmic organelles (includes endoplasmic reticulum, Golgi complex, lysosomes, endosomes, peroxisomes, and glyoxisomes)
- Specialized cytoplasmic organelles for aerobic respiration (mitochondria) and photosynthesis (chloroplasts)
- Complex cytoskeletal system (including microfilaments, intermediate filaments, and microtubules) and associated motor proteins
- Complex flagella and cilia
- Ability to ingest fluid and particulate material by enclosure within plasma membrane vesicles (endocytosis and phagocytosis)
- Cellulose-containing cell walls (in plants)
- Cell division using a microtubule-containing mitotic spindle that separates chromosomes
- Presence of two copies of genes per cell (diploidy), one from each parent
- Presence of three different RNA synthesizing enzymes (RNA polymerases)
- Sexual reproduction requiring meiosis and fertilization

of similar construction that serve as a selectively permeable barrier between the living and nonliving worlds. Both types of cells may be surrounded by a rigid, nonliving *cell wall* that protects the delicate life form within. Although the cell walls of prokaryotes and eukaryotes may have similar functions, their chemical composition is very different.

Internally, eukaryotic cells are much more complex—both structurally and functionally—than prokaryotic cells (Figure 1.8). The difference in structural complexity is evident in the electron micrographs of a bacterial and an animal cell shown in Figures 1.18*a* and 1.10, respectively. Both contain a nuclear region, which houses the cell's genetic material, surrounded by cytoplasm. The genetic material of a prokaryotic cell is present in a **nucleoid**: a poorly demarcated region of the cell that lacks a boundary membrane to separate it from the surrounding cytoplasm. In contrast, eukaryotic cells possess a **nucleus**: a region bounded by a complex membranous structure called the *nuclear envelope*. This difference in nuclear structure is the basis for the terms *prokaryotic* (*pro* = before, *karyon* = nucleus) and *eukaryotic* (*eu* = true, *karyon* = nucleus). Prokaryotic cells contain relatively small amounts of DNA; the DNA content of bacteria ranges from about 600,000 base pairs to nearly 8 million base pairs and encodes between about 500 and several thousand proteins.[2] Although a "simple" baker's yeast cell

has only slightly more DNA (12 million base pairs encoding about 6200 proteins) than the most complex prokaryotes, most eukaryotic cells contain considerably more genetic information. Both prokaryotic and eukaryotic cells have DNA-containing chromosomes. Eukaryotic cells possess a number of separate chromosomes, each containing a single linear molecule of DNA. In contrast, nearly all prokaryotes that have been studied contain a single, circular chromosome. More importantly, the chromosomal DNA of eukaryotes, unlike that of prokaryotes, is tightly associated with proteins to form a complex nucleoprotein material known as **chromatin**.

The cytoplasm of the two types of cells is also very different. The cytoplasm of a eukaryotic cell is filled with a great diversity of structures, as is readily apparent by examining an electron micrograph of nearly any plant or animal cell

---

[2]Eight million base pairs is equivalent to a DNA molecule nearly 3 mm long.
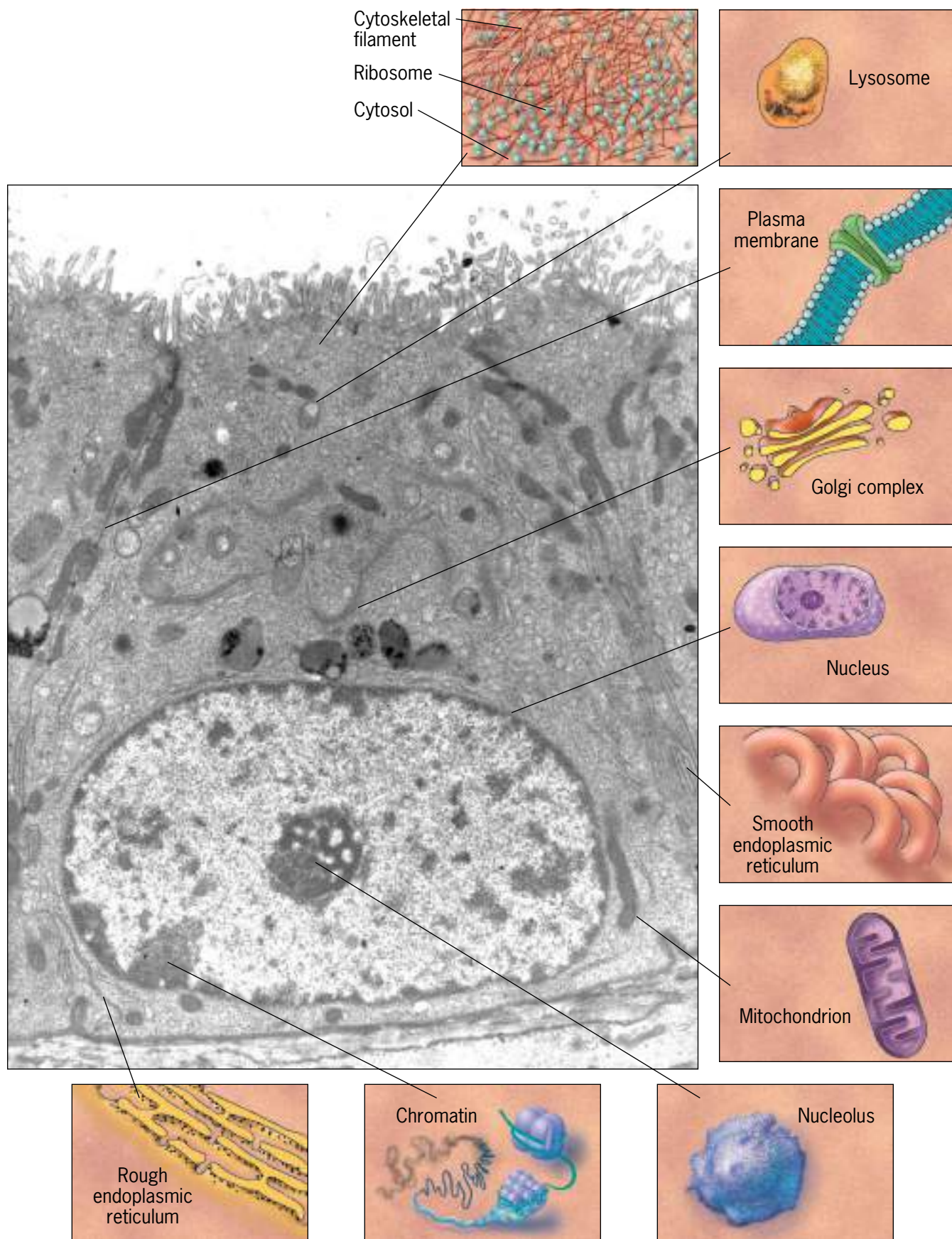
FIGURE 1.10 The structure of a eukaryotic cell. This epithelial cell lines the male reproductive tract in the rat. A number of different organelles are indicated and depicted in schematic diagrams around the border of the figure. (DAVID PHILLIPS/VISUALS UNLIMITED.)

(Figure 1.10). Even yeast, the simplest eukaryote, is much more complex structurally than an average bacterium (compare Figures 1.18*a* and *b*), even though these two organisms have a similar number of genes. Eukaryotic cells contain an array of membrane-bound organelles. Eukaryotic organelles include mitochondria, where chemical energy is made available to fuel cellular activities; an endoplasmic reticulum, where many of a cell's proteins and lipids are manufactured; Golgi complexes, where materials are sorted, modified, and transported to specific cellular destinations; and a variety of simple membrane-bound vesicles of varying dimension. Plant cells contain additional membranous organelles, including chloroplasts, which are the sites of photosynthesis, and often a single large vacuole that can occupy most of the volume of the cell. Taken as a group, the membranes of the eukaryotic cell serve to divide the cytoplasm into compartments within which specialized activities can take place. In contrast, the cytoplasm of prokaryotic cells is essentially devoid of membranous structures. The complex photosynthetic membranes of the cyanobacteria are a major exception to this generalization (see Figure 1.15).

The cytoplasmic membranes of eukaryotic cells form a system of interconnecting channels and vesicles that function in the transport of substances from one part of a cell to another, as well as between the inside of the cell and its environment. Because of their small size, directed intracytoplasmic communication is less important in prokaryotic cells, where the necessary movement of materials can be accomplished by simple diffusion.

Eukaryotic cells also contain numerous structures lacking a surrounding membrane. Included in this group are the elongated tubules and filaments of the cytoskeleton, which participate in cell contractility, movement, and support. It was thought until recently that prokaryotic cells lacked any trace of a cytoskeleton, but primitive cytoskeletal filaments have been found in bacteria. It is still fair to say that the prokaryotic cytoskeleton is much simpler, both structurally and functionally, than that of eukaryotes. Both eukaryotic and prokaryotic cells possess ribosomes, which are nonmembranous particles that function as "workbenches" on which the proteins of the cell are manufactured. Even though ribosomes of prokaryotic and eukaryotic cells have considerably different dimensions (those of prokaryotes are smaller and contain fewer components), these structures participate in the assembly of proteins by a similar mechanism in both types of cells. Figure 1.11 is a colorized electron micrograph of a portion of the cytoplasm near the thin edge of a single-celled eukaryotic organism. This is a region of the cell where membrane-bound organelles tend to be absent. The micrograph shows individual filaments of the cytoskeleton (red) and other large macromolecular complexes of the cytoplasm (green). Most of these complexes are ribosomes. It is evident from this type of image that the cytoplasm of a eukaryotic cell is extremely crowded, leaving very little space for the soluble phase of the cytoplasm, which is called the **cytosol**.

Other major differences between eukaryotic and prokaryotic cells can be noted. Eukaryotic cells divide by a complex process of mitosis in which duplicated chromosomes condense into compact structures that are segregated by an elaborate microtubule-containing apparatus (Figure 1.12). This apparatus, which is called a *mitotic spindle*, allows each daugh-
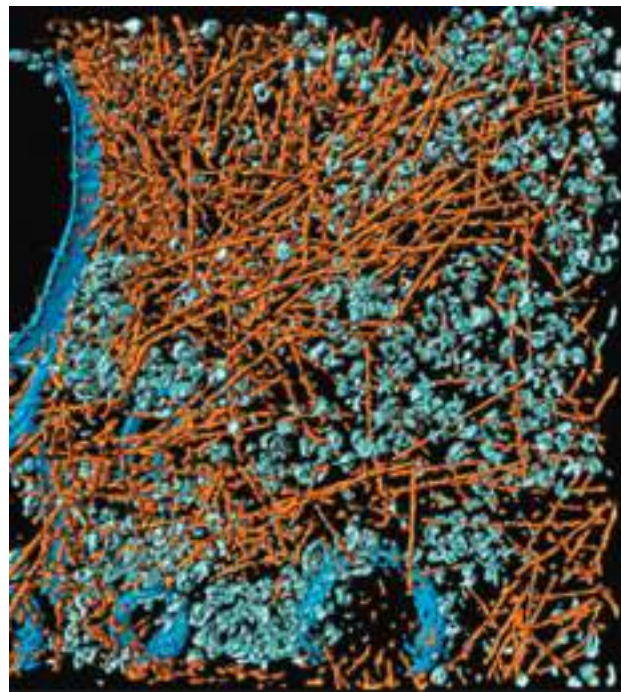
FIGURE 1.11 The cytoplasm of a eukaryotic cell is a crowded compartment. This colorized electron micrographic image shows a small region near the edge of a single-celled eukaryotic organism that had been quickly frozen prior to microscopic examination. The three-dimensional appearance is made possible by capturing two-dimensional digital images of the specimen at different angles and merging the individual frames using a computer. Cytoskeletal filaments are shown in red, macromolecular complexes (primarily ribosomes) are green, and portions of cell membranes are blue. (REPRINTED WITH PERMISSION FROM OHAD MEDALIA ET AL., SCIENCE 298:1211, 2002, COURTESY OF WOLFGANG BAUMEISTER, COPYRIGHT © 2002 AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

4 μm
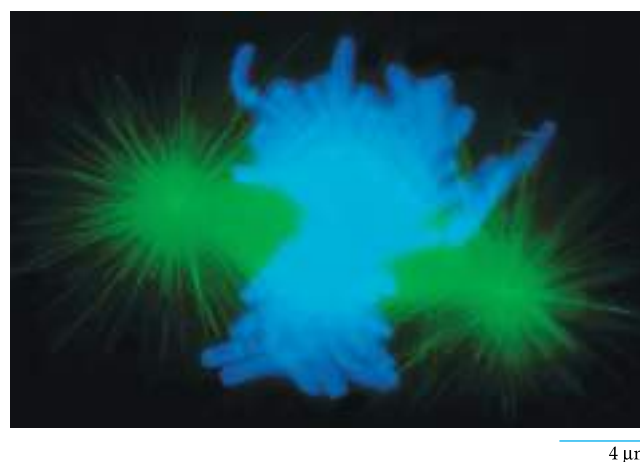
FIGURE 1.12 Cell division in eukaryotes requires the assembly of an elaborate chromosome-separating apparatus called the mitotic spindle, which is constructed primarily of microtubules. The microtubules in this micrograph appear green because they are bound by an antibody that is linked to a green fluorescent dye. The chromosomes, which were about to be separated into two daughter cells when this cell was fixed, are stained blue. (COURTESY OF CONLY L. RIEDER.)

ter cell to receive an equivalent array of genetic material. In prokaryotes, there is no compaction of the chromosome and no mitotic spindle. The DNA is duplicated, and the two copies are separated accurately by the growth of an intervening cell membrane.

For the most part, prokaryotes are nonsexual organisms. They contain only one copy of their single chromosome and have no processes comparable to meiosis, gamete formation, or true fertilization. Even though true sexual reproduction is lacking among prokaryotes, some are capable of *conjugation*, in which a piece of DNA is passed from one cell to another (Figure 1.13). However, the recipient almost never receives a whole chromosome from the donor, and the condition in which the recipient cell contains both its own and its partner's DNA is fleeting. The cell soon reverts back to possession of a single chromosome. Although prokaryotes may not be as efficient as eukaryotes in exchanging DNA with other members of their own species, they are more adept than eukaryotes at picking up and incorporating foreign DNA from their environment, which has had considerable impact on microbial evolution (page 28).
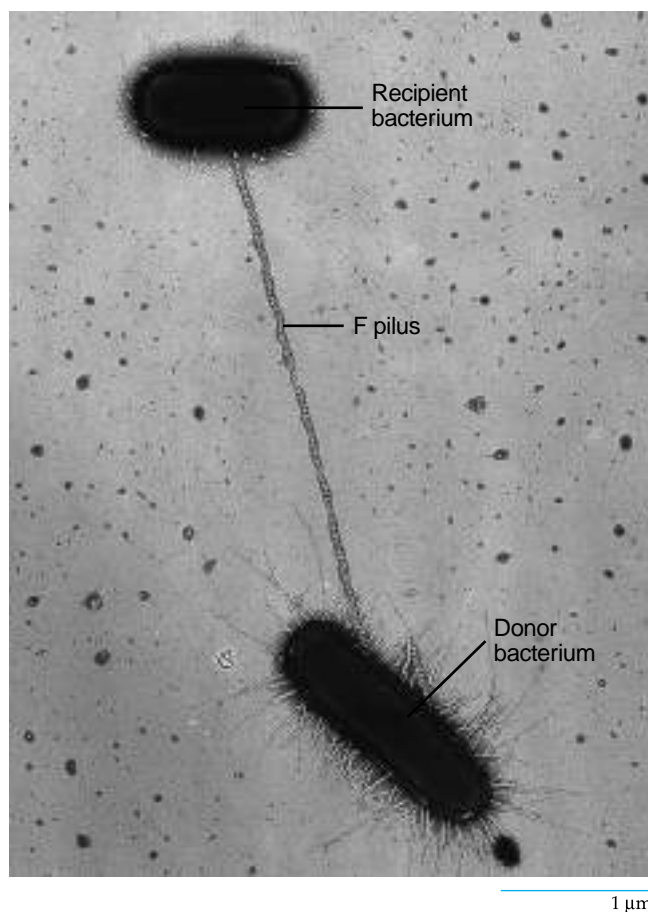
FIGURE 1.13 Bacterial conjugation. Electron micrograph showing a conjugating pair of bacteria joined by a structure of the donor cell, termed the F pilus, through which DNA is thought to be passed. (Courtesy of Charles C. Brinton, Jr., and Judith Carnahan.)

Eukaryotic cells possess a variety of complex locomotor mechanisms, whereas those of prokaryotes are relatively simple. The movement of a prokaryotic cell may be accomplished by a thin protein filament, called a *flagellum*, which protrudes from the cell and rotates (Figure 1.14*a*). The rotations of the flagellum, which can exceed 1000 times per second, exert pressure against the surrounding fluid, propelling the cell through the medium. Certain eukaryotic cells, including many protists and sperm cells, also possess flagella, but the eukaryotic versions are much more complex than the simple protein filaments of bacteria (Figure 1.14*b*), and they generate movement by a different mechanism.

In the preceding paragraphs, many of the most important differences between the prokaryotic and eukaryotic levels of cellular organization were mentioned. We will elaborate on many of these points in later chapters. Before you dismiss prokaryotes as inferior, keep in mind that these organisms have remained on Earth for more than three billion years, and at this very moment, trillions of them are clinging to the outer surface of your body and feasting on the nutrients within your digestive tract. We think of these organisms as individual, solitary creatures, but recent insights have shown that they live in complex, multispecies communities called *biofilms*. The layer of plaque that grows on our teeth is an example of a biofilm. Different cells in a biofilm may carry out different specialized activities, not unlike the cells in a plant or an animal. Consider also that, metabolically, prokaryotes are very sophisticated, highly evolved organisms. For example, a bacterium, such as *Escherichia coli*, a common inhabitant of both the human digestive tract and the laboratory culture dish, has the ability to live and prosper in a medium containing one or two low-molecular-weight organic compounds and a few inorganic ions. Other bacteria are able to live on a diet consisting solely of inorganic substances. One species of bacteria has been found in wells more than a thousand meters below the Earth's surface living on basalt rock and molecular hydrogen (H_2) produced by inorganic reactions. In contrast, even the most metabolically talented cells in your body require a variety of organic compounds, including a number of vitamins and other essential substances they cannot make on their own. In fact, many of these essential dietary ingredients are produced by the bacteria that normally live in the large intestine.

Types of Prokaryotic Cells

The distinction between prokaryotic and eukaryotic cells is based on structural complexity (as detailed in Table 1.1) and not on phylogenetic relationship. Prokaryotes are divided into two major taxonomic groups, or domains: the Archaea (or archaebacteria) and the Bacteria (or eubacteria). Members of the Archaea are more closely related to eukaryotes than they are to the other group of prokaryotes (the Bacteria). The experiments that led to the discovery that life is represented by three distinct branches are discussed in the Experimental Pathways at the end of the chapter.

The domain Archaea includes several groups of organisms whose evolutionary ties to one another are revealed by

(a)

1 μm

30 nm

(b)

0.12 μm
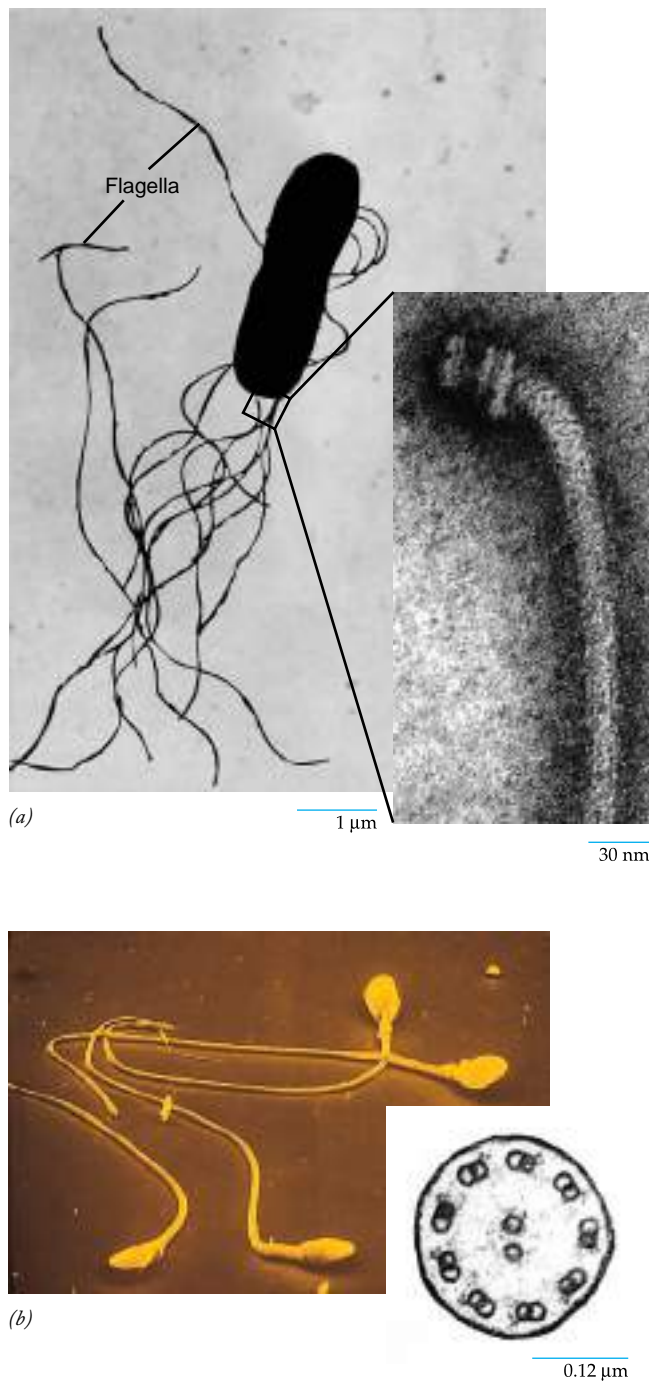
FIGURE 1.14 The difference between prokaryotic and eukaryotic flagella. (*a*) The bacterium *Salmonella* with its numerous flagella. Inset shows a high-magnification view of a portion of a single bacterial flagellum, which consists largely of a single protein called flagellin. (*b*) Each of these human sperm cells is powered by the undulatory movements of a single flagellum. The inset shows a cross section through a mammalian sperm flagellum near its tip. The flagella of eukaryotic cells are so similar that this cross section could just as well have been taken of a flagellum from a protist or green alga. (A: FROM BERNARD R. GERBER, LEWIS M. ROUTLEDGE, AND SHIRO TAKASHIMA, J. MOL. BIOL. 71:322, 1972. COPYRIGHT © 1972 BY PERMISSION OF THE PUBLISHER ACADEMIC PRESS; ELSEVIER SCIENCE, INSET COURTESY OF JULIUS ADLER AND M. L. DEPAMPHILIS; B: DAVID M. PHILLIPS/VISUALS UNLIMITED; (INSET) DON W. FAWCETT/VISUALS UNLIMITED.)

similarities in the nucleotide sequences of their nucleic acids. The best known Archaea are species that live in extremely inhospitable environments; they are often referred to as "extremophiles." Included among the Archaea are the methanogens [prokaryotes capable of converting CO_2 and H_2 gases into methane (CH_4) gas]; the halophiles (prokaryotes that live in extremely salty environments, such as the Dead Sea or certain deep sea basins that possess a salinity equivalent to 5M $MgCl_2$); acidophiles (acid-loving prokaryotes that thrive at a pH as low as 0, such as that found in the drainage fluids of abandoned mine shafts); and thermophiles (prokaryotes that live at very high temperatures). Included in this last-named group are hyperthermophiles, which live in the hydrothermal vents of the ocean floor. The latest record holder among this group has been named "strain 121" because it is able to grow and divide in superheated water at a temperature of 121°C, which just happens to be the temperature used to sterilize surgical instruments in an autoclave.

All other prokaryotes are classified in the domain Bacteria. This domain includes the smallest known cells, the mycoplasma (0.2 μm diameter), which are the only known prokaryotes to lack a cell wall and to contain a genome with as few as 500 genes. Bacteria are present in every conceivable habitat on Earth, from the permanent ice shelf of the Antarctic to the driest African deserts, to the internal confines of plants and animals. Bacteria have even been found living in rock layers situated several kilometers beneath the Earth's surface. Some of these bacterial communities are thought to have been cut off from life on the surface for more than one hundred million years. The most complex prokaryotes are the cyanobacteria. Cyanobacteria contain elaborate arrays of cytoplasmic membranes, which serve as sites of photosynthesis (Figure 1.15*a*). The membranes of cyanobacteria are very similar to the photosynthetic membranes present within the chloroplasts of plant cells. As in eukaryotic plants, photosynthesis in cyanobacteria is accomplished by splitting water molecules, which releases molecular oxygen.

Many cyanobacteria are capable not only of photosynthesis, but also of **nitrogen fixation**, the conversion of nitrogen (N_2) gas into reduced forms of nitrogen (such as ammonia, NH_3) that can be used by cells in the synthesis of nitrogen-containing organic compounds, including amino acids and nucleotides. Those species capable of both photosynthesis and nitrogen fixation can survive on the barest of resources—light, N_2, CO_2, and H_2O. It is not surprising, therefore, that cyanobacteria are usually the first organisms to colonize the bare rocks rendered lifeless by a scorching volcanic eruption. Another unusual habitat occupied by cyanobacteria is illustrated in Figure 1.15*b*.

Prokaryotic Diversity For the most part, microbiologists are familiar only with those microorganisms they are able to grow in a culture medium. When a patient suffering from a respiratory or urinary tract infection sees his or her physician, one of the first steps often taken is to culture the pathogen. Once it has been cultured, the organism can be

(a)

(b)

FIGURE 1.15 Cyanobacteria. (*a*) Electron micrograph of a cyanobacterium showing the cytoplasmic membranes that carry out photosynthesis. These concentric membranes are very similar to the thylakoid membranes present within the chloroplasts of plant cells, a reminder that chloroplasts evolved from symbiotic cyanobacteria. (*b*) Cyanobacteria living inside the hairs of these polar bears are responsible for the unusual greenish color of their coats. (A: COURTESY OF NORMA J. LANG; B: COURTESY ZOOLOGICAL SOCIETY OF SAN DIEGO.)

identified and the proper treatment prescribed. It has proven relatively easy to culture *most* disease-causing prokaryotes, but the same is not true for those living free in nature. The problem is compounded by the fact that prokaryotes are barely visible in a light microscope and their morphology is often not very distinctive. To date, roughly 6000 species of prokaryotes have been identified by traditional techniques, which is less than one-tenth of 1 percent of the millions of prokaryotic species thought to exist on Earth! Our appreciation for the diversity of prokaryotic communities has increased dramatically in recent years with the use of molecular techniques that do not require the isolation of a particular organism.

Suppose one wanted to learn about the diversity of prokaryotes that live in the upper layers of the Pacific Ocean off the coast of California. Rather than trying to culture such organisms, which would prove largely futile, a researcher could concentrate the cells from a sample of ocean water, extract the DNA, and analyze certain DNA sequences present in the preparation. All organisms share certain genes, such as those that code for the RNAs present in ribosomes or the enzymes of certain metabolic pathways. Even though all organisms may share such genes, the sequences of the nucleotides that make up the genes vary considerably from one species to another. This is the basis of biological evolution. By using techniques that reveal the variety of DNA sequences of a particular gene in a particular habitat, one learns directly about the diversity of species that live in that habitat. Recent sequencing techniques have become so rapid and cost-efficient that virtually all of the genes present in the microbes of a given habitat can be sequenced, generating a collective genome, or *metagenome*. This approach can provide information about the types of proteins these organisms manufacture and thus about many of the metabolic activities in which they engage.

These same molecular strategies are being used to explore the remarkable diversity of the "unseen passengers" that live on or within our own bodies, in habitats such as the intestinal tract, mouth, vagina, and skin. This collection of microbes, which is known as the human *microbiome*, is the subject of several international research efforts aimed at identifying and characterizing these organisms in people of different age, diet, geography, and state of health. It has already been demonstrated, for example, that obese and lean humans have markedly different populations of bacteria in their digestive tracts. As obese individuals lose weight, their bacterial profile shifts toward that of the leaner individuals. Studies on mice suggest that some of the bacterial species that predominate in obese individuals may release more calories from digested food than their counterparts in the digestive tracts of the lean population and thus contribute to weight gain.

By using sequence-based molecular techniques, biologists have found that most habitats on Earth are teeming with previously unrecognized prokaryotic life. One estimate of the sheer numbers of prokaryotes in the major habitats of the Earth is given in Table 1.2. It is noteworthy that more than 90 percent of these organisms are now thought to live in the subsurface

TABLE 1.2 Number and Biomass of Prokaryotes in the World

Environment	No. of prokaryotic cells, $\times 10^{28}$	Pg of C in prokaryotes*
Aquatic habitats	12	2.2
Oceanic subsurface	355	303
Soil	26	26
Terrestrial subsurface	25–250	22–215
Total	415–640	353–546

*1 Petagram (Pg) = 10^{15} g.
Source: W. B. Whitman et al., *Proc. Nat'l. Acad. Sci. U.S.A.* 95:6581, 1998.

sediments well beneath the oceans and upper soil layers. It was only a decade or so ago that these deeper sediments were thought to be only sparsely populated by living organisms. Table 1.2 also provides an estimate of the amount of carbon that is sequestered in the world's prokaryotic cells. To put this number into more familiar terms, it is roughly comparable to the total amount of carbon present in all of the world's plant life.

Types of Eukaryotic Cells: Cell Specialization

In many regards, the most complex eukaryotic cells are not found inside of plants or animals, but rather among the single-celled (*unicellular*) protists, such as those pictured in Figure 1.16. All of the machinery required for the complex activities in which this organism engages—sensing the environment, trapping food, expelling excess fluid, evading predators—is housed within the confines of a single cell.
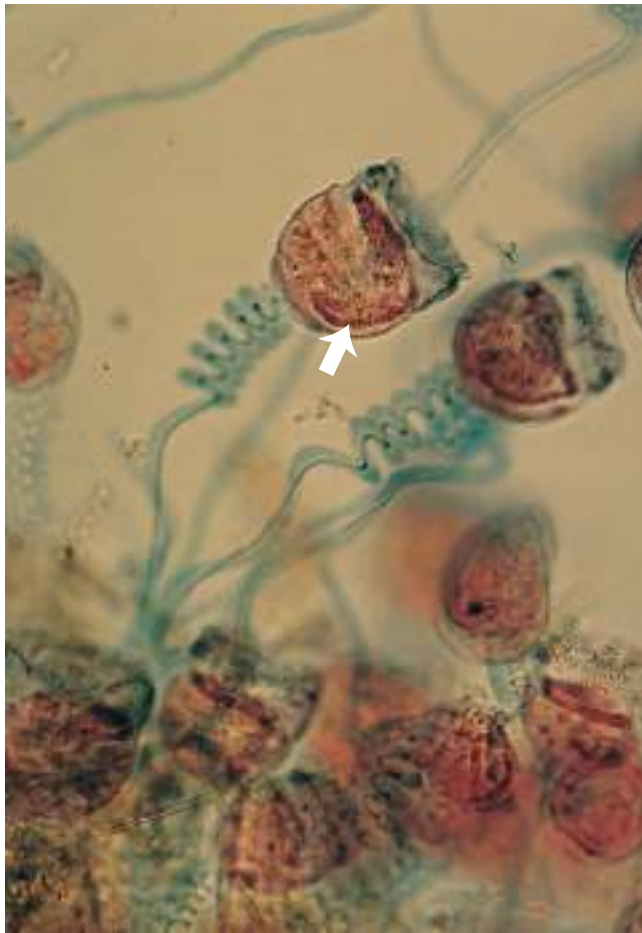
FIGURE 1.16 **Vorticella, a complex ciliated protist.** A number of these unicellular organisms are seen here; most have withdrawn their "heads" due to shortening of the blue-stained contractile ribbon in the stalk. Each cell has a single, large nucleus, called a macronucleus (arrow), which contains many copies of the genes. (CAROLINA BIOLOGICAL SUPPLY CO./PHOTOTAKE.)

Complex unicellular organisms represent one evolutionary pathway. An alternate pathway has led to the evolution of multicellular organisms in which different activities are conducted by different types of specialized cells. Specialized cells are formed by a process called **differentiation**. A fertilized human egg, for example, will progress through a course of embryonic development that leads to the formation of approximately 250 distinct types of differentiated cells. Some cells become part of a particular digestive gland, others part of a large skeletal muscle, others part of a bone, and so forth (Figure 1.17). The pathway of differentiation followed by each embryonic cell depends primarily on the signals it receives from the surrounding environment; these signals in turn depend on the position of that cell within the embryo. As discussed in the accompanying Human Perspective, researchers are learning how to control the process of differentiation in the culture dish and applying this knowledge to the treatment of complex human diseases.

As a result of differentiation, different types of cells acquire a distinctive appearance and contain unique materials. Skeletal muscle cells contain a network of precisely aligned filaments composed of unique contractile proteins; cartilage cells become surrounded by a characteristic matrix containing polysaccharides and the protein collagen, which together provide mechanical support; red blood cells become disk-shaped sacks filled with a single protein, hemoglobin, which transports oxygen; and so forth. Despite their many differences, the various cells of a multicellular plant or animal are composed of similar organelles. Mitochondria, for example, are found in essentially all types of cells. In one type, however, they may have a rounded shape, whereas in another they may be highly elongated and thread-like. In each case, the number, appearance, and location of the various organelles can be correlated with the activities of the particular cell type. An analogy might be made to a variety of orchestral pieces: all are composed of the same notes, but varying arrangement gives each its unique character and beauty.

Model Organisms   Living organisms are highly diverse, and the results obtained from a particular experimental analysis may depend on the particular organism being studied. As a result, cell and molecular biologists have focused considerable research activities on a small number of "representative" or **model organisms**. It is hoped that a comprehensive body of knowledge built on these studies will provide a framework to understand those basic processes that are shared by most organisms, especially humans. This is not to suggest that many other organisms are not widely used in the study of cell and molecular biology. Nevertheless, six model organisms—one prokaryote and five eukaryotes—have captured much of the attention: a bacterium, *E. coli*; a budding yeast, *Saccharomyces cerevisiae*; a flowering plant, *Arabidopsis thaliana*; a nematode, *Caenorhabditis elegans*; a fruit fly, *Drosophila melanogaster*; and a mouse, *Mus musculus*. Each of these organisms has specific advantages that make it particularly useful as a research subject for answering certain types of questions. Each of these organisms
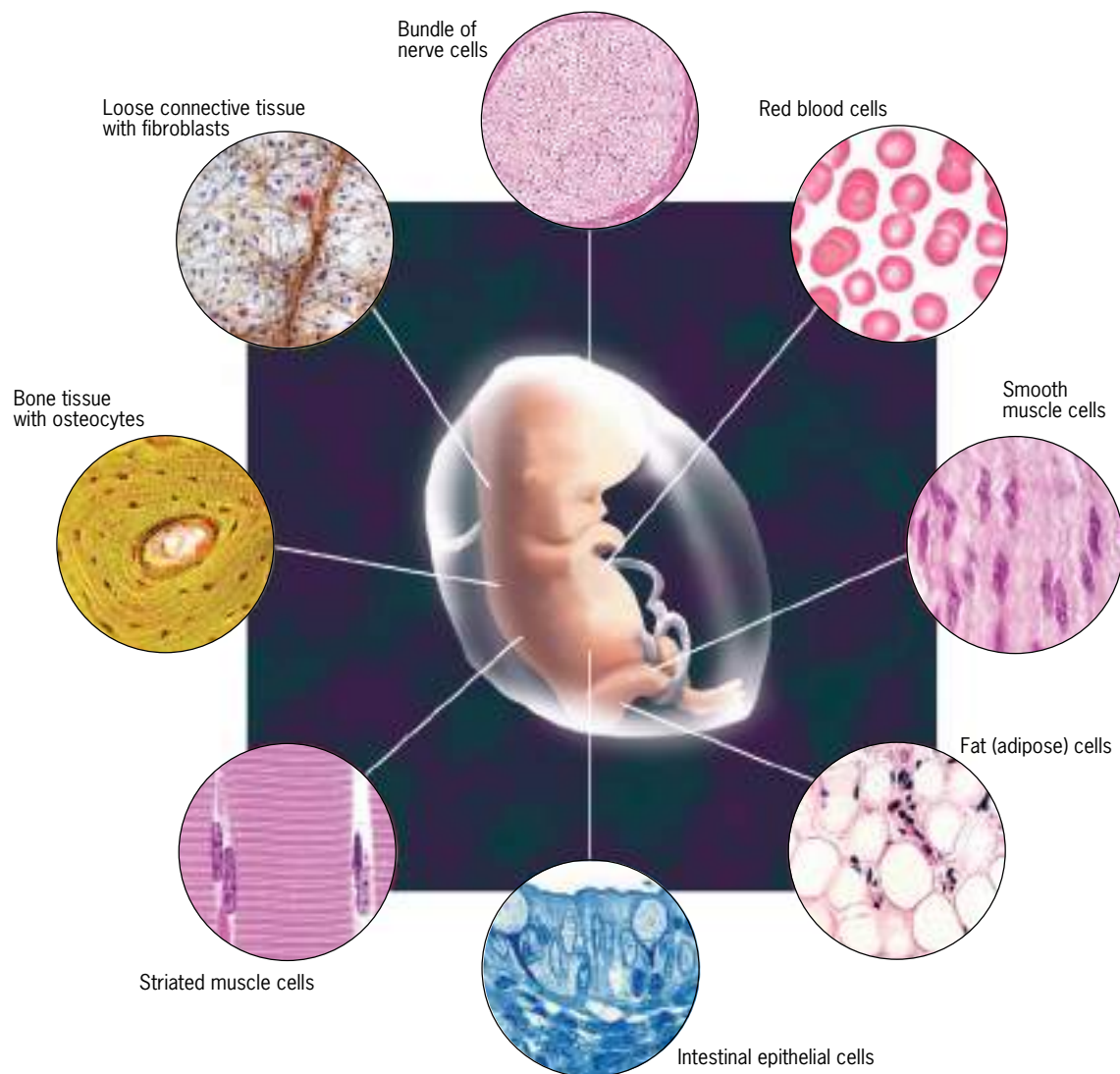
FIGURE 1.17   **Pathways of cell differentiation.** A few of the types of differentiated cells present in a human fetus. (MICROGRAPHS COURTESY OF MICHAEL ROSS, UNIVERSITY OF FLORIDA.)

is pictured in Figure 1.18, and a few of their advantages as research systems are described in the accompanying legend. We will concentrate in this text on results obtained from studies on mammalian systems—mostly on the mouse and on cultured mammalian cells—because these findings are most applicable to humans. Even so, we will have many occasions to describe research carried out on the cells of other species. You may be surprised to discover how similar you are at the cell and molecular level to these much smaller and simpler organisms.

The Sizes of Cells and Their Components

Figure 1.19 shows the relative size of a number of structures of interest in cell biology. Two units of linear measure are most commonly used to describe structures within a cell: the **micrometer** (μm) and the **nanometer** (nm). One μm is equal to 10^{-6} meters, and one nm is equal to 10^{-9} meters. The

angstrom (Å), which is equal to one-tenth of a nm, is commonly employed by molecular biologists for atomic dimensions. One angstrom is roughly equivalent to the diameter of a hydrogen atom. Large biological molecules (i.e., macromolecules) are described in either angstroms or nanometers. Myoglobin, a typical globular protein, is approximately 4.5 nm $\times$ 3.5 nm $\times$ 2.5 nm; highly elongated proteins (such as collagen or myosin) are over 100 nm in length; and DNA is approximately 2.0 nm in width. Complexes of macromolecules, such as ribosomes, microtubules, and microfilaments, are between 5 and 25 nm in diameter. Despite their tiny dimensions, these macromolecular complexes constitute remarkably sophisticated "nanomachines" capable of performing a diverse array of mechanical, chemical, and electrical activities.

   Cells and their organelles are more easily defined in micrometers. Nuclei, for example, are approximately 5–10 μm in diameter, and mitochondria are approximately 2 μm in length. Prokaryotic cells typically range in length from about 1 to 5 μm,

(a)

(b)

(c)

(d)

(e)

(f)
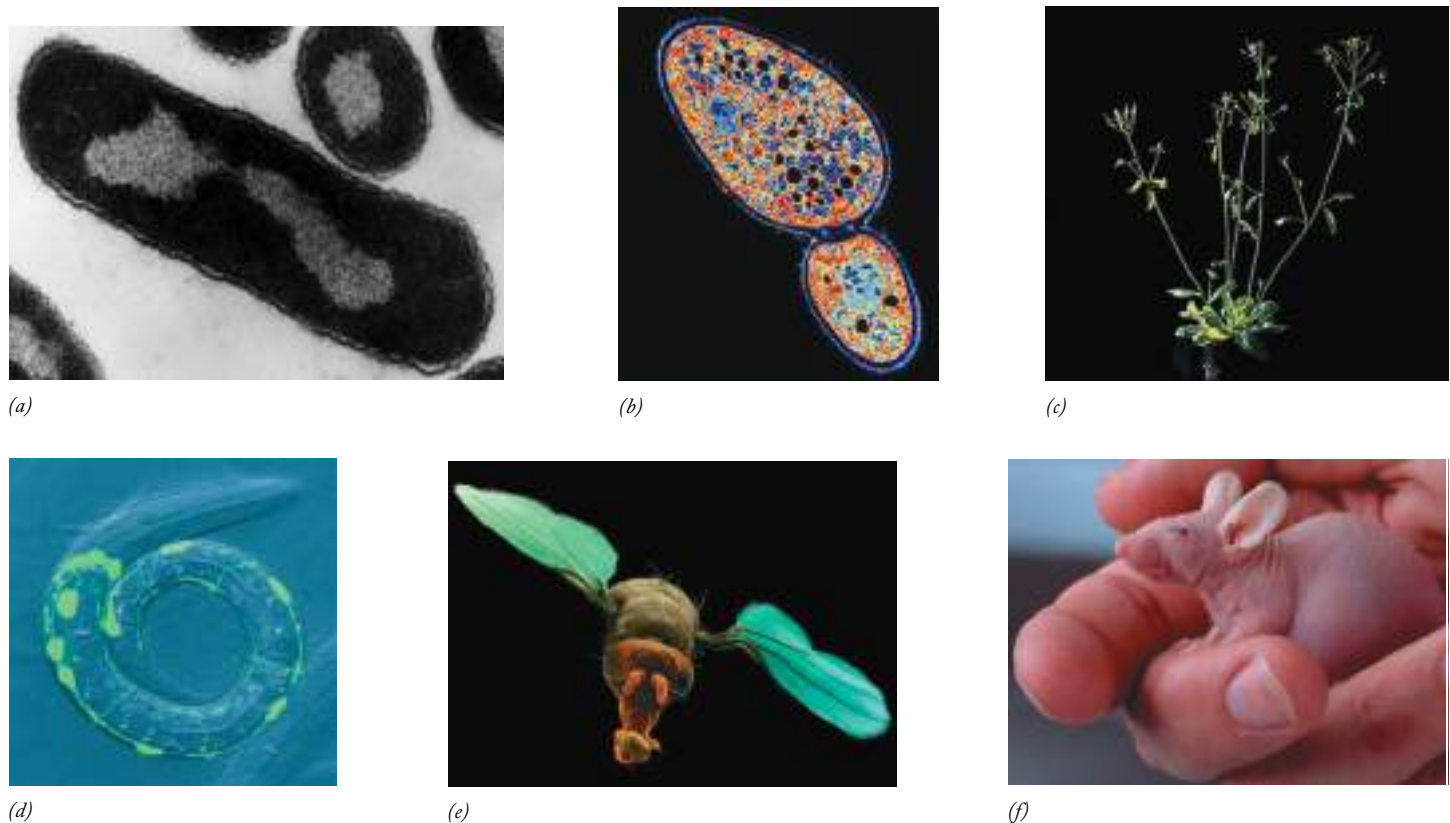
FIGURE 1.18 Six model organisms. (*a*) *Escherichia coli* is a rod-shaped bacterium that lives in the digestive tract of humans and other mammals. Much of what we will discuss about the basic molecular biology of the cell, including the mechanisms of replication, transcription, and translation, was originally worked out on this one prokaryotic organism. The relatively simple organization of a prokaryotic cell is illustrated in this electron micrograph (compare to part *b* of a eukaryotic cell). (*b*) *Saccharomyces cerevisiae*, more commonly known as baker's yeast or brewer's yeast. It is the least complex of the eukaryotes commonly studied, yet it contains a surprising number of proteins that are homologous to proteins in human cells. Such proteins typically have a conserved function in the two organisms. The species has a small genome encoding about 6200 proteins; it can be grown in a haploid state (one copy of each gene per cell rather than two as in most eukaryotic cells); and it can be grown under either aerobic (O_2-containing) or anaerobic (O_2-lacking) conditions. It is ideal for the identification of genes through the use of mutants. (*c*) *Arabidopsis thaliana*, a member of a genus of mustard plants, has an unusually small genome (120 million base pairs) for a flowering plant, a rapid generation time, and large seed production, and it grows to a height of only a few inches. (*d*) *Caenorhabditis elegans*, a microscopic-sized nematode, consists of a defined number of cells (roughly 1000),

each of which develops according to a precise pattern of cell divisions. The animal is easily cultured, has a transparent body wall, a short generation time, and facility for genetic analysis. This micrograph shows the larval nervous system, which has been labeled with the green fluorescent protein (GFP). The 2002 Nobel Prize was awarded to the researchers who pioneered its study. (*e*) *Drosophila melanogaster*, the fruit fly, is a small but complex eukaryote that has been a favored animal for genetic study for nearly 100 years. The organism is also well suited for the study of the molecular biology of development and the neurological basis of simple behavior. Certain larval cells have giant chromosomes, whose individual genes can be identified for studies of evolution and gene expression. (*f*) *Mus musculus*, the common house mouse, is easily kept and bred in the laboratory. Thousands of different genetic strains have been developed, many of which are stored simply as frozen embryos due to lack of space to house the adult animals. The "nude mouse" pictured here develops without a thymus gland and, therefore, is able to accept human tissue grafts that are not rejected. (A&B: BIOPHOTO ASSOCIATES/PHOTO RESEARCHERS; C: JEAN CLAUDE REVY/PHOTOTAKE; D: FROM KARLA KNOBEL, KIM SCHUSKE, AND ERIK JORGENSEN, TRENDS GENETICS, VOL. 14, COVER #12, 1998; E: DENNIS KUNKEL/VISUALS UNLIMITED; F: TED SPIEGEL/CORBIS IMAGES.)

eukaryotic cells from about 10 to 30 μm. There are a number of reasons most cells are so small. Consider the following.

■ Most eukaryotic cells possess a single nucleus that contains only two copies of most genes. Because genes serve as templates for the production of information-carrying messenger RNAs, a cell can only produce a limited number of these messenger RNAs in a given amount of time. The greater a cell's cytoplasmic volume, the longer it will take to synthesize the number of messages required by that cell.

■ As a cell increases in size, the surface area/volume ratio decreases.[3] The ability of a cell to exchange substances with its environment is proportional to its surface area. If a cell were to grow beyond a certain size, its surface would not be sufficient to take up the substances (e.g., oxygen,

[3]You can verify this statement by calculating the surface area and volume of a cube whose sides are 1 cm in length versus a cube whose sides are 10 cm in length. The surface area/volume ratio of the smaller cube is considerably greater than that of the larger cube.

FIGURE 1.19 Relative sizes of cells and cell components. These structures differ in size by more than seven orders of magnitude.

Synthetic Biology

A goal of one field of biological research, often referred to as *synthetic biology*, is to create a living cell in the laboratory, essentially from "scratch." One motivation of these researchers is simply to accomplish the feat and, in the process, demonstrate that life at the cellular level emerges spontaneously when the proper constituents are brought together from chemically synthesized materials. At this point in time, biologists are nowhere near accomplishing this feat, and many members of society would argue that it should never be allowed to take place. A more modest goal of synthetic biology is to develop novel life forms, beginning with existing organisms, that have a unique value in medicine and industry, or in cleaning up the environment. It might be possible, for example, to "custom-build" a species of bacteria that could convert cellulose, the bulk constituent of plant cell walls, into a biofuel such as the hydrocarbons present in gasoline. If, as most biologists would argue, the properties and activities of a cell spring from the genetic blueprint of that cell, then it should be possibe to create a new type of cell by introducing a new genetic blueprint into the cytoplasm of an existing cell. The feasibility of accomplishing this type of feat was demonstrated in 2007 when the genome of one bacterium was replaced with that of the genome of a closely related species, effectively transforming one species into the other.

In 2008 another important achievement in the field of synthetic biology was reported with the chemical synthesis of the complete genome of the bacterium *Mycoplasma genitalium*. The genome of this bacterium, the smallest of any organism that can be cultured in the laboratory, consists of a circular DNA molecule approximately 580,000 base pairs in length, containing roughly 500 genes. To accomplish this feat, the researchers began with the chemical synthesis of small segments of DNA approximately 100 bases in length, which is about the maximum allowed with current techniques. The base sequence of each of these small segments was precisely determined by the researchers to match that of the natural sequence, with a few intentional alterations. The small synthetic segments were then assembled into larger DNA fragments, which were eventually stitched together to create the complete bacterial genome. At the time of this writing, this synthetic genome has not been introduced into a living bacterial cell, but that is not a major barrier given the genome-replacement experiment described above. Once this is accomplished, the research team will have produced cells containing a "genetic skeleton" to which they can add new genes taken from other organisms. Ultimately, this line of research carries with it the prospect of creating new life forms possessing novel properties.

REVIEW

1. Compare a prokaryotic and eukaryotic cell on the basis of structural, functional, and metabolic differences.
2. What is the importance of cell differentiation?
3. Why are cells almost always microscopic?
4. If a mitochondrion were 2 μm in length, how many angstroms would it be? How many nanometers? How many millimeters?

nutrients) needed to support its metabolic activities. Cells that are specialized for absorption of solutes, such as those of the intestinal epithelium, typically possess microvilli, which greatly increase the surface area available for exchange (see Figure 1.3). The interior of a large plant cell is typically filled by a large, fluid-filled vacuole rather than metabolically active cytoplasm (see Figure 8.36*b*).

■ A cell depends to a large degree on the random movement of molecules (*diffusion*). Oxygen, for example, must diffuse from the cell's surface through the cytoplasm to the interior of its mitochondria. The time required for diffusion is proportional to the square of the distance to be traversed. For example, O_2 requires only 100 microseconds to diffuse a distance of 1 μm, but requires 10^6 times as long to diffuse a distance of 1 mm. As a cell becomes larger, and the distance from the surface to the interior becomes greater, the time required for diffusion to move substances in and out of a metabolically active cell becomes prohibitively long.

THE HUMAN PERSPECTIVE

The Prospect of Cell Replacement Therapy

To a person whose heart or liver is failing, an organ transplant is the best hope for survival and return to a normal life. Organ transplantation is one of the great successes of modern medicine, but its scope is greatly limited by the availability of donor organs and the high risk of immunologic rejection. Imagine the possibilities if we could grow cells and organs in the laboratory and use them to replace damaged or nonfunctional parts within our bodies. Recent studies have given researchers hope that one day this type of therapy will be commonplace. To better understand the concept of cell replacement therapy, we can consider a present-day procedure known as *bone marrow transplantation* in which cells are extracted from the interior of the pelvic bones of a donor and infused into the body of a recipient.

Bone marrow transplantation is used most often to treat lymphomas and leukemias, which are cancers that affect the nature and number of white blood cells. To carry out the procedure, the patient is exposed to a high level of radiation and/or toxic chemicals, which kills the cancerous cells, but also kills all of the cells involved in the formation of red and white blood cells. This treatment has this effect because blood-forming cells are particularly sensitive to radiation and toxic chemicals. Once a person's blood-forming cells have been destroyed, they are replaced by bone marrow cells transplanted from a healthy donor. Bone marrow can regenerate the blood tissue of the transplant recipient because it contains a small percentage of cells that can proliferate and restock the patient's blood-forming bone marrow tissue.[1] These blood-forming cells in the bone marrow are termed **hematopoietic stem cells** (or **HSCs**), and they are normally responsible for replacing the millions of red and white blood cells that age and die every minute in our bodies (see Figure 17.6). Amazingly, a *single* HSC is capable of reconstituting the entire hematopoietic (blood-forming) system of an irradiated mouse. An increasing number of parents are saving the blood from the umbilical cord of their newborn baby as a type of "stem-cell insurance policy" in case that child should ever develop a disease that might be treated by administration of HSCs.

Stem cells are defined as undifferentiated cells that (1) are capable of self-renewal, that is, production of more cells like themselves, and (2) are multipotent, that is, are capable of differentiating into two or more mature cell types. HSCs of the bone marrow are only one type of stem cell. Most, if not all, of the organs in a human adult contain stem cells that are capable of replacing the particular cells of the tissue in which they are found. Even the adult brain, which is not known for its ability to regenerate, contains stem cells that can generate new neurons and glial cells (the supportive cells of the brain). Figure 1*a* shows an isolated stem cell present in adult skeletal muscle; these "satellite cells," as they are called, are thought to divide and differentiate as needed for the repair of injured muscle tissue. Figure 1*b* shows a culture of adipose (fat) cells that have differentiated in vitro from adult stem cells that are present within fat tissue.

An interesting series of studies has recently been performed on a strain of golden retrievers that suffer from an inherited disease very similar to the human skeletal-muscle disorder muscular dystrophy. Researchers have isolated stem cells from the muscles of these dogs,

(a)   *(b)*
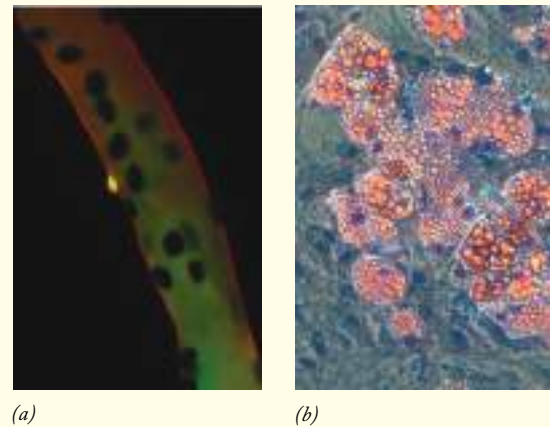
FIGURE 1 **An adult muscle stem cell.** (*a*) A portion of a muscle fiber, with its many nuclei stained blue. A single stem cell (yellow) is seen to be lodged between the outer surface of the muscle fiber and an extracellular layer (or basement membrane), which is stained red. The undifferentiated stem cell exhibits this yellow color because it expresses a protein that is not present in the differentiated muscle fiber. (*b*) Adult stem cells undergoing differentiation into adipose (fat) cells in culture. Stem cells capable of this process are present in adult fat tissue and also bone marrow. (A: FROM CHARLOTTE A. COLLINS; ET AL., CELL 122:291, 2005; BY PERMISSION OF CELL PRESS. B: COURTESY OF THERMO FISHER SCIENTIFIC, FROM NATURE 451:855, 2008.)

corrected the genetic disorder in the isolated cells, and expanded the number of genetically modified cells they have to work with by growing them in culture. When these stem cells are injected back into the diseased animals, many of them return to a skeletal muscle where they take up residence. Once back in muscle tissue, the corrected satellite cells divide and differentiate into new muscle cells and, in doing so, contribute to a marked improvement in the mobility and gait of the diseased animals. There is considerable optimism that similar types of therapeutic approaches could be used for humans. The human heart, for example, contains cardiac stem cells that are capable of differentiating into the cells that form both the muscle tissue of the heart (the cardiomyocytes of the myocardium) and the heart's blood vessels. These stem cells might have the potential to regenerate healthy heart tissue in a patient who has experienced a serious heart attack or is suffering from congestive heart failure. Adult stem cells are an ideal system for cell replacement therapies because they can be isolated directly from the patient, grown in culture, and reintroduced back into the same patient.

Although adult stem cells may ultimately prove to be an invaluable resource in cell replacement therapy, clinical studies carried out to date have been disappointing. Much of the excitement that has been generated in the field over the past decade has come from studies on **embryonic stem (ES) cells**, which are a type of stem cell isolated from very young mammalian embryos (see Figure 2). These are the cells in the early embryo that give rise to all of the various structures of the mammalian fetus. Unlike adult stem cells, ES cells are *pluripotent*; that is, they are capable of differentiating into every

---

[1]Bone marrow transplantation can be contrasted to a simple blood transfusion where the recipient receives differentiated blood cells (especially red blood cells and platelets) present in the circulation.
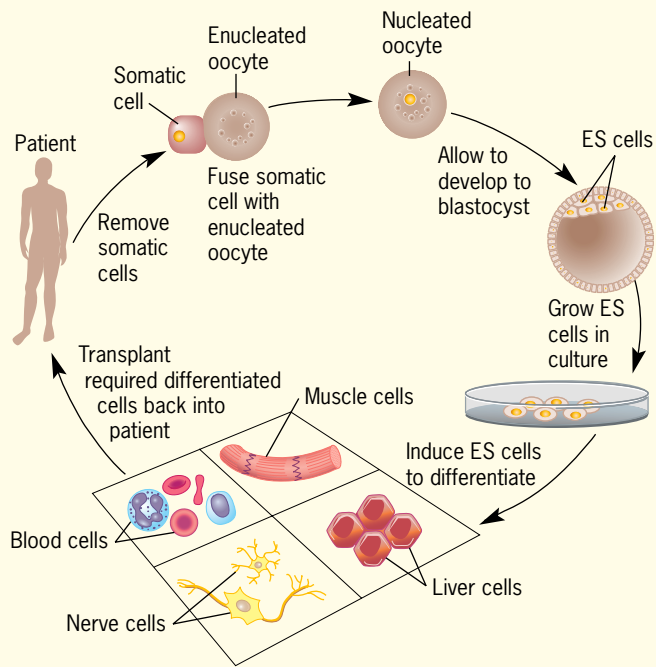
FIGURE 2   A procedure for obtaining differentiated cells for use in cell replacement therapy. A small piece of tissue is taken from the patient, and one of the somatic cells is fused with a donor oocyte whose own nucleus had been previously removed. The resulting oocyte (egg), with the patient's cell nucleus, is allowed to develop into an early embryo, and the ES cells are harvested and grown in culture. A population of ES cells are induced to differentiate into the required cells, which are subsequently transplanted into the patient to restore organ function.

type of cell in the body. In most cases, human ES cells have been isolated from embryos provided by in vitro fertilization clinics. Worldwide, dozens of genetically distinct human ES cell lines, each derived from a single embryo, are available for experimental investigation.

The long-range goal of clinical researchers is to learn how to coax ES cells to differentiate in culture into each of the many cell types that might be used for cell replacement therapy. Considerable progress has been made in this pursuit, and numerous studies have shown that transplants of differentiated, ES-derived cells can improve the condition of animals with diseased or damaged organs. The first trials in humans are likely to utilize cells, called oligodendrocytes, that produce the myelin sheaths that become wrapped around nerve cells (see Figure 4.5). It was found, by trial and error, that pure colonies of oligodendrocytes would differentiate from human ES cells that were cultured in a medium containing insulin, thyroid hormone, and a combination of certain growth factors. When implants of these human oligodendrocytes were transplanted into rats with paralyzing spinal cord injuries, the animals regained considerable motility. In 2009, the FDA approved the first clinical trials in which these ES-derived oligodendrocytes would be implanted into patients who have sustained recent damage to their spinal cord. Clinical trials are also planned for the treatment of type 1 diabetes and the eye disease macular degeneration. The primary risk with this type of procedure is the unnoticed presence of undifferentiated ES cells among the differentiated cell population. Undifferentiated ES cells are capable of forming a type of benign tumor, called a teratoma, which may contain

a bizarre mass of various differentiated tissues, including hair and teeth. The formation of a teratoma within the central nervous system could have severe consequences. In addition, the culture of ES cells at the present time involves the use of nonhuman biological materials, which also poses potential risks of causing disease.

Although adult stem cells lack the unlimited differentiation capacity that is characteristic of ES cells, they do have an advantage over ES cells in that they can be isolated from the individual who is being treated and, thus, will not face immunologic rejection when used in subsequent cell replacement. However, it may be possible to "customize" ES cells so that they possess the same genetic makeup of the individual who is being treated, and thus would not be subject to attack by the recipient's immune system. This can likely be accomplished by a roundabout procedure called *somatic cell nuclear transfer* (*SCNT*), shown in Figure 2, that begins with an unfertilized egg—a cell that is obtained from the ovaries of an unrelated woman donor. In this approach, the nucleus of the unfertilized egg would be replaced by the nucleus of a cell from the patient to be treated, which would cause the egg to have the same chromosome composition as that of the patient. The egg would then be allowed to develop to an early embryonic stage, and the ES cells would be removed, cultured, and induced to differentiate into the type of cells needed by the patient. Because this procedure involves the formation of a human embryo that is used only as a source of ES cells, there are major ethical questions that must be settled before it could be routinely practiced. In addition, the process of SCNT is so expensive and technically demanding that it is highly improbable that it could ever be practiced as part of any routine medical treatment. It is more likely that, if ES cell-based therapy is ever practiced, it would depend on the use of a bank of hundreds or thousands of different ES cells. Such a bank could contain cells that are close enough as a tissue match to be suitable for use in the majority of patients.

It had long been thought that the process of cell differentiation in mammals was irreversible; once a cell had become a fibroblast, or white blood cell, or cartilage cell, it could never again revert to any other cell type. This concept was shattered in 2006 when Shinya Yamanaka and coworkers of Kyoto University announced a stunning discovery; his lab had succeeded in reprogramming a fully differentiated mouse cell—in this case a type of connective tissue fibroblast—into a pluripotent stem cell. They accomplished the feat by introducing, into the mouse fibroblast, the genes that encoded four key proteins that are characteristic of ES cells. These genes (*Oct4, Sox2, Klf4,* and *Myc*) are thought to play a key role in maintaining the cells in an undifferentiated state and allowing them to continue to self-renew. The genes were introduced into cultured fibroblasts using gene-carrying viruses, and those rare cells that became reprogrammed were selected from the others in the culture by specialized techniques. They called this new type of cells *induced pluripotent cells* (*iPS cells*) and demonstrated that they were indeed pluripotent by injecting them into a mouse blastocyst and finding that they participated in the differentiation of all the cells of the body, including eggs and sperm. Within the next year or so the same reprogramming feat had been accomplished in several labs with human cells, and it was demonstrated that the human iPS cells are virtually indistinguishable from authentic human ES cells by numerous criteria. What this means is that researchers now have available to them an unlimited supply of pluripotent cells that can be directed to differentiate into various types of body cells using similar experimental protocols to those already developed for ES cells. Indeed, iPS cells have already been used to correct certain disease conditions in experimental animals, including sickle cell anemia in mice as depicted in Figure 3.

iPS cells have also been prepared from adult cells taken from patients with genetic disorders, such as Huntington's disease and type 1 diabetes. Researchers will be able to follow the differentiation of these iPS cells into their diseased cell types in culture. It is hoped that such studies will reveal the mechanisms of disease formation as it unfolds in a culture dish just as it would normally occur in an unobservable way deep within the body. Such "diseased cells" might also serve as targets for testing the effects of newly developed drugs in halting disease progression.

Unlike ES cells, the generation of iPS cells does not require the use of an embryo. This feature removes all of the ethical reservations that accompany work with ES cells and also makes it much easier to generate these cells in the lab. In fact, almost any laboratory that works with cultured mammalian cells should be able to jump into this exciting field—and many have. Just as with ES cells, there are difficulties to overcome before iPS cells can be used as a

source of cells for human therapy. It will be important to develop efficient cell reprogramming techniques that do not use gene-carrying viruses because such cells carry the potential of developing into cancers. Recent studies suggest that this goal can be achieved. Like ES cells, undifferentiated iPS cells also give rise to teratomas, so it is essential that only fully differentiated cells are transplanted into human subjects. Also like ES cells, the iPS cells in current use have the same tissue antigens as the donors who originally provided them, so they would stimulate an immune attack if they were to be transplanted into other human recipients. Unlike the formation of ES cells, however, it will be much easier to generate personalized, tissue-compatible iPS cells, because they can be derived from a simple skin biopsy from each patient. Still, it does take considerable time, expense, and technical expertise to generate a population of iPS cells from a specific donor. Consequently, if iPS cells are ever developed for therapeutic use, they would likely come from a large cell bank that could provide cells that are close tissue matches to most potential recipients. It may also be possible to remove all of the genes from iPS cells that normally prevent them from being transplanted into random recipients. If this could be achieved, it might be possible to develop a single "universal donor" iPS cell line that is invisible to the recipient's immune system. Cells of this type could be used as the basis for cell replacement in everyone.

In 2008 the field of cellular reprogramming took another unexpected turn with the announcement that one type of differentiated cell had been converted directly into another type of differentiated cell, a case of "transdifferentiation." In this report, the acinar cells of the pancreas, which produce enzymes responsible for digestion of food in the intestine, were transformed into pancreatic beta cells, which synthesize and secrete the hormone insulin. The reprogramming process occurred directly, in a matter of a few days, without the cells passing through an intermediate stem cell state—and it occurred while the cells remained in their normal residence within the pancreas of a live mouse. This feat was accomplished by injection into the animals of viruses that carried three genes known to be important in differentiation of beta cells in the embryo. In this case, the recipients of the injection were diabetic mice, and the transdifferentiation of a significant number of acinar cells into beta cells allowed the animals to regulate their blood sugar levels with much lower doses of insulin. It is also noteworthy that the adenoviruses used to deliver the genes in this experiment do not become a permanent part of the recipient cell, which removes some of the concerns about the use of viruses as gene carriers in humans. It is too early to speculate on whether this type of direct reprogramming strategy has real therapeutic potential, but it certainly raises the prospect that diseased cells that need to be replaced might be formed directly from other types of cells within the same organ.

FIGURE 3 **Steps taken to generate induced pluripotent stem (iPS) cells for use in correcting the inherited disease sickle cell anemia in mice.** Skin cells are collected from the diseased animal, reprogrammed in culture by introducing the four required genes that are ferried into the cells by viruses, and allowed to develop into undifferentiated pluripotent iPS cells. The iPS cells are then treated so as to replace the defective (globin) gene with a normal copy, and the corrected iPS cells are caused to differentiate into normal blood stem cells in culture. These blood stem cells are then injected back into the diseased mouse, where they proliferate and differentiate into normal blood cells, thereby curing the disorder. (BASED ON AN ILLUSTRATION BY RUDOLF JAENISCH, CELL 132:5, 2008.)

1.4 VIRUSES

By the end of the nineteenth century, the work of Louis Pasteur and others had convinced the scientific world that infectious diseases of plants and animals were due to bacteria. But studies of tobacco mosaic disease in tobacco plants and hoof-and-mouth disease in cattle pointed to the existence of another type of infectious agent. It was found, for example, that sap from a diseased tobacco plant could transmit mosaic disease to a healthy plant, even when the sap

showed no evidence of bacteria in the light microscope. To gain further insight into the size and nature of the infectious agent, Dmitri Ivanovsky, a Russian biologist, forced the sap from a diseased plant through filters whose pores were so small that they retarded the passage of the smallest known bacterium. The filtrate was still infective, causing Ivanovsky to conclude that certain diseases were caused by pathogens that were even smaller, and presumably simpler, than the smallest known bacteria. These pathogens became known as **viruses**.

In 1935, Wendell Stanley of the Rockefeller Institute reported that the virus responsible for tobacco mosaic disease could be crystallized and that the crystals were infective. Substances that form crystals have a highly ordered, well-defined structure and are vastly less complex than the simplest cells. Stanley mistakenly concluded that tobacco mosaic virus (TMV) was a protein. In fact, TMV is a rod-shaped particle consisting of a single molecule of RNA surrounded by a helical shell composed of protein subunits (Figure 1.20).

Viruses are responsible for dozens of human diseases, including AIDS, polio, influenza, cold sores, measles, and a few types of cancer (see Section 16.2). Viruses occur in a wide variety of very different shapes, sizes, and constructions, but all of them share certain common properties. All viruses are obligatory intracellular parasites; that is, they cannot reproduce unless present within a host cell. Depending on the specific virus, the host may be a plant, animal, or bacterial cell. Outside of a living cell, the virus exists as a particle, or **virion**, which is little more than a macromolecular package. The virion contains a small amount of genetic material that, depending on the virus, can be single-stranded or double-stranded, RNA or DNA. Remarkably, some viruses have as few as three or four different genes, but others may have as many as several hundred. The genetic material of the virion is surrounded by a protein capsule, or *capsid*. Virions are macromolecular aggregates, inanimate particles that by themselves are unable to reproduce, metabolize, or carry on any of the other activities associated with life. For this reason, viruses are not considered to be organisms and are not described as being alive.

Viral capsids are generally made up of a specific number of subunits. There are numerous advantages to construction by subunit, one of the most apparent being an economy of genetic information. If a viral coat is made of many copies of a single protein, as is that of TMV, or a few proteins, as are the coats of many other viruses, the virus needs only one or a few genes to code for its protein container. Many viruses have a capsid whose subunits are organized into a polyhedron, that is, a structure having planar faces. A particularly common polyhedral shape of viruses is the 20-sided *icosahedron*. For example, adenovirus, which causes respiratory infections in mammals, has an icosahedral capsid (Figure 1.21a). In many animal viruses, including the *human immunodeficiency virus* (*HIV*) responsible for AIDS, the protein capsid is surrounded by a lipid-containing outer envelope that is derived from the modified plasma membrane of the host cell as the virus buds from the host-cell surface (Figure 1.21b). Bacterial viruses, or *bacteriophages*, are among the most complex viruses (Figure 1.21c). They are also the most abundant biological entities on Earth. The T bacteriophages (which were used in key experiments that revealed the structure and properties of the genetic material) consist of a polyhedral head containing DNA, a cylindrical stalk through which the DNA is injected into the bacterial cell, and tail fibers, which together cause the particle to resemble a landing module for the moon.

Each virus has on its surface a protein that is able to bind to a particular surface component of its host cell. For example, the protein that projects from the surface of the HIV particle

Protein coat (capsid)   Nucleic acid

(a)

(b)   60 nm
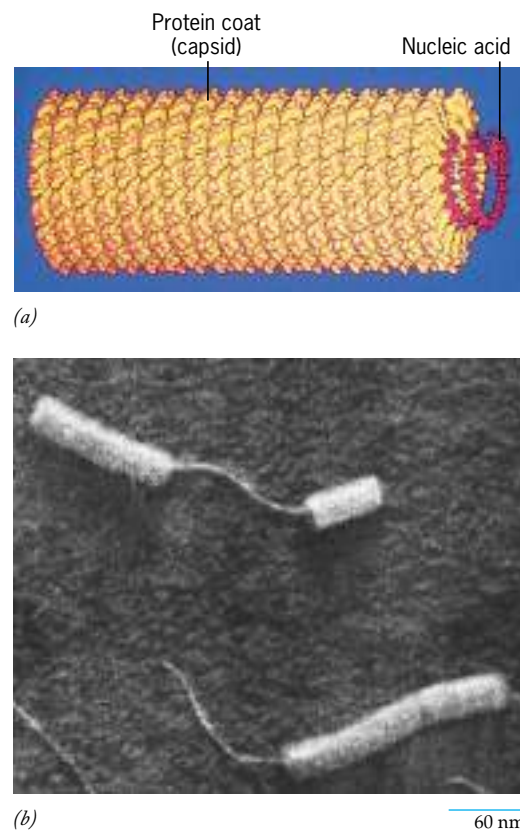
FIGURE 1.20   Tobacco mosaic virus (TMV). (*a*) Model of a portion of a TMV particle. The protein subunits, which are identical along the entire rod-shaped particle, enclose a single helical RNA molecule (red). (*b*) Electron micrograph of TMV particles taken after treatment with phenol has removed the protein subunits from the middle part of the upper particle and the ends of the lower particle. Intact rods are approximately 300 nm long and 18 nm in diameter. (A: COURTESY OF GERALD STUBBS, KEIICHI NAMBA, AND DONALD CASPAR; B: COURTESY OF M. K. CORBETT.)

(labeled gp120 in Figure 1.21*b*, which stands for glycoprotein of molecular mass 120,000 daltons[4]) interacts with a specific protein (called CD4) on the surface of certain white blood cells, facilitating entry of the virus into its host cell. The interaction between viral and host proteins determines the specificity of the virus, that is, the types of host cells that the virus can enter and infect. Some viruses have a wide *host range*, being able to infect cells from a variety of different organs or host species. The virus that causes rabies, for example, is able to infect many different types of mammalian hosts, including dogs, bats, and humans. Most viruses, however, have a relatively narrow host range. This is true, for example, of human cold and influenza viruses, which are generally able to infect only the respiratory epithelial cells of human hosts.

A change in host-cell specificity can have striking consequences. This point is dramatically illustrated by the 1918 influenza pandemic, which killed more than 30 million people

---

[4]One dalton is equivalent to one unit of atomic mass, the mass of a single hydrogen (^{1}H) atom.
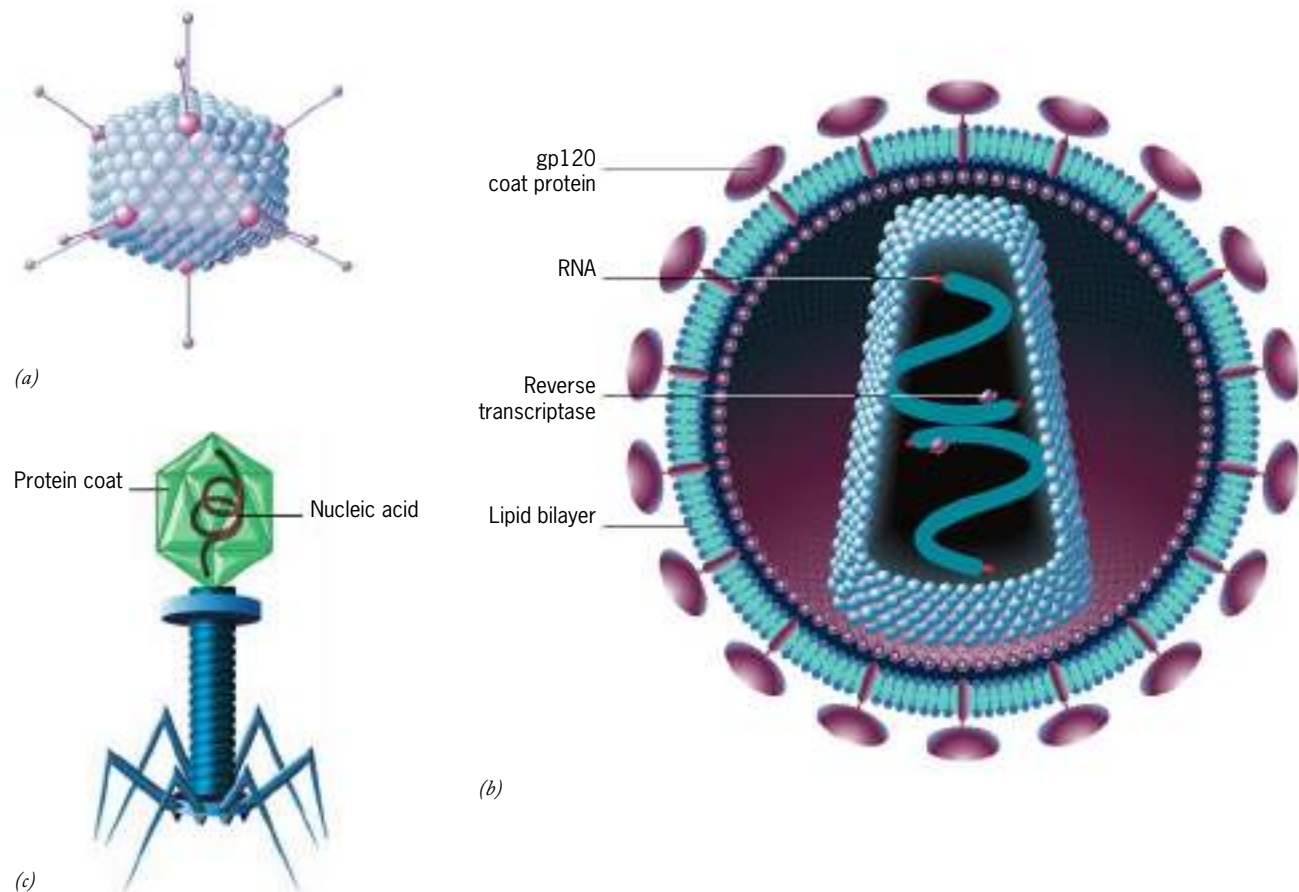
FIGURE 1.21 Virus diversity. The structures of (*a*) an adenovirus, (*b*) a human immunodeficiency virus (HIV), and (*c*) a T-even bacteriophage. *Note:* These viruses are not drawn to the same scale.

worldwide. The virus was especially lethal in young adults, who do not normally fall victim to influenza. In fact, the 675,000 deaths from this virus in the United States temporarily lowered average life expectancy by several years. In one of the most acclaimed—and controversial—feats of the past few years, researchers have been able to determine the genomic sequence of the virus responsible for this pandemic and to reconstitute the virus in its full virulent state. This was accomplished by isolating the viral genes (which are part of a genome consisting of 8 separate RNA molecules encoding 11 different proteins) from the preserved tissues of victims who had died from the infection 90 years earlier. The best preserved samples were obtained from a Native American woman who had been buried in the Alaskan permafrost. The sequence of the "1918 virus" suggested that the pathogen had jumped from birds to humans. Although the virus had accumulated a considerable number of mutations, which adapted it to a mammalian host, it had never exchanged genetic material with that of a human influenza virus as had been thought likely.

Analysis of the sequence of the 1918 virus has provided some clues to explain why it was so deadly and how it spread so efficiently from one human to another. Using the genomic sequence, researchers reconstituted the 1918 virus into infectious particles, which were found to be exceptionally virulent

in laboratory tests. Whereas laboratory mice normally survive infection by modern human influenza viruses, the reconstituted 1918 strain killed 100 percent of infected mice and produced enormous numbers of viral particles in the animals' lungs. Because of the potential risk to public health, publication of the full sequence of the 1918 virus and its reconstitution went forward only after approval by governmental safety panels and the demonstration that existing influenza vaccines and drugs protect mice from the reconstituted virus.

There are two basic types of viral infection. (1) In most cases, the virus arrests the normal synthetic activities of the host and redirects the cell to use its available materials to manufacture viral nucleic acids and proteins, which assemble into new virions. Viruses, in other words, do not grow like cells; they are assembled from components directly into the mature-sized virions. Ultimately, the infected cell ruptures (*lyses*) and releases a new generation of viral particles capable of infecting neighboring cells. An example of this type of *lytic* infection is shown in Figure 1.22*a*. (2) In other cases, the infecting virus does not lead to the death of the host cell, but instead inserts (*integrates*) its DNA into the DNA of the host cell's chromosomes. The integrated viral DNA is called a **provirus**. An integrated provirus can have different effects depending on the type of virus and host cell. For example,

- Bacterial cells containing a provirus behave normally until exposed to a stimulus, such as ultraviolet radiation, that activates the dormant viral DNA, leading to the lysis of the cell and release of viral progeny.

- Some animal cells containing a provirus produce new viral progeny that bud at the cell surface without lysing the infected cell. Human immunodeficiency virus (HIV) acts in this way; an infected cell may remain alive for a period, acting as a factory for the production of new virions (Figure 1.22*b*).

Phage being assembled

Empty phage

(a)   0.2 μm

Virus budding from cell
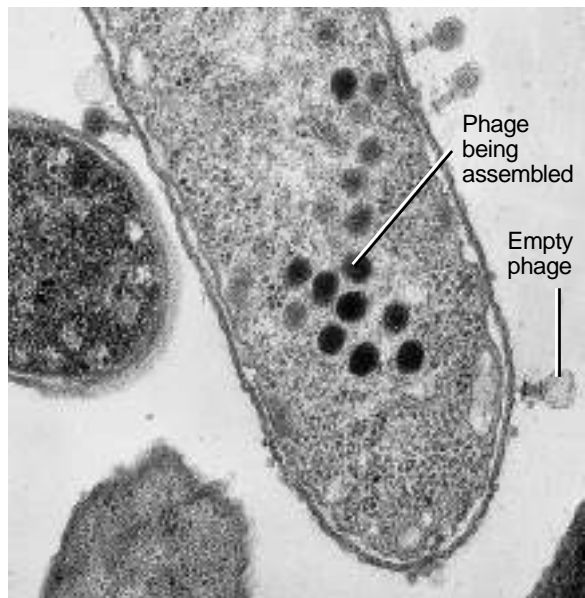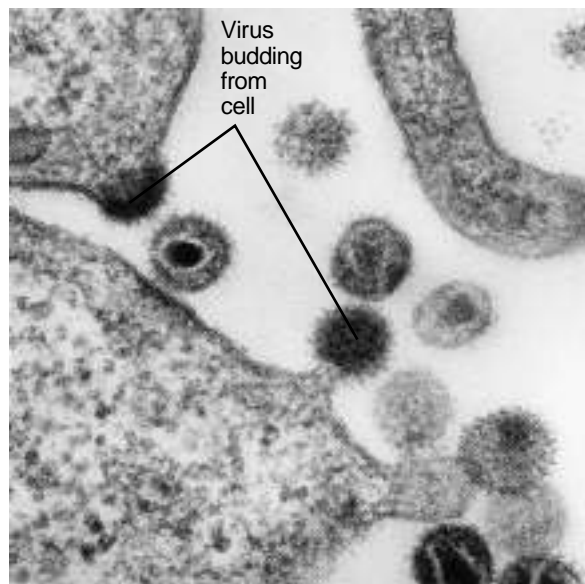
(b)   0.1 μm

FIGURE 1.22 **A virus infection.** (*a*) Micrograph showing a late stage in the infection of a bacterial cell by a bacteriophage. Virus particles are being assembled within the cell, and empty phage coats are still present on the cell surface. (*b*) Micrograph showing HIV particles budding from an infected human lymphocyte. (A: COURTESY OF JONATHAN KING, AND ERIKA HARTWIG; B: COURTESY OF HANS GELDERBLOM.)

- Some animal cells containing a provirus lose control over their own growth and division and become malignant. This phenomenon is readily studied in the laboratory by infecting cultured cells with the appropriate tumor virus.

Viruses are not without their virtues. Because the activities of viral genes mimic those of host genes, investigators have used viruses for decades as a research tool to study the mechanism of DNA replication and gene expression in their much more complex hosts. In addition, viruses are now being used as a means to introduce foreign genes into human cells, a technique that will likely serve as the basis for the treatment of human diseases by gene therapy. Lastly, insect- and bacteria-killing viruses may play an increasing role in the war against insect pests and bacterial pathogens. Bacteriophages have been used for decades to treat bacterial infections in Eastern Europe and Russia, while physicians in the West have relied on antibiotics. Given the rise in antibiotic-resistant bacteria, bacteriophages may be making a comeback on the heels of promising studies on infected mice. Several biotechnology companies are now producing bacteriophages intended to combat bacterial infections and to protect certain foods from bacterial contamination.

Viroids

It came as a surprise in 1971 to discover that viruses are not the simplest types of infectious agents. In that year, T. O. Diener of the U.S. Department of Agriculture reported that potato spindle-tuber disease, which causes potatoes to become gnarled and cracked, is caused by an infectious agent consisting of a small circular RNA molecule that totally lacks a protein coat. Diener named the pathogen a **viroid**. The RNAs of viroids range in size from about 240 to 600 nucleotides, one-tenth the size of the smaller viruses. No evidence has been found that the naked viroid RNA codes for any proteins. Rather, any biochemical activities in which viroids engage take place using host-cell proteins. For example, duplication of the viroid RNA within an infected cell utilizes the host's RNA polymerase II, an enzyme that normally transcribes the host's DNA into messenger RNAs. Viroids are thought to cause disease by interfering with the cell's normal path of gene expression. The effect on crops can be serious: a viroid disease called cadang-cadang has devastated the coconut palm groves of the Philippines, and another viroid has wreaked havoc on the chrysanthemum industry in the United States. The discovery of a different type of infectious agent even simpler than a viroid is described in the Human Perspective in Chapter 2.

REVIEW ?

1. What properties distinguish a virus from a bacterium?
2. What types of infections are viruses able to cause? How is it possible to study viruses under the light microscope?
3. Compare and contrast: nucleoid and nucleus; the flagellum of a bacterium and a sperm; an archaebacterium and a cyanobacterium; nitrogen fixation and photosynthesis; bacteriophages and tobacco mosaic virus; a provirus and a virion.

EXPERIMENTAL PATHWAYS

The Origin of Eukaryotic Cells

We have seen in this chapter that cells can be divided into two groups: prokaryotic cells and eukaryotic cells. Almost from the time this division of cellular life was proposed, biologists have been fascinated by the question: What is the origin of the eukaryotic cell? It is generally (but not universally) agreed that prokaryotic cells (1) arose before eukaryotic cells and (2) gave rise to eukaryotic cells. The first point can be verified directly from the fossil record, which shows that prokaryotic cells were present in rocks approximately 2.7 billion years old (page 7), which is roughly one billion years before any evidence is seen of eukaryotes. The second point follows from the fact that the two types of cells have to be related to one another because they share many complex traits (e.g., very similar genetic codes, enzymes, metabolic pathways, and plasma membranes) that could not have evolved independently in different organisms.

Up until about 1970, it was generally believed that eukaryotic cells evolved from prokaryotic cells by a process of gradual evolution in which the organelles of the eukaryotic cell became progressively more complex. Acceptance of this concept changed dramatically about that time largely through the work of Lynn Margulis, then at Boston University. Margulis resurrected an idea that had been proposed earlier, and dismissed, that certain organelles of a eukaryotic cell—most notably the mitochondria and chloroplasts—had evolved from smaller prokaryotic cells that had taken up residence in the cytoplasm of a larger host cell.[1,2] This hypothesis is referred to as the **endosymbiont theory** because it describes how a single "composite" cell of greater complexity could evolve from two or more separate, simpler cells living in a symbiotic relationship with one another.

Our earliest prokaryotic ancestors were presumed to have been anaerobic heterotrophic cells: *anaerobic* meaning they derived their energy from food matter without employing molecular oxygen (O_2), and *heterotrophic* meaning they were unable to synthesize organic compounds from inorganic precursors (such as CO_2 and water), but instead had to obtain preformed organic compounds from their environment.

According to one version of the endosymbiont theory, a large, anaerobic, heterotrophic prokaryote ingested a small, aerobic prokaryote (step 1, Figure 1). Resisting digestion within the cytoplasm, the small aerobic prokaryote took up residence as a permanent

FIGURE 1 **A model depicting possible steps in the evolution of eukaryotic cells, including the origin of mitochondria and chloroplasts by endosymbiosis.** In step 1, a large anaerobic, heterotrophic prokaryote takes in a small aerobic prokaryote. Evidence strongly indicates that the engulfed prokaryote was an ancestor of modern-day rickettsia, a group of bacteria that causes typhus and other diseases. In step 2, the aerobic endosymbiont has evolved into a mitochondrion. In step 3, a portion of the plasma membrane has invaginated and is seen in the process of evolving into a nuclear envelope and associated endoplasmic reticulum. The primitive eukaryote depicted in step 3 gives rise to two major groups of eukaryotes. In one path (step 4), the primitive eukaryote evolves into nonphotosynthetic protist, fungal, and animal cells. In the other path (step 5), the primitive eukaryote takes in a photosynthetic prokaryote, which will become an endosymbiont and evolve into a chloroplast. (*Note:* The engulfment of the symbiont shown in step 1 could have occurred after development of some of the internal membranes, but evidence suggests it was a relatively early step in the evolution of eukaryotes.)

endosymbiont. As the host cell reproduced, so did the endosymbiont, so that a colony of these composite cells was soon produced. Over many generations, endosymbionts lost many of the traits that were no longer required for survival, and the once-independent oxygen-respiring microbes evolved into precursors of modern-day mitochondria (step 2, Figure 1).

A cell whose ancestors had formed through the sequence of symbiotic events just described could have given rise to a line of cells that evolved other basic characteristics of eukaryotic cells, including a system of membranes (a nuclear membrane, endoplasmic reticulum, Golgi complex, lysosomes), a complex cytoskeleton, and a mitotic type of cell division. These characteristics are proposed to have arisen by a gradual process of evolution, rather than in a single step as might occur through acquisition of an endosymbiont. The endoplasmic reticulum and nuclear membranes, for example, might have been derived from a portion of the cell's outer plasma membrane that became internalized and then modified into a different type of membrane (step 3, Figure 1). A cell that possessed these various internal compartments would have been an ancestor of a heterotrophic eukaryotic cell, such as a fungal cell or a protist (step 4, Figure 1). The oldest fossils thought to be the remains of eukaryotes date back about 1.8 billion years.

It is proposed that the acquisition of another endosymbiont, specifically a cyanobacterium, could have converted an early heterotrophic eukaryote into an ancestor of photosynthetic eukaryotes: the green algae and plants (step 5, Figure 1). The acquisition of chloroplasts (roughly one billion years ago) must have been one of the last steps in the sequence of endosymbioses because these organelles are only present in plants and algae. In contrast, all known groups of eukaryotes either (1) possess mitochondria or (2) show definitive evidence they have evolved from organisms that possessed these organelles.[a]

The division of all living organisms into two categories, prokaryotes and eukaryotes, reflects a basic dichotomy in the structures of cells, but it is not necessarily an accurate phylogenetic distinction, that is, one that reflects the evolutionary relationships among living organisms. How does one determine evolutionary relationships among organisms that have been separated in time for billions of years, such as prokaryotes and eukaryotes? Most taxonomic schemes that attempt to classify organisms are based heavily on anatomic or physiologic characteristics. In 1965, Emile Zuckerkandl and Linus Pauling suggested an alternate approach based on comparisons of the structure of informational molecules (proteins and nucleic acids) of living organisms.[3] Differences between organisms in the sequence of amino acids that make up a protein or the sequence of nucleotides that make up a nucleic acid are the result of mutations in DNA that have been transmitted to offspring. Mutations can accumulate in a given gene at a relatively constant rate over long periods of time. Consequently, comparisons of amino acid or nucleotide sequences can be used to determine how closely organisms are related to one another. For example, two organisms that are closely related, that is, have diverged only recently from a common ancestor, should have fewer sequence differences in a particular gene

than two organisms that are distantly related, that is, do not have a recent common ancestor. Using this type of sequence information as an "evolutionary clock," researchers can construct phylogenetic trees showing proposed pathways by which different groups of living organisms may have diverged from one another during the course of evolution.

Beginning in the mid-1970s, Carl Woese and his colleagues at the University of Illinois began a series of studies that compared the nucleotide sequence in different organisms of the RNA molecule that resides in the small subunit of the ribosome. This RNA— which is called the 16S rRNA in prokaryotes or the 18S rRNA in eukaryotes—was chosen because it is present in large quantities in all cells, it is easy to purify, and it tends to change only slowly over long periods of evolutionary time, which means that it could be used to study relationships of very distantly related organisms. There was one major disadvantage: nucleic acid sequencing at the time required laborious, time-consuming methods. In their approach, they purified the 16S rRNA from a particular source, then subjected the preparation to an enzyme, T1 ribonuclease, that digests the molecule into short fragments called oligonucleotides. The oligonucleotides in the mixture were then separated from one another by two-dimensional electrophoresis to produce a two-dimensional "fingerprint" as shown in Figure 2. Once they were separated, the nucleotide sequence of each of the oligonucleotides could be determined and the sequences from various organisms compared. In one of their first studies, Woese and his colleagues analyzed the 16S rRNA present in the ribosomes of chloroplasts from the photosynthetic protist *Euglena*.[4] They found that the sequence of this chloroplast rRNA molecule was much more similar to that of the 16S rRNA found in ribosomes of cyanobacteria than it was to its counterpart in the ribosomes from eukaryotic cytoplasm. This finding provided strong evidence for the symbiotic origin of chloroplasts from cyanobacteria.

In 1977, Woese and George Fox published a landmark paper in the study of molecular evolution.[5] They compared the nucleotide sequences of small-subunit rRNAs that had been purified from 13 different prokaryotic and eukaryotic species. The data from a comparison of all possible pairs of these organisms are shown in Table 1. The numbers along the top identify the organisms by the same numbers used along the left margin of the table. Each value in the table reflects the similarity in sequence between rRNAs from the two

---

[a]There are a number of anaerobic unicellular eukaryotes (e.g., the intestinal parasite *Giardia*) that lack mitochondria. For years, these organisms formed the basis for a proposal that mitochondrial endosymbiosis was a late event that took place after the evolution of these mitochondria-lacking groups. However, recent analysis of the nuclear DNA of these organisms indicates the presence of genes that were likely transferred to the nucleus from mitochondria, suggesting that the ancestors of these organisms lost their mitochondria during the course of evolution.

FIGURE 2 **Two-dimensional electrophoretic fingerprint of a T1 digest of chloroplast 16S ribosomal RNA.** The RNA fragments were electrophoresed in one direction at pH 3.5 and then in a second direction at pH 2.3. (FROM L. B. ZABLEN ET AL., PROC. NAT'L. ACAD. SCI. U.S.A. 72:2419, 1975.)

TABLE 1 Nucleotide Sequence Similarities between Representative Members of the Three Primary Kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.08
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Cholorbium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diptheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (Lemna)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanebacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium sp.*, Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

Source: C. R. Woese and G. E. Fox, *Proc. Nat'l. Acad. Sci. U.S.A.* 74:5089, 1977.

organisms being compared: the lower the number, the less similar the two sequences. They found that the sequences clustered into three distinct groups, as indicated in the table. It is evident that the rRNAs within each group (numbers 1–3, 4–9, and 10–13) are much more similar to one another than they are to rRNAs of the other two groups. The first of the groups shown in the table contains only eukaryotes; the second group contains the "typical" bacteria (gram-positive, gram-negative, and cyanobacteria); and the third group contains several species of methanogenic (methane-producing) "bacteria." Woese and Fox concluded, to their surprise, that the methanogenic organisms "appear to be no more related to typical bacteria than they are to eukaryotic cytoplasms." These results suggested that the members of these three groups represent three distinct evolutionary lines that branched apart from one another at a very early stage in the evolution of cellular organisms. Consequently, they assigned these organisms to three different kingdoms, which they named the Urkaryotes, Eubacteria, and Archaebacteria, a terminology that divided the prokaryotes into two fundamentally distinct groups.

Subsequent research provided support for the concept that prokaryotes could be divided into two distantly related lineages, and it expanded the ranks of the archaebacteria to include at least two other groups, the thermophiles, which live in hot springs and ocean vents, and the halophiles, which live in very salty lakes and seas. In 1989, two published reports rooted the tree of life and suggested that the archaebacteria were actually more closely related to eukaryotes than they were to eubacteria.[6,7] Both groups of researchers compared the amino acid sequences of several proteins that were present in a wide variety of different prokaryotes, eukaryotes, mitochondria, and chloroplasts. A phylogenetic tree constructed from sequences of ribosomal RNAs, which comes to the same conclusion, is shown in Figure 3.[8] In this latter paper, Woese and colleagues proposed a revised taxonomic scheme, which has been widely accepted. In this scheme, the archaebacteria, eubacteria, and eukaryotes are assigned to separate domains, which are named Archaea, Bacteria, and Eucarya, respectively.[b] Each domain can then be divided into one or

---

[b]Many biologists dislike the terms *archaebacteria* and *eubacteria*. Although these terms have gradually faded from the literature, being replaced simply by *archaea* and *bacteria*, many researchers in this field continue to use the former terms in published articles. Given that this is an introductory chapter in an introductory text, I have continued to refer to these organisms as archaebacteria and eubacteria to avoid possible confusion over meaning of *bacterial*.
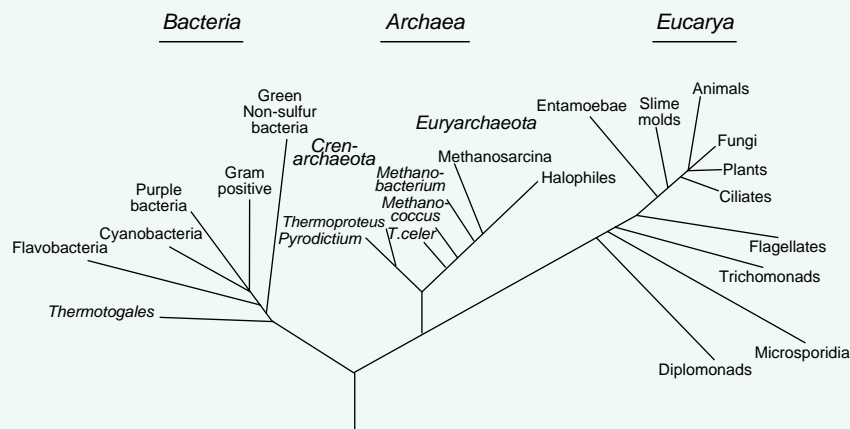
FIGURE 3 A phylogenetic tree based on rRNA sequence comparisons showing the three domains of life. The Archaea are divided into two subgroups as indicated. (FROM C. R. WOESE ET AL., PROC. NAT'L. ACAD. SCI. U.S.A. 87:4578, 1990.)

more kingdoms; the Eucarya, for example, can be divided into the traditional kingdoms containing fungi, protists, plants, and animals.

According to the model in Figure 3, the first major split in the tree of life produced two separate lineages, one leading to the Bacteria and the other leading to both the Archaea and the Eucarya. If this view is correct, it was an archaebacterium, not a eubacterium, that took in a symbiont and gave rise to the lineage that led to the first eukaryotic cells. Although the host prokaryote was presumably an archaebacterium, the symbionts that evolved into mitochondria and chloroplasts were almost certainly eubacteria, as indicated by their close relationship with modern members of this group.

Up until 1995, phylogenetic trees of the type shown in Figure 3 were based primarily on the analysis of the gene encoding the 16S–18S rRNA. By then, phylogenetic comparisons of a number of other genes were suggesting that the scheme depicted in Figure 3 might be oversimplified. Questions about the origin of prokaryotic and eukaryotic cells came into sharp focus between 1995 and 1997 with the publication of the entire sequences of a number of prokaryotic genomes, both archaebacterial and eubacterial, and the genome of a eukaryote, the yeast *Saccharomyces cerevisiae*. Researchers could now compare the sequences of hundreds of genes simultaneously, and this analysis raised a number of puzzling questions and blurred the lines of distinction between the three domains.[9] For example, the genomes of several archaebacteria showed the presence of a significant number of eubacterial genes. For the most part, those genes in archaebacteria whose products are involved with informational processes (chromosome structure, transcription, translation, and replication) were very different from their counterparts in eubacterial cells and, in fact, resembled the corresponding genes in eukaryotic cells. This observation fit nicely with the scheme in Figure 3. In contrast, many of the genes in archaebacteria that encode the enzymes of metabolism exhibited an unmistakable eubacterial character.[10,11] The genomes of eubacterial species also showed evidence of a mixed origin, often containing a significant number of genes that bore an archaebacterial character.[12]

Most investigators who study the origin of ancient organisms have held on to the basic outline of the phylogenetic tree as demarcated in Figure 3 and argue that the presence of eubacteria-like genes in archaebacteria, and vice versa, is the result of the transfer of genes from one species to another, a phenomenon referred to as *lateral gene transfer* (*LGT*).[13] According to the original premise that led to the phylogenetic tree of Figure 3, genes are inherited from one's parents, not from one's neighbors. This is the premise that allows an investigator to conclude that two species are closely related when they both possess a gene (e.g., the rRNA gene) of similar nucleotide sequence. If, however, cells can pick up genes from other species in their environment, then two species that are actually unrelated may possess genes of very similar sequence. An early measure of the importance of lateral gene transfer in the evolution of prokaryotes came from a study that compared the genomes of two related eubacteria, *Escherichia* and *Salmonella*. It was found that 755 genes or nearly 20 percent of the *E. coli* genome is derived from "foreign" genes transferred into the *E. coli* genome over the past 100 million years, which is the time when the two eubacteria diverged. These 755 genes were acquired as the result of at least 234 separate lateral transfers from many different sources.[14] (The effect of lateral gene transfer on antibiotic resistance in pathogenic bacteria is discussed in the Human Perspective of Chapter 3.)

If genomes are a mosaic composed of genes from diverse sources, how does one choose which genes to use in determining phylogenetic relationships? According to one viewpoint, genes that are involved in informational activities (transcription, translation, replication) make the best subjects for determining phylogenetic relationships, because such genes are less likely to be transferred laterally than genes involved in metabolic reactions.[15] These authors argue that the products of informational genes (e.g., rRNAs) are parts of large complexes whose components must interact with many other molecules. It is unlikely that a foreign gene product could become integrated into the existing machinery. When "informational genes" are used as the subjects of comparison, archaebacteria and eubacteria tend to separate into distinctly different groups, whereas archaebacteria and eukaryotes tend to group together as evolutionary relatives, just as they do in Figure 3.[See reference 16 for further discussion.]

Analysis of eukaryotic genomes have produced similar evidence of a mixed heritage. Studies of the yeast genome show unmistakable presence of genes derived from both archaebacteria and eubacteria. The "informational genes" tend to have an archaeal character and the "metabolic genes" a eubacterial character.[17] There are several possible explanations for the mixed character of the eukaryotic genome. Eukaryotic cells may have evolved from archaebacterial ancestors and then picked up genes from eubacteria with which they shared environments. In addition, some of the genes in the nucleus of a eukaryotic cell are clearly derived from eubacterial genes that have been transferred from the genome of the symbionts that evolved into mitochondria and chloroplasts.[18] A number of researchers have taken a more radical position and proposed that the eukaryote genome was originally derived from the fusion of an archaebacterial and a eubacterial cell followed by the integration of their two genomes.[e.g.,19] Given these various routes of gene acquisition, it is evident that no simple phylogenetic tree, such as that depicted in Figure 3, can represent the evolutionary history of the entire genome of an organism.[Reviewed in 20–22] Instead, each gene or group of genes of a particular genome may have its own unique evolutionary tree, which can be a disconcerting thought to scientists seeking to determine the origin of our earliest eukaryotic ancestors.

References

1. SAGAN (MARGULIS), L. 1967. On the origin of mitosing cells. *J. Theor. Biol.* 14:225–274.

2. MARGULIS, L. 1970. *Origin of Eukaryotic Cells.* Yale University Press.

3. ZUCKERKANDL, E. & PAULING, L. 1965. Molecules as documents of evolutionary history. *J. Theor. Biol.* 8:357–365.

4. ZABLEN, L. B., ET AL. 1975. Phylogenetic origin of the chloroplast and prokaryotic nature of its ribosomal RNA. *Proc. Nat'l. Acad. Sci. U.S.A.* 72:2418–2422.

5. WOESE, C. R. & FOX, G. E. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Nat'l. Acad. Sci. U.S.A.* 74:5088–5090.

6. IWABE, N., ET AL. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Nat'l. Acad. Sci. U.S.A.* 86:9355–9359.

7. GOGARTEN, J. P., ET AL. 1989. Evolution of the vacuolar H^+-ATPase: Implications for the origin of eukaryotes. *Proc. Nat'l. Acad. Sci. U.S.A.* 86:6661–6665.

8. WOESE, C., ET AL. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Nat'l. Acad. Sci. U.S.A.* 87:4576–4579.

9. DOOLITTLE, W. F. 1999. Lateral genomics. *Trends Biochem. Sci.* 24:M5–M8 (Dec.)

10. BULT, C. J., ET AL. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273:1058–1073.

11. KOONIN, E. V., ET AL. 1997. Comparison of archaeal and bacterial genomes. *Mol. Microbiol.* 25:619–637.

12. NELSON, K. E., ET AL., 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399:323–329.

13. OCHMAN, H., ET AL. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304.

14. LAWRENCE, J. G. & OCHMAN, H. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proc. Nat'l. Acad. Sci. U.S.A.* 95:9413–9417.

15. JAIN, R., ET AL. 1999. Horizontal gene transfer among genomes: The complexity hypothesis. *Proc. Nat'l. Acad. Sci. U.S.A.* 96:3801–3806.

16. MCINERNEY, J. O. & PISANI, D. 2007. Paradigm for life. *Science* 318:1390–1391.

17. RIVERA, M. C., ET AL. 1998. Genomic evidence for two functionally distinct gene classes. *Proc. Nat'l. Acad. Sci. U.S.A.* 95:6239–6244.

18. TIMMIS, J. N., ET AL. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Rev. Gen.* 5:123–135.

19. MARTIN, W. & MÜLLER, M. 1998. The hydrogen hypothesis for the first eukaryote. *Nature* 392:37–41.

20. WALSH, D. A. & DOOLITTLE, W. F. 2005. The real "domains" of life. *Curr. Biol.* 15:R237–R240.

21. ESSER, C. & MARTIN, W. 2007. Supertrees and symbiosis in eukaryote genome evolution. *Trends Microbiol.* 15:435–437.

22. ARCHIBALD, J. M. 2008. The eocyte hypothesis and the origin of eukaryotic cells. *Proc. Nat'l. Acad. Sci. U.S.A.* 105:20049–20050.

SYNOPSIS

The cell theory has three tenets. (1) All organisms are composed of one or more cells; (2) the cell is the basic organizational unit of life; and (3) all cells arise from preexisting cells. (p. 2)

The properties of life, as exhibited by cells, can be described by a collection of properties. Cells are very complex and their substructure is highly organized and predictable. The information to build a cell is encoded in its genes. Cells reproduce by cell division; their activities are fueled by chemical energy; they carry out enzymatically controlled chemical reactions; they engage in numerous mechanical activities; they respond to stimuli; and they are capable of a remarkable level of self-regulation. (p. 3)

Cells are either prokaryotic or eukaryotic. Prokaryotic cells are found only among archaebacteria and eubacteria, whereas all other types of organisms—protists, fungi, plants, and animals—are composed of eukaryotic cells. Prokaryotic and eukaryotic cells share many common features, including a similar cellular membrane, a common system for storing and using genetic information, and similar metabolic pathways. Prokaryotic cells are the simpler type, lacking the complex membranous organelles (e.g., endoplasmic reticulum, Golgi complex, mitochondria, and chloroplasts), chromosomes, and cytoskeleton characteristic of the cells of eukaryotes. The two cell types can also be distinguished by their mechanism of cell division, their locomotor structures, and the type of cell wall they produce (if a cell wall is present). Complex plants and animals contain many different types of cells, each specialized for particular activities. (p. 7)

Cells are almost always microscopic in size. Bacterial cells are typically 1 to 5 µm in length, whereas eukaryotic cells are typically 10 to 30 µm. Cells are microscopic in size for a number of reasons: their nuclei possess a limited number of copies of each gene; the surface area (which serves as the cell's exchange surface) becomes limiting as a cell increases in size; and the distance between the cell surface and interior becomes too great for the cell's needs to be met by simple diffusion. (p. 16)

Viruses are noncellular pathogens that can only reproduce when present within a living cell. Outside of the cell, the virus exists as a macromolecular package, or virion. Virions occur in a variety of shapes and sizes, but all of them consist of viral nucleic acid enclosed in a wrapper containing viral proteins. Viral infections may lead to either (1) the destruction of the host cell with accompanying production of viral progeny, or (2) the integration of viral nucleic acid into the DNA of the host cell, which often alters the activities of that cell. Viruses are not considered to be living organisms. (p. 21)

ANALYTIC QUESTIONS

1. Consider some question about cell structure or function that you would be interested in answering. Would the data required to answer the question be easier to collect by working on an entire plant or animal or on a population of cultured cells? What might be the advantages and disadvantages of working on a whole organism versus a cell culture?

2. Figure 1.3 shows an intestinal epithelial cell with large numbers of microvilli. What is the advantage to the organism of having these microvilli? What do you expect would happen to an individual that lacked such microvilli as the result of an inherited mutation?

3. The first human cells to be successfully cultured were derived from a malignant tumor. Do you think this simply reflects the availability of cancer cells, or might such cells be better subjects for cell culture? Why?

4. The drawings of plant and animal cells in Figure 1.8*b,c* include certain structures that are present in plant cells but absent in animal cells. How do you think each of these structures affects the life of the plant?

5. It was noted that cells possess receptors on their surface that allow them to respond to specific stimuli. Many cells in the human body possess receptors that allow them to bind specific hormones that circulate in the blood. Why do you think these hormone receptors are important? What would be the effect on the physiological activities of the body if cells lacked these receptors, or if all cells had the same receptors?

6. If you were to argue that viruses are living organisms, what features of viral structure and function might you use in your argument?

7. If we presume that activities within cells do occur in a manner analogous to that shown in the Rube Goldberg cartoon of Fig-

ure 1.7, how would this differ from a human activity, such as building a car on an assembly line or shooting a free throw in a basketball game?

8. Unlike bacterial cells, the nucleus of a eukaryotic cell is bounded by a double-layered membrane studded by complex pores. How do you think this might affect traffic between the DNA and cytoplasm of a eukaryotic cell compared to that of a prokaryotic cell?

9. Examine the photograph of the ciliated protist in Figure 1.16 and consider some of the activities in which this cell engages that a muscle or nerve cell in your body does not.

10. Which type of cell would you expect to achieve the largest volume: a highly flattened cell or a spherical cell? Why?

11. Suppose you were a scientist living in the 1890s and were studying a disease of tobacco crops that stunted the growth of the plants and mottled their leaves. You find that the sap from a diseased plant, when added to a healthy plant, is capable of transmitting the disease to that plant. You examine the sap in the best light microscopes of the period and see no evidence of bacteria. You force the sap through filters whose pores are so small that they retard the passage of the smallest known bacteria, yet the fluid that passes through the filters is still able to transmit the disease. Like Dimitri Ivanovsky, who conducted these experiments more than a hundred years ago, you would probably conclude that the infectious agent was an unknown type of unusually small bacterium. What kinds of experiments might you perform today to test this hypothesis?

12. Most evolutionary biologists believe that all mitochondria have evolved from a single ancestral mitochondrion and all chloroplasts have evolved from a single ancestral chloroplast. In other words, the symbiotic event that gave rise to each of these organelles occurred only once. If this is the case, where on the phylogenetic tree of Figure 3, page 27, would you place the acquisition of each of these organelles?

13. Publication of the complete sequence of the 1918 flu virus and reconstitution of active viral particles was met with great controversy. Those who favored publication of the work argued that this type of information can help to better understand the virulence of influenza viruses and help develop better therapeutics against them. Those opposed to its publication argued that the virus could be reconstituted by bioterrorists or that another pandemic could be created by the accidental release of the virus by a careless investigator. What is your opinion on the merits of conducting this type of work?

2
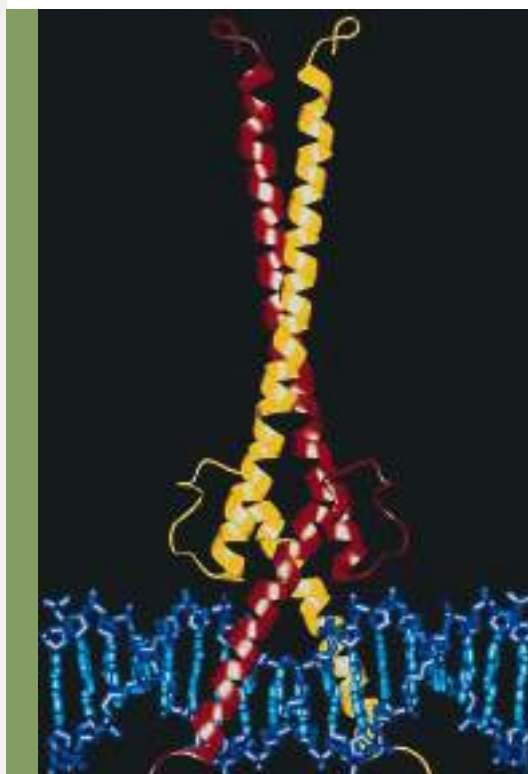
The Chemical Basis of Life

W e will begin this chapter with a brief examination of the atomic basis of matter, a subject that may seem out of place in a biology textbook. Yet life is based on the properties of atoms and is governed by the same principles of chemistry and physics as all other types of matter. The cellular level of organization is only a small step from the atomic level, as will become evident when we examine the importance of the movement of a few atoms of a molecule during such activities as muscle contraction or the transport of substances across cell membranes. The properties of cells and their organelles derive directly from the activities of the molecules of which they are composed. Consider a process such as cell division, which can be followed in considerable detail under a simple light microscope. To understand the activities that occur when a cell divides, one needs to know, for example, about the interactions between DNA and protein molecules that cause the chromosomes to condense into rod-shaped packages that can be separated into different cells; the molecular construction of protein-containing microtubules that allows them to disassemble at one moment in the cell and reassemble the next moment in an entirely different cellular location; and the properties of lipid molecules that make the outer cell membrane deformable so that it can be pulled into the middle of a cell, thereby pinching the cell in two. It is impossible even to begin to understand cellular function without a reasonable knowledge of the structure and properties of the major types of biological molecules. This is the goal of the present chapter: to provide the necessary information about the chemistry of life to allow the reader to understand the basis of life. We will begin by considering the types of bonds that atoms can form with one another. ■

A complex between two different macromolecules. A portion of a DNA molecule (shown in blue) is complexed to a protein that consists of two polypeptide subunits, one red and the other yellow. Those parts of the protein that are seen to be inserted into the grooves of the DNA have recognized and bound to a specific sequence of nucleotides in the nucleic acid molecule. (COURTESY OF A. R. FERRÉ-D'AMARÉ AND STEPHEN K. BURLEY.)

2.1 COVALENT BONDS

The atoms that make up a molecule are joined together by **covalent bonds** in which pairs of electrons are shared between pairs of atoms. The formation of a covalent bond between two atoms is governed by the fundamental principle that an atom is most stable when its outermost electron shell is filled. Consequently, the number of bonds an atom can form depends on the number of electrons needed to fill its outer shell.

The electronic structure of a number of atoms is shown in Figure 2.1. The outer (and only) shell of a hydrogen or helium atom is filled when it contains two electrons; the outer shells of the other atoms in Figure 2.1 are filled when they contain eight electrons. Thus, an oxygen atom, with six outer-shell electrons, can fill its outer shell by combining with two hydro-

gen atoms, forming a molecule of water. The oxygen atom is linked to each hydrogen atom by a *single* covalent bond (denoted as H:O or H—O). The formation of a covalent bond is accompanied by the release of energy, which must be reabsorbed at some later time if the bond is to be broken. The energy required to cleave C—H, C—C, or C—O covalent bonds is quite large—typically between 80 and 100 kilocalories per mole (kcal/mol)[1] of molecules—making these bonds stable under most conditions.

In many cases, two atoms can become joined by bonds in which more than one pair of electrons are shared. If two electron pairs are shared, as occurs in molecular oxygen (O_2), the covalent bond is a *double bond*, and if three pairs of electrons are shared (as in molecular nitrogen, N_2), it is a *triple bond*. Quadruple bonds are not known to occur. The type of bond

---

[1]One calorie is the amount of thermal energy required to raise the temperature of one gram of water one degree centigrade. One kilocalorie (kcal) equals 1000 calories (or one large Calorie). In addition to calories, energy can also be expressed in Joules, which is a term that was used historically to measure energy in the form of work. One kilocalorie is equivalent to 4186 Joules. Conversely, 1 Joule = 0.239 calories. A mole is equal to Avogadro's number (6×10^{23}) of molecules. A mole of a substance is its molecular weight expressed in grams.

FIGURE 2.1 A representation of the arrangement of electrons in a number of common atoms. Electrons are present around an atom's nucleus in "clouds" or *orbitals* that are roughly defined by their boundaries, which may have a spherical or dumbbell shape. Each orbital contains a maximum of two electrons, which is why the electrons (dark dots in the drawing) are grouped in pairs. The innermost shell contains a single orbital (thus two electrons), the second shell contains four orbitals (thus eight electrons), the third shell also contains four orbitals, and so forth. The number of outer-shell electrons is a primary determinant of the chemical properties of an element. Atoms with a similar number of outer-shell electrons have similar properties. Lithium (Li) and sodium (Na), for example, have one outer-shell electron, and both are highly reactive metals. Carbon (C) and silicon (Si) atoms can each bond with four different atoms. Because of its size, however, a carbon atom can bond to other carbon atoms, forming long-chained organic molecules, whereas silicon is unable to form comparable molecules. Neon (Ne) and argon (Ar) have filled outer shells, making these atoms highly nonreactive; they are referred to as inert gases.

between atoms has important consequences in determining the shapes of molecules. For example, atoms joined by a single bond are able to rotate relative to one another, whereas the atoms of double (and triple) bonds lack this ability. As illustrated in Figure 6.6, double bonds can function as energy-capturing centers, driving such vital processes as respiration and photosynthesis.

When atoms of the same element bond to one another, as in H_2, the electron pairs of the outer shell are equally shared between the two bonded atoms. When two unlike atoms are covalently bonded, however, the positively charged nucleus of one atom exerts a greater attractive force on the outer electrons than the other. Consequently, the shared electrons tend to be located more closely to the atom with the greater attractive force, that is, the more **electronegative atom**. Among the atoms most commonly present in biological molecules, nitrogen and oxygen are strongly electronegative.

Polar and Nonpolar Molecules

Let's examine a molecule of water. Water's single oxygen atom attracts electrons much more forcefully than do either of its hydrogen atoms. As a result, the O—H bonds of a water molecule are said to be *polarized*, such that one of the atoms has a partial negative charge and the other a partial positive charge. This is generally denoted in the following manner:
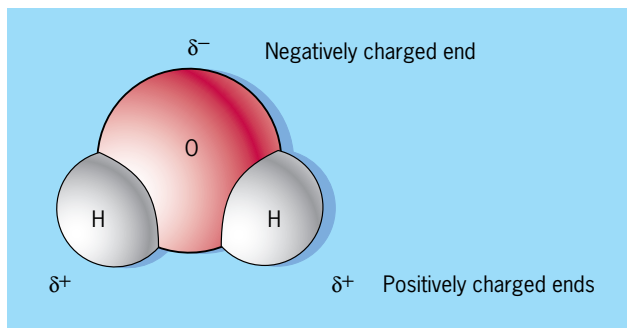
Molecules, such as water, that have an asymmetric distribution of charge (or *dipole*) are referred to as **polar** molecules. Polar molecules of biological importance contain one or more electronegative atoms, usually O, N, and/or S. Molecules that lack electronegative atoms and strongly polarized bonds, such as molecules that consist entirely of carbon and hydrogen atoms, are said to be **nonpolar**. The presence of strongly polarized bonds is of utmost importance in determining the reactivity of molecules. Large nonpolar molecules, such as waxes and fats, are relatively inert. Some of the more interesting biological molecules, including proteins and phospholipids, contain both polar and nonpolar regions, which behave very differently.

Ionization

Some atoms are so strongly electronegative that they can capture electrons from other atoms during a chemical reaction. For example, when the elements sodium (a silver-colored metal) and chlorine (a toxic gas) are mixed, the single electron in the outer shell of each sodium atom migrates to the electron-deficient chlorine atom. As a result, these two atoms are transformed into charged **ions**.

$$2\,\mathrm{Na}\cdot + \,:\!\ddot{\mathrm{C}}\mathrm{l}\!:\!\ddot{\mathrm{C}}\mathrm{l}\!: \rightarrow 2\,\mathrm{Na}\!:\!\ddot{\mathrm{C}}\mathrm{l}\!: \rightarrow 2\,\mathrm{Na^+} + 2\,:\!\ddot{\mathrm{C}}\mathrm{l}\!:^-$$
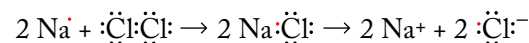
Because the chloride ion has an extra electron (relative to the number of protons in its nucleus), it has a negative charge (Cl^-) and is termed an **anion**. The sodium atom, which has lost an electron, has an extra positive charge (Na^+) and is termed a **cation**. When present in crystals, these two ions form sodium chloride, or table salt.

The Na^+ and Cl^- ions depicted above are relatively stable because they possess filled outer shells. A different arrangement of electrons within an atom can produce a highly reactive species, called a *free radical*. The structure of free radicals and their importance in biology are considered in the accompanying Human Perspective.

REVIEW ?

1. Oxygen atoms have eight protons in their nucleus. How many electrons do they have? How many orbitals are in the inner electron shell? How many electrons are in the outer shell? How many more electrons can the outer shell hold before it is filled?

2. Compare and contrast: a sodium atom and a sodium ion; a double bond and a triple bond; an atom of weak and strong electronegativity; the electron distribution around an oxygen atom bound to another oxygen atom and an oxygen atom bound to two hydrogen atoms.

2.2 NONCOVALENT BONDS

Covalent bonds are strong bonds between the atoms that make up a molecule. Interactions between molecules (or between different parts of a large biological molecule) are governed by a variety of weaker linkages called noncovalent bonds. **Noncovalent bonds** do not depend on shared electrons but rather on attractive forces between atoms having an opposite charge. Individual noncovalent bonds are weak (about 1 to 5 kcal/mol) and are thus readily broken and reformed. As will be evident throughout this book, this feature allows noncovalent bonds to mediate the dynamic interactions among molecules in the cell.

Even though individual noncovalent bonds are weak, when large numbers of them act in concert, as between the two strands of a DNA molecule or between different parts of a large protein, their attractive forces are additive. Taken as a whole, they provide the structure with considerable stability. We will examine several types of noncovalent bonds that are important in cells.

Ionic Bonds: Attractions between Charged Atoms

A crystal of table salt is held together by an electrostatic attraction between positively charged Na^+ and negatively charged Cl^- ions. This type of attraction between fully

THE HUMAN PERSPECTIVE

Free Radicals as a Cause of Aging

Many biologists believe that aging results from the gradual accumulation of damage to our body's tissues. The most destructive damage probably occurs to DNA. Alterations in DNA lead to the production of faulty genetic messages that promote gradual cellular deterioration. How does cellular damage occur, and why should it occur more rapidly in a shorter lived animal, such as a chimpanzee, than a human? The answer may reside at the atomic level.

Atoms are stabilized when their shells are filled with electrons. Electron shells consist of orbitals, each of which can hold a maximum of two electrons. Atoms or molecules that have orbitals containing a single unpaired electron tend to be highly unstable—they are called **free radicals**. Free radicals may be formed when a covalent bond is broken such that each portion keeps one-half of the shared electrons, or they may be formed when an atom or molecule accepts a single electron transferred during an oxidation–reduction reaction. For example, water can be converted into free radicals when exposed to radiation from the sun:

$$H_2O \rightarrow HO\cdot + H\cdot$$

<div align="center">hydroxyl
radical
("·" indicates a free radical)</div>
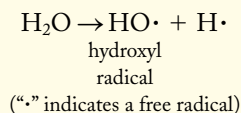
Free radicals are extremely reactive and capable of chemically altering many types of molecules, including proteins, nucleic acids, and lipids. The formation of hydroxyl radicals is probably a major reason that sunlight is so damaging to skin.

In 1956, Denham Harman of the University of Nebraska proposed that aging results from tissue damage caused by free radicals. Because the subject of free radicals was not one with which biologists and physicians were familiar, the proposal failed to generate significant interest. Then, in 1969, Joe McCord and Irwin Fridovich of Duke University discovered an enzyme, superoxide dismutase (SOD), whose sole function was the destruction of the superoxide radical ($O_2\cdot^-$), a type of free radical formed when molecular oxygen picks up an extra electron. SOD catalyzes the following reaction:

$$O_2\cdot^- + O_2\cdot^- + 2H^+ \rightarrow H_2O_2 + O_2$$
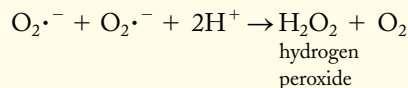
<div align="center">hydrogen
peroxide</div>

Hydrogen peroxide is also a potentially reactive oxidizing agent, which is why it is often used as a disinfectant and bleaching agent. If it is not rapidly destroyed, H_2O_2 can break down to form hydroxyl radicals that attack the cell's macromolecules. Hydrogen peroxide is normally destroyed in the cell by the enzymes catalase or glutathione peroxidase.

Subsequent research has revealed that superoxide radicals are formed within cells during normal oxidative metabolism and that a superoxide dismutase is present in the cells of diverse organisms, from bacteria to humans. In fact, animals possess three different versions (isoforms) of SOD: a cytosolic, mitochondrial, and extracellular isoform. It is estimated that as much as 1–2 percent of the oxygen taken into human mitochondria can be converted to hydrogen peroxide rather than to water, the normal end product of respiration. The importance of SOD is most clearly revealed in studies of mutant bacteria and yeast that lack the enzyme; these cells are unable to grow in the presence of oxygen. Similarly, mice that are lacking the mitochondrial version of the enzyme (SOD2) are not able to survive more than a week or so after birth. Conversely, mice that have been genetically engineered so that their mitochondria contain elevated levels of the H_2O_2-destroying enzyme catalase live 20 percent longer than untreated controls. This finding, reported in 2005, marked the first demonstration that enhanced antioxidant defenses can increase the life span of a mammal. Although the destructive potential of free radicals, such as superoxide and hydroxyl radicals, is unquestioned, the importance of these agents as a factor in aging remains controversial.

The life spans of animals can be increased by restricting the calories present in the diet. As first shown in the 1930s, mice that are maintained on very strict diets typically live 30 to 40 percent longer than their littermates who are fed diets of normal caloric content. Studies of the metabolic rates of these mice have produced contradictory data, but there is general agreement that animals fed calorie-restricted diets exhibit a marked decrease in $O_2\cdot^-$ and H_2O_2 production, which could explain their increased longevity.

As reported in numerous television news shows, a growing number of humans are hoping to extend their life span by practicing calorie restriction, which in essence means that they are willing to subject themselves to an extremely limited, but balanced, diet. The National Institutes of Aging has also begun a study (named CALERIE) on human subjects who are overweight (but not obese) that are kept on diets containing about 25 percent fewer calories than would be required to maintain their initial body weight. After a period of six months of calorie restriction, these individuals show remarkable metabolic changes; they have a lower body temperature, their blood insulin and LDL-cholesterol levels are lower, they have lost weight as would be expected, and their energy expenditure is reduced beyond that expected due simply to their lower body mass. In addition, the level of DNA damage experienced by the cells of these individuals is reduced, which suggests a decrease in production of reactive oxygen species. Long-term studies are currently in progress on rhesus monkeys to see if they too live longer and healthier lives when maintained on calorie-restricted diets. Although these studies have not been conducted long enough to determine if their *maximum* life span (normally about 40 years) is increased, these animals also have lower blood levels of glucose, insulin, and triglycerides, which makes them less prone to age-related disorders such as diabetes and coronary artery disease. Lowered blood-insulin levels may be particularly important in promoting longevity, as studies on nematodes and fruit flies suggest that reducing the activity of insulin-like hormones can dramatically increase life span in these invertebrates.

A related area of research concerns the study of substances called antioxidants that are able to destroy free radicals in the test tube. The sale of these substances provides a major source of revenue for the vitamin/supplements industry. Common antioxidants found in the body include glutathione, vitamins E and C, and beta-carotene (the orange pigment in carrots and other vegetables). Although these substances may prove beneficial in the diet because of their ability to destroy free radicals, studies with rats and mice have failed to provide convincing evidence that they retard the aging process or increase maximum life span. An antioxidant receiving considerable interest is resveratrol, a polyphenolic compound found at high concentration in the skin of red grapes. It is widely believed that resveratrol is responsible for the health-related benefits attributed to red wine. Rather than scavenging for free radicals, resveratrol appears to act by stimulating an enzyme (Sir2) that serves as a key player in promoting longevity in animal studies.
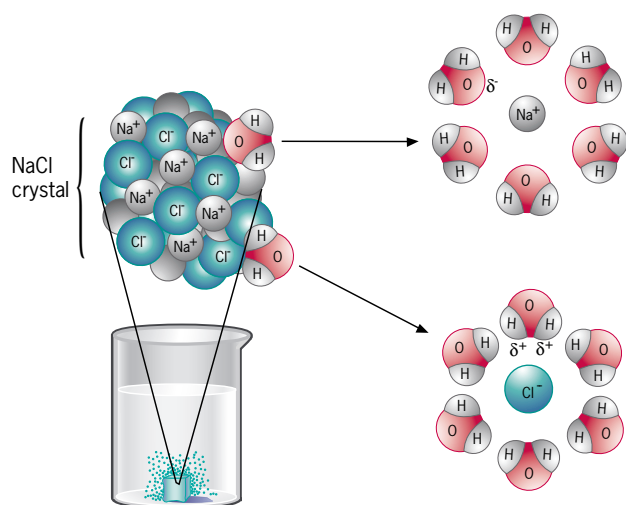
FIGURE 2.2 The dissolution of a salt crystal. When placed in water, the Na$^+$ and Cl$^-$ ions of a salt crystal become surrounded by water molecules, breaking the ionic bonds between the two ions. As the salt dissolves, the negatively charged oxygen atoms of the water molecules associate with the positively charged sodium ions, and the positively charged hydrogen atoms of the water molecules associate with the negatively charged chloride ions.

charged components is called an **ionic bond** (or a *salt bridge*). Ionic bonds within a salt crystal may be quite strong. However, if a crystal of salt is dissolved in water, each of the individual ions becomes surrounded by water molecules, which inhibit oppositely charged ions from approaching one another closely enough to form ionic bonds (Figure 2.2). Because cells are composed primarily of water, bonds between *free* ions are of little importance. In contrast, weak ionic bonds between oppositely charged groups of large biological molecules are of considerable importance. For example, when negatively charged phosphate atoms in a DNA molecule are closely associated with positively charged groups on the surface of a protein (Figure 2.3), ionic bonds between them help hold the complex together. The strength of ionic bonds in a cell is generally weak (about 3 kcal/mol) due to the presence of water, but deep within the core of a protein, where water is often excluded, such bonds can be much stronger.

Hydrogen Bonds

When a hydrogen atom is covalently bonded to an electronegative atom, particularly an oxygen or a nitrogen atom, the single pair of shared electrons is greatly displaced toward the nucleus of the electronegative atom, leaving the hydrogen atom with a partial positive charge. As a result, the bare, positively charged nucleus of the hydrogen atom can approach near enough to an unshared pair of outer electrons of a second electronegative atom to form an attractive interaction (Figure 2.4). This weak attractive interaction is called a **hydrogen bond**.

Hydrogen bonds occur between most polar molecules and are particularly important in determining the structure and properties of water (discussed later). Hydrogen bonds also
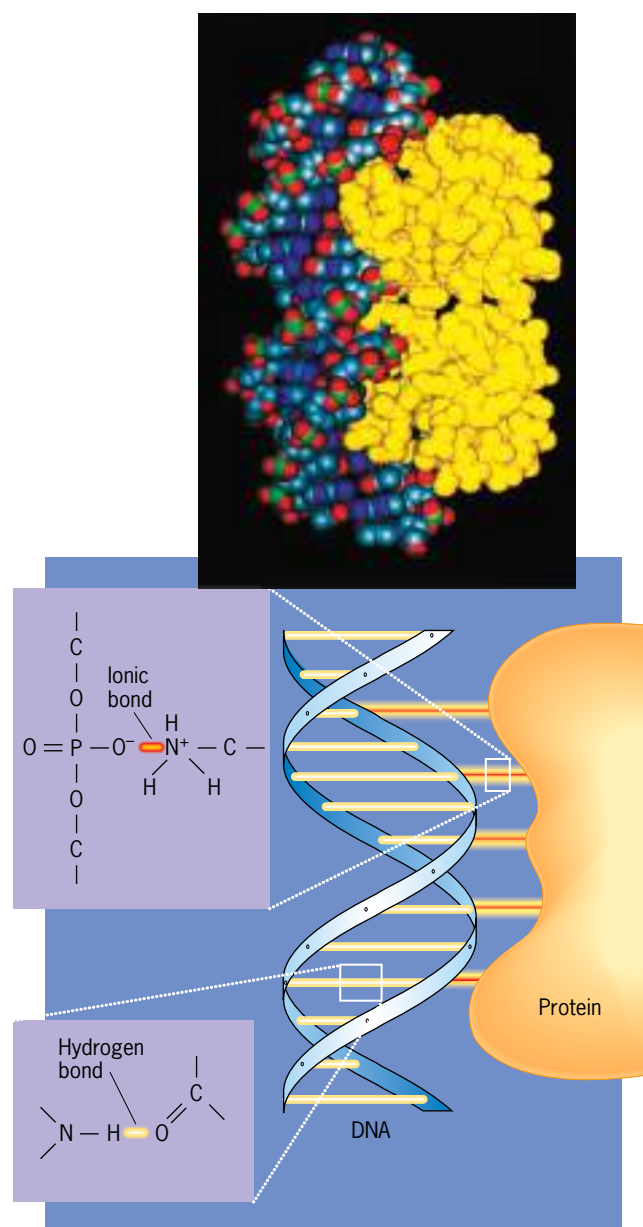
FIGURE 2.3 Noncovalent ionic bonds play an important role in holding the protein molecule on the right (yellow atoms) to the DNA molecule on the left. Ionic bonds form between positively charged nitrogen atoms in the protein and negatively charged oxygen atoms in the DNA. The DNA molecule itself consists of two separate strands held together by noncovalent hydrogen bonds. Although a single noncovalent bond is relatively weak and easily broken, large numbers of these bonds between two molecules, as between two strands of DNA, make the overall complex quite stable. (Top image courtesy of Stephen Harrison.)

form between polar groups present in large biological molecules, as occurs between the two strands of a DNA molecule (see Figure 2.3). Because their strength is additive, the large number of hydrogen bonds between the strands makes the DNA duplex a stable structure. However, because individual hydrogen bonds are weak (2–5 kcal/mol), the two strands can be partially separated to allow enzymes access to individual strands of the DNA molecule.
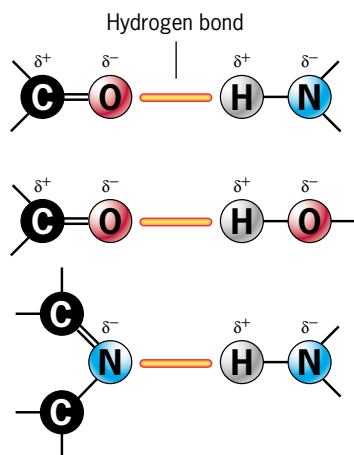
Hydrogen bond

FIGURE 2.4 **Hydrogen bonds** form between a bonded electronegative atom, such as nitrogen or oxygen, which bears a partial negative charge, and a bonded hydrogen atom, which bears a partial positive charge. Hydrogen bonds (about 0.18 nm) are typically about twice as long as the much stronger covalent bonds.

Hydrophobic Interactions and van der Waals Forces

Because of their ability to interact with water, polar molecules, such as sugars and amino acids (described shortly), are said to be **hydrophilic**, or "water loving." Nonpolar molecules, such as steroid or fat molecules, are essentially insoluble in water because they lack the charged regions that would attract them to the poles of water molecules. When nonpolar compounds are mixed with water, the nonpolar, **hydrophobic** ("water fearing") molecules are forced into aggregates, which minimizes their exposure to the polar surroundings (Figure 2.5). This association of nonpolar molecules is called a **hydrophobic interaction**. This is why droplets of fat molecules rapidly reappear on the surface of beef or chicken soup even after the liquid is stirred with a spoon. This is also the reason that nonpolar groups tend to localize within the interior of most soluble proteins away from the surrounding water molecules.

Hydrophobic interactions of the type just described are not classified as true bonds because they do not result from an attraction between hydrophobic molecules.[2] In addition to this type of interaction, hydrophobic groups can form weak bonds with one another based on electrostatic attractions. Polar molecules associate because they contain permanent asymmetric charge distributions within their structure. Closer examination of the covalent bonds that make up a nonpolar molecule (such as H_2 or CH_4) reveals that electron distributions are not always symmetric. The distribution of electrons around an atom at any given instant is a statistical matter and,

[2]This statement reflects an accepted hypothesis that hydrophobic interactions are driven by increased entropy (disorder). When a hydrophobic group projects into an aqueous solvent, the water molecules become ordered in a cage around the hydrophobic group. These solvent molecules become disordered when the hydrophobic group withdraws from the surrounding solvent. A discussion of this and other views can be found in *Nature* 437:640, 2005 and *Curr. Opin. Struct. Biol.* 16:152, 2006.

Hydrogen bond

Hydrophobic interactions
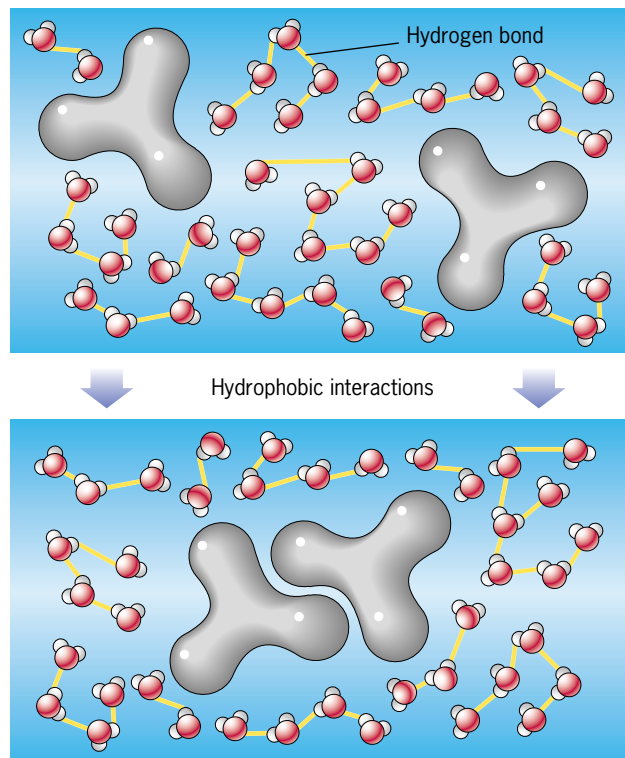
FIGURE 2.5 **In a hydrophobic interaction,** the nonpolar (hydrophobic) molecules are forced into aggregates, which minimizes their exposure to the surrounding water molecules.

therefore, varies from one instant to the next. Consequently, at any given time, the electron density may happen to be greater on one side of an atom, even though the atom shares the electrons equally with some other atom. These transient asymmetries in electron distribution result in momentary separations of charge (*dipoles*) within the molecule. If two molecules with transitory dipoles are very close to one another and oriented in the appropriate manner, they experience a weak attractive force, called a **van der Waals force**, that bonds them together. Moreover, the formation of a temporary separation of charge in one molecule can *induce* a similar separation in an adjacent molecule. In this way, additional attractive forces can be generated between nonpolar molecules. A single van der Waals force is very weak (0.1 to 0.3 kcal/mol) and very sensitive to the distance that separates the two atoms (Figure 2.6*a*). As we will see in later chapters, however, biological molecules that interact with one another, for example, an antibody and a protein on the surface of a virus, often possess complementary shapes. As a result, many atoms of both interactants may have the opportunity to approach each other very closely (Figure 2.6*b*), making van der Waals forces important in biological interactions.

The Life-Supporting Properties of Water

Life on Earth is totally dependent on water, and water may be essential to the existence of life anywhere in the universe. Even though it contains only three atoms, a molecule of water

Separation between centers of atoms (Å)
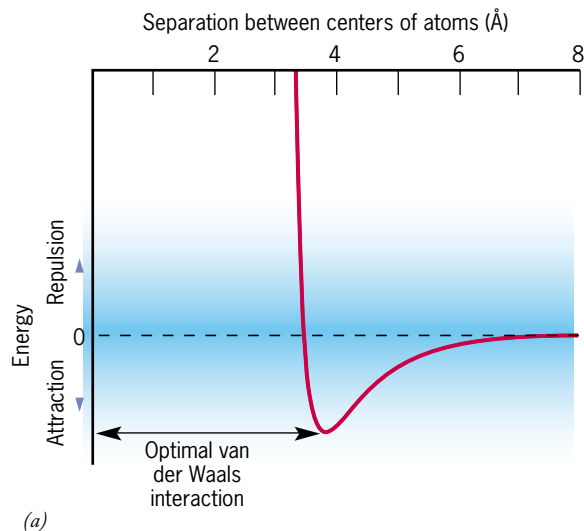
(a)
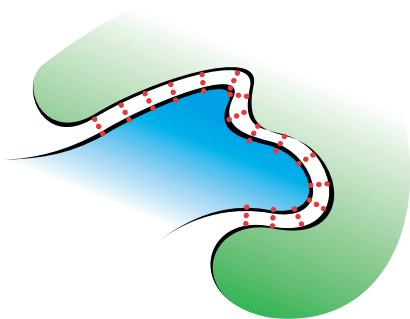
(b)

FIGURE 2.6 Van der Waals forces. (*a*) As two atoms approach each other, they experience a weak attractive force that increases up to a specific distance, typically about 4 Å. If the atoms approach more closely, their electron clouds repel one another, causing the atoms to be forced apart. (*b*) Although individual van der Waals forces are very weak, large numbers of such attractive forces can be formed if two macromolecules have a complementary surface, as is indicated schematically in this figure (see Figure 2.40 for an example).

has a unique structure that gives the molecule extraordinary properties.[3] Most importantly,

1. Water is a highly asymmetric molecule with the O atom at one end and the two H atoms at the opposite end.

2. Each of the two covalent bonds in the molecule is highly polarized.

3. All three atoms in a water molecule are adept at forming hydrogen bonds.

The life-supporting attributes of water stem from these properties.

  Each molecule of water can form hydrogen bonds with as many as four other water molecules, producing a highly inter-

_____
[3]One way to appreciate the structure of water is by comparing it to H_2S. Like oxygen, sulfur has six outer-shell electrons and forms single bonds with two hydrogen atoms. But because sulfur is a larger atom, it is less electronegative than oxygen, and its ability to form hydrogen bonds is greatly reduced. At room temperature, H_2S is a gas, not a liquid. In fact, the temperature has to drop to $-86°C$ before H_2S freezes into a solid.
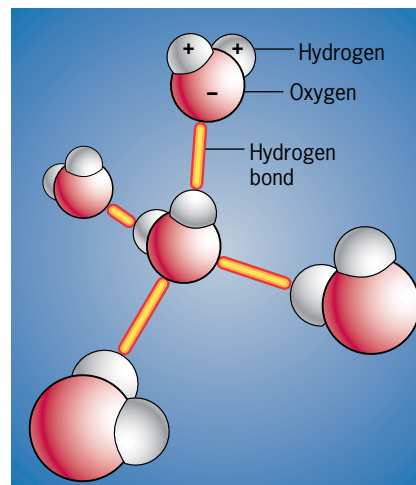
FIGURE 2.7 Hydrogen bond formation between neighboring water molecules. Each H atom of the molecule has about four-tenths of a full positive charge, and the single O atom has about eight-tenths of a full negative charge.

connected network of molecules (Figure 2.7). Each hydrogen bond is formed when the partially positive-charged hydrogen of one water molecule becomes aligned next to a partially negative-charged oxygen atom of another water molecule. Because of their extensive hydrogen bonding, water molecules have an unusually strong tendency to adhere to one another. This feature is most evident in the thermal properties of water. For example, when water is heated, most of the thermal energy is consumed in disrupting hydrogen bonds rather than contributing to molecular motion (which is measured as an increased temperature). Similarly, evaporation from the liquid to the gaseous state requires that water molecules break the hydrogen bonds holding them to their neighbors, which is why it takes so much energy to convert water to steam. Mammals take advantage of this property when they sweat because the heat required to evaporate the water is absorbed from the body, which thus becomes cooler.

  The small volume of aqueous fluid present within a cell contains a remarkably complex mixture of dissolved substances, or *solutes*. In fact, water is able to dissolve more types of substances than any other solvent. But water is more than just a solvent; it determines the structure of biological molecules and the types of interactions in which they can engage. Water is the fluid matrix around which the insoluble fabric of the cell is constructed. It is also the medium through which materials move from one compartment of the cell to another; it is a reactant or product in many cellular reactions; and it protects the cell in many ways—from excessive heat, cold, or damaging radiation.

  Water is such an important factor in a cell because it is able to form weak interactions with so many different types of chemical groups. Recall from page 35 how water molecules, with their strongly polarized O—H bonds, form a shell around ions, separating the ions from one another. Similarly, water molecules form hydrogen bonds with organic molecules that contain polar groups, such as amino acids and sugars, which ensures their solubility within the cell. Water also plays
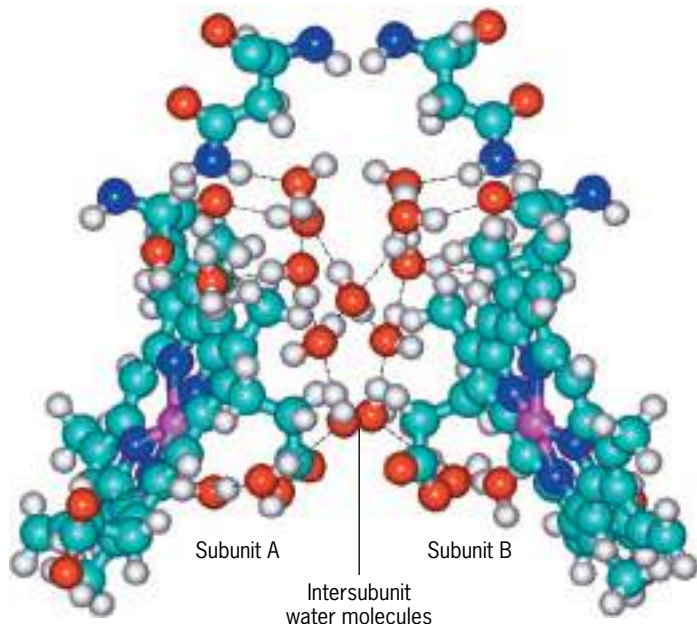
FIGURE 2.8 The importance of water in protein structure. The water molecules (each with a single red oxygen atom and two smaller gray hydrogen atoms) are shown in their ordered locations between the two subunits of a clam hemoglobin molecule. (FROM MARTIN CHAPLIN, NATURE REVS. MOL. CELL BIOL. 7:864, 2006, © COPYRIGHT 2006, BY MACMILLAN MAGAZINES LIMITED.)

a key role in maintaining the structure and function of macromolecules and the complexes that they form (such as membranes). Figure 2.8 shows the ordered arrangement of water molecules between two subunits of a protein molecule. The water molecules are hydrogen bonded to each other and to specific amino acids of the protein.

REVIEW

1. Describe some of the properties that distinguish covalent and noncovalent bonds.
2. Why do polar molecules, such as table sugar, dissolve so readily in water? Why do fat droplets form on the surface of an aqueous solution? Why does sweating help cool the body?

2.3 ACIDS, BASES, AND BUFFERS

Protons are not only found within atomic nuclei, they are also released into the medium whenever a hydrogen atom loses a shared electron. Consider acetic acid—the distinctive ingredient of vinegar—which can undergo the following reaction, described as a *dissociation*.

| Acetic | Acetate | Proton |
| acid | ion | (hydrogen ion) |
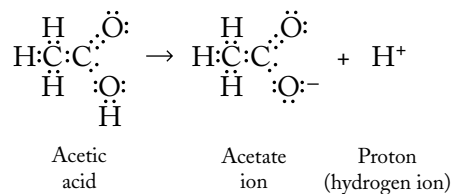
A molecule that is capable of releasing (donating) a hydrogen ion is termed an **acid**. The proton released by the acetic acid molecule in the previous reaction does not remain in the free state; instead, it combines with another molecule. Possible reactions involving a proton include

- Combination with a water molecule to form a hydronium ion (H_3O^+).

$$H^+ + H_2O \rightarrow H_3O^+$$
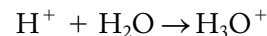
- Combination with a hydroxyl ion (OH^-) to form a molecule of water.

$$H^+ + OH^- \rightarrow H_2O$$

- Combination with an amino group ($-NH_2$) in a protein to form a charged amine.

$$H^+ + -NH_2 \rightarrow -NH_3^+$$

Any molecule that is capable of accepting a proton is defined as a **base**. Acids and bases exist in pairs, or *couples*. When the acid loses a proton (as when acetic acid gives up a hydrogen ion), it becomes a base (in this case, acetate ion), which is termed the *conjugate base* of the acid. Similarly, when a base (such as an $-NH_2$ group) accepts a proton, it forms an acid (in this case $-NH_3^+$), which is termed the *conjugate acid* of that base. Thus, the acid always contains one more positive charge than its conjugate base. Water is an example of an *amphoteric* molecule, that is, one that can serve both as an acid and a base:

$$\underset{\text{Acid}}{H_3O^+} \rightleftharpoons H^+ + \underset{\substack{\text{Amphoteric} \\ \text{molecule}}}{H_2O} \rightleftharpoons \underset{\text{Base}}{OH^-} + H^+$$
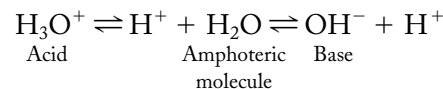
We will discuss another important group of amphoteric molecules, the amino acids, on page 49.

   Acids vary markedly with respect to the ease with which the molecule gives up a proton. The more readily the proton is lost, that is, the less strong the attraction of a conjugate base for its proton, the stronger the acid. Hydrogen chloride is a very strong acid, one that will readily transfer its proton to water molecules. The conjugate base of a strong acid, such as HCl, is a weak base (Table 2.1). Acetic acid, in contrast, is a relatively weak acid because for the most part it remains undissociated when dissolved in water. In a sense, one can consider the degree of dissociation of an acid in terms of the competition for protons among the components of a solution. Water is a better competitor, that is, a stronger base, than chloride ion, so HCl completely dissociates. In contrast, acetate ion is a stronger base than water, so it remains largely as undissociated acetic acid.

   The acidity of a solution is measured by the concentration of hydrogen ions[4] and is expressed in terms of **pH**.

$$pH = -\log[H^+]$$

where $[H^+]$ is the molar concentration of protons. For example, a solution having a pH of 5 contains a hydrogen ion concentration of 10^{-5} M. Because the pH scale is logarithmic, an increase of one pH unit corresponds to a tenfold decrease in

---

[4]In aqueous solutions, protons do not exist in the free state, but rather as H_3O^+ or $H_5O_2^+$. For the sake of simplicity, we will refer to them simply as protons or hydrogen ions.

TABLE 2.1 Strengths of Acids and Bases

	Acids	Bases	
Very weak	H_2O	OH^-	Strong
Weak	NH_4^+	NH_3	Weak
	H_2S	S^{2-}	
	CH_3COOH	CH_3COO^-	
	H_2CO_3	HCO_3^-	
Strong	H_3O^+	H_2O	Very
	HCl	Cl^-	weak
	H_2SO_4	SO_4^{2-}	

REVIEW ?

1. If you were to add hydrochloric acid to water, what effect would this have on the hydrogen ion concentration? on the pH? on the ionic charge of any proteins in solution?

2. What is the relationship between a base and its conjugate acid?

H^+ concentration (or a tenfold increase in OH^- concentration). Stomach juice (pH 1.8), for example, has nearly one million times the H^+ concentration of blood (pH 7.4).

When a water molecule dissociates into a hydroxyl ion and a proton, $H_2O \rightarrow H^+ + OH^-$, the equilibrium constant for the reaction can be expressed as:

$$K_{eq} = \frac{[H^+][OH^-]}{H_2O}$$

Because the concentration of pure water is always 55.51 M, we can generate a new constant, K_W, the *ion-product constant* for water,

$$K_w = [H^+][OH^-]$$

which is equal to 10^{-14} at 25°C. In pure water, the concentration of both H^+ and OH^- is approximately 10^{-7} M. The extremely low level of dissociation of water indicates that it is a very weak acid. In the presence of an acid, the concentration of hydrogen ions rises and the concentration of hydroxyl ions drops (as a result of combination with protons to form water), so that the ion product remains at 10^{-14}.

Most biological processes are acutely sensitive to pH because changes in hydrogen ion concentration affect the ionic state of biological molecules. For example, as the hydrogen ion concentration increases, the $-NH_2$ group of the amino acid arginine becomes protonated to form $-NH_3^+$, which can disrupt the activity of the entire protein. Even slight changes in pH can impede biological reactions. Organisms, and the cells they comprise, are protected from pH fluctuations by **buffers**—compounds that react with free hydrogen or hydroxyl ions, thereby resisting changes in pH. Buffer solutions usually contain a weak acid together with its conjugate base. Blood, for example, is buffered by carbonic acid and bicarbonate ions, which normally hold blood pH at about 7.4.

$$\underset{\substack{\text{Bicarbonate} \\ \text{ion}}}{HCO_3^-} + \underset{\substack{\text{Hydrogen} \\ \text{ion}}}{H^+} \rightleftharpoons \underset{\substack{\text{Carbonic} \\ \text{acid}}}{H_2CO_3}$$
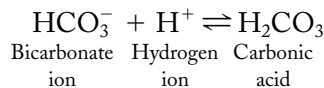
If the hydrogen ion concentration rises (as occurs during exercise), the bicarbonate ions combine with the excess protons, removing them from solution. Conversely, excess OH^- ions (which are generated during hyperventilation) are neutralized by protons derived from carbonic acid. The pH of the fluid within the cell is regulated in a similar manner by a phosphate buffer system consisting of $H_2PO_4^-$ and HPO_4^{2-}.

2.4 THE NATURE OF BIOLOGICAL MOLECULES

The bulk of an organism is water. If the water is evaporated away, most of the remaining dry weight consists of molecules containing atoms of carbon. When first discovered, it was thought that carbon-containing molecules were present only in living organisms and thus were referred to as *organic molecules* to distinguish them from *inorganic molecules* found in the inanimate world. As chemists learned to synthesize more and more of these carbon-containing molecules in the lab, the mystique associated with organic compounds disappeared. The compounds produced by living organisms are called **biochemicals**.

The chemistry of life centers around the chemistry of the carbon atom. The essential quality of carbon that has allowed it to play this role is the incredible number of molecules it can form. Having four outer-shell electrons, a carbon atom can bond with up to four other atoms. Most importantly, each carbon atom is able to bond with other carbon atoms so as to construct molecules with backbones containing long chains of carbon atoms. Carbon-containing backbones may be linear, branched, or cyclic.

Linear Cyclic Branched
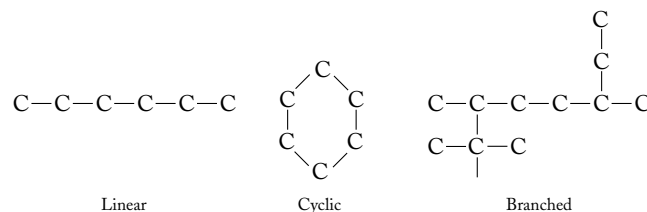
Cholesterol, whose structure is depicted in Figure 2.9, illustrates various arrangements of carbon atoms.
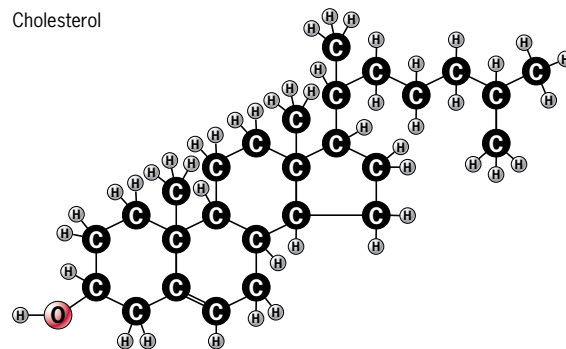
Cholesterol

FIGURE 2.9 Cholesterol, whose structure illustrates how carbon atoms (represented by the black balls) are able to form covalent bonds with as many as four other carbon atoms. As a result, carbon atoms can be linked together to form the backbones of a virtually unlimited variety of organic molecules. The carbon backbone of a cholesterol molecule includes four rings, which is characteristic of steroids (e.g., estrogen, testosterone, cortisol). The cholesterol molecule shown here is drawn as a ball-and-stick model, which is another way that molecular structure is depicted.

Both the size and electronic structure of carbon make it uniquely suited for generating large numbers of molecules, several hundred thousand of which are known. In contrast, silicon, which is just below carbon in the periodic table and also has four outer-shell electrons (see Figure 2.1), is too large for its positively charged nucleus to attract the outer-shell electrons of neighboring atoms with sufficient force to hold such large molecules together. We can best understand the nature of biological molecules by starting with the simplest group of organic molecules, the *hydrocarbons*, which contain only carbon and hydrogen atoms. The molecule ethane (C_2H_6) is a simple hydrocarbon

$$
\begin{array}{ccc}
 & H & H \\
 & | & | \\
H- & C- & C-H \\
 & | & | \\
 & H & H
\end{array}
$$

Ethane
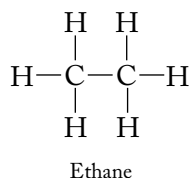
consisting of two atoms of carbon in which each carbon is bonded to the other carbon as well as to three atoms of hydrogen. As more carbons are added, the skeletons of organic molecules increase in length and their structure becomes more complex.

Functional Groups

Hydrocarbons do not occur in significant amounts within most living cells (though they constitute the bulk of the fossil fuels formed from the remains of ancient plants and animals). Many of the organic molecules that are important in biology contain chains of carbon atoms, like hydrocarbons, but certain hydrogen atoms are replaced by various **functional groups**. Functional groups are particular groupings of atoms that often behave as a unit and give organic molecules their physical properties, chemical reactivity, and solubility in aqueous solution. Some of the more common functional groups are listed in Table 2.2. Two of the most common linkages between functional groups are **ester bonds**, which form between carboxylic acids and alcohols, and **amide bonds**, which form between carboxylic acids and amines.

$$
\underset{\text{Acid}}{\overset{\overset{\text{O}}{\parallel}}{-C-OH}} + \underset{\text{Alcohol}}{HO-\overset{|}{\underset{|}{C}}-} \rightarrow \underset{\text{Ester}}{-\overset{\overset{\text{O}}{\parallel}}{C}-O-\overset{|}{\underset{|}{C}}-}
$$

$$
\underset{\text{Acid}}{\overset{\overset{\text{O}}{\parallel}}{-C-OH}} + \underset{\text{Amine}}{HN-\overset{|}{\underset{|}{C}}-} \rightarrow \underset{\text{Amide}}{-\overset{\overset{\text{O}}{\parallel}}{C}-N-\overset{|}{\underset{|}{C}}-}
$$
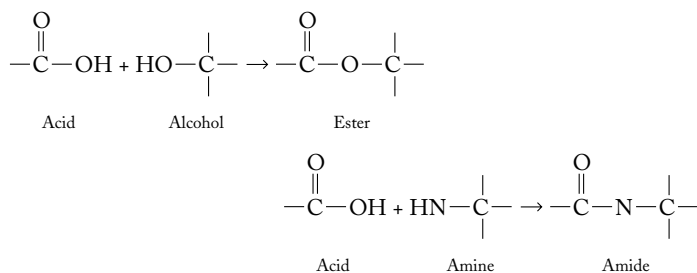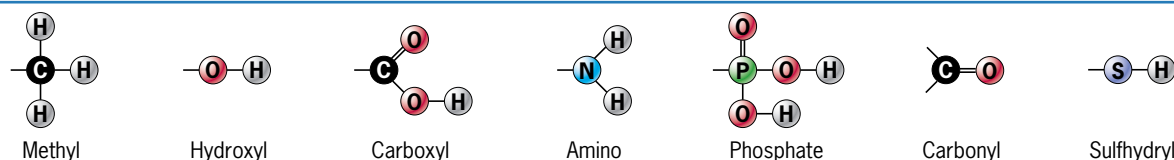
Most of the groups in Table 2.2 contain one or more electronegative atoms (N, P, O, and/or S) and make organic molecules more polar, more water soluble, and more reactive. Several of these functional groups can ionize and become positively or negatively charged. The effect on molecules by the substitution of various functional groups is readily demonstrated. The hydrocarbon ethane (CH_3CH_3) depicted above is a toxic, flammable gas. Replace one of the hydrogens with a hydroxyl group (—OH) and the molecule (CH_3CH_2OH) becomes palatable—it is ethyl alcohol. Substitute a carboxyl group (—COOH) and the molecule becomes acetic acid (CH_3COOH), the strong-tasting ingredient in vinegar. Substitute a sulfhydryl group (—SH), and you have formed CH_3CH_2SH, a strong, foul-smelling agent, ethyl mercaptan, used by biochemists in studying enzyme reactions.

A Classification of Biological Molecules by Function

The organic molecules commonly found within living cells can be divided into several categories based on their role in metabolism.

1. *Macromolecules.* The molecules that form the structure and carry out the activities of cells are huge, highly organized molecules called **macromolecules**, which contain anywhere from dozens to millions of carbon atoms. Because of their size and the intricate shapes that macromolecules can assume, some of these molecular giants can perform complex tasks with great precision and efficiency. The presence of macromolecules, more than any other characteristic, endows organisms with the properties of life and sets them apart chemically from the inanimate world.

    Macromolecules can be divided into four major categories: proteins, nucleic acids, polysaccharides, and certain lipids. The first three types are *polymers* composed of a large number of low-molecular-weight building blocks, or *monomers*. These macromolecules are constructed from monomers by a process of *polymerization* that resembles coupling railroad cars onto a train (Figure 2.10). The basic structure and function of each type of macromolecule are similar in all organisms. It is not until you look at the specific sequences of monomers that make up individual macromolecules that the diversity among organisms becomes apparent.

2. *The building blocks of macromolecules.* Most of the macromolecules within a cell have a short lifetime compared with the cell itself; with the exception of the cell's DNA, they are continually broken down and replaced by new

TABLE 2.2 Functional Groups

| Methyl | Hydroxyl | Carboxyl | Amino | Phosphate | Carbonyl | Sulfhydryl |
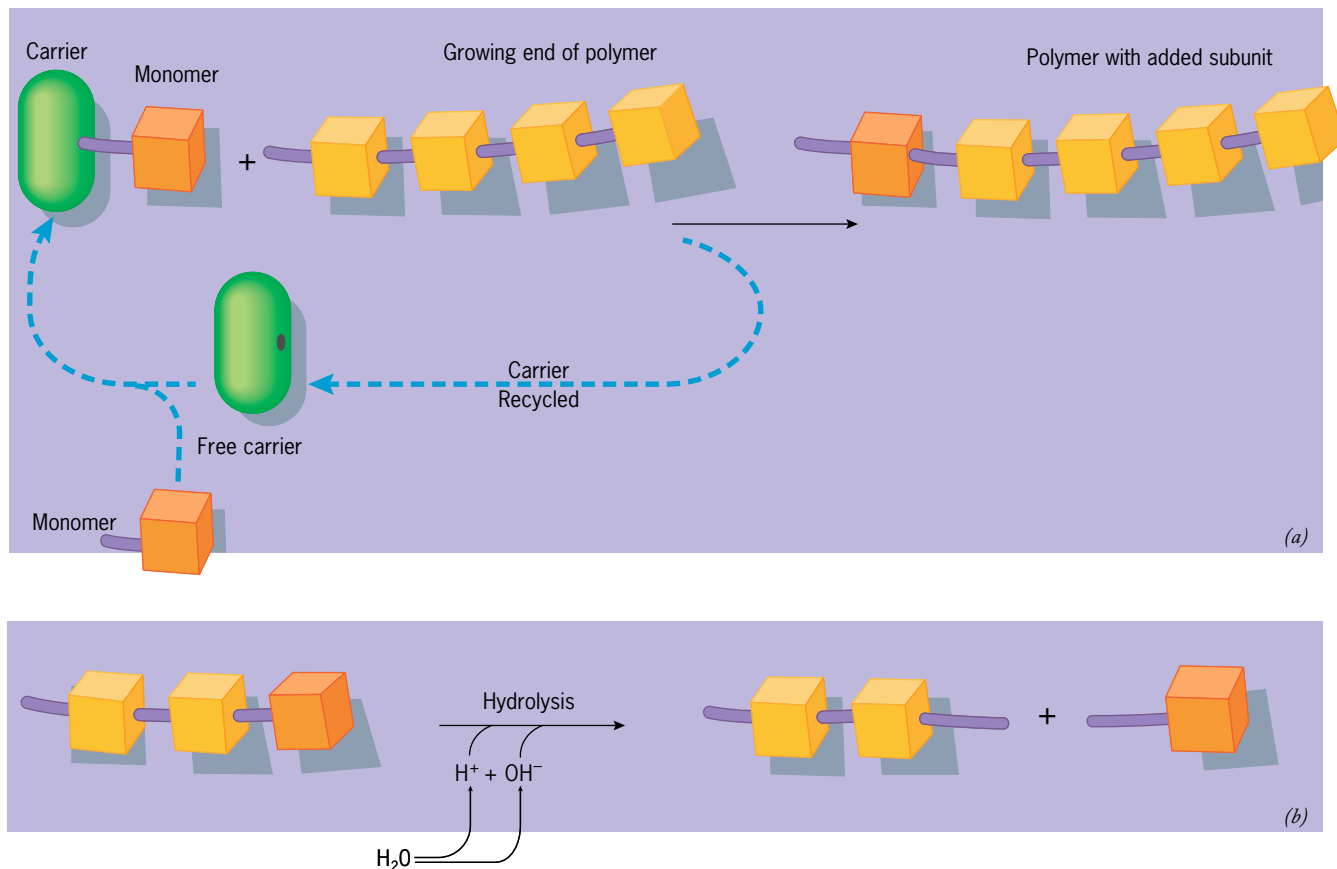
FIGURE 2.10 Monomers and polymers; polymerization and hydrolysis. (*a*) Polysaccharides, proteins, and nucleic acids consist of monomers (subunits) linked together by covalent bonds. Free monomers do not simply react with each other to become macromolecules. Rather, each monomer is first activated by attachment to a carrier molecule that subsequently transfers the monomer to the end of the growing macromolecule. (*b*) A macromolecule is disassembled by hydrolysis of the bonds that join the monomers together. Hydrolysis is the splitting of a bond by water. All of these reactions are catalyzed by specific enzymes.

macromolecules. Consequently, most cells contain a supply (or *pool*) of low-molecular-weight precursors that are ready to be incorporated into macromolecules. These include sugars, which are the precursors of polysaccharides; amino acids, which are the precursors of proteins; nucleotides, which are the precursors of nucleic acids; and fatty acids, which are incorporated into lipids.

3. *Metabolic intermediates* (*metabolites*). The molecules in a cell have complex chemical structures and must be synthesized in a step-by-step sequence beginning with specific starting materials. In the cell, each series of chemical reactions is termed a **metabolic pathway**. The cell starts with compound A and converts it to compound B, then to compound C, and so on, until some functional end product (such as an amino acid building block of a protein) is produced. The compounds formed along the pathways leading to the end products might have no function per se and are called **metabolic intermediates**.

4. *Molecules of miscellaneous function.* This is obviously a broad category of molecules but not as large as you might expect; the vast bulk of the dry weight of a cell is made up of macromolecules and their direct precursors. The mole-

cules of miscellaneous function include such substances as vitamins, which function primarily as adjuncts to proteins; certain steroid or amino acid hormones; molecules involved in energy storage, such as ATP; regulatory molecules such as cyclic AMP; and metabolic waste products such as urea.

REVIEW   **?**

1. What properties of a carbon atom are critical to life?
2. Draw the structures of four different functional groups. How would each of these groups alter the solubility of a molecule in water?

2.5 FOUR TYPES OF BIOLOGICAL MOLECULES

The macromolecules just described can be divided into four types of organic molecules: carbohydrates, lipids, proteins, and nucleic acids. The localization of these molecules in a number of cellular structures is shown in an overview in Figure 2.11.
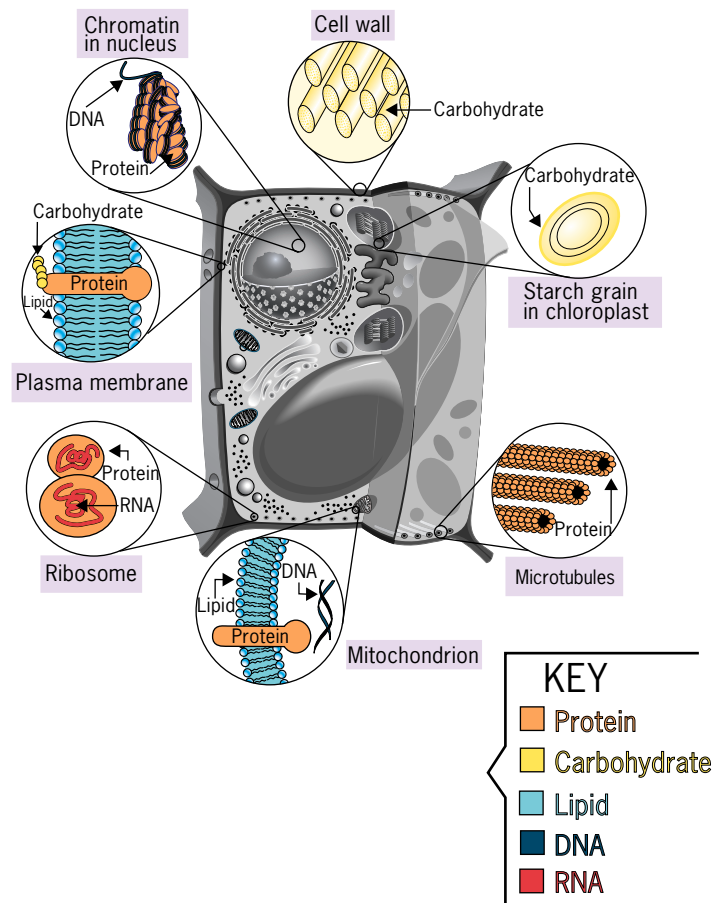
FIGURE 2.11 An overview of the types of biological molecules that make up various cellular structures.

Carbohydrates

Carbohydrates (or **glycans** as they are often called) include simple sugars (or *monosaccharides*) and all larger molecules constructed of sugar building blocks. Carbohydrates function primarily as stores of chemical energy and as durable building materials for biological construction. Most sugars have the general formula $(CH_2O)_n$. The sugars of importance in cellular metabolism have values of *n* that range from 3 to 7. Sugars containing three carbons are known as *trioses*, those with four carbons as *tetroses*, those with five carbons as *pentoses*, those with six carbons as *hexoses*, and those with seven carbons as *heptoses*.

The Structure of Simple Sugars Each sugar molecule consists of a backbone of carbon atoms linked together in a linear array by single bonds. Each of the carbon atoms of the backbone is linked to a single hydroxyl group, except for one that bears a *carbonyl* ($C = O$) group. If the carbonyl group is located at an internal position (to form a ketone group), the sugar is a *ketose*, such as fructose, which is shown in Figure 2.12a. If the carbonyl is located at one end of the sugar, it forms an aldehyde group and the molecule is known as an *aldose*, as exemplified by glucose, which is shown in Figure 2.12b–f. Because of their large numbers of hydroxyl groups, sugars tend to be highly water soluble.

Although the straight-chained formulas shown in Figure 2.12a,b are useful for comparing the structures of various sugars, they do not reflect the fact that sugars with five or more carbon atoms undergo a self-reaction (Figure 2.12c) that converts them into a closed, or ring-containing, molecule. The
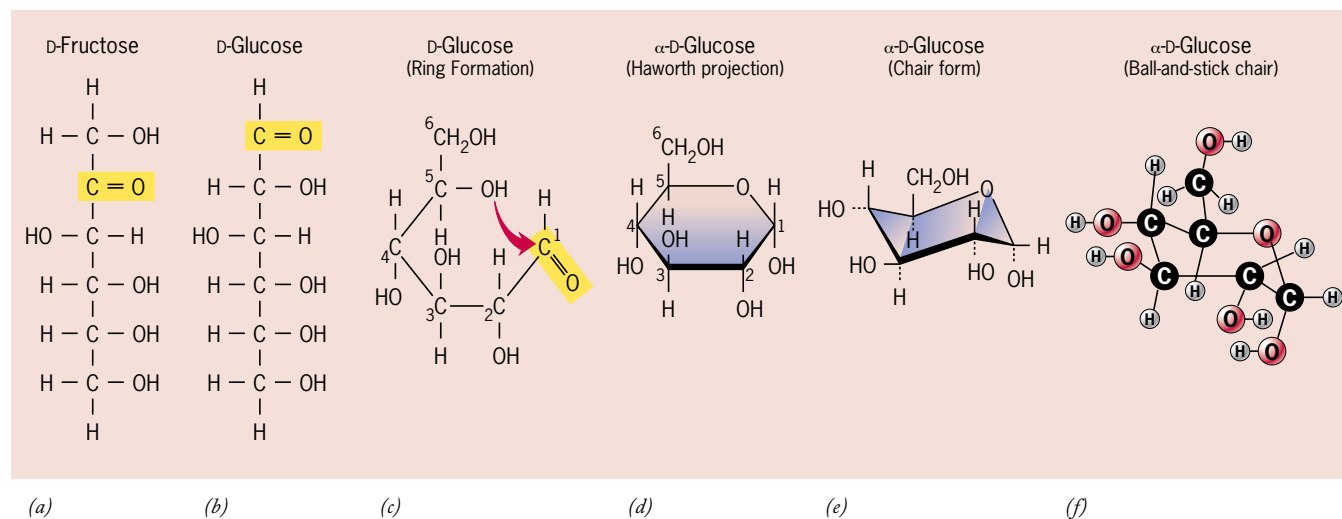
FIGURE 2.12 The structures of sugars. (*a*) Straight-chain formula of fructose, a ketohexose [keto, indicating the carbonyl (yellow), is located internally, and hexose because it consists of six carbons]. (*b*) Straight-chain formula of glucose, an aldohexose (aldo because the carbonyl is located at the end of the molecule). (*c*) Self-reaction in which glucose is converted from an open chain to a closed ring (a pyranose ring). (*d*) Glucose is commonly depicted in the form of a flat (planar) ring lying perpendicular to the page with the thickened line situated closest to the reader and the H and OH groups projecting either above or below the ring. The basis for the designation α-D-glucose is discussed in the following section. (*e*) The chair conformation of glucose, which depicts its three-dimensional structure more accurately than the flattened ring of part *d*. (*f*) A ball-and-stick model of the chair conformation of glucose.

ring forms of sugars are usually depicted as flat (*planar*) structures (Figure 2.12*d*) lying perpendicular to the plane of the paper with the thickened line situated closest to the reader. The H and OH groups lie parallel to the plane of the paper, projecting either above or below the ring of the sugar. In actual fact, the sugar ring is not a planar structure, but usually exists in a three-dimensional conformation resembling a chair (Figure 2.12*e,f*).

Stereoisomerism   As noted earlier, a carbon atom can bond with four other atoms. The arrangement of the groups around a carbon atom can be depicted as in Figure 2.13*a* with the carbon placed in the center of a tetrahedron and the bonded groups

(a)

(b)

(c)
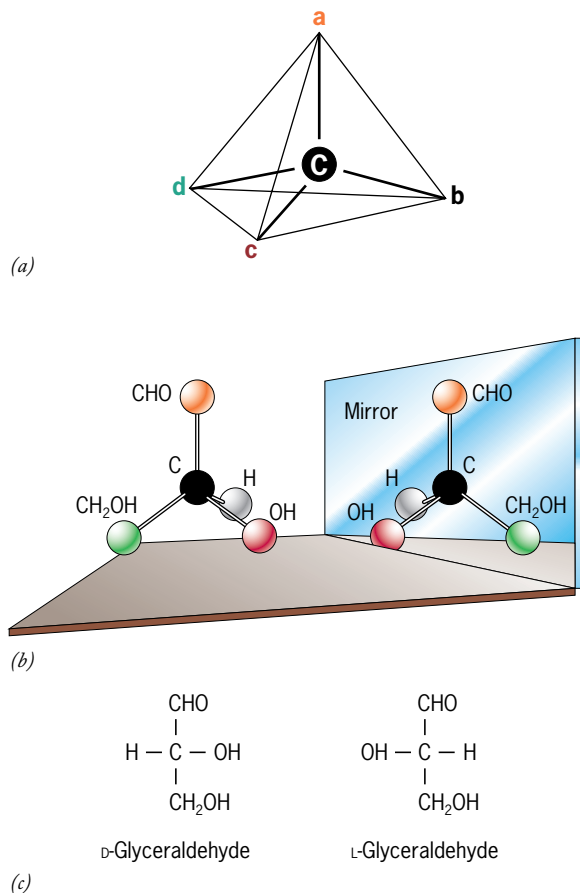
FIGURE 2.13 Stereoisomerism of glyceraldehyde. (*a*) The four groups bonded to a carbon atom (labeled a, b, c, and d) occupy the four corners of a tetrahedron with the carbon atom at its center. (*b*) Glyceraldehyde is the only three-carbon aldose; its second carbon atom is bonded to four different groups (—H, —OH, —CHO, and —CH₂OH). As a result, glyceraldehyde can exist in two possible configurations that are not superimposable, but instead are mirror images of each other as indicated. These two stereoisomers (or enantiomers) can be distinguished by the configuration of the four groups around the asymmetric (or *chiral*) carbon atom. Solutions of these two isomers rotate plane-polarized light in opposite directions and, thus, are said to be optically active. (*c*) Straight-chain formulas of glyceraldehyde. By convention, the D-isomer is shown with the OH group on the right.

projecting into its four corners. Figure 2.13*b* depicts a molecule of glyceraldehyde, which is the only aldotriose. The second carbon atom of glyceraldehyde is linked to four different groups (—H, —OH, —CHO, and —CH₂OH). If the four groups bonded to a carbon atom are all different, as in glyceraldehyde, then two possible configurations exist that cannot be superimposed on one another. These two molecules (termed *stereoisomers* or *enantiomers*) have essentially the same chemical reactivities, but their structures are mirror images (not unlike a pair of right and left human hands). By convention, the molecule is called D-glyceraldehyde if the hydroxyl group of carbon 2 projects to the right, and L-glyceraldehyde if it projects to the left (Figure 2.13*c*). Because it acts as a site of stereoisomerism, carbon 2 is referred to as an *asymmetric* carbon atom.

As the backbone of sugar molecules increases in length, so too does the number of asymmetric carbon atoms and, consequently, the number of stereoisomers. Aldotetroses have two asymmetric carbons and thus can exist in four different configurations (Figure 2.14). Similarly, there are 8 different aldopentoses and 16 different aldohexoses. The designation of each of these sugars as D or L is based by convention on the arrangement of groups attached to the asymmetric carbon atom farthest from the aldehyde (the carbon associated with the aldehyde is designated C1). If the hydroxyl group of this carbon projects to the right, the aldose is a D-sugar; if it projects to the left, it is an L-sugar. The enzymes present in living cells can distinguish between the D and L forms of a sugar. Typically, only one of the stereoisomers (such as D-glucose and L-fucose) is used by cells.

The self-reaction in which a straight-chain glucose molecule is converted into a six-membered (*pyranose*) ring was shown in Figure 2.12*c*. Unlike its precursor in the open chain, the C1 of the ring bears four different groups and thus becomes a new center of asymmetry within the sugar molecule. Because of this extra asymmetric carbon atom, each type of pyranose exists as α and β stereoisomers (Figure 2.15). By convention, the molecule is an α-pyranose when the OH group of the first carbon projects below the plane of the ring, and a β-pyranose when the hydroxyl projects upward. The difference between the two forms has important biological consequences, resulting, for example, in the compact shape of glycogen and starch molecules and the extended conformation of cellulose (discussed later).
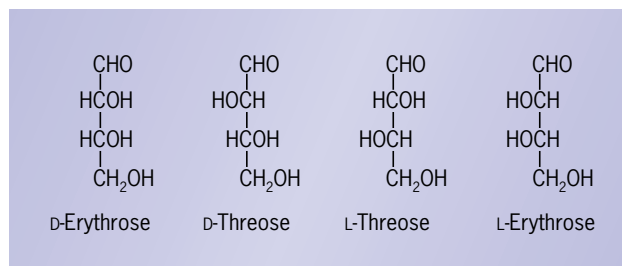
FIGURE 2.14 Aldotetroses. Because they have two asymmetric carbon atoms, aldotetroses can exist in four configurations.
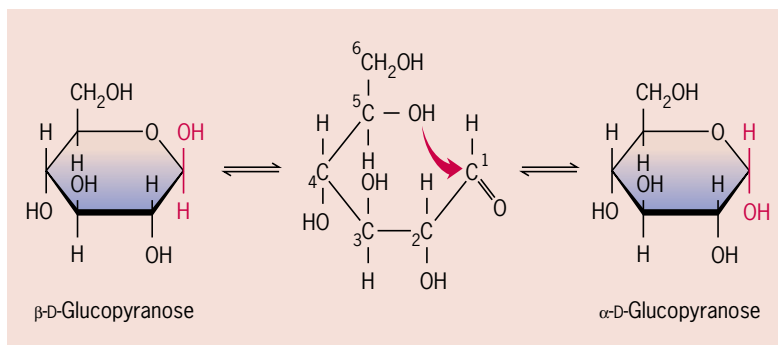
FIGURE 2.15 Formation of an α- and β-pyranose. When a molecule of glucose undergoes self-reaction to form a pyranose ring (i.e., a six-membered ring), two stereoisomers are generated. The two isomers are in equilibrium with each other through the open-chain form of the molecule. By convention, the molecule is an α-pyranose when the OH group of the first carbon projects below the plane of the ring, and a β-pyranose when the hydroxyl group projects upward.

Linking Sugars Together   Sugars can be joined to one another by covalent **glycosidic bonds** to form larger molecules. Glycosidic bonds form by reaction between carbon atom C1 of one sugar and the hydroxyl group of another sugar, generating a —C—O—C— linkage between the two sugars. As discussed below (and indicated in Figures 2.16 and 2.17), sugars can be joined by quite a variety of different glycosidic bonds. Molecules composed of only two sugar units are *disaccharides* (Figure 2.16). Disaccharides serve primarily as readily available energy stores. Sucrose, or table sugar, is a major component of plant sap, which carries chemical energy from one part of the plant to another. Lactose, present in the milk of most mammals, supplies newborn mammals with fuel for early growth and development. Lactose in the diet is hy-drolyzed by the enzyme lactase, which is present in the plasma membranes of the cells that line the intestine. Many people lose this enzyme after childhood and find that eating dairy products causes digestive discomfort.

Sugars may also be linked together to form small chains called **oligosaccharides** (*oligo* = few). Most often such chains are found covalently attached to lipids and proteins, converting them into glycolipids and glycoproteins, respectively. Oligosaccharides are particularly important on the glycolipids and glycoproteins of the plasma membrane, where they project from the cell surface (see Figure 4.4c). Because oligosaccharides may be composed of many different combinations of sugar units, these carbohydrates can play an informational role; that is, they can serve to distinguish one type of cell from another and help mediate specific interactions of a cell with its surroundings.

Polysaccharides   By the middle of the nineteenth century, it was known that the blood of people suffering from diabetes had a sweet taste due to an elevated level of glucose, the key sugar in energy metabolism. Claude Bernard, a prominent French physiologist of the period, was looking for the cause of diabetes by investigating the source of blood sugar. It was assumed at the time that any sugar present in a human or an animal had to have been previously consumed in the diet. Working with dogs, Bernard found that, even if the animals were placed on a diet totally lacking carbohydrates, their blood still contained a normal amount of glucose. Clearly, glucose could be formed in the body from other types of compounds.

After further investigation, Bernard found that glucose enters the blood from the liver. Liver tissue, he found, contains an insoluble polymer of glucose he named **glycogen**. Bernard concluded that various food materials (such as proteins) were carried to the liver where they were chemically converted to glucose and stored as glycogen. Then, as the body needed sugar for fuel, the glycogen in the liver was transformed to glucose, which was released into the bloodstream to satisfy glucose-depleted tissues. In Bernard's hypothesis, the balance between glycogen formation and glycogen breakdown in the liver was the prime determinant in maintaining the relatively constant (*homeostatic*) level of glucose in the blood.

FIGURE 2.16 Disaccharides. Sucrose and lactose are two of the most common disaccharides. Sucrose is composed of glucose and fructose joined by an α(1 → 2) linkage, whereas lactose is composed of glucose and galactose joined by a β(1 → 4) linkage.
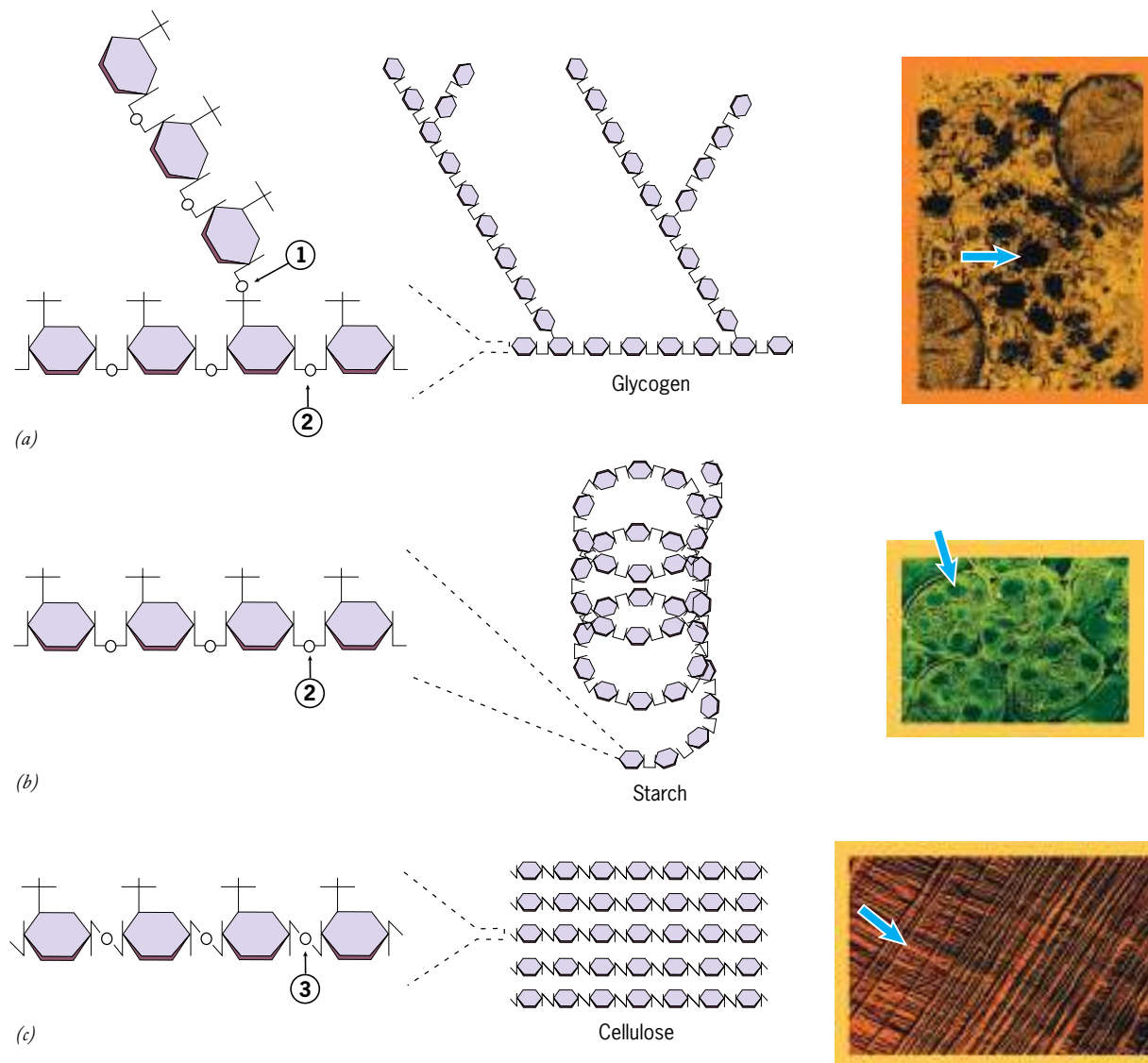
FIGURE 2.17 Three polysaccharides with identical sugar monomers but dramatically different properties. Glycogen (*a*), starch (*b*), and cellulose (*c*) are each composed entirely of glucose subunits, yet their chemical and physical properties are very different due to the distinct ways that the monomers are linked together (three different types of linkages are indicated by the circled numbers). Glycogen molecules are the most highly branched, starch molecules assume a helical arrangement, and cellulose molecules are unbranched and highly extended. Whereas glycogen and starch are energy stores, cellulose molecules are bundled together into tough fibers that are suited for their structural role. Colorized electron micrographs show glycogen granules in a liver cell, starch grains (amyloplasts) in a plant seed, and cellulose fibers in a plant cell wall; each is indicated by an arrow. [PHOTO INSETS: (TOP) DON FAWCETT/VISUALS UNLIMITED; (CENTER) JEREMY BURGESS/PHOTO RESEARCHERS; (BOTTOM) CABISCO/VISUALS UNLIMITED.]

Bernard's hypothesis proved to be correct. The molecule he named glycogen is a type of **polysaccharide**—a polymer of sugar units joined by glycosidic bonds.

Glycogen and Starch: Nutritional Polysaccharides   Glycogen is a branched polymer containing only one type of monomer: glucose (Figure 2.17*a*). Most of the sugar units of a glycogen molecule are joined to one another by $\alpha(1 \to 4)$ glycosidic bonds (type 2 bond in Figure 2.17*a*). Branch points contain a sugar joined to three neighboring units rather than to two, as in the unbranched segments of the polymer. The extra neighbor, which forms the branch, is linked by an $\alpha(1 \to 6)$ glycosidic bond (type 1 bond in Figure 2.17*a*).

Glycogen serves as a storehouse of surplus chemical energy in most animals. Human skeletal muscles, for example, typically contain enough glycogen to fuel about 30 minutes of moderate activity. Depending on various factors, glycogen typically ranges in molecular weight from about one to four million daltons. When stored in cells, glycogen is highly concentrated in what appears as dark-staining, irregular granules in electron micrographs (Figure 2.17*a*, right).

Most plants bank their surplus chemical energy in the form of **starch**, which like glycogen is also a polymer of glucose. Potatoes and cereals, for example, consist primarily of starch. Starch is actually a mixture of two different polymers, amylose and amylopectin. Amylose is an unbranched, helical

molecule whose sugars are joined by $\alpha(1 \rightarrow 4)$ linkages (Figure 2.17*b*), whereas amylopectin is branched. Amylopectin differs from glycogen in being much less branched and having an irregular branching pattern. Starch is stored as densely packed granules, or *starch grains*, which are enclosed in membrane-bound organelles (*plastids*) within the plant cell (Figure 2.17*b*, right). Although animals don't synthesize starch, they possess an enzyme (*amylase*) that readily hydrolyzes it.

Cellulose, Chitin, and Glycosaminoglycans: Structural Polysaccharides   Whereas some polysaccharides constitute easily digested energy stores, others form tough, durable structural materials. Cotton and linen, for example, consist largely of **cellulose**, which is the major component of plant cell walls. Cotton textiles owe their durability to the long, unbranched cellulose molecules, which are ordered into side-by-side aggregates to form molecular cables (photo, Figure 2.17*c*) that are ideally constructed to resist pulling (tensile) forces. Like glycogen and starch, cellulose consists solely of glucose monomers; its properties differ dramatically from these other polysaccharides because the glucose units are joined by $\beta(1 \rightarrow 4)$ linkages (bond 3 in Figure 2.17*c*) rather than $\alpha(1 \rightarrow 4)$ linkages. Ironically, multicellular animals (with rare exception) lack the enzyme needed to degrade cellulose, which happens to be the most abundant organic material on Earth and rich in chemical energy. Animals that "make a living" by digesting cellulose, such as termites and sheep, do so by harboring bacteria and protozoa that synthesize the necessary enzyme, cellulase.

Not all biological polysaccharides consist of glucose monomers. *Chitin* is an unbranched polymer of the sugar *N*-acetylglucosamine, which is similar in structure to glucose but has an acetyl amino group instead of a hydroxyl group bonded to the second carbon atom of the ring.

N-Acetylglucosamine
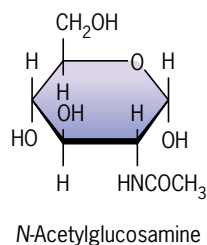
Chitin occurs widely as a structural material among invertebrates, particularly in the outer covering of insects, spiders, and crustaceans. Chitin is a tough, resilient, yet flexible material not unlike certain plastics. Insects owe much of their success to this highly adaptive polysaccharide (Figure 2.18).

Another group of polysaccharides that has a more complex structure is the **glycosaminoglycans** (or **GAGs**). Unlike other polysaccharides, they have the structure —A—B—A—B—, where A and B represent two different sugars. The best-studied GAG is heparin, which is secreted by cells in the lungs and other tissues in response to tissue injury. Heparin inhibits blood coagulation, thereby preventing the formation of clots that can block the flow of blood to the heart or lungs. Heparin accomplishes this feat by activating an inhibitor (antithrombin) of a key enzyme (thrombin) that is

FIGURE 2.18 **Chitin is the primary component of the glistening outer skeleton of this grasshopper.** (FROM ROBERT AND LINDA MITCHELL.)

required for blood coagulation. Heparin, which is normally extracted from pig tissue, has been used for decades to prevent blood clots in patients following major surgery. Unlike heparin, most GAGs are found in the spaces that surround cells, and their structure and function are discussed in Section 7.1. The most complex polysaccharides are found in plant cell walls (Section 7.6).

Lipids

Lipids are a diverse group of nonpolar biological molecules whose common properties are their ability to dissolve in organic solvents, such as chloroform or benzene, and their inability to dissolve in water—a property that explains many of their varied biological functions. Lipids of importance in cellular function include fats, steroids, and phospholipids.

Fats   **Fats** consist of a glycerol molecule linked by ester bonds to three fatty acids; the composite molecule is termed a **triacylglycerol** (Figure 2.19*a*). We will begin by considering the structure of **fatty acids**. Fatty acids are long, unbranched hydrocarbon chains with a single carboxyl group at one end (Figure 2.19*b*). Because the two ends of a fatty acid molecule have a very different structure, they also have different properties. The hydrocarbon chain is hydrophobic, whereas the carboxyl group (—COOH), which bears a negative

(a)

(b)

Stearic acid

Tristearate

(c)

Linseed oil

(d)
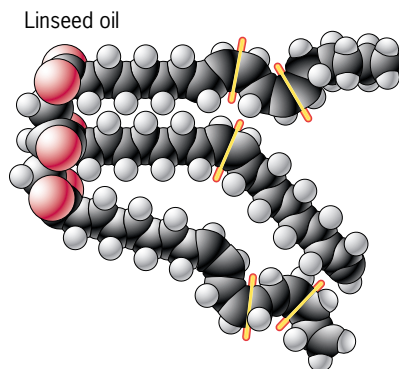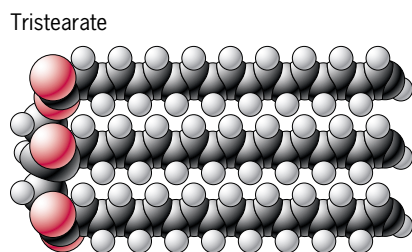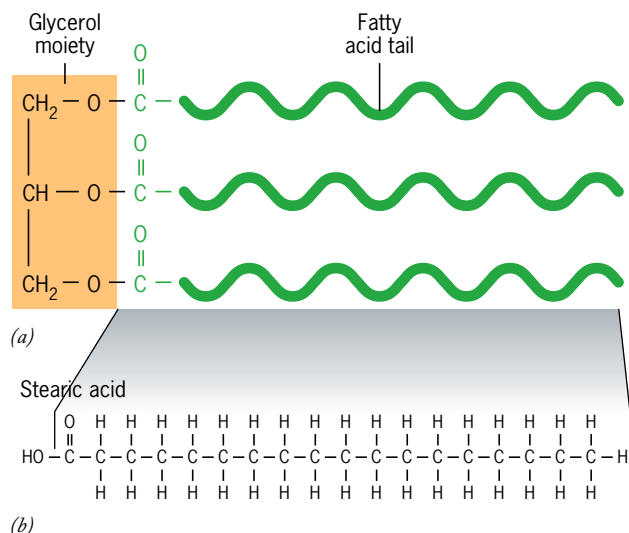
FIGURE 2.19 Fats and fatty acids. (*a*) The basic structure of a triacyl-glycerol (also called a triglyceride or a neutral fat). The glycerol moiety, indicated in orange, is linked by three ester bonds to the carboxyl groups of three fatty acids whose tails are indicated in green. (*b*) Stearic acid, an 18-carbon saturated fatty acid that is common in animal fats. (*c*) Space-filling model of tristearate, a triacylglycerol containing three identical stearic acid chains. (*d*) Space-filling model of linseed oil, a triacylglycerol derived from flax seeds that contains three unsaturated fatty acids (linoleic, oleic, and linolenic acids). The sites of unsaturation, which pro-duce kinks in the molecule, are indicated by the yellow-orange bars.
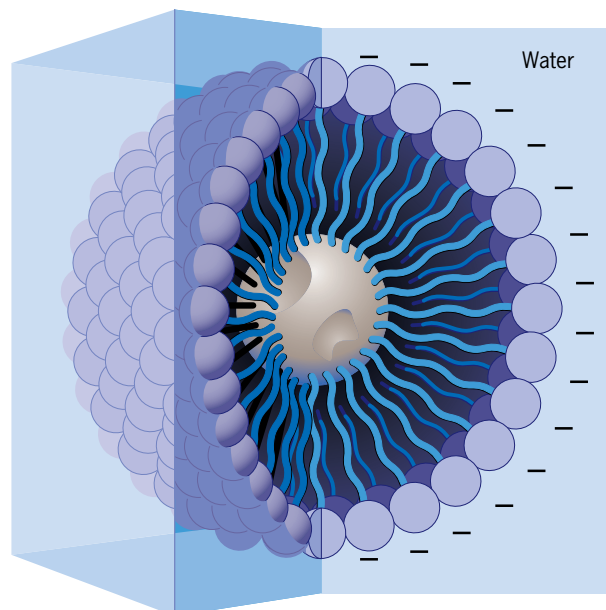
FIGURE 2.20 Soaps consist of fatty acids. In this schematic drawing of a soap micelle, the nonpolar tails of the fatty acids are directed inward, where they interact with the greasy matter to be dissolved. The negatively charged heads are located at the surface of the micelle, where they interact with the surrounding water. Membrane proteins, which also tend to be insoluble in water, can also be solubilized in this way by extraction of membranes with detergents.

which consists of fatty acids. In past centuries, soaps were made by heating animal fat in strong alkali (NaOH or KOH) to break the bonds between the fatty acids and the glycerol. Today, most soaps are made synthetically. Soaps owe their grease-dissolving capability to the fact that the hydrophobic end of each fatty acid can embed itself in the grease, whereas the hydrophilic end can interact with the surrounding water. As a result, greasy materials are converted into complexes (*micelles*) that can be dispersed by water (Figure 2.20).

Fatty acids differ from one another in the length of their hydrocarbon chain and the presence or absence of double bonds. Fatty acids present in cells typically vary in length from 14 to 20 carbons. Fatty acids that lack double bonds, such as stearic acid (Figure 2.19*b*), are described as **saturated**; those possessing double bonds are **unsaturated**. Naturally occurring fatty acids have double bonds in the *cis* configuration. Double bonds (of the *cis* configuration)

cis   as opposed to   *trans*
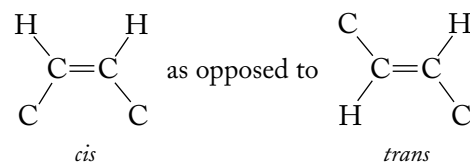
produce kinks in a fatty acid chain. Consequently, the more double bonds that fatty acid chains possess, the less effectively these long chains can be packed together. This lowers the temperature at which a fatty acid-containing lipid melts. Tri-stearate, whose fatty acids lack double bonds (Figure 2.19*c*), is a common component of animal fats and remains in a solid

charge at physiological pH, is hydrophilic. Molecules having both hydrophobic and hydrophilic regions are said to be **amphipathic**; such molecules have unusual and biologically important properties. The properties of fatty acids can be appreciated by considering the use of a familiar product: soap,

state well above room temperature. In contrast, the profusion of double bonds in vegetable fats accounts for their liquid state—both in the plant cell and on the grocery shelf—and for their being labeled as "polyunsaturated." Fats that are liquid at room temperature are described as **oils**. Figure 2.19*d* shows the structure of linseed oil, a highly volatile lipid extracted from flax seeds, that remains a liquid at a much lower temperature than does tristearate. Solid shortenings, such as margarine, are formed from unsaturated vegetable oils by chemically reducing the double bonds with hydrogen atoms (a process termed *hydrogenation*). The hydrogenation process also converts some of the *cis* double bonds into *trans* double bonds, which are straight rather than kinked. This process generates partially hydrogenated or trans-fats.

A molecule of fat can contain three identical fatty acids (as in Figure 2.19*c*), or it can be a *mixed fat*, containing more than one fatty acid species (as in Figure 2.19*d*). Most natural fats, such as olive oil or butterfat, are mixtures of molecules having different fatty acid species.

Fats are very rich in chemical energy; a gram of fat contains over twice the energy content of a gram of carbohydrate (for reasons discussed in Section 3.1). Carbohydrates function primarily as a short-term, rapidly available energy source, whereas fat reserves store energy on a long-term basis. It is estimated that a person of average size contains about 0.5 kilograms (kg) of carbohydrate, primarily in the form of glycogen. This amount of carbohydrate provides approximately 2000 kcal of total energy. During the course of a strenuous day's exercise, a person can virtually deplete his or her body's entire store of carbohydrate. In contrast, the average person contains approximately 16 kg of fat (equivalent to 144,000 kcal of energy), and as we all know, it can take a very long time to deplete our store of this material.

Because they lack polar groups, fats are extremely insoluble in water and are stored in cells in the form of dry lipid droplets. Since lipid droplets do not contain water as do glycogen granules, they represent an extremely concentrated storage fuel. In many animals, fats are stored in special cells (*adipocytes*) whose cytoplasm is filled with one or a few large lipid droplets. Adipocytes exhibit a remarkable ability to change their volume to accommodate varying quantities of fat.

Steroids   Steroids are built around a characteristic four-ringed hydrocarbon skeleton. One of the most important steroids is *cholesterol*, a component of animal cell membranes and a precursor for the synthesis of a number of steroid hormones, such as testosterone, progesterone, and estrogen (Figure 2.21). Cholesterol is largely absent from plant cells, which is why vegetable oils are considered "cholesterol-free," but plant cells may contain large quantities of related compounds.

Phospholipids   The chemical structure of a common phospholipid is shown in Figure 2.22. The molecule resembles a fat (triacylglycerol), but has only two fatty acid chains rather than three; it is a *diacylglycerol*. The third hydroxyl of the glycerol backbone is covalently bonded to a phosphate group, which in turn is covalently bonded to a small polar group, such
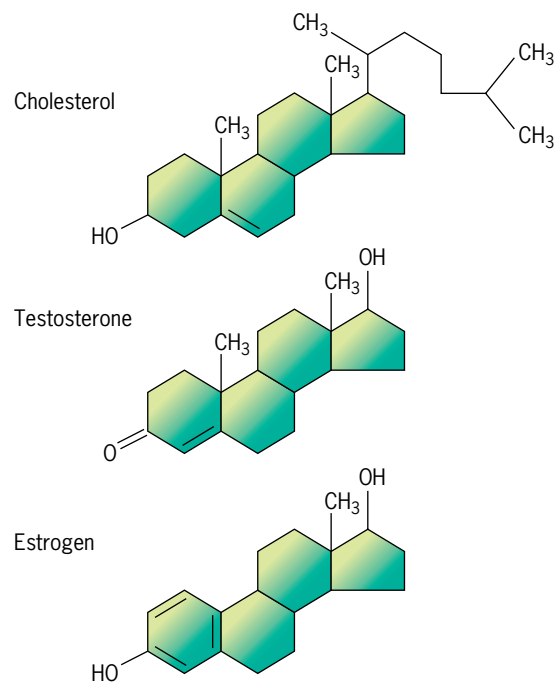
FIGURE 2.21 **The structure of steroids.** All steroids share the basic four-ring skeleton. The seemingly minor differences in chemical structure between cholesterol, testosterone, and estrogen generate profound biological differences.

as choline, as shown in Figure 2.22. Thus, unlike fat molecules, phospholipids contain two ends that have very different properties: the end containing the phosphate group has a distinctly hydrophilic character; the other end composed of the two fatty acid tails has a distinctly hydrophobic character. Because phospholipids function primarily in cell membranes,
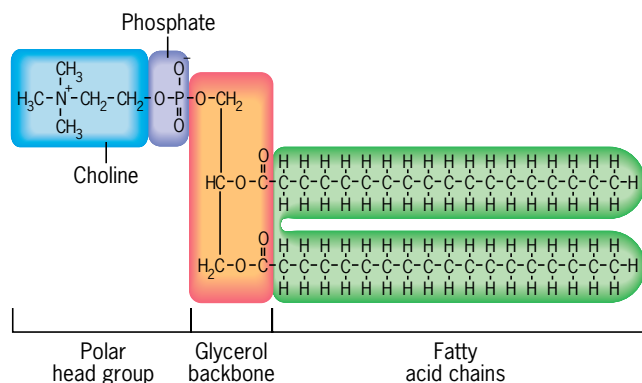
FIGURE 2.22 **The phospholipid phosphatidylcholine.** The molecule consists of a glycerol backbone whose hydroxyl groups are covalently bonded to two fatty acids and a phosphate group. The negatively charged phosphate is also bonded to a small, positively charged choline group. The end of the molecule that contains the phosphorylcholine is hydrophilic, whereas the opposite end, consisting of the fatty acid tail, is hydrophobic. The structure and function of phosphatidylcholine and other phospholipids are discussed at length in Section 4.3.

and because the properties of cell membranes depend on their phospholipid components, they will be discussed further in Sections 4.3 and 15.2 in connection with cell membranes.

Proteins

Proteins are the macromolecules that carry out virtually all of a cell's activities; they are the molecular tools and machines that make things happen. As enzymes, proteins vastly accelerate the rate of metabolic reactions; as structural cables, proteins provide mechanical support both within cells and outside their perimeters (Figure 2.23*a*); as hormones, growth factors, and gene activators, proteins perform a wide variety of regulatory functions; as membrane receptors and transporters, proteins determine what a cell reacts to and what types of substances enter or leave the cell; as contractile filaments and molecular motors, proteins constitute the machinery for biological movements. Among their many other functions, proteins act as antibodies, serve as toxins, form blood clots, absorb or refract light (Figure 2.23*b*), and transport substances from one part of the body to another.

How can one type of molecule have so many varied functions? The explanation resides in the virtually unlimited molecular structures that proteins, *as a group*, can assume. Each protein, however, has a unique and defined structure that enables it to carry out a particular function. Most importantly, proteins have shapes and surfaces that allow them to interact *selectively* with other molecules. Proteins, in other words, exhibit a high degree of **specificity**. It is possible, for example, for a particular DNA-cutting enzyme to recognize a segment of DNA containing one specific sequence of eight nucleotides, while ignoring all the other 65,535 possible sequences composed of this number of nucleotides.

The Building Blocks of Proteins Proteins are polymers made of amino acid monomers. Each protein has a unique sequence of amino acids that gives the molecule its unique properties. Many of the capabilities of a protein can be understood by examining the chemical properties of its constituent amino acids. Twenty different amino acids are commonly used in the construction of proteins, whether from a virus or a human. There are two aspects of amino acid structure to consider: that which is common to all of them and that which is unique to each. We will begin with the shared properties.

The Structures of Amino Acids All amino acids have a carboxyl group and an amino group, which are separated from each other by a single carbon atom, the α-carbon (Figure 2.24*a,b*). In a neutral aqueous solution, the α-carboxyl group loses its proton and exists in a negatively charged state ($-COO^-$), and the α-amino group accepts a proton and exists in a positively charged state (NH_3^+) (Figure 2.24*b*). We saw on page 43 that carbon atoms bonded to four different groups can exist in two configurations (*stereoisomers*) that cannot be superimposed on one another. Amino acids also have asymmetric carbon atoms. With the exception of glycine, the α-carbon of amino acids bonds to four different groups so that each amino acid can exist in either a D or an L form (Figure 2.25). Amino acids used in the synthesis of a protein on a ribosome are always L-amino acids. The "selection" of L-amino acids must have occurred very early in cellular evolution and has been conserved for billions of years. Microorgan-

(a)

(b)

FIGURE 2.23 **Two examples of the thousands of biological structures composed predominantly of protein.** These include (*a*) feathers, which are adaptations in birds for thermal insulation, flight, and sex recognition; and (*b*) the lenses of eyes, as in this net-casting spider, which focus light rays. (A: Darrell Gulin/Getty Images; B: Mantis Wildlife Films/Oxford Scientific Films/Animals Animals.)

(a)

Side Chain

Amino group   Carboxyl group

(b)

H_2O

Peptide bond

(c)
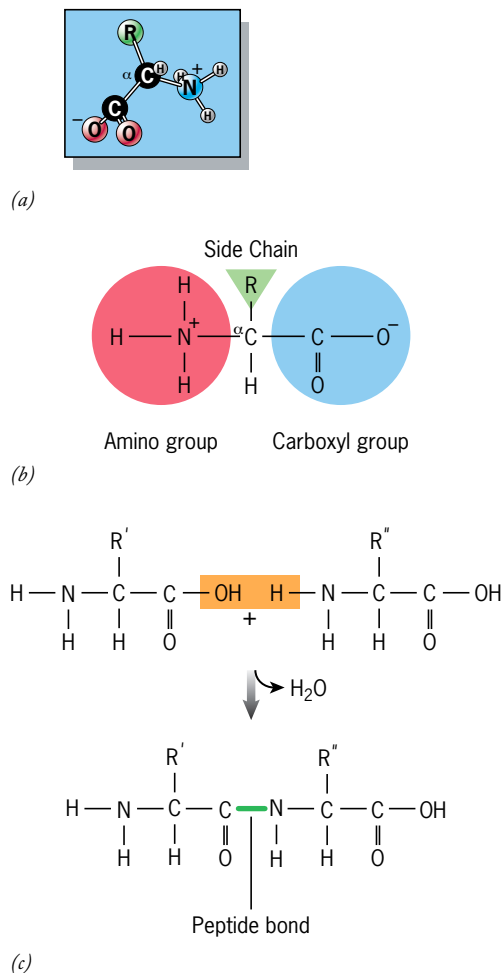
FIGURE 2.24 Amino acid structure. Ball-and-stick model (*a*) and chemical formula (*b*) of a generalized amino acid in which R can be any of a number of chemical groups (see Figure 2.26). (*c*) The formation of a peptide bond occurs by the condensation of two amino acids, drawn here in the uncharged state. In the cell, this reaction occurs on a ribosome as an amino acid is transferred from a carrier (a tRNA molecule) onto the end of the growing polypeptide chain (see Figure 11.49).
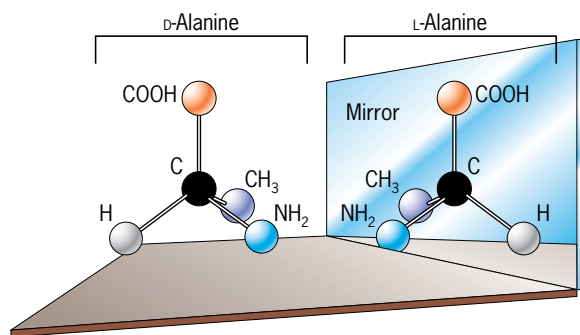
FIGURE 2.25 Amino acid stereoisomerism. Because the α-carbon of all amino acids except glycine is bonded to four different groups, two stereoisomers can exist. The D and L forms of alanine are shown.

isms, however, use D-amino acids in the synthesis of certain small peptides, including those of the cell wall and several antibiotics (e.g., gramicidin A).

During the process of protein synthesis, each amino acid becomes joined to two other amino acids, forming a long, continuous, unbranched polymer called a **polypeptide chain**. The amino acids that make up a polypeptide chain are joined by **peptide bonds** that result from the linkage of the carboxyl group of one amino acid to the amino group of its neighbor, with the elimination of a molecule of water (Figure 2.24*c*). A polypeptide chain composed of a string of amino acids joined by peptide bonds has the following backbone:

Peptide bond
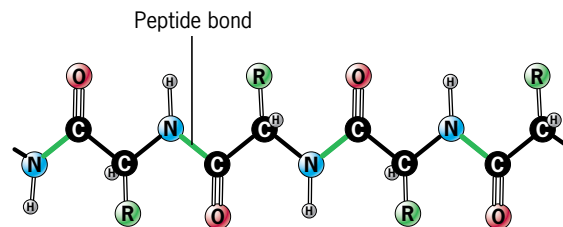
The "average" polypeptide chain contains about 450 amino acids. The longest known polypeptide, found in the muscle protein titin, contains more than 30,000 amino acids. Once incorporated into a polypeptide chain, amino acids are termed *residues*. The residue on one end of the chain, the *N-terminus*, contains an amino acid with a free (unbonded) α-amino group, whereas the residue at the opposite end, the *C-terminus*, has a free α-carboxyl group. In addition to amino acids, many proteins contain other types of components that are added after the polypeptide is synthesized. These include carbohydrates (to form glycoproteins), metal-containing groups (to form metalloproteins) and organic groups (e.g., flavoproteins).

The Properties of the Side Chains   The backbone, or main chain, of the polypeptide is composed of that part of each amino acid that is common to all of them. The **side chain** or **R group** (Figure 2.24), bonded to the α-carbon, is highly variable among the 20 building blocks, and it is this variability that ultimately gives proteins their diverse structures and activities. If the various amino acid side chains are considered together, they exhibit a large variety of structural features, ranging from fully charged to hydrophobic, and they can participate in a wide variety of covalent and noncovalent bonds. As discussed in the following chapter, the side chains of the "active sites" of enzymes can facilitate (catalyze) many different organic reactions. The assorted characteristics of the side chains of the amino acids are important in both *intra*molecular interactions, which determine the structure and activity of the molecule, and *inter*molecular interactions, which determine the relationship of a polypeptide with other molecules, including other polypeptides (page 59).

Amino acids are classified on the character of their side chains. They fall roughly into four categories: polar and charged, polar and uncharged, nonpolar, and those with unique properties (Figure 2.26).

1. *Polar, charged.* Amino acids of this group include aspartic acid, glutamic acid, lysine, and arginine. These four amino acids contain side chains that become fully charged; that

Polar charged

Aspartic acid
(Asp or D)

Glutamic acid
(Glu or E)

Lysine
(Lys or K)

Arginine
(Arg or R)

Histidine
(His or H)

Properties of side chains (R groups):

Hydrophilic side chains act as acids or bases which tend to be fully charged (+ or −) under physiologic conditions.
Side chains form ionic bonds and are often involved in chemical reactions.

Polar uncharged

Serine
(Ser or S)

Threonine
(Thr or T)

Glutamine
(Gln or Q)

Asparagine
(Asn or N)

Tyrosine
(Tyr or Y)

Properties of side chains:

Hydrophilic side chains tend to have partial + or − charge allowing them to participate in chemical reactions, form
H-bonds, and associate with water.

Nonpolar

Alanine
(Ala or A)

Valine
(Val or V)

Leucine
(Leu or L)

Isoleucine
(Ile or I)

Methionine
(Met or M)

Phenylalanine
(Phe or F)

Tryptophan
(Trp or W)

Properties of side chains:

Hydrophobic side chain consists almost entirely of C and H atoms. These amino acids tend to form the inner core of
soluble proteins, buried away from the aqueous medium. They play an important role in membranes by associating
with the lipid bilayer.

Side chains with unique properties

Glycine
(Gly or G)

Cysteine
(Cys or C)

Proline
(Pro or P)

Side chain consists only of hydrogen
atom and can fit into either a hydrophilic
or hydrophobic environment. Glycine often
resides at sites where two polypeptides
come into close contact.

Though side chain has polar, uncharged
character, it has the unique property of
forming a covalent bond with another
cysteine to form a disulfide link.

Though side chain has hydrophobic
character, it has the unique property
of creating kinks in polypeptide
chains and disrupting ordered
secondary structure.

FIGURE 2.26 **The chemical structure of amino acids.** These 20 amino acids represent those most commonly found in proteins and, more specifically, those encoded by DNA. Other amino acids occur as the result of a modification to one of those shown here. The amino acids are arranged into four groups based on the character of their side chains, as described in the text. All molecules are depicted as free amino acids in their ionized state as they would exist in solution at neutral pH.

is, the side chains contain relatively strong organic acids and bases. The ionization reactions of glutamic acid and lysine are shown in Figure 2.27. At physiologic pH, the side chains of these amino acids are almost always present in the fully charged state. Consequently, they are able to form ionic bonds with other charged species in the cell. For example, the positively charged arginine residues of histone proteins are linked by ionic bonds to the negatively charged phosphate groups of DNA (see Figure 2.3). Histidine is also considered a polar, charged amino acid, though in most cases it is only partially charged at physiologic pH. In fact, because of its ability to gain or lose a proton in physiologic pH ranges, histidine is a particularly important residue in the active site of many proteins (as in Figure 3.13).

(a)

(b)
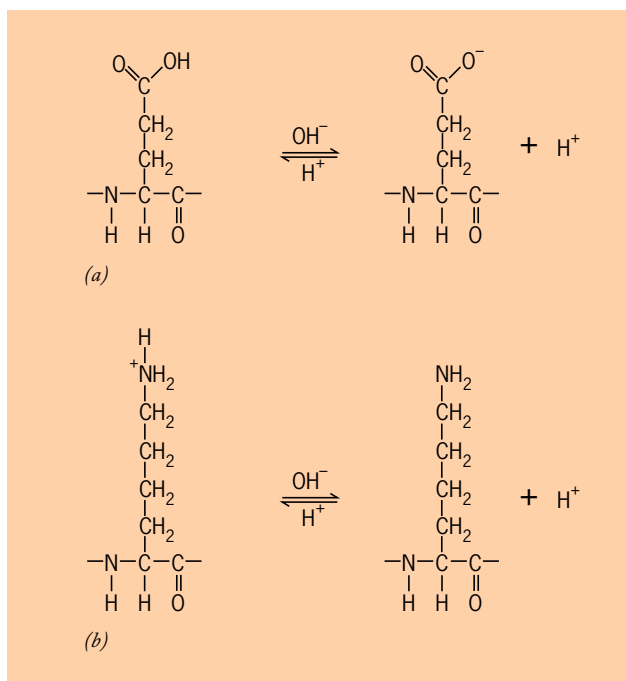
FIGURE 2.27 The ionization of charged, polar amino acids. (a) The side chain of glutamic acid loses a proton when its carboxylic acid group ionizes. The degree of ionization of the carboxyl group depends on the pH of the medium: the greater the hydrogen ion concentration (the lower the pH), the smaller the percentage of carboxyl groups that are present in the ionized state. Conversely, a rise in pH leads to an increased ionization of the proton from the carboxyl group, increasing the percentage of negatively charged glutamic acid side chains. The pH at which 50 percent of the side chains are ionized and 50 percent are unionized is called the pK, which is 4.4 for the side chain of free glutamic acid. At physiologic pH, virtually all of the glutamic acid residues of a polypeptide are negatively charged. (b) The side chain of lysine becomes ionized when its amino group gains a proton. The greater the hydroxyl ion concentration (the higher the pH), the smaller the percentage of amino groups that are positively charged. The pH at which 50 percent of the side chains of lysine are charged and 50 percent are uncharged is 10.0, which is the pK for the side chain of free lysine. At physiologic pH, virtually all of the lysine residues of a polypeptide are positively charged. Once incorporated into a polypeptide, the pK of a charged group can be greatly influenced by the surrounding environment.

2. *Polar, uncharged.* The side chains of these amino acids have a partial negative or positive charge and thus can form hydrogen bonds with other molecules including water. These amino acids are often quite reactive. Included in this category are asparagine and glutamine (the amides of aspartic acid and glutamic acid), threonine, serine, and tyrosine.

3. *Nonpolar.* The side chains of these amino acids are hydrophobic and are unable to form electrostatic bonds or interact with water. The amino acids of this category are alanine, valine, leucine, isoleucine, tryptophan, phenylalanine, and methionine. The side chains of the nonpolar amino acids generally lack oxygen and nitrogen. They vary primarily in size and shape, which allows one or another of them to pack tightly into a particular space within the core of a protein, associating with one another as the result of van der Waals forces and hydrophobic interactions.

4. *The other three amino acids*—glycine, proline, and cysteine—have unique properties that separate them from the others. The side chain of glycine consists of only a hydrogen atom, and glycine is a very important amino acid for just this reason. Owing to its lack of a side chain, glycine residues provide a site where the backbones of two polypeptides (or two segments of the same polypeptide) can approach one another very closely. In addition, glycine is more flexible than other amino acids and allows parts of the backbone to move or form a hinge. Proline is unique in having its α-amino group as part of a ring (making it an imino acid). Proline is a hydrophobic amino acid that does not readily fit into an ordered secondary structure, such as an α helix (page 54), often producing kinks or hinges. Cysteine contains a reactive sulfhydryl (—SH) group and is often covalently linked to another cysteine residue, as a **disulfide (—SS—) bridge**.
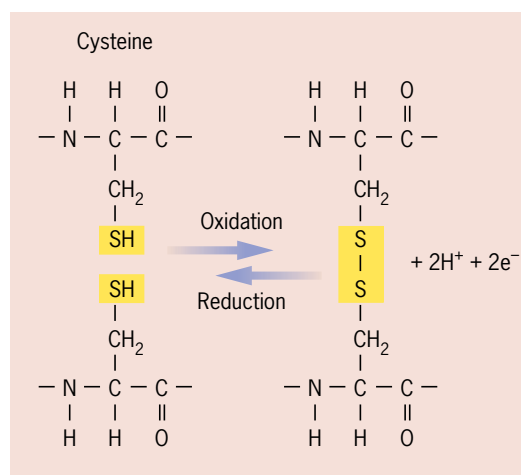
Disulfide bridges often form between two cysteines that are distant from one another in the polypeptide backbone or even in two separate polypeptides. Disulfide bridges help stabilize the intricate shapes of proteins, particularly those present outside of cells where they are subjected to added physical and chemical stress.

Not all of the amino acids described in this section are found in all proteins, nor are the various amino acids distributed in an equivalent manner. A number of other amino acids are also found in proteins, but they arise by alterations to the side chains of the 20 basic amino acids *after* their incorporation into a polypeptide chain. For this reason they are called **posttranslational modifications (PTMs)**. Dozens of different types of PTMs have been documented. The most widespread and important PTM is the reversible addition of a phosphate group to a serine, threonine, or tyrosine residue. PTMs can generate dramatic changes in the properties and function of a protein, most notably by modifying its three-dimensional structure, level of activity, localization within the cell, life span, and/or its interactions with other molecules. The presence or absence of a single phosphate group on a key regulatory protein has the potential to determine whether or not a cell will behave as a cancer cell or a normal cell. Because of PTMs, a single polypeptide can exist as a number of distinct biological molecules.

The ionic, polar, or nonpolar character of amino acid side chains is very important in protein structure and function. Most soluble (i.e., nonmembrane) proteins are constructed so that the polar residues are situated at the surface of the molecule where they can associate with the surrounding water and contribute to the protein's solubility in aqueous solution

(Figure 2.28*a*). In contrast, the nonpolar residues are situated predominantly in the core of the molecule (Figure 2.28*b*). The hydrophobic residues of the protein interior are often tightly packed together, creating a type of three-dimensional jigsaw puzzle in which water molecules are generally excluded. Hydrophobic interactions among the nonpolar side chains of these residues are a driving force during protein folding (page 62) and contribute substantially to the overall stability of the protein. In many enzymes, reactive polar groups project into the nonpolar interior, giving the protein its catalytic activity. For example, a nonpolar environment can enhance ionic interactions between charged groups that would be lessened by competition with water in an aqueous environment. Some reactions that might proceed at an imperceptibly slow rate in water can occur in millionths of a second within the protein.

The Structure of Proteins   Nowhere in biology is the intimate relationship between form and function better illustrated than with proteins. The structure of most proteins is completely defined and predictable. Each amino acid in one of these giant macromolecules is located at a specific site within the structure, giving the protein the precise shape and reactivity required for the job at hand. Protein structure can be described at several levels of organization, each emphasizing a
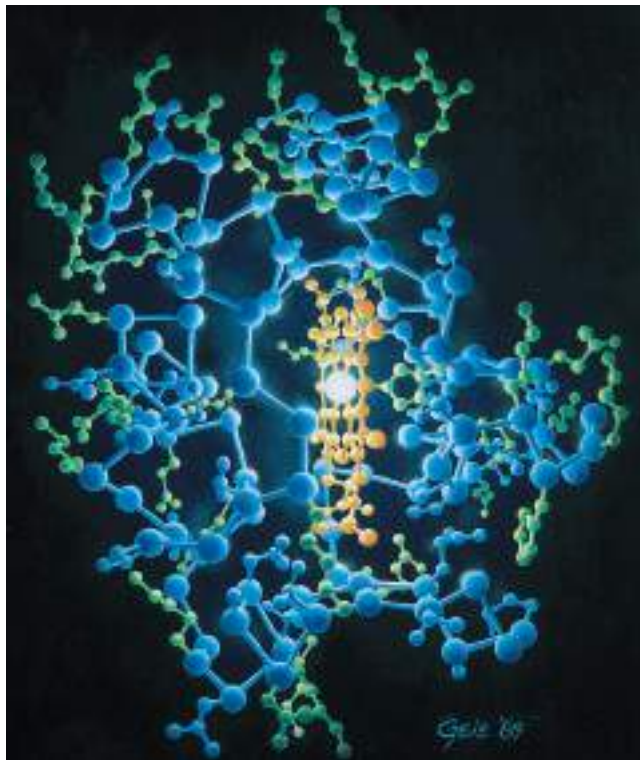
(a)

(b)

FIGURE 2.28 Disposition of hydrophilic and hydrophobic amino acid residues in the soluble protein cytochrome c. (*a*) The hydrophilic side chains, which are shown in green, are located primarily at the surface of the protein where they contact the surrounding aqueous medium. (*b*) The hydrophobic residues, which are shown in red, are located primarily within the center of the protein, particularly in the vicinity of the central heme group. (ILLUSTRATION, IRVING GEIS. IMAGE FROM IRVING GEIS COLLECTION/HOWARD HUGHES MEDICAL INSTITUTE. RIGHTS OWNED BY HHMI. REPRODUCED BY PERMISSION ONLY.)

different aspect and each dependent on different types of interactions. Customarily, four such levels are described: *primary*, *secondary*, *tertiary*, and *quaternary*. The first, primary structure, concerns the amino acid sequence of a protein, whereas the latter three levels concern the organization of the molecule in space. To understand the mechanism of action and biological function of a protein it is essential to know how that protein is constructed.

Primary Structure   The primary structure of a polypeptide is the specific linear sequence of amino acids that constitute the chain. With 20 different building blocks, the number of different polypeptides that can be formed is 20^n, where *n* is the number of amino acids in the chain. Because most polypeptides contain well over 100 amino acids, the variety of possible sequences is essentially unlimited. The information for the precise order of amino acids in every protein that an organism can produce is encoded within the genome of that organism.

   As we will see later, the amino acid sequence provides the information required to determine a protein's three-dimensional shape and thus its function. The sequence of amino acids, therefore, is all-important, and changes that arise in the sequence as a result of genetic mutations in the DNA may not be readily tolerated. The earliest and best-studied example of this relationship is the change in the amino acid sequence of hemoglobin that causes the disease *sickle cell anemia*. This severe, inherited anemia results solely from a single change in amino acid sequence within the hemoglobin molecule: a nonpolar valine residue is present where a charged glutamic acid is normally located. This change in hemoglobin structure can have a dramatic effect on the shape of red blood cells, converting them from disk-shaped cells to sickle-shaped cells (Figure 2.29), which tend to clog small blood vessels, causing pain and life-threatening crises. Not all amino acid changes have such a dramatic effect, as evidenced by the differences in amino acid sequence in the same protein among related organisms. The degree to which changes in the primary sequence are toler-

1 μm
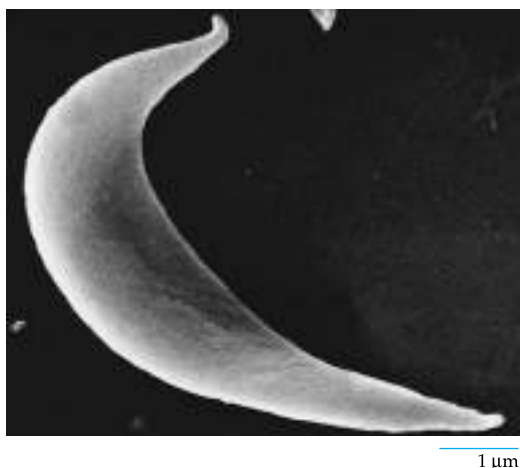
FIGURE 2.29 **Scanning electron micrograph** of a red blood cell from a person with sickle cell anemia. Compare with the micrograph of a normal red blood cell of Figure 4.32*a*. (COURTESY OF J. T. THORNWAITE, B. F. CAMERON, AND R. C. LEIF.)

ated depends on the degree to which the shape of the protein or the critical functional residues are disturbed.

   The first amino acid sequence of a protein was determined by Frederick Sanger and co-workers at Cambridge University in the early 1950s. Beef insulin was chosen for the study because of its availability and its small size—two polypeptide chains of 21 and 30 amino acids each. The *sequencing* of insulin was a momentous feat in the newly emerging field of molecular biology. It revealed that proteins, the most complex molecules in cells, have a definable substructure that is neither regular nor repeating, unlike those of polysaccharides. Each particular polypeptide, whether insulin or some other species, has a precise sequence of amino acids that does not vary from one molecule to another. With the advent of techniques for rapid DNA sequencing (see Section 18.15), the primary structure of a polypeptide can be deduced from the nucleotide sequence of the encoding gene. In the past few years, the complete sequences of the genomes of hundreds of organisms, including humans, have been determined. This information will eventually allow researchers to learn about every protein that an organism can manufacture. However, translating information about primary sequence into knowledge of higher levels of protein structure remains a formidable challenge.

Secondary Structure   All matter exists in space and therefore has a three-dimensional expression. Proteins are formed by linkages among vast numbers of atoms; consequently their shape is complex. The term **conformation** refers to the three-dimensional arrangement of the atoms of a molecule, that is, to their spatial organization. Secondary structure describes the conformation of portions of the polypeptide chain. Early studies on secondary structure were carried out by Linus Pauling and Robert Corey of the California Institute of Technology. By studying the structure of simple peptides consisting of a few amino acids linked together, Pauling and Corey concluded that polypeptide chains exist in preferred conformations that provide the maximum possible number of hydrogen bonds between neighboring amino acids.

   Two conformations were proposed. In one conformation, the backbone of the polypeptide assumed the form of a cylindrical, twisting spiral called the **alpha (α) helix** (Figure 2.30*a,b*). The backbone lies on the inside of the helix, and the side chains project outward. The helical structure is stabilized by hydrogen bonds between the atoms of one peptide bond and those situated just above and below it along the spiral (Figure 2.30*c*). The X-ray diffraction patterns of actual proteins produced during the 1950s bore out the existence of the α helix, first in the protein keratin found in hair and later in various oxygen-binding proteins, such as myoglobin and hemoglobin (see Figure 2.34). Surfaces on opposite sides of an α helix may have contrasting properties. In water-soluble proteins, the outer surface of an α helix often contains polar residues in contact with the solvent, whereas the surface facing inward typically contains nonpolar side chains.

   The second conformation proposed by Pauling and Corey was the **beta (β)-pleated sheet**, which consists of several segments of a polypeptide lying side by side. Unlike the coiled, cylindrical form of the α helix, the backbone of each segment of

(a)   (b)   (c)

FIGURE 2.30 The alpha helix. (*a*) Linus Pauling (*left*) and Robert Corey with a wooden model of the alpha helix. The model has a scale of 1 inch per Å, which represents a 254 million-fold enlargement. (*b*) The helical path around a central axis taken by the polypeptide backbone in a region of α helix. Each complete (360°) turn of the helix corresponds to 3.6 amino acid residues. The distance along the axis between adjacent residues is 1.5 Å. (*c*) The arrangement of the atoms of the backbone of the α helix and the hydrogen bonds that form between amino acids.
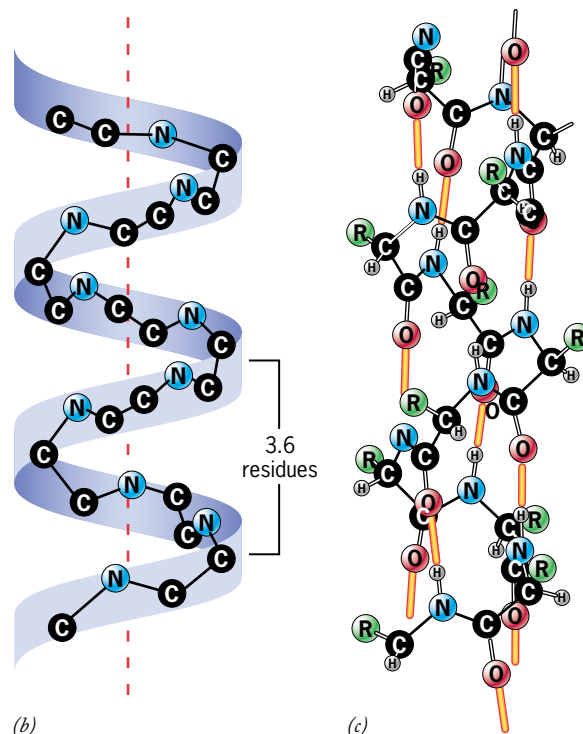
Because of the helical rotation, the peptide bonds of every fourth amino acid come into close proximity. The approach of the carbonyl group (C=O) of one peptide bond to the imine group (H—N) of another peptide bond results in the formation of hydrogen bonds between them. The hydrogen bonds (orange bars) are essentially parallel to the axis of the cylinder and thus hold the turns of the chain together. (A: COURTESY OF THE ARCHIVES, CALIFORNIA INSTITUTE OF TECHNOLOGY.)

polypeptide (or **β strand**) in a β sheet assumes a folded or pleated conformation (Figure 2.31*a*). Like the α helix, the β sheet is also characterized by a large number of hydrogen bonds, but these are oriented perpendicular to the long axis of the polypeptide chain and project across from one part of the chain to another (Figure 2.31*b*). Like the α helix, the β sheet has also been found in many different proteins. Because β strands are highly extended, the β sheet resists pulling (tensile) forces. Silk

is composed of a protein containing an extensive amount of β sheet; silk fibers are thought to owe their strength to this architectural feature. Remarkably, a single fiber of spider silk, which may be a tenth the thickness of a human hair, is roughly five times stronger than a steel fiber of comparable weight.

Those portions of a polypeptide chain not organized into an α helix or a β sheet may consist of hinges, turns, loops, or finger-like extensions. Often, these are the most flexible portions of a

(a)

(b)
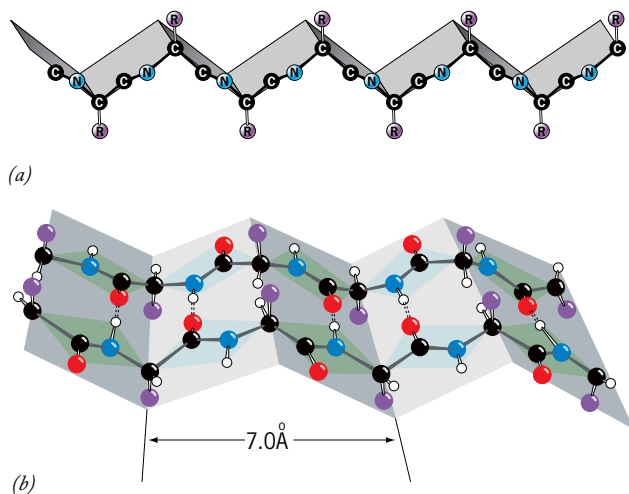
FIGURE 2.31 The β-pleated sheet. (*a*) Each polypeptide of a β sheet assumes an extended but pleated conformation referred to as a β strand. The pleats result from the location of the α-carbons above and below the plane of the sheet. Successive side chains (R groups in the figure) project upward and downward from the backbone. The distance along the axis between adjacent residues is 3.5 Å (*b*) A β-pleated sheet consists of a number of β strands that lie parallel to one another and are joined together by a regular array of hydrogen bonds between the carbonyl and imine groups of the neighboring backbones. Neighboring segments of the polypeptide backbone may lie either parallel (in the same N-terminal → C-terminal direction) or antiparallel (in the opposite N-terminal → C-terminal direction). (ILLUSTRATION, IRVING GEIS. IMAGE FROM THE IRVING GEIS COLLECTION/HOWARD HUGHES MEDICAL INSTITUTE. RIGHTS OWNED BY HHMI. REPRODUCTION BY PERMISSION ONLY.)
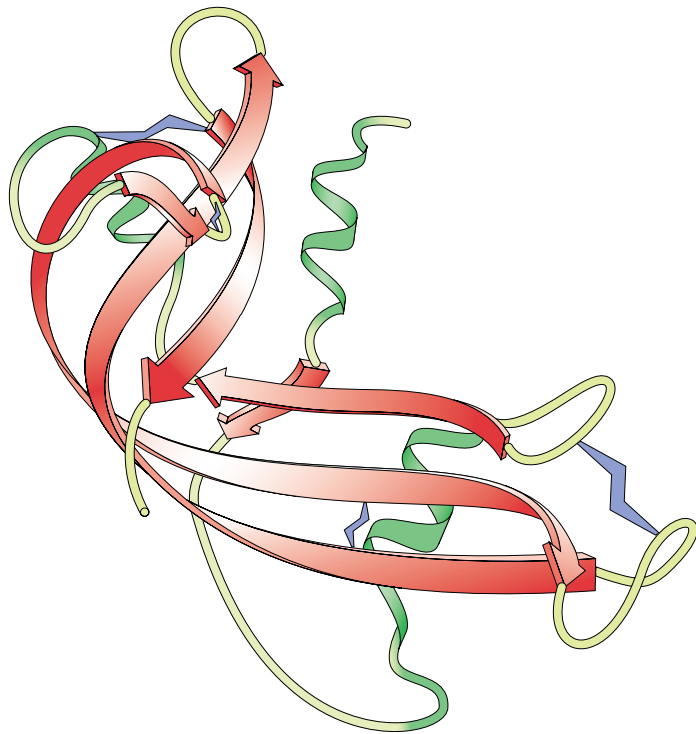
FIGURE 2.32 **A ribbon model of ribonuclease.** The regions of α helix are depicted as spirals and β strands as flattened ribbons with the arrows indicating the N-terminal → C-terminal direction of the polypeptide. Those segments of the chain that do not adopt a regular secondary structure (i.e., an α helix or β strand) consist largely of loops and turns and are shown in lime green. Disulfide bonds are shown in blue. (AFTER A DRAWING BY JANE S. RICHARDSON.)

FIGURE 2.33 **An X-ray diffraction pattern of myoglobin.** The pattern of spots is produced as a beam of X-rays is diffracted by the atoms in the protein crystal, causing the X-rays to strike the film at specific sites. Information derived from the position and intensity (darkness) of the spots can be used to calculate the positions of the atoms in the protein that diffracted the beam, leading to complex structures such as that shown in Figure 2.34. (COURTESY OF JOHN C. KENDREW.)

polypeptide chain and the sites of greatest biological activity. For example, antibody molecules are known for their specific interactions with other molecules (antigens); these interactions are mediated by a series of loops at one end of the antibody molecule (see Figures 17.15 and 17.16). The various types of secondary structures are most simply depicted as shown in Figure 2.32: α helices are represented by helical ribbons, β strands as flattened arrows, and connecting segments as thinner strands.

Tertiary Structure   The next level above secondary structure is tertiary structure, which describes the conformation of the entire polypeptide. Whereas secondary structure is stabilized primarily by hydrogen bonds between atoms that form the peptide bonds of the backbone, tertiary structure is stabilized by an array of noncovalent bonds between the diverse side chains of the protein. Secondary structure is largely limited to a small number of conformations, but tertiary structure is virtually unlimited.

The detailed tertiary structure of a protein is usually determined using the technique of **X-ray crystallography**.[5]

---

[5]The three-dimensional structure of small proteins can also be determined by nuclear magnetic resonance (NMR) spectroscopy, which is not discussed in this text (see the supplement to the July issue of *Nature Struct. Biol.*, 1998, *Nature Struct. Biol.* 7:982, 2000, and *Chem. Rev.* 104:3517–3704, 2004 for reviews of this technology). Figure 1*a*, page 65, shows an NMR-derived structure.

In this technique (which is described in more detail in Sections 3.2 and 18.8), a crystal of the protein is bombarded by a thin beam of X-rays, and the radiation that is scattered (diffracted) by the electrons of the protein's atoms is allowed to strike a radiation-sensitive plate or detector, forming an image of spots, such as those of Figure 2.33. When these diffraction patterns are subjected to complex mathematical analysis, an investigator can work backward to derive the structure responsible for producing the pattern.

It has become evident in recent years, as more and more protein structures have been solved, that a surprising number of proteins contain sizable segments that lack a defined conformation. Examples of proteins containing these types of unstructured (or *disordered*) segments can be seen in the models of the PrP protein in Figure 1 on page 65 and the histone tails in Figure 12.9*c*. The disordered regions in these proteins are depicted as dashed lines in the images, conveying the fact that these segments of the polypeptide can be present in many different positions and, thus, cannot be studied by X-ray crystallography. Disordered segments tend to have a predictable amino acid composition, being enriched in charged and polar residues and deficient in hydrophobic residues. You might be wondering whether proteins lacking a fully defined structure could be engaged in a useful function. In fact, these disordered regions play key roles in vital cellular processes, often binding to DNA or to other proteins. Remarkably, these segments

often undergo a physical transformation once they bind to an appropriate partner and are then seen to possess a defined, folded structure.

Most proteins can be categorized on the basis of their overall conformation as being either **fibrous proteins**, which have an elongated shape, or **globular proteins**, which have a compact shape. Most proteins that act as structural materials outside living cells are fibrous proteins, such as collagens and elastins of connective tissues, keratins of hair and skin, and silk. These proteins resist pulling or shearing forces to which they are exposed. In contrast, most proteins within the cell are globular proteins.

Myoglobin: The First Globular Protein Whose Tertiary Structure Was Determined   The polypeptide chains of globular proteins are folded and twisted into complex shapes. Distant points on the linear sequence of amino acids are brought next to each other and linked by various types of bonds. The first glimpse at the tertiary structure of a globular protein came in 1957 through the X-ray crystallographic studies of John Kendrew and his colleagues at Cambridge University using X-ray diffraction patterns such as that shown in Figure 2.33. The protein they reported on was myoglobin.

Myoglobin functions in muscle tissue as a storage site for oxygen; the oxygen molecule is bound to an iron atom in the center of a heme group. (The heme is an example of a *prosthetic group*, i.e., a portion of the protein that is not composed of amino acids, which is joined to the polypeptide chain after its assembly on the ribosome.) It is the heme group of myoglobin that gives most muscle tissue its reddish

color. The first report on the structure of myoglobin provided a low-resolution profile sufficient to reveal that the molecule was compact (globular) and that the polypeptide chain was folded back on itself in a complex arrangement. There was no evidence of regularity or symmetry within the molecule, such as that revealed in the earlier description of the DNA double helix. This was not surprising considering the singular function of DNA and the diverse functions of protein molecules.

The earliest crude profile of myoglobin revealed eight rod-like stretches of α helix ranging from 7 to 24 amino acids in length. Altogether, approximately 75 percent of the 153 amino acids in the polypeptide chain is in the α-helical conformation. This is an unusually high percentage compared with that for other proteins that have since been examined. No β-pleated sheet was found. Subsequent analyses of myoglobin using additional X-ray diffraction data provided a much more detailed picture of the molecule (Figures 2.34*a* and 3.16). For example, it was shown that the heme group is situated within a pocket of hydrophobic side chains that promotes the binding of oxygen without the oxidation (loss of electrons) of the iron atom. Myoglobin contains no disulfide bonds; the tertiary structure of the protein is held together exclusively by noncovalent interactions. All of the noncovalent bonds thought to occur between side chains within proteins—hydrogen bonds, ionic bonds, and van der Waals forces—have been found (Figure 2.35). Unlike myoglobin, most globular proteins contain both α helices and β sheets. Most importantly, these early landmark studies revealed that each protein has a unique tertiary structure that
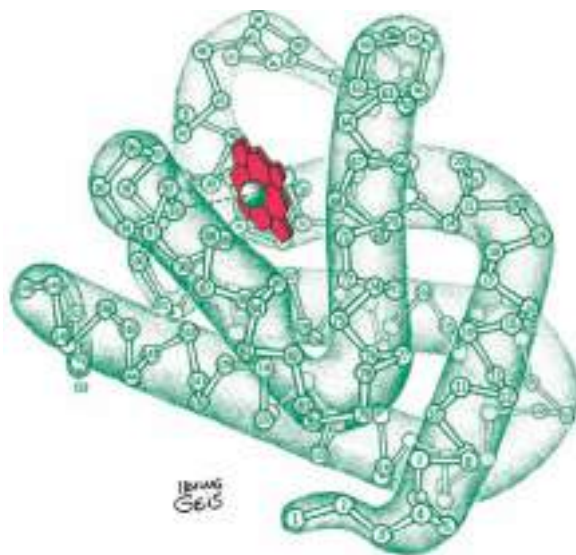
(a)
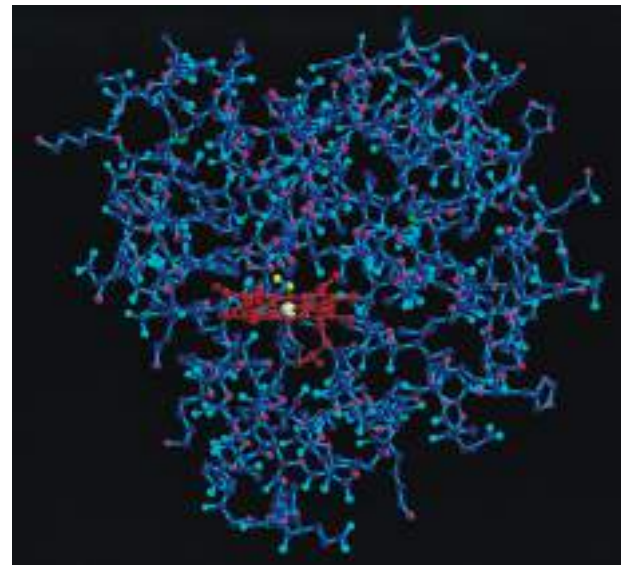
(b)

FIGURE 2.34 The three-dimensional structure of myoglobin. (*a*) The tertiary structure of whale myoglobin. Most of the amino acids are part of α helices. The nonhelical regions occur primarily as turns, where the polypeptide chain changes direction. The position of the heme is indicated in red. (*b*) The three-dimensional structure of myoglobin (heme indicated in red). The positions of all of the molecule's atoms, other than hydrogen, are shown. (A: ILLUSTRATION, IRVING GEIS. IMAGE FROM IRVING GEIS COLLECTION/HOWARD HUGHES MEDICAL INSTITUTE. RIGHTS OWNED BY HHMI. REPRODUCED BY PERMISSION ONLY; B: KEN EWARD/PHOTO RESEARCHERS.)
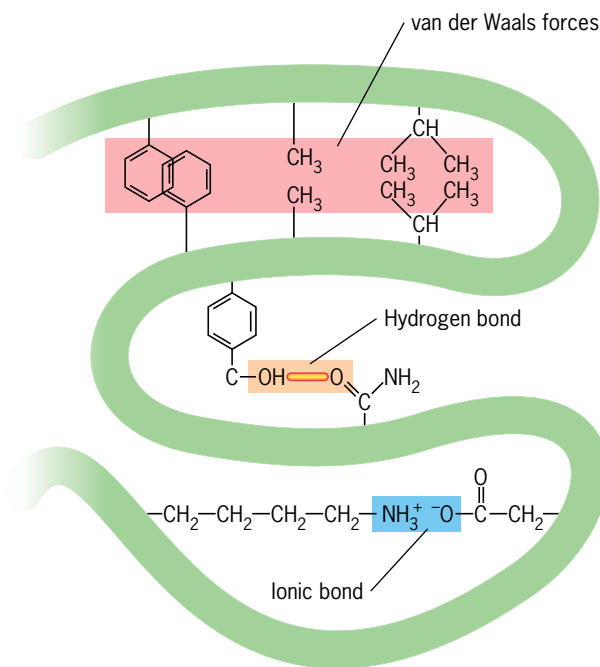
FIGURE 2.35 Types of noncovalent bonds maintaining the conformation of proteins.

FIGURE 2.36 Proteins are built of structural units, or domains. The mammalian enzyme phospholipase C is constructed of four domains, indicated in different colors. The catalytic domain of the enzyme is shown in blue. Each of the domains of this enzyme can be found independently in other proteins as indicated by the matching color. (FROM LIISA HOLM AND CHRIS SANDER, STRUCTURE 5:167, 1997.)

can be correlated with its amino acid sequence and its biological function.

Protein Domains   Unlike myoglobin, most eukaryotic proteins are composed of two or more spatially distinct modules, or **domains**, that fold independent of one another. For example, the mammalian enzyme phospholipase C, shown in the central part of Figure 2.36, consists of four distinct domains, colored differently in the drawing. The different domains of a polypeptide often represent parts that function in a semi-independent manner. For example, they might bind different factors, such as a coenzyme and a substrate or a DNA strand and another protein, or they might move relatively independent of one another. Protein domains are often identified with a specific function. For example, proteins containing a PH domain bind to membranes containing a specific phospholipid, whereas proteins containing a chromodomain bind to a methylated lysine residue in another protein. The functions of a newly identified protein can usually be predicted by the domains of which it is made.

Many polypeptides containing more than one domain are thought to have arisen during evolution by the fusion of genes that encoded different ancestral proteins, with each domain representing a part that was once a separate molecule. Each domain of the mammalian phospholipase C molecule, for example, has been identified as a homologous unit in another protein (Figure 2.36). Some domains have been found only in one or a few proteins. Other domains have been shuffled widely about during evolution, appearing in a variety of proteins whose other regions show little or no evidence of an evolutionary relationship. Shuffling of domains creates pro-

teins with unique combinations of activities. On average, mammalian proteins tend to be larger and contain more domains than proteins of less complex organisms, such as fruit flies and yeast.

Dynamic Changes within Proteins   Although X-ray crystallographic structures possess exquisite detail, they are static images frozen in time. Proteins, in contrast, are not static and inflexible, but capable of considerable internal movements. Proteins are, in other words, molecules with "moving parts." Because they are tiny, nanometer-sized objects, proteins can be greatly influenced by the energy of their environment. Random, small-scale fluctuations in the arrangement of the bonds within a protein create an incessant thermal motion within the molecule. Spectroscopic techniques, such as nuclear magnetic resonance (NMR), can monitor dynamic movements within proteins, and they reveal shifts in hydrogen bonds, waving movements of external side chains, and the full rotation of the aromatic rings of tyrosine and phenylalanine residues about one of the single bonds. The important role

that such movements can play in a protein's function is illustrated in studies of acetylcholinesterase, the enzyme responsible for degrading acetylcholine molecules that are left behind following the transmission of an impulse from one nerve cell to another (Section 4.8). When the tertiary structure of acetylcholinesterase was first revealed by X-ray crystallography, there was no obvious pathway for acetylcholine molecules to enter the enzyme's catalytic site, which was situated at the bottom of a deep gorge in the molecule. In fact, the narrow entrance to the site was completely blocked by the presence of a number of bulky amino acid side chains. Using high-speed computers, researchers have been able to simulate the random movements of thousands of atoms within the enzyme. These simulations indicated that dynamic movements of side chains within the protein would lead to the rapid opening and closing of a "gate" that would allow acetylcholine molecules to diffuse into the enzyme's catalytic site (Figure 2.37).

Predictable (nonrandom) movements within a protein that are triggered by the binding of a specific molecule are described as **conformational changes**. Conformational changes typically involve the coordinated movements of various parts of the molecule. A comparison of the polypeptides depicted in Figures 3*a* and 3*b* on page 79 shows the dramatic conformational change that occurs in a bacterial protein (GroEL) when it interacts with another protein (GroES). Virtually every activity in which

a protein takes part is accompanied by conformational changes within the molecule (see http://molmovdb.org to watch examples). The conformational change in the protein myosin that occurs during muscle contraction is shown in Figures 9.60 and 9.61. In this case, binding of myosin to an actin molecule leads to a small (20°) rotation of the head of the myosin, which results in a 50 to 100 Å movement of the adjacent actin filament. The importance of this dynamic event can be appreciated if you consider that the movements of your body result from the additive effect of millions of conformational changes taking place within the contractile proteins of your muscles.

Quaternary Structure   Whereas many proteins such as myoglobin are composed of only one polypeptide chain, most are made up of more than one chain, or **subunit**. The subunits may be linked by covalent disulfide bonds, but most often they are held together by noncovalent bonds as occur typically between hydrophobic "patches" on the complementary surfaces of neighboring polypeptides. Proteins composed of subunits are said to have **quaternary structure**. Depending on the protein, the polypeptide chains may be identical or nonidentical. A protein composed of two identical subunits is described as a *homodimer*, whereas a protein composed of two nonidentical subunits is a *heterodimer*. A ribbon drawing of a homodimeric protein is depicted in Figure 2.38*a*. The two subunits of the protein are drawn in different colors, and the hydrophobic residues that form the complementary sites of contact are indicated. The best-studied multisubunit protein is hemoglobin, the O_2-carrying protein of red blood cells. A molecule of human hemoglobin consists of two α-globin and two β-globin polypeptides (Figure 2.38*b*), each of which binds a single molecule of oxygen. Elucidation of the three-dimensional structure of hemoglobin by Max Perutz of Cambridge University in 1959 was one of the early landmarks in the study of molecular biology. Perutz demonstrated that each of the four globin polypeptides of a hemoglobin molecule has a tertiary structure similar to that of myoglobin, a fact that strongly suggested the two proteins had evolved from a common ancestral polypeptide with a common O_2-binding mechanism. Perutz also compared the structure of the oxygenated and deoxygenated versions of hemoglobin. In doing so, he discovered that the binding of oxygen was accompanied by the movement of the bound iron atom closer to the plane of the heme group. This seemingly inconsequential shift in position of a single atom pulled on an α helix to which the iron is connected, which in turn led to a series of increasingly larger movements within and between the subunits. This finding revealed for the first time that the complex functions of proteins may be carried out by means of small changes in their conformation.

FIGURE 2.37 Dynamic movements within the enzyme acetylcholinesterase. A portion of the enzyme is depicted here in two different conformations: (1) a closed conformation (left) in which the entrance to the catalytic site is blocked by the presence of aromatic rings that are part of the side chains of tyrosine and phenylalanine residues (shown in purple) and (2) an open conformation (right) in which the aromatic rings of these side chains have swung out of the way, opening the "gate" to allow acetylcholine molecules to enter the catalytic site. These images are constructed using computer programs that take into account a host of information about the atoms that make up the molecule, including bond lengths, bond angles, electrostatic attraction and repulsion, van der Waals forces, etc. Using this information, researchers are able to simulate the movements of the various atoms over a very short time period, which provides images of the conformations that the protein can assume. An animation of this image can be found on the Web at http://mccammon.ucsd.edu. (COURTESY OF J. ANDREW MCCAMMON.)

Protein–Protein Interactions   Even though hemoglobin consists of four subunits, it is still considered a single protein with a single function. Many examples are known in which different proteins, each with a specific function, become physically associated to form a much larger **multiprotein complex**. One of the first multiprotein complexes to be discovered and studied was the pyruvate dehydrogenase complex of the bac-

(a)
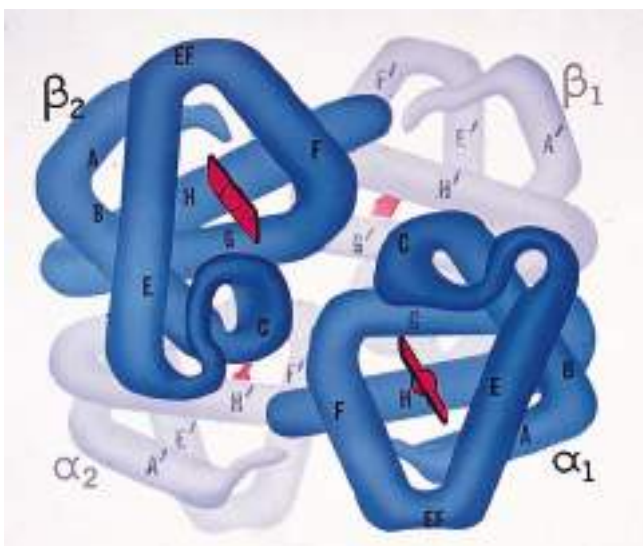
(b)

FIGURE 2.38   Proteins with quaternary structure. (*a*) Drawing of transforming growth factor-β2 (TGF-β2), a protein that is a dimer composed of two identical subunits. The two subunits are colored yellow and blue. Shown in white are the cysteine side chains and disulfide bonds. The spheres shown in yellow and blue are the hydrophobic residues that form the interface between the two subunits. (*b*) Drawing of a hemoglobin molecule, which consists of two α-globin chains and two β-globin chains (a heterotetramer) joined by noncovalent bonds. When the four globin polypeptides are assembled into a complete hemoglobin molecule, the kinetics of O_2 binding and release are quite different from those exhibited by isolated polypeptides. This is because the binding of O_2 to one polypeptide causes a conformational change in the other polypeptides that alters their affinity for O_2 molecules. (A: REPRINTED WITH PERMISSION FROM S. DAOPIN, ET AL., SCIENCE 257:372, 1992, COURTESY OF DAVID R. DAVIES. COPYRIGHT © 1992 AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE; B: ILLUSTRATION, IRVING GEIS. IMAGE FROM IRVING GEIS COLLECTION/HOWARD HUGHES MEDICAL INSTITUTE. RIGHTS OWNED BY HHMI. REPRODUCED BY PERMISSION ONLY.)
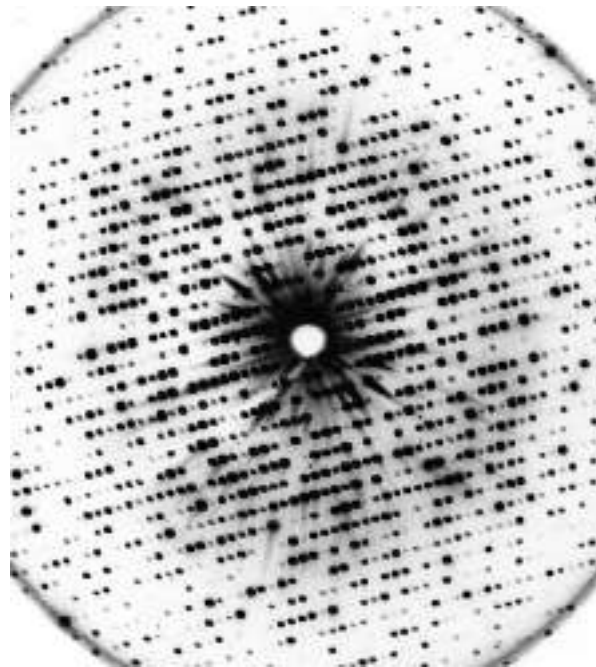
terium *Escherichia coli*, which consists of 60 polypeptide chains constituting three different enzymes (Figure 2.39). The enzymes that make up this complex catalyze a series of reactions connecting two metabolic pathways, glycolysis and the TCA cycle (see Figure 5.7). Because the enzymes are so closely associated, the product of one enzyme can be channeled directly

(a)   20 nm *(b)*
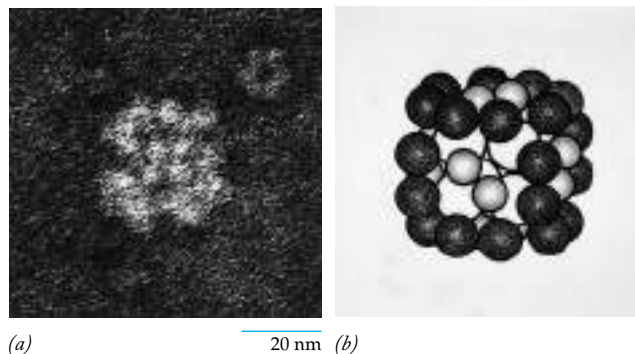
FIGURE   2.39   Pyruvate dehydrogenase: a multiprotein complex. (*a*) Electron micrograph of a negatively stained pyruvate dehydrogenase complex isolated from *E. coli*. Each complex contains 60 polypeptide chains constituting three different enzymes. Its molecular mass approaches 5 million daltons. (*b*) A model of the pyruvate dehydrogenase complex. The core of the complex consists of a cube-like cluster of dihydrolipoyl transacetylase molecules. Pyruvate dehydrogenase dimers (black spheres) are distributed symmetrically along the edges of the cube, and dihydrolipoyl dehydrogenase dimers (small gray spheres) are positioned in the faces of the cube. (COURTESY OF LESTER J. REED.)

to the next enzyme in the sequence without becoming diluted in the aqueous medium of the cell.

Multiprotein complexes that form within the cell are not necessarily stable assemblies, such as the pyruvate dehydrogenase complex. In fact, most proteins interact with other proteins in highly dynamic patterns, associating and dissociating depending on conditions within the cell at any given time. Interacting proteins tend to have complementary surfaces. Often a projecting portion of one molecule fits into a pocket within its partner. Once the two molecules have come into close contact, their interaction is stabilized by noncovalent bonds.

The reddish-colored object in Figure 2.40*a* is called an SH3 domain, and it is found as part of more than 200 different proteins involved in molecular signaling. The surface of an SH3 domain contains shallow hydrophobic "pockets" that become filled by complementary "knobs" projecting from another protein (Figure 2.40*b*). A large number of different structural domains have been identified that, like SH3, act as adaptors to mediate interactions between proteins. In many cases, protein–protein interactions are regulated by modifications, such as the addition of a phosphate group to a key amino acid, which act as a switch to turn on or off the protein's ability to bind a protein partner. As more and more complex molecular activities have been discovered, the importance of interactions among proteins has become increasingly apparent. For example, such diverse processes as DNA synthesis, ATP formation, and RNA processing are all accomplished by "molecular machines" that consist of a large number of interacting proteins, some of which form stable relationships and others transient liasons. Several hundred different protein complexes have been purified in large-scale studies on yeast.

Most investigators who study protein–protein interactions want to know whether one protein they are working with, call it protein X, interacts physically with another protein, call it protein Y. This type of question can be answered using a tech-

nique called the yeast two-hybrid (Y2H) assay, which is discussed in Section 18.7 and illustrated in Figure 18.27. In this technique, *genes* encoding the two proteins being studied are introduced into the same yeast cell. If the yeast cell tests positive for a particular reporter protein, which is indicated by an obvious color change in the yeast cell, then the two proteins in question had to have interacted within the yeast cell nucleus.

Interactions between proteins are generally studied one at a time, producing the type of data seen in Figure 2.40. In recent years, a number of research teams have set out to study protein–protein interactions on a global scale. For example, one might want to know all of the interactions that occur among the 14,000 or so proteins encoded by the genome of the fruit fly *Drosophila melanogaster*. Now that the entire genome of this insect has been sequenced, virtually every gene

within the genome is available as an individual DNA segment that can be cloned and used as desired. Consequently, it should be possible to test the proteins encoded by the fly genome, two at a time, for possible interactions in a Y2H assay. One study of this type reported that, of the millions of possible combinations, more than 20,000 interactions were detected among 7048 fruit-fly proteins tested.

Although the Y2H assay has been the mainstay in the study of protein–protein interactions for more than 15 years, it is an indirect assay (see Figure 18.27) and is fraught with uncertainties. On one hand, a large percentage of interactions known to occur between specific proteins fail to be detected in these types of experiments. The Y2H assay, in other words, gives a significant number of false negatives. The Y2H assay is also known to generate a large number of false positives; that is, it indicates that two proteins are capable of interacting when it is known from other studies that they do not do so under normal conditions within cells. In the study of fruit-fly proteins described above, the authors used computer analyses to narrow the findings from the original 20,000 interactions to approximately 5000 interactions in which they had high confidence. Overall, it is estimated that, on average, each protein encoded in the genome of a eukaryotic organism interacts with about five different protein partners. According to this estimate, human proteins would engage in roughly 100,000 different interactions.

The results from large-scale protein–protein interaction studies can be presented in the form of a network, such as that shown in Figure 2.41. This figure displays the potential binding partners of the various yeast proteins that contain an SH3 domain (see Figure 2.40*a*) and illustrates the complexities of such interactions at the level of an entire organism. In this particular figure, only those interactions detected by two entirely different assays (Y2H and another technique) are displayed,

(a)

(b)

FIGURE 2.40 Protein–protein interactions. (*a*) A model illustrating the complementary molecular surfaces of portions of two interacting proteins. The reddish-colored molecule is an SH3 domain of the enzyme PI3K, whose function is discussed in Chapter 15. This domain binds specifically to a variety of proline-containing peptides, such as the one shown in the space-filling model at the top of the figure. The proline residues in the peptide, which fit into hydrophobic pockets on the surface of the enzyme, are indicated. The polypeptide backbone of the peptide is colored yellow, and the side chains are colored green. (*b*) Schematic model of the interaction between an SH3 domain and a peptide showing the manner in which certain residues of the peptide fit into hydrophobic pockets in the SH3 domain. (A: FROM HONGTAO YU AND STUART SCHREIBER, CELL 76:940, 1994, BY PERMISSION OF CELL PRESS.)

FIGURE 2.41 A network of protein–protein interactions. Each red line represents an interaction between two yeast proteins, which are indicated by the named black dots. In each case, the arrow points from an SH3 domain protein to a target protein with which it can bind. The 59 interactions depicted here were detected using two different types of techniques that measure protein–protein interactions. (See *Trends Biochem. Sci.* 34:1, 2009, for discussion of the validity of protein–protein interaction studies.) (FROM A.H.Y. TONG, ET AL., SCIENCE 295:323, 2002. COPYRIGHT © 2002 AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

(a)   *(b)*
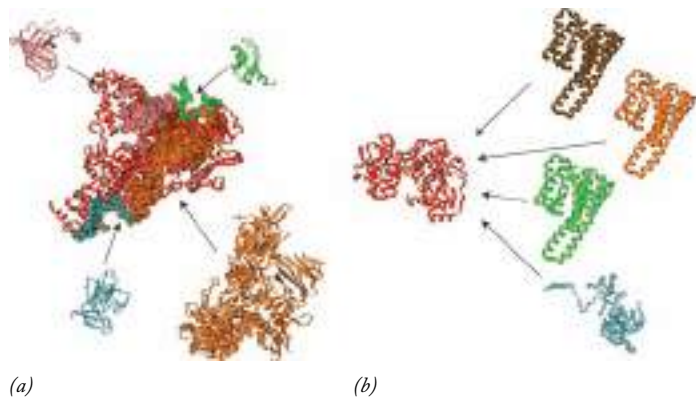
FIGURE 2.42 **Protein–protein interactions of hub proteins.** (*a*) The enzyme RNA polymerase II, which synthesizes messenger RNAs in the cell, binds a multitude of other proteins simultaneously using multiple interfaces. (*b*) The enzyme Cdc28, which phosphorylates other proteins as it regulates the cell division cycle of budding yeast. Cdc28 binds a number of different proteins (Cln1-Cln3) at the same interface, which allows only one of these partners to bind at a time. (FROM DAMIEN DEVOS AND ROBERT B. RUSSELL, CURR. OPIN. STRUCT. BIOL. 17:373, 2007. COPYRIGHT 2007, WITH PERMISSION OF ELSEVIER SCIENCE.)

FIGURE 2.43 **Denaturation and refolding of ribonuclease.** A native ribonuclease molecule (with intramolecular disulfide bonds indicated) is reduced and unfolded with β-mercaptoethanol and 8 M urea. After removal of these reagents, the protein undergoes spontaneous refolding. (FROM C. J. EPSTEIN, R. F. GOLDBERGER, AND C. B. ANFINSEN, COLD SPRING HARBOR SYMP. QUANT. BIOL. 28:439, 1963.)

which makes the conclusions much more reliable than those based on a single technology. Those proteins that have multiple binding partners, such as Las17 situated near the center of Figure 2.41, are referred to as *hubs*. Some hub proteins have several different binding interfaces and are capable of binding a number of different binding partners at the same time. In contrast, other hubs have a single binding interface, which is capable of binding several different partners, but only one at a time. Examples of each of these types of hub proteins are illustrated in Figure 2.42. The hub protein depicted in Figure 2.42*a* plays a central role in the process of gene expression, while that shown in Figure 2.42*b* plays an equally important role in the process of cell division.

Aside from obtaining a lengthy list of *potential* interactions, what do we learn about cellular activities from these types of large-scale studies? Most importantly, they provide a guideline for further investigation. Genome-sequencing projects have provided scientists with the amino acid sequences of a huge number of proteins whose very existence was previously unknown. What do these proteins do? One approach to determining a protein's function is to identify the proteins with which it associates. If, for example, a known protein has been shown to be involved in DNA replication, and an unknown protein is found to interact with the known protein, then it is likely that the unknown protein is also part of the cell's DNA-replication machinery. Thus, regardless of their limitations, these large-scale Y2H studies (and others using different assays not discussed) provide the starting point to explore a myriad of previously unknown protein–protein interactions, each of which has the potential to lead investigators to a previously unknown biological process.

Protein Folding   The elucidation of the tertiary structure of myoglobin in the late 1950s led to an appreciation for the com-

plexity of protein architecture. An important question immediately arose: How does such a complex, folded, asymmetric organization arise in the cell? The first insight into this problem was gained through a serendipitous observation in 1956 by Christian Anfinsen at the National Institutes of Health. Anfinsen was studying the properties of ribonuclease A, a small enzyme that consists of a single polypeptide chain of 124 amino acids with 4 disulfide bonds linking various parts of the chain. The disulfide bonds of a protein are typically broken (reduced) by adding a reducing agent, such as mercaptoethanol, which converts each disulfide bridge to a pair of sulfhydryl (—SH) groups (see drawing, page 52). To make all of the disulfide bonds accessible to the reducing agent, Anfinsen found that the molecule had to first be partially unfolded. The unfolding or disorganization of a protein is termed **denaturation**, and it can be brought about by a variety of agents, including detergents, organic solvents, radiation, heat, and compounds such as urea and guanidine chloride, all of which interfere with the various interactions that stabilize a protein's tertiary structure.

When Anfinsen treated ribonuclease molecules with mercaptoethanol and concentrated urea, he found that the preparation lost all of its enzymatic activity, which would be expected if the protein molecules had become totally unfolded. When he removed the urea and mercaptoethanol from the preparation, he found, to his surprise, that the molecules regained their normal enzymatic activity. The active ribonuclease molecules that had re-formed from the unfolded protein were indistinguishable both structurally and functionally from the correctly folded (i.e., **native**) molecules present at the beginning of the experiment (Figure 2.43). After extensive study, Anfinsen concluded that the linear sequence of amino acids contained all of the in-

(a)

Unfolded   Secondary structure   Native

(b)

Unfolded   Secondary structure   Native
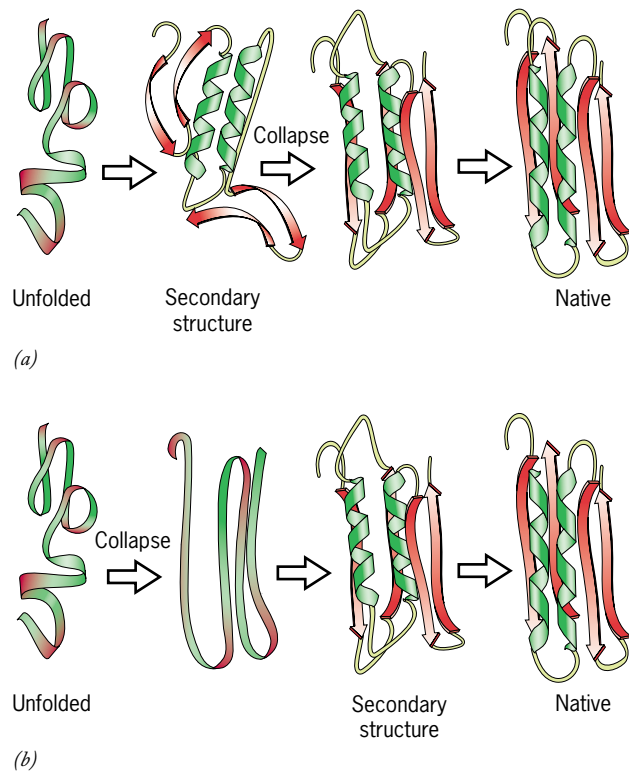
FIGURE 2.44 **Two alternate pathways** by which a newly synthesized or denatured protein could achieve its native conformation. Curled segments represent α helices, and arrows represent β strands.

FIGURE 2.45 **Along the folding pathway.** The image on the left shows the native tertiary structure of the enzyme acyl-phosphatase. The image on the right is the transition structure that appears very rapidly during the folding of this enzyme. The transition structure consists of numerous individual lines because it is a set (ensemble) of closely related structures. The overall architecture of the transition structure is similar to that of the native protein, but many of the finer structural features of the fully folded protein have yet to emerge. Conversion of the transition state to the native protein involves completing secondary structure formation, tighter packing of the side chains, and finalizing the burial of hydrophobic side chains from the aqueous solvent. (FROM K. LINDORFF-LARSEN, ET AL, TRENDS BIOCHEM. SCI. 30:14, 2005, COURTESY OF CHRISTOPHER M. DOBSON. COPYRIGHT 2005, WITH PERMISSION OF ELSEVIER SCIENCE.)

formation required for the formation of the polypeptide's three-dimensional conformation. Ribonuclease, in other words, is capable of **self-assembly**. As discussed in Chapter 3, events tend to progress toward states of lower energy. According to this concept, the tertiary structure that a polypeptide chain assumes after folding is the accessible structure with the lowest energy, which makes it the most thermodynamically stable structure that can be formed by that chain. It would appear that evolution selects for those amino acid sequences that generate a polypeptide chain capable of spontaneously arriving at a meaningful native state in a biologically reasonable time period.

There have been numerous controversies in the study of protein folding, one of which concerns the types of events that occur at various stages during the folding process. For the sake of simplicity, we will restrict the discussion to "simple" proteins, such as ribonuclease, that consist of a single domain. In the course depicted in Figure 2.44a, protein folding is initiated by interactions among neighboring residues that lead to the formation of much of the secondary structure of the molecule. Once the α helices and β sheets are formed, subsequent folding is driven by hydrophobic interactions that force nonpolar residues together in the central core of the protein. According to an alternate scheme shown in Figure 2.44b, the first major event in protein folding is the collapse of the polypeptide to form a compact structure, and only then does significant secondary structure develop. Recent studies indicate that the two pathways depicted in Figure 2.44 lie at opposite extremes, and that most proteins probably fold by a middle-of-the-road

scheme in which secondary structure formation and compaction occur simultaneously. These early folding events lead to the formation of a partially folded, transient structure that resembles the native protein but lacks many of the specific interactions between amino acid side chains that are present in the fully folded molecule (Figure 2.45).

If the information that governs folding is embedded in a protein's amino acid sequence, then alterations in this sequence have the potential to change the way a protein folds, leading to an abnormal tertiary structure. In fact, many mutations responsible for inherited disorders have been found to alter a protein's three-dimensional structure. In some cases, the consequences of protein misfolding can be fatal. Two examples of fatal neurodegenerative diseases that result from abnormal protein folding are discussed in the accompanying Human Perspective.

The Role of Molecular Chaperones   Not all proteins are able to assume their final tertiary structure by a simple process of self-assembly. This is not because the primary structure of these proteins lacks the required information for proper folding, but rather because proteins undergoing folding have to be prevented from interacting nonselectively with other molecules in the crowded compartments of the cell. Several families of proteins have evolved whose function is to help unfolded or misfolded proteins achieve their proper three-dimensional conformation. These "helper proteins" are called **molecular chaperones**, and they selectively bind to short stretches of hydrophobic amino acids that tend to be exposed in nonnative proteins but buried in proteins having a native conformation.

THE HUMAN PERSPECTIVE

Protein Misfolding Can Have Deadly Consequences

In April 1996 a paper was published in the medical journal *Lancet* that generated widespread alarm in the populations of Europe. The paper described a study of 10 persons afflicted with Creutzfeld-Jakob disease (CJD), a rare, fatal disorder that attacks the brain, causing a loss of motor coordination and dementia. Like numerous other diseases, CJD can occur as an inherited disease that runs in certain families or as a sporadic form that appears in individuals who have no family history of the disease. Unlike virtually every other inheritable disease, however, CJD can also be *acquired*. Until recently, persons who had acquired CJD had been recipients of organs or organ products that were donated by a person with undiagnosed CJD. The cases described in the 1996 *Lancet* paper had also been acquired, but the apparent source of the disease was contaminated beef that the infected individuals had eaten years earlier. The contaminated beef was derived from cattle raised in England that had contracted a neurodegenerative disease that caused the animals to lose motor coordination and develop demented behavior. The disease became commonly known as "mad cow disease." Patients who have acquired CJD from eating contaminated beef can be distinguished by several criteria from those who suffer from the classical forms of the disease. To date, roughly 200 people have died of CJD acquired from contaminated beef, and the numbers of such deaths have been declining.[1]

A disease that runs in families can invariably be traced to a faulty gene, whereas diseases that are acquired from a contaminated source can invariably be traced to an infectious agent. How can the same disease be both inherited and infectious? The answer to this question has emerged gradually over the past several decades, beginning with observations by D. Carleton Gajdusek in the 1960s on a strange malady that once afflicted the native population of Papua, New Guinea. Gajdusek showed that these islanders were contracting a fatal neurodegenerative disease—which they called "kuru"—during a funeral ritual in which they ate the brain tissue of a recently deceased relative. Autopsies of the brains of patients who had died of kuru showed a distinct pathology, referred to as *spongiform encephalopathy*, in which certain brain regions were riddled with microscopic holes (vacuolations), causing the tissue to resemble a sponge.

It was soon shown that the brains of islanders suffering from kuru were strikingly similar in microscopic appearance to the brains of persons afflicted with CJD. This observation raised an important question: Did the brain of a person suffering from CJD, which was known to be an inherited disease, contain an infectious agent? In 1968, Gajdusek showed that when extracts prepared from a biopsy of the brain of a person who had died from CJD were injected into a suitable laboratory animal, that animal did indeed develop a spongi-

form encephalopathy similar to that of kuru or CJD. Clearly, the extracts contained an infectious agent, which at the time was presumed to be a virus.

In 1982, Stanley Prusiner of the University of California, San Francisco, published a paper suggesting that, unlike viruses, the infectious agent responsible for CJD lacked nucleic acid and instead was composed solely of protein. He called the protein a *prion*. This "protein only" hypothesis, as it is called, was originally met with considerable skepticism, but subsequent studies by Prusiner and others have provided overwhelming support for the proposal. It was presumed initially that the prion protein was an external agent—some type of virus-like particle lacking nucleic acid. Contrary to this expectation, the prion protein was soon shown to be encoded by a gene (called *PRNP*) within the cell's own chromosomes. The gene is expressed within *normal* brain tissue and encodes a protein designated PrP^C (standing for prion protein cellular) that resides at the surface of nerve cells. The precise function of PrP^C remains a mystery. A modified version of the protein (designated PrP^{Sc}, standing for prion protein scrapie) is present in the brains of humans with CJD. Unlike the normal PrP^C, the modified version of the protein accumulates within nerve cells, forming aggregates that kill the cells.

In their purified states, PrP^C and PrP^{Sc} have very different physical properties. PrP^C remains as a monomeric molecule that is soluble in salt solutions and is readily destroyed by protein-digesting enzymes. In contrast, PrP^{Sc} molecules interact with one another to form insoluble fibrils that are resistant to enzymatic digestion. Based on these differences, one might expect these two forms of the PrP protein to be composed of distinctly different sequences of amino acids, but this is not the case. The two forms can have identical amino acid sequences, but they differ in the way the polypeptide chain folds to form the three-dimensional protein molecule (Figure 1). Whereas a PrP^C molecule consists largely of α-helical segments and interconnecting coils, the core of a PrP^{Sc} molecule consists largely of β sheet.

It is not hard to understand how a mutant polypeptide might be less stable and more likely to fold into the abnormal PrP^{Sc} conformation, but how is such a protein able to act as an infectious agent? According to the prevailing hypothesis, an abnormal prion molecule (PrP^{Sc}) can bind to a normal protein molecule (PrP^C) and cause the normal protein to fold into the abnormal form. This conversion can be shown to occur in the test tube: addition of PrP^{Sc} to a preparation of PrP^C can convert the PrP^C molecules into the PrP^{Sc} conformation. According to this hypothesis, the appearance of the abnormal protein in the body—whether as a result of a rare misfolding event in the case of sporadic disease or by exposure to contaminated beef—starts a chain reaction in which normal protein molecules in the cells are gradually converted to the abnormal prion form. The precise mechanism by which prions lead to neurodegeneration remains unclear.

CJD is a rare disease caused by a protein with unique infective properties. Alzheimer's disease (AD), on the other hand, is a common disorder that strikes as many as 10 percent of individuals who are at least 65 years of age and perhaps 40 percent of individuals who are 80 years or older. Persons with AD exhibit memory loss, confusion, and a loss of reasoning ability. CJD and AD share a number of important features. Both are fatal neurodegenerative diseases that

---

[1]On the surface, this would suggest that the epidemic has run its course, but there are several reasons for public health officials to remain concerned. For one, studies of tissues that had been removed during surgeries in England indicate that thousands of people are likely to be infected with the disease without exhibiting symptoms. Even if these individuals never develop clinical disease, they remain potential carriers who could pass CJD on to others through blood transfusions. In fact, at least two individuals are believed to have contracted CJD after receiving blood from a donor harboring the disease. These findings underscore the need to test blood for the presence of the responsible agent (whose nature is discussed below).

(a)   *(b)*
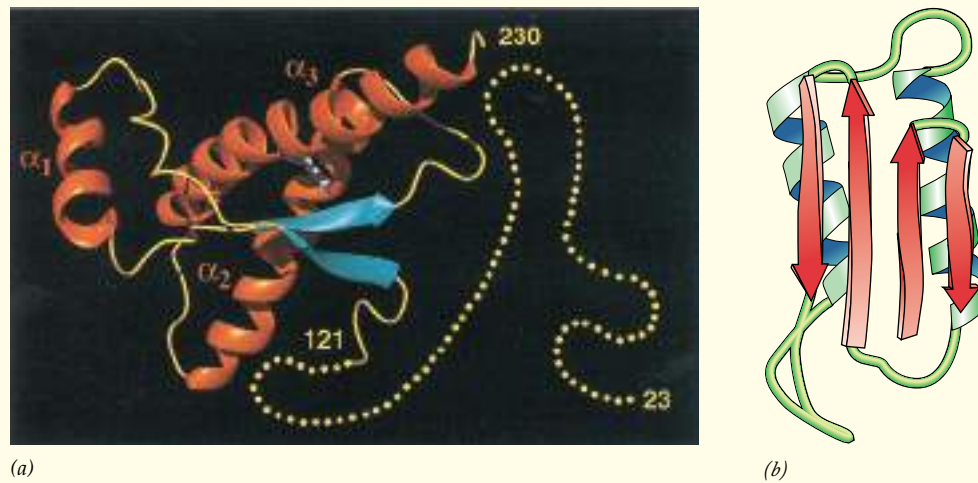
FIGURE 1 **A contrast in structure.** (*a*) Tertiary structure of the normal (PrP^C) protein as determined by NMR spectroscopy. The orange portions represent α-helical segments, and the blue portions are short β strands. The yellow dotted line represents the N-terminal portion of the polypeptide, which lacks defined structure. (*b*) A proposed model of the abnormal, infectious (PrP^Sc) prion protein, which consists largely of β-sheet. The actual tertiary structure of the prion protein has not been determined. The two molecules shown in this figure are formed by polypeptide chains that can be identical in amino acid sequence but fold very differently. As a result of the differences in folding, PrP^C remains soluble, whereas PrP^Sc produces aggregates that kill the cell. (The two molecules shown in this figure are called *conformers* because they differ only in conformation.) (A: COURTESY OF KÜRT WÜTHRICH; B: AFTER S. B. PRUSINER, TRENDS BIOCHEM. SCI. 21:483, 1996. COPYRIGHT 1996, WITH PERMISSION OF ELSEVIER SCIENCE.)

can occur in either an inherited or sporadic form. Like CJD, the brain of a person with Alzheimer's disease contains fibrillar deposits of an insoluble material referred to as *amyloid* (Figure 2). In both diseases, the fibrillar deposits result from the self-association of a polypeptide composed predominantly of β sheet. There are also many basic differences between the two diseases: the proteins that form the disease-causing aggregates are unrelated, the parts of the brain that are affected are distinct, and the protein responsible for AD is not an infectious agent (i.e., AD is *nontransmissable*).

Over the past two decades, research on AD has been dominated by the *amyloid hypothesis*, which contends that the disease is caused by the production of a molecule, called the *amyloid β-peptide* (*Aβ*).

Amyloid plaque   Neurofibrillary tangle (NFT)

(a)

Normal   Alzheimer's

Neurofibrillary tangles

Neuron

Amyloid Plaques

(b)
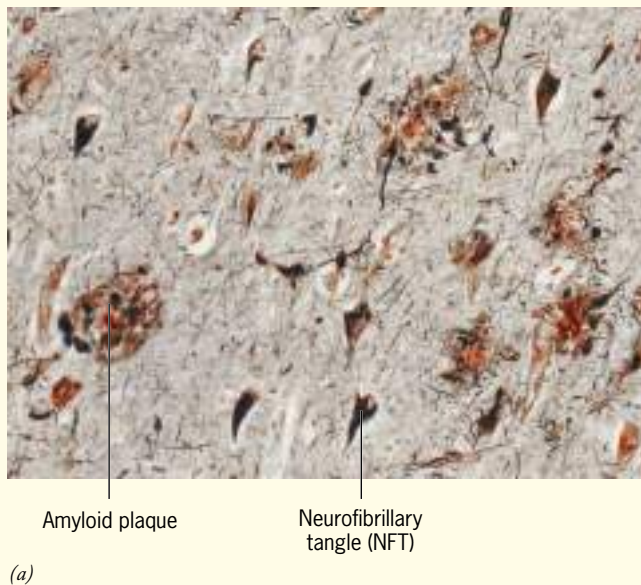
FIGURE 2 **Alzheimer's disease.** (*a*) The defining characteristics of brain tissue from a person who died of Alzheimer's disease. (*b*) Amyloid plaques containing aggregates of the Aβ peptide appear extracellularly (between nerve cells), whereas neurofibrillary tangles (NFTs) appear within the cells themselves. NFTs, which are discussed at the end of the Human Perspective, are composed of misfolded tangles of a protein called tau that is involved in maintaining the microtubule organization of the nerve cell. Both the plaques and tangles have been implicated as a cause of the disease. (A: © THOMAS DEERINCK, NCMIR/PHOTO RESEARCHERS, INC. B: © AMERICAN HEALTH ASSISTANCE FOUNDATION.)
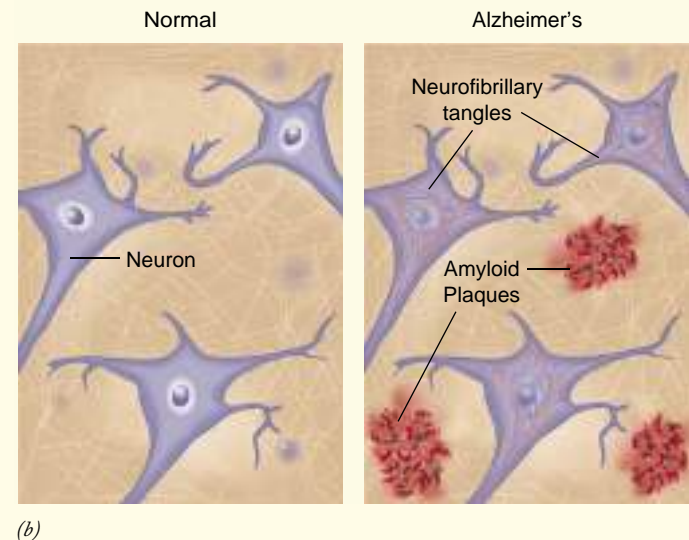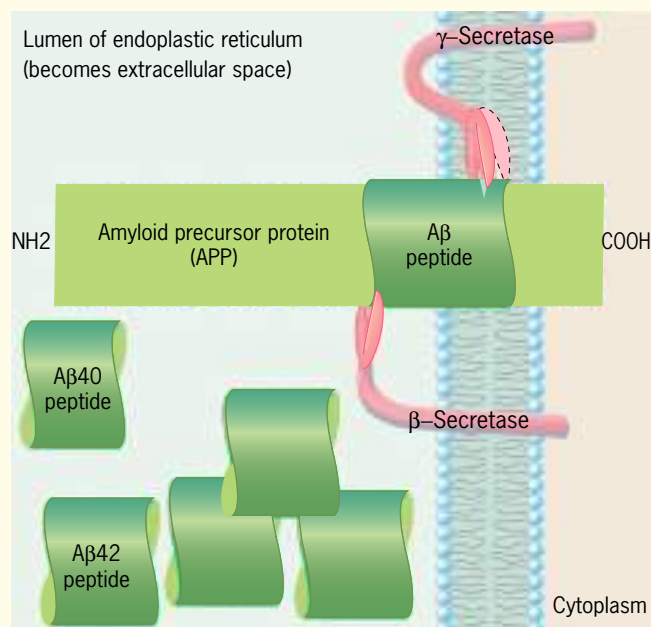
FIGURE 3 **Formation of the Aβ peptide.** The Aβ peptide is carved from the amyloid precursor protein (APP) as the result of cleavage by two enzymes, β-secretase and γ-secretase. It is interesting that APP and the two secretases are all proteins that span the membrane. Cleavage of APP occurs inside the cell (probably in the endoplasmic reticulum), and the Aβ product is ultimately secreted into the space outside of the cell. The γ-secretase can cut at either of two sites in the APP molecule, producing either Aβ40 or Aβ42 peptides, the latter of which is primarily responsible for production of the amyloid plaques seen in Figure 2. γ-Secretase is a multisubunit enzyme that cleaves its substrate at a site within the membrane.

Aβ is originally part of a larger protein called the *amyloid precursor protein* (*APP*), which spans the nerve cell membrane. The Aβ peptide is released from the APP molecule following cleavage by two specific enzymes, β-secretase and γ-secretase (Figure 3). The length of the Aβ peptide is somewhat variable. The predominant species has a length of 40 amino acids (designated as Aβ40), but a minor species with two additional hydrophobic residues (designated as Aβ42) is also produced. Both of these peptides can exist in a soluble form that consists predominantly of α helices, but Aβ42 has a tendency to spontaneously refold into a very different conformation that contains considerable β-pleated sheet. It is the misfolded Aβ42 version of the molecule that has the greatest potential to cause damage to the brain. Aβ42 tends to self-associate to form small complexes (oligomers) as well as large aggregates that are visible as fibrils in the electron microscope. Although the issue is far from settled, a body of evidence suggests that it is the soluble oligomers that are most toxic to nerve cells, rather than the insoluble aggregates. Cultured nerve cells, for example, are much more likely to be damaged by the presence of soluble Aβ oligomers than by either Aβ monomers or fibrillar aggregates. In the brain, the Aβ oligomers appear to attack the synapses that connect one nerve cell to another and eventually lead to the death of the nerve cells. Persons who suffer from an inherited form of AD carry a mutation that leads to an increased production of the Aβ42 peptide. Overproduction of Aβ42 can be caused by possession of extra copies (duplications) of

the *APP* gene, by mutations in the *APP* gene, or by mutations in genes (*PS1*, *PS2*) that encode subunits of γ-secretase. Individuals with such mutations exhibit symptoms of the disease at an early age, typically in their 50s. The fact that AD can be caused by mutations in genes that increase amyloid formation is the strongest argument favoring amyloid formation as the underlying basis of the disease. The strongest argument against the amyloid hypothesis is the weak correlation that can exist between the number and size of amyloid plaques in the brain and the severity of the disease. Elderly persons who show little or no sign of memory loss or dementia can have relatively high levels of amyloid deposits in their brain and those with severe disease can have little or no amyloid deposition.

All of the drugs currently on the market for the treatment of AD are aimed only at management of symptoms; none has any effect on stopping disease progression. With the amyloid hypothesis as the guiding influence, researchers have followed three basic strategies in the pursuit of new drugs for the prevention and/or reversal of mental decline associated with AD. These strategies are (1) to prevent the formation of the Aβ42 peptide in the first place; (2) to remove the Aβ42 peptide (or the amyloid deposits it produces) once it has been formed; and (3) to prevent the interaction between Aβ molecules, thereby preventing the formation of both oligomers and fibrillar aggregates. Before examining each of these strategies, we can consider how investigators can determine what type of drugs might be successful in the prevention or treatment of AD.

One of the best approaches to the development of treatments for human diseases is to find laboratory animals, particularly mice, that develop similar diseases, and use these animals to test the effectiveness of potential therapies. Animals that exhibit a disease that mimics a human disease are termed *animal models*. For whatever reason, the brains of aging mice show no evidence of the amyloid deposits found in humans, and, up until 1995, there was no animal model for AD. Then, in that year, researchers found that they could create a strain of mice that developed amyloid plaques in their brain and performed poorly at tasks that required memory. They created this strain by genetically engineering the mice to carry a mutant human *APP* gene, one responsible for causing AD in families. These genetically engineered (*transgenic*) mice have proven invaluable for testing potential therapies for AD. The greatest excitement in the field of AD therapeutics has centered on the second strategy mentioned above, and we can use these investigations to illustrate some of the steps required in the development of a new drug.

In 1999, Dale Schenk and his colleagues at Elan Pharmaceuticals published an extraordinary finding. They had discovered that the formation of amyloid plaques in mice carrying the mutant human *APP* gene could be blocked by repeatedly injecting the animals with the very same substance that causes the problem, the aggregated Aβ42 peptide. In effect, the researchers had immunized (i.e., vaccinated) the mice against the disease. When young (6-week-old) mice were immunized with Aβ42, they failed to develop the amyloid brain deposits as they grew older. When older (13-month-old) mice whose brains already contained extensive amyloid deposits were immunized with the Aβ42, a significant fraction of the fibrillar deposits was cleared out of the nervous system. Even more importantly, the immunized mice performed better than their nonimmunized littermates on memory-based tests.

The dramatic success of these experiments on mice, combined with the fact that the animals showed no ill effects from the immunization procedure, led government regulators to quickly approve a Phase I clinical trial of the Aβ42 vaccine. A Phase I clinical trial is the first step in testing a new drug or procedure in humans and usu-

ally comes after years of preclinical testing on cultured cells and animal models. Phase I tests are carried out on a small number of subjects and are designed to monitor the safety of the procedure rather than its effectiveness against the disease.

None of the subjects in two separate Phase I trials of the Aβ vaccine showed any ill-effects from the injection of the amyloid peptide. As a result, the investigators were allowed to proceed to a Phase II clinical trial, which involves a larger group of subjects and is designed to obtain a measure of the effectiveness of the procedure (or drug). This particular Phase II trial was carried out as a randomized, double-blind, placebo-controlled study. In this type of study:

1. the patients are *randomly* divided into two groups that are treated similarly except that one group is given the curative factor (protein, antibodies, drugs, etc.) being investigated and the other group is given a *placebo* (an inactive substance that has no therapeutic value), and
2. the study is *double-blinded*, which means that neither the researchers nor patients know who is receiving treatment and who is receiving the placebo.

The Phase II trial for the Aβ vaccine began in 2001 and enrolled more than 350 individuals in the United States and Europe who had been diagnosed with mild to moderate AD. After receiving two injections of synthetic β-amyloid (or a placebo), 6 percent of the subjects experienced a potentially life-threatening inflammation of the brain. Most of these patients were successfully treated with steroids, but the trial was discontinued.

Once it had become apparent that vaccination of patients with Aβ42 had inherent risks, it was decided to pursue a safer form of immunization therapy, which is to administer antibodies directed against Aβ that have been produced outside the body. This type of approach is known as *passive immunization* because the person does not produce the therapeutic antibodies themselves. Passive immunization with an anti-Aβ42 antibody (called bapineuzumab) had already proven capable of restoring memory function in transgenic mice and was quickly shown to be safe, and apparently effective, in Phase I and II clinical trials. The last step before government approval is a Phase III trial, which typically employs large numbers of subjects (a thousand or more at several research centers) and compares the effectiveness of the new treatment against standard approaches. The first results of the Phase III trials on bapineuzumab were reported in 2008 and were disappointing; there was little or no evidence that the antibody provided benefits in preventing the progression of the disease. However, another trial on a subset of AD patients is continuing.

Meanwhile, a comprehensive analysis of some of the patients who had been vaccinated with Aβ42 in the original immunization trial from 2001 was also reported in 2008. Analysis of this patient group indicated that the Aβ42 vaccination had had no effect on preventing disease progression. It was particularly striking that in several of these patients who had died of severe dementia, there were virtually no amyloid plaques left in their brains. This finding strongly suggests that removal of amyloid deposits in a patient already suffering the symptoms of mild-to-moderate dementia does not stop disease progression. These results can be interpreted in more than one way. One interpretation is that the amyloid deposits are not the cause of the symptoms of dementia. An alternate interpretation is that irreversible toxic effects of the deposits had already occurred by the time immunization had begun and it was too late to reverse the disease course using treatments that remove existing amyloid deposits. It is important to note, in this regard, that the formation of amyloid deposits in the

brain begins 10 or more years before any clinical symptoms of AD are reported. It is possible that, if these treatments had started earlier, the symptoms of the disease might never have appeared. Recent advances in brain-imaging procedures now allow clinicians to observe amyloid deposits in the brains of individuals long before any symptoms of AD have developed. Because of these advances, it may be possible to begin preventive treatments in persons who are at very high risk of developing AD before they develop symptoms.

Drugs have also been developed that follow the other two strategies outlined above. Two drugs that had reached the most advanced stages of development were Alzhemed and Flurizan. Alzhemed is a small molecule designed to bind to Aβ peptides and block certain interactions, thereby stopping molecular aggregation and fibril formation. A large Phase III study on Alzhemed has failed to demonstrate that the drug is effective in stopping disease progression. Flurizan is a nonsteroidal anti-inflammatory drug, or NSAID, like ibuprofen. NSAIDs first came to the attention of Alzheimer's researchers when it was reported that arthritis patients who were taking certain types of NSAIDs were much less likely to develop AD. Subsequent research revealed that those NSAIDs with apparent activity in preventing AD were also ones that altered the activity of γ-secretase (Figure 3). In preclinical studies, Flurizan had been shown to inhibit the production of toxic Aβ peptides, both in cultured nerve cells and in transgenic AD mice. However, like the other drugs discussed in this section, a large Phase III study of Fluirizan has failed to show any benefit in stopping AD progression.

Taken collectively, the apparent failure of these three promising drugs, each aimed at a different step in the AD pathway, has left the field of AD therapeutics without a clear pathway. Some pharmaceutical companies are continuing to develop new drugs aimed at blocking the formation of amyloid aggregates, whereas others are moving in different directions. These findings also raise a more basic question: Is the Aβ peptide even part of the underlying mechanism that leads to AD? It hasn't been mentioned, but Aβ is not the only misfolded protein found in the brains of persons with AD. Another protein called tau, which functions as part of a nerve cell's cytoskeleton (Section 9.3), develops into bundles of tangled cellular filaments called neurofibrillary tangles (or NFTs) (Figure 2). NFTs have been shown to cause certain types of dementia, but they have been largely ignored as a causative factor in AD pathogenesis, due primarily to the fact that the transgenic AD mouse models discussed above do not develop NFTs. If one extrapolates the results of these mouse studies to humans, they suggest that NFTs are not required for the cognitive decline that occurs in patients with AD. A great deal of work on AD has been based on transgenic mice carrying human AD genes. These animals have served as the primary preclinical subjects for testing AD drugs, and they have been used extensively in basic research that aims to understand the disease mechanisms responsible for the development of AD. But many questions have been raised as to how accurately these animal models mimic the disease in humans, particularly the sporadic human cases in which affected individuals lack the mutant genes that cause the animals to develop the corresponding disorder. In fact, the most promising new drug at the time of this writing is one that acts on NFTs rather than β-amyloid. In this case, the drug methylthioninium chloride (brand name remberTM), which dissolves NFTs, was tested on a group of more than 300 patients with mild to moderate AD in a Phase II trial. The drug was found to reduce mental decline over a period of one year by an average of 81 percent compared to patients receiving a placebo. The drug is now being tested in a larger Phase III study. (The results of studies on these and other treatments can be examined at www.alzforum.org/dis/tre/drc)
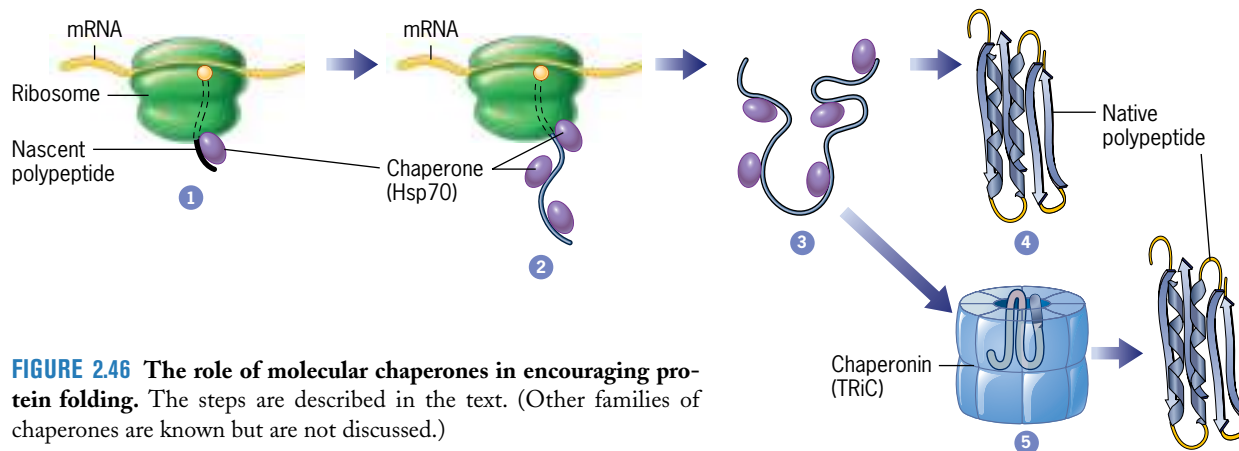
FIGURE 2.46 **The role of molecular chaperones in encouraging protein folding.** The steps are described in the text. (Other families of chaperones are known but are not discussed.)

Figure 2.46 depicts the activities of two families of molecular chaperones that operate in the cytosol of eukaryotic cells. Polypeptide chains are synthesized on ribosomes by the addition of amino acids, one at a time, beginning at the chain's N-terminus (step 1, Figure 2.46). Chaperones of the Hsp70 family bind to elongating polypeptide chains as they emerge from an exit channel within the large subunit of the ribosome (step 2). Hsp70 chaperones are thought to prevent these partially formed polypeptides (i.e., *nascent* polypeptides) from binding to other proteins in the cytosol, which would cause them either to aggregate or misfold. Once their synthesis has been completed (step 3), many of these proteins are simply released by the chaperone into the cytosol where they spontaneously fold into their native state (step 4). Many of the larger polypeptides are transferred from Hsp70 proteins to a different type of chaperone called a *chaperonin* (step 5). Chaperonins are cylindrical protein complexes that contain chambers in which newly synthesized polypeptides can fold without interference from other macromolecules in the cell. TRiC is a chaperonin thought to assist in the folding of up to 15 percent of the polypeptides synthesized in mammalian cells. The discovery and mechanism of action of Hsp70 and chaperonins are discussed in the Experimental Pathways on page 78.

The Emerging Field of Proteomics With all of the attention on genome sequencing in recent years, it is easy to lose sight of the fact that genes are primarily information storage units, whereas proteins orchestrate cellular activities. Genome sequencing provides a kind of "parts list." The human genome probably contains between 20,000 and 22,000 genes, each of which can potentially give rise to a variety of different proteins.[6] To date, only a fraction of these molecules have been characterized.

The entire inventory of proteins that is produced by an organism, whether human or otherwise, is known as that organism's **proteome**. The term *proteome* is also applied to the inventory of all proteins that are present in a particular tissue, cell, or cellular organelle. Because of the sheer numbers of proteins that are currently being studied, investigators have sought to develop techniques that allow them to determine the properties or activities of a large number of proteins in a single experiment. A new term—**proteomics**—was coined to describe the expanding field of protein biochemistry. This term carries with it the concept that advanced technologies and high-speed computers are used to perform large-scale studies on diverse arrays of proteins. This is the same basic approach that has proven so successful over the past decade in the study of genomes. But the study of proteomics is inherently more difficult than the study of genomics because proteins are more difficult to work with than DNA. In physical terms, one gene is pretty much the same as all other genes, whereas each protein has unique chemical properties and handling requirements. In addition, small quantities of a particular DNA segment can be expanded greatly using readily available enzymes, whereas protein quantities cannot be increased. This is particularly troublesome when one considers that many of the proteins regulating important cellular processes are present in only a handful of copies per cell.

Traditionally, protein biochemists have sought to answer a number of questions about particular proteins. These include: What specific activity does the protein demonstrate in vitro, and how does this activity help a cell carry out a particular function such as cell locomotion or DNA replication? What is the protein's three-dimensional structure? When does the protein appear in the development of the organism and in which types of cells? Where in the cell is it localized? Is the protein modified after synthesis by the addition of chemical groups (e.g., phosphates or sugars) and, if so, how does this modify its activity? How much of the protein is present, and how long does it survive before being degraded? Does the level of the protein change during physiologic activities or as the result of disease? Which other proteins in the cell does it interact with? Biologists have been attempting to answer these questions for decades but, for the most part,

---

[6]There are a number of ways that a single gene can give rise to more than one polypeptide. Two of the most prominent mechanisms, alternative splicing and posttranslational modification, are discussed in other sections of the text. It can also be noted that many proteins have more than one distinct function. Even myoglobin, which has long been studied as an oxygen-storage protein, has recently been shown to be involved in the conversion of nitric oxide (NO) to nitrate (NO_3^-).

FIGURE 2.47 (Art © Joe Sutliff.)

they've been doing it one protein at a time. Proteomics researchers attempt to answer similar questions on a more comprehensive scale using automated, large-scale techniques (Figure 2.47). We have already seen how a systematic approach can be used to investigate protein-protein interactions (page 61). Let us turn our attention to the seemingly daunting task of identifying the vast array of proteins produced by a particular cell.

Figure 2.48 shows portions of two gels that were used to fractionate (separate) a mixture of proteins extracted from the same part of the brain of a human (Figure 2.48*a*) or a chimpanzee (Figure 2.48*b*). The proteins have been fractionated by a procedure called two-dimensional polyacrylamide gel electrophoresis (described in detail in Section 18.7). The gels contain hundreds of different stained spots, each of which comprises a single protein, or at most a few proteins that have very similar physical properties and are therefore difficult to separate from one another. This technique, which was invented in the mid-1970s, was the first and is still one of the best ways to separate large numbers of proteins in a mixture. It is evident that virtually all of the spots on one gel have a counterpart in the other gel; those spots that are present in both gels correspond to homologous proteins shared by the two species. Certain spots are labeled by red numbers. Numbered spots correspond to proteins that exhibit differences in the two gels, either because (1) the proteins have migrated to a slightly different position (as the result of a difference in modification or amino acid sequence in the protein between the two organisms) (indicated by the double-headed arrows) or (2) the proteins are present in noticeably different amount in the brains of the two organisms (indicated by the green arrows). In either case, this type of proteomic experiment provides an overview of the difference in protein expression in our brain compared to that of our closest living relative.

It is one thing to separate proteins from one another, but quite another to identify each of the separated molecules. In the past few years, two technologies—mass spectrometry and high-speed computation—have come together to make it possible to identify any or all of the proteins present in a gel such as those shown in Figure 2.48. As discussed in further detail in Section 18.7, mass spectrometry is a technique to determine the precise mass of a molecule or fragment of a molecule, which can then be used to identify that molecule. Suppose that we wanted to identify the protein present in the spot that is indicated by the red number 1. The protein

(a)

(b)

FIGURE 2.48 The study of proteomics often requires the separation of complex mixtures of proteins. The two electrophoretic gels shown here contain proteins extracted from the frontal cortex of humans (*a*) or chimpanzees (*b*). The numbered spots represent homologous proteins that show distinct differences in the two gels as discussed in the text. (*Note:* This is only a small portion of a much larger gel that contains approximately 8500 spots that represent proteins synthesized by a primate brain.) (*Note:* The animals in this study had died of natural causes.) (from W. Enard et al., Science 296:342, 2002, courtesy of Svante Pääbo. Copyright © 2002 American Association for the Advancement of Science.)

that makes up the spot in question can be removed from the gel and digested into peptides with the enzyme trypsin. When these peptides are introduced into a mass spectrometer, they are converted into gaseous ions and separated according to their mass/charge (m/z) ratio. The results are displayed as a series of peaks of known m/z ratio, such as that shown in Figure 2.49. The pattern of peaks constitutes a highly characteristic *peptide mass fingerprint* of that protein. But how can a protein be identified based on its peptide mass fingerprint?

The generation of peptide mass fingerprints is one element of the new proteomic technology. Another is based on advances in computer technology. Once a genome has been sequenced, the amino acid sequences of encoded proteins can be predicted. This list of "virtual proteins" can then be subjected to a theoretical trypsin digestion and the masses of the
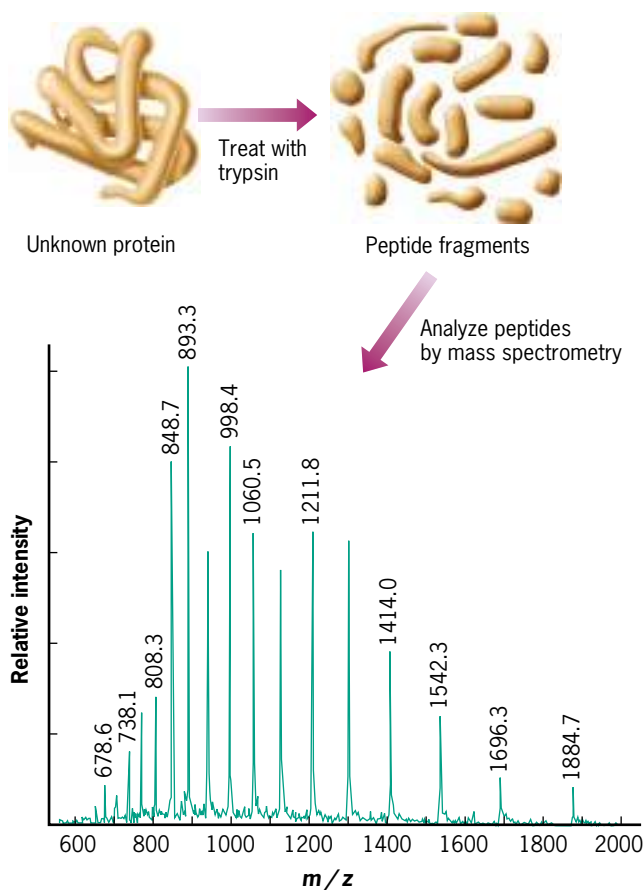
FIGURE 2.49 **Identifying proteins by mass spectrometry.** A protein is isolated from a source (such as one of the spots on one of the gels of Figure 2.48) and subjected to digestion by the enzyme trypsin. The peptide fragments are then introduced into a mass spectrometer where they are ionized and separated according to their mass/charge (m/z) ratio. The separated peptides appear as a pattern of peaks whose precise m/z ratio is indicated. A comparison of these ratios to those obtained by a theoretical digest of virtual proteins encoded by the genome allows researchers to identify the protein being studied. In this case, the MS spectrum is that of horse myoglobin lacking its heme group. (DATA REPRINTED FROM J. R. YATES, METHODS ENZYMOL. 271:353, 1996. COPYRIGHT 1996, WITH PERMISSION FROM ELSEVIER.)

resulting virtual peptides calculated and entered into a database. Once this has been done, the actual peptide masses of a purified protein obtained by the mass spectrometer can be compared using high-speed supercomputers to the masses predicted by theoretical digests of all polypeptides encoded by the genome. In most cases, the protein that has been isolated and subjected to mass spectrometry can be directly identified based on this type of database search. The protein labeled number 1 in the gels of Figure 2.48, for example, happens to be the enzyme aldose reductase. Mass spectrometers are not restricted to handling one purified protein at a time, but are also capable of analyzing proteins present in complex mixtures (Section 18.7). Mass spectrometry is particularly useful in revealing how the protein complement of a cell or tissue changes over time, as might occur, for example, after the secretion of a hormone within the body, or after taking a drug, or during a particular disease.

Many clinical researchers believe that proteomics will play an important role in advancing the practice of medicine. It is thought that most human diseases leave telltale patterns (or *biomarkers*) among the thousands of proteins present in the blood or other bodily fluids. Many efforts have been made to compare the proteins present in the blood of healthy individuals with those present in the blood of persons suffering from various diseases, especially cancer. For the most part, the results of these biomarker searches have proven generally unreliable in that the findings of one research group cannot be duplicated by the efforts of other groups. The primary difficulty stems from the fact that human blood serum is such a complex solution containing thousands of proteins that range in abundance over 9 or 10 orders of magnitude. To date, the level of technical and computational sophistication has not been up to the daunting task at hand, but that may be changing. It is hoped that, one day, it will be possible to use a single blood test to reveal the existence of early-stage heart, liver, or kidney disease that can be treated before it becomes a life-threatening condition.

Protein separation and mass spectrometric techniques don't tell us anything about a protein's function. Researchers have been working to devise techniques that allow protein function to be determined on a large scale, rather than one protein at a time. Several new technologies have been developed to accomplish this mission; we will consider only one—the use of *protein microarrays* (or *protein chips*). A protein microarray uses a solid surface, typically a glass microscope slide, which is covered by microscopic-sized spots, each containing an individual protein sample. Protein microarrays are constructed by the application of tiny volumes of individual proteins to specific sites on the slide, generating an array of proteins such as that shown in Figure 2.50a. The 6000 or so yeast proteins encoded by the yeast genome can fit comfortably as individual spots on a single glass slide.

Once a protein microarray has been created, the proteins that it contains can be screened for various types of activities. Let's consider the role of Ca^{2+} ions for a moment. Ca^{2+} ions play a key role in many activities, such as the formation of nerve impulses, the release of hormones into the blood, and
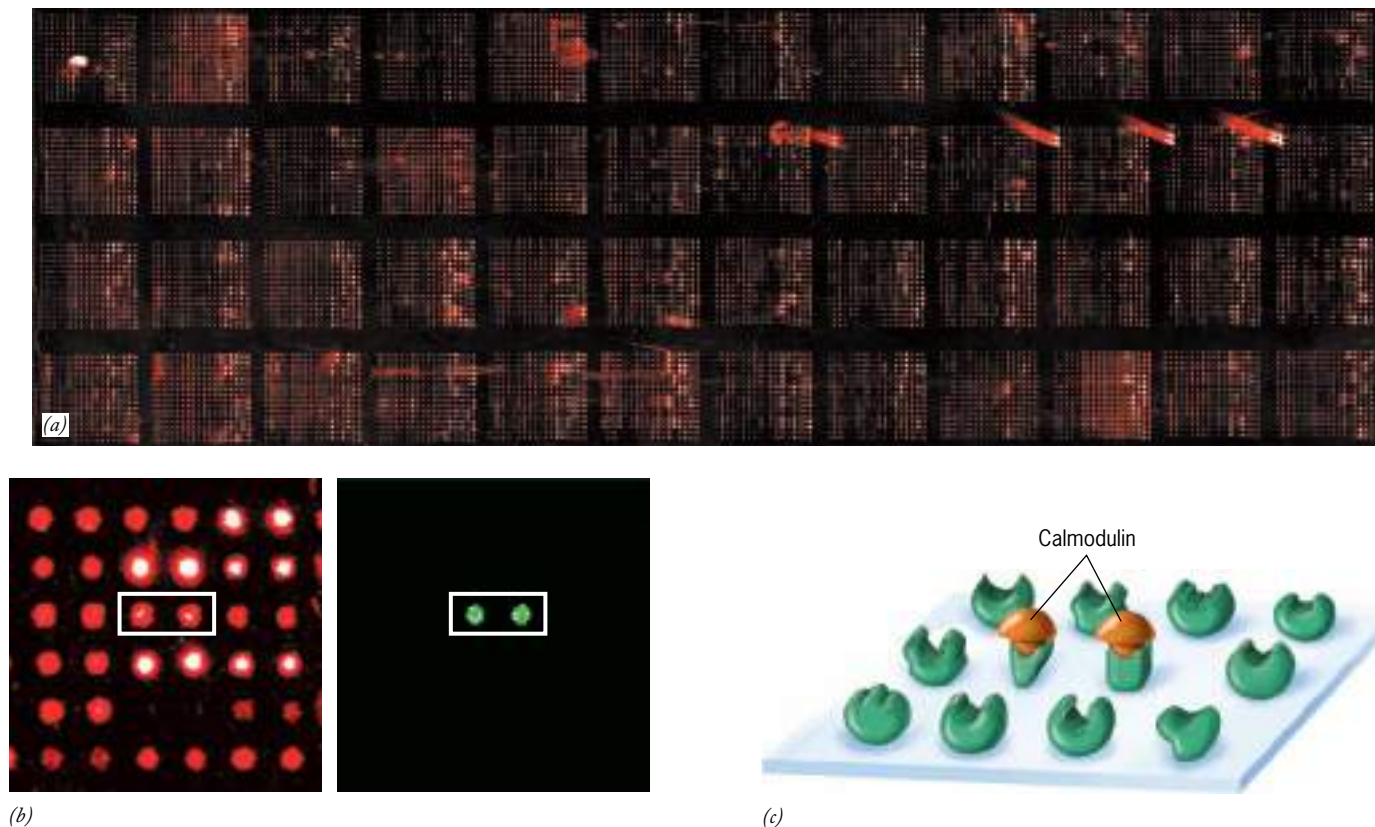
FIGURE 2.50 Global analysis of protein activities using protein chips.
(*a*) This single microscope slide contains 5800 different yeast proteins spotted in duplicate. The proteins spotted on the slide were synthesized in genetically engineered cells. The spots display red fluorescence because they have been incubated with a fluorescent antibody that can bind to all of the proteins in the array. (*b*) The left image shows a small portion of the protein array depicted in part *a*. The right image shows the same portion of the array following incubation with calmodulin in the presence of calcium ions. The two proteins exhibiting green fluorescence are calmodulin-binding proteins. (*c*) Schematic illustration of the events occurring in part *b* where calmodulin has bound to two proteins of the microarray that have complementary binding sites. (A,B: FROM H. ZHU, ET AL., SCIENCE 293:2101, 2001, COURTESY OF MICHAEL SNYDER. COPYRIGHT © 2001 AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

muscle contraction. In each of these cases, Ca^{2+} ions act by attaching to a calcium-binding protein. As discussed in Chapter 15, calmodulin is an important calcium-binding protein, even in single-celled yeast. Figure 2.50*b* shows a small portion of a yeast-protein microarray that had been incubated with the protein calmodulin in the presence of Ca^{2+} ions. Those proteins in the array that display green fluorescence are ones that had bound calmodulin during the incubation and thus are likely to be involved in calcium signaling activity. Thirty-three new calmodulin-binding proteins were discovered in this particular protein screening experiment. These types of studies open the door to the analysis of the properties of large numbers of proteins of unknown function.

Protein chips are also expected to be used one day by clinical laboratories to screen for proteins that are characteristic of particular disorders. The simplest way to determine whether a particular protein characteristic of a disease is present in a blood or urine sample, and how much of that protein is present, is to measure the protein's interaction with a specific antibody. This is the basis, for example, of the PSA test used in routine screening of men for prostate cancer. PSA is a protein that is found in the blood of normal men, but is pres-

ent at elevated levels in individuals with prostate cancer. PSA levels are determined by measuring the amount of protein in the blood that binds to anti-PSA antibodies. It is expected in the near future that biotechnology companies will be manufacturing microarrays that contain antibodies capable of binding a number of different blood proteins. The presence and/or amount of these proteins would indicate that a person may be suffering from one or another of a wide variety of diseases.

Protein Engineering Advances in molecular biology have created the opportunity to design and mass-produce novel proteins that are different from those made by living organisms. It is possible with current DNA-synthesizing techniques to create an artificial gene that can be used in the production of a protein having any desired sequence of amino acids. Polypeptides can also be synthesized from "scratch" using chemical techniques. This latter strategy allows researchers to incorporate building blocks other than the 20 amino acids that normally occur in nature.

The problem with these types of engineering efforts is in knowing which of the virtually infinite variety of possible proteins one could manufacture might have some useful function.

Consider, for example, a pharmaceutical company that wanted to manufacture a protein that would bind to the surface of the AIDS virus and remove it from an aqueous solution, such as the bloodstream. Assume that computer simulation programs could predict the shape such a protein should have to bind to the viral surface. What sequence of amino acids strung together would produce such a protein? The answer requires detailed insight into the rules governing the complex relationship between a protein's primary structure and its tertiary structure.

Given the magnitude of the task, it came as a surprise when researchers reported in 2008 that they had successfully designed and produced artificial proteins that were capable of catalyzing two different organic reactions, neither of which was catalyzed by any known natural enzyme. One of the reactions involved breaking a carbon–carbon bond, and the other the transfer of a proton from a carbon atom. These protein architects began by choosing a catalytic mechanism that might accelerate each chosen reaction and then used computer-based calculations to construct an idealized space in which amino acid side chains were positioned (forming an *active site*) to accomplish the task. They then searched among known protein structures to find ones that might serve as a framework or scaffold that could hold the active site they had designed. To transform the computer models into an actual protein, they used computational techniques to generate DNA sequences that had the potential to encode such a protein. The proposed DNA molecules were synthesized and introduced into bacterial cells where the proteins were manufactured. The catalytic activities of the proteins were then tested. Those proteins that showed the greatest promise were then subjected to a process of test-tube evolution; the proteins were mutated to create a new generation of altered proteins, which could in turn be screened for enhanced activity. Eventually, the team obtained proteins that could accelerate the rates of reaction as much as one million times the uncatalyzed reaction. While this is not a rate of enhancement that would fill a natural enzyme with pride, it is a remarkable accomplishment for a team of biochemists. It suggests, in fact, that scientists will ultimately be able to construct proteins from scratch that will be capable of catalyzing virtually any chemical reaction.

An alternate approach to the production of novel proteins has been to modify those that are already produced by cells. Recent advances in DNA technology have allowed investigators to isolate an individual gene from human chromosomes, to alter its information content in a precisely determined way, and to synthesize the modified protein with its altered amino acid sequence. This technique, which is called **site-directed mutagenesis** (Section 18.18), has many different uses, both in basic research and in applied biology. If, for example, an investigator wants to know about the role of a particular residue in the folding or function of a polypeptide, the gene can be mutated in a way that substitutes an amino acid with different charge, hydrophobic character, or hydrogen-bonding properties. The effect of the substitution on the structure and function of the modified protein can then be determined. As

we will see throughout this textbook, site-directed mutagenesis is proving invaluable in the analysis of the specific functions of minute parts of virtually every protein of interest to biologists.

Site-directed mutagenesis is also used to modify the structure of clinically useful proteins to bring about various physiological effects. For example, the drug Somavert, which was approved by the FDA in 2003, is a modified version of human growth hormone (GH) containing several alterations. GH normally acts by binding to a receptor on the surface of target cells, which triggers a physiological response. Somavert competes with GH in binding to the GH receptor, but interaction between drug and receptor fails to trigger the cellular response. Somavert is prescribed for the treatment of acromegaly, a disorder that results from excess production of growth hormone.

Structure-Based Drug Design   The production of new proteins is one clinical application of recent advances in molecular biology; another is the development of new drugs that act by binding to known proteins, thereby inhibiting their activity. Drug companies have access to chemical "libraries" that contain millions of different organic compounds that have been either isolated from plants or microorganisms or chemically synthesized. One way to search for potential drugs is to expose the protein being targeted to combinations of these compounds and determine which compounds, if any, happen to bind to the protein with reasonable affinity. An alternate approach, which can be employed if the tertiary structure of a protein has been determined, is to use computers to design "virtual" drug molecules whose size and shape might allow them to fit into the apparent cracks and crevices of the protein, rendering it inactive.

We can illustrate both of these technologies by considering the development of the drug Gleevec, as depicted in Figure 2.51*a*. Introduction of Gleevec into the clinic has revolutionized the treatment of a number of relatively rare cancers, most notably that of chronic myelogenous leukemia (CML). As discussed at length in Chapters 15 and 16, a group of enzymes called tyrosine kinases are often involved in the transformation of normal cells into cancer cells. Tyrosine kinases catalyze a reaction in which a phosphate group is added to specific tyrosine residues within a target protein, an event that may activate or inhibit the target protein. The development of CML is driven almost single-handedly by the presence of an overactive tyrosine kinase called ABL.

During the 1980s researchers identified a compound called 2-phenylaminopyrimidine that was capable of inhibiting tyrosine kinases. This compound was discovered by randomly screening a large chemical library for compounds that exhibited this particular activity (Figure 2.51*a*). As is usually the case in these types of blind screening experiments, 2-phenylaminopyrimidine would not have made a very effective drug. For one reason, it was only a weak enzyme inhibitor, which meant it would have had to be used in very large quantities. 2-Phenylaminopyrimidine is described as a *lead* compound, a starting point from which usable drugs might be

(a)

(b)

(c)   *(d)*
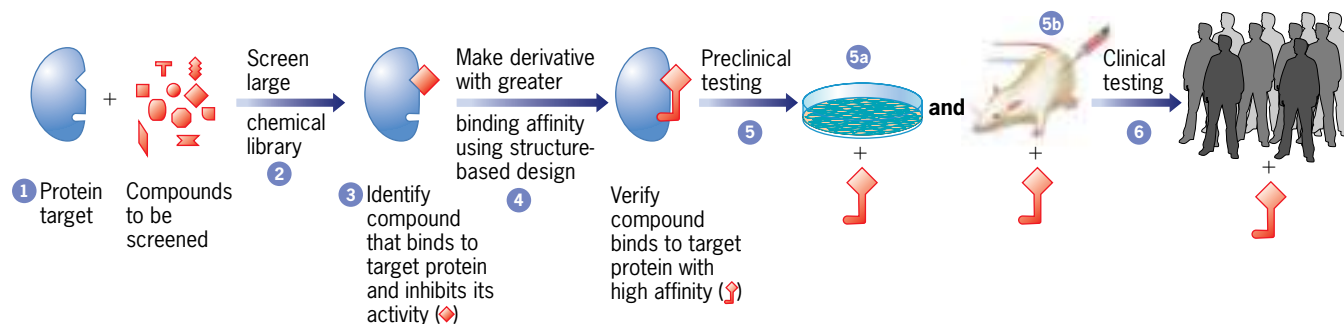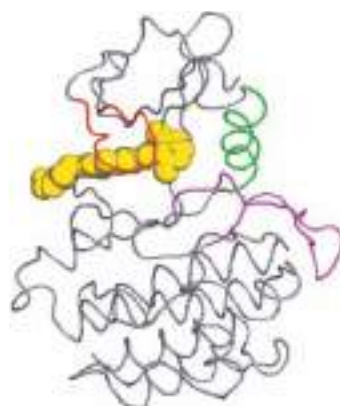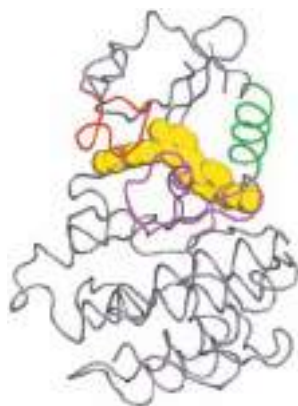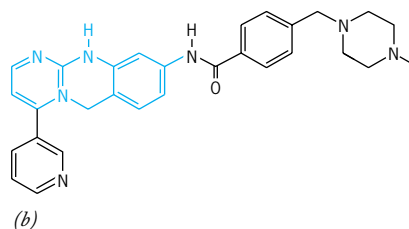
FIGURE 2.51 Development of a protein-targeting drug, such as Gleevec. (*a*) Typical steps in drug development. In step 1, a protein (e.g., ABL) has been identified that plays a causative role in the disease. This protein is a likely target for a drug that inhibits its enzymatic activity. In step 2, the protein is incubated with thousands of compounds in a search for ones that bind with reasonable affinity and inhibit its activity. In step 3, one such compound (e.g, 2-phenylaminopyrimidine in the case of ABL) has been identified. In step 4, knowledge of the structure of the target protein is used to make derivatives of the compound (e.g., Gleevec) that have greater binding affinity and thus can be used at lower concentrations. In step 5, the compound in question is tested in preclinical experiments for toxicity and efficacy (level of effectiveness) in vivo. Preclinical experiments are typically carried out on cultured human cells (step 5a) (e.g., those from patients with CML) and laboratory animals (step 5b) (e.g., mice carrying transplants of human CML cells). If the drug appears safe and effective in animals, the drug is tested in clinical trials (step 6) as discussed on page 67. (*b*) The structure of Gleevec. The blue portion of the molecule indicates the structure of the compound 2-phenylaminopyrimidine that was initially identified as an ABL kinase inhibitor. (*c,d*) The structure of the catalytic domain of ABL in complex (*c*) with Gleevec (shown in yellow) and (*d*) with a second-generation inhibitor called Sprycel. Gleevec binds to the inactive conformation of the protein, whereas Sprycel binds to the active conformation. Both binding events block the activity that is required for the cell's cancerous phenotype. Sprycel is effective against cancer cells that have become resistant to the action of Gleevec. (C,D: REPRINTED WITH PERMISSION FROM ELLEN WEISBERG ET AL., COURTESY OF JAMES D. GRIFFIN, NATURE REVS. CANCER 7:353, 2007 © COPYRIGHT 2007, BY MACMILLAN MAGAZINES LIMITED.)

developed. Beginning with this lead molecule, compounds of greater potency and specificity were synthesized using structure-based drug design. One of the compounds to emerge from this process was Gleevec (Figure 2.51*b*), which was found to bind tightly to the inactive form of the ABL tyrosine kinase and prevent the enzyme from becoming activated, which is a necessary step if the cell is to become cancerous. The complementary nature of the interaction between the drug and its enzyme target is shown in Figure 2.51*c*. Preclinical studies demonstrated that Gleevec strongly inhibited the growth in the laboratory of cells from CML patients and that the compound showed no harmful effects in tests in animals. In the very first clinical trial of Gleevec, all 31 CML patients went into remission after taking once-daily doses of the compound. Gleevec has gone on to become the primary drug prescribed for treatment of CML, but this is not the end of the story. Many patients taking Gleevec experience

a recurrence of their cancer as the ABL kinase becomes resistant to the drug. In such cases, the cancer can continue to be suppressed by treatment with more recently designed drugs that are capable of inhibiting Gleevec-resistant forms of the ABL kinase. One of these newer (second-generation) ABL kinase inhibitors is shown bound to the protein in Figure 2.51*d*.

Protein Adaptation and Evolution   Adaptations are traits that improve the likelihood that an organism will survive in a particular environment. Proteins are biochemical adaptations that are subject to natural selection and evolutionary change in the same way as other types of characteristics, such as eyes or skeletons. This is best revealed by comparing evolutionarily related (*homologous*) proteins in organisms living in very different environments. For example, the proteins of halophilic (salt-loving) archaebacteria possess amino acid sub-
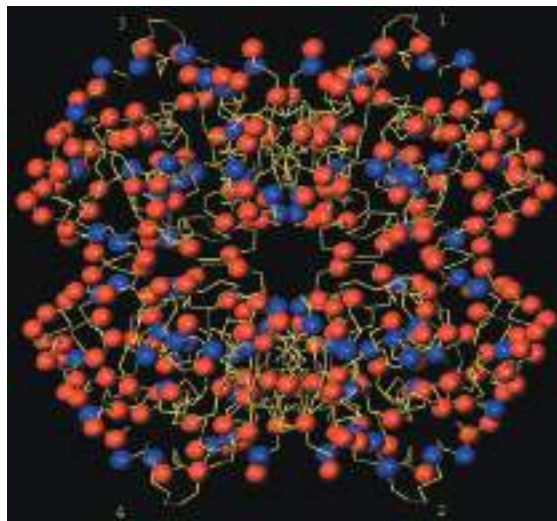
FIGURE 2.52 **Distribution of polar, charged amino acid residues in the enzyme malate dehydrogenase from a halophilic archaebacterium.** Red balls represent acidic residues, and blue balls represent basic residues. The surface of the enzyme is seen to be covered with acidic residues, which gives the protein a net charge of -156, and promotes its solubility in extremely salty environments. For comparison, a homologous protein from the dogfish, an ocean-dwelling shark, has a net charge of $+16$. (REPRINTED WITH PERMISSION FROM O. DYM, M. MEVARECH, AND J. L. SUSSMAN, SCIENCE 267:1345, 1995. COPYRIGHT © 1995 AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

stitutions that allow them to maintain their solubility and function at very high cytosolic salt concentrations (up to 4 M KCl). Unlike its counterpart in other organisms, the surface of the halophilic version of the protein malate dehydrogenase, for example, is coated with aspartic and glutamic acid residues whose carboxyl groups can compete with the salt for water molecules (Figure 2.52).

Homologous proteins isolated from different organisms can exhibit virtually identical shapes and folding patterns, but show strikingly divergent amino acid sequences. The greater the evolutionary distance between two organisms, the greater the difference in the amino acid sequences of their proteins. In some cases, only a few key amino acids located in a critical portion of the protein will be present in all of the organisms from which that protein has been studied. In one comparison of 226 globin sequences, only two residues were found to be absolutely conserved in all of these polypeptides; one is a histidine residue that plays a key role in the binding and release of O_2. These observations indicate that the secondary and tertiary structures of proteins change much more slowly during evolution than their primary structures.

We have seen how evolution has produced different versions of proteins in different organisms, but it has also produced different versions of proteins in individual organisms. Take a particular protein with a given function, such as globin or collagen. Several different versions of each of these proteins are encoded by the human genome. In most cases, different versions of a protein, which are known as **isoforms**, are

adapted to function in different tissues or at different stages of development. For example, humans possess six different genes encoding isoforms of the contractile protein actin. Two of these isoforms are found in smooth muscle, one in skeletal muscle, one in heart muscle, and two in virtually all other types of cells.

As more and more amino acid sequences and tertiary structures of proteins were reported, it became apparent that most proteins are members of much larger **families** (or *super-families*) of related molecules. The genes that encode the various members of a protein family are thought to have arisen from a single ancestral gene that underwent a series of duplications during the course of evolution (see Figure 10.23). Over long periods of time, the nucleotide sequences of the various copies diverge from one another to generate proteins with related structures. Many protein families contain a remarkable variety of proteins that have evolved diverse functions. The expansion of protein families is responsible for much of the protein diversity encoded in the genomes of today's complex plants and animals.

Nucleic Acids

Nucleic acids are macromolecules constructed out of long chains (**strands**) of monomers called **nucleotides**. Nucleic acids function primarily in the storage and transmission of genetic information, but they may also have structural or catalytic roles. There are two types of nucleic acids found in living organisms, **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. DNA serves as the genetic material of all cellular organisms, though RNA carries out that role for many viruses. In cells, information stored in DNA is used to govern cellular activities through the formation of RNA messages. In the present discussion, we will examine the basic structure of nucleic acids using single-stranded RNA as the representative molecule. We will look at the structure of DNA in Chapter 10, where it can be tied to its central role in the chemical basis of life.

Each nucleotide in a strand of RNA consists of three parts (Figure 2.53*a*): (1) a five-carbon sugar, ribose; (2) a nitrogenous base (so called because nitrogen atoms form part of the rings of the molecule); and (3) a phosphate group. The sugar and nitrogenous base together form a *nucleoside*, so that the nucleotides of an RNA strand are also known as ribonucleoside monophosphates. The phosphate is linked to the 5′ carbon of the sugar, and the nitrogenous base is attached to the sugar's 1′ carbon. During the assembly of a nucleic acid strand, the hydroxyl group attached to the 3′ carbon of the sugar of one nucleotide becomes linked by an ester bond to the phosphate group attached to the 5′ carbon of the next nucleotide in the chain. Thus the nucleotides of an RNA (or DNA) strand are connected by sugar–phosphate linkages (Figure 2.53*b*), which are described as *3′–5′-phosphodiester bonds* because the phosphate atom is esterified to two oxygen atoms, one from each of the two adjoining sugars.

A strand of RNA (or DNA) contains four different types of nucleotides distinguished by their nitrogenous base. Two
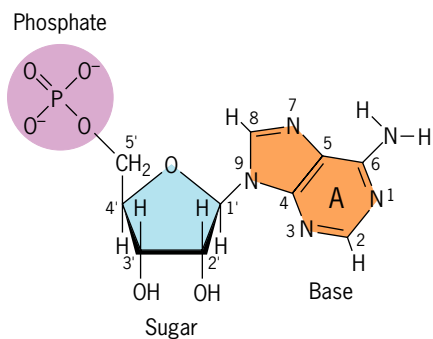
(a)
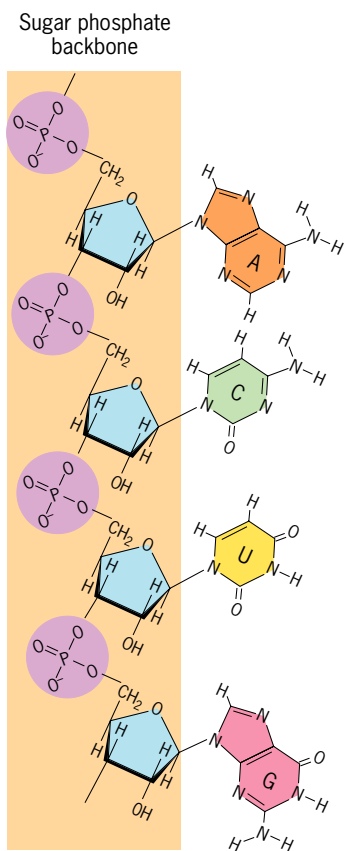
(b)

FIGURE 2.53 Nucleotides and nucleotide strands of RNA. (*a*) Nucleotides are the monomers from which strands of nucleic acid are constructed. A nucleotide consists of three parts: a sugar, a nitrogenous base, and a phosphate. The nucleotides of RNA contain the sugar ribose, which has a hydroxyl group bonded to the second carbon atom. In contrast, the nucleotides of DNA contain the sugar deoxyribose, which has a hydrogen atom rather than a hydroxyl group attached to the second carbon atom. Each nucleotide is polarized, having a 5′ end (corresponding to the 5′ side of the sugar) and a 3′ end. (*b*) Nucleotides are joined together to form a strand by covalent bonds that link the 3′ hydroxyl group of one sugar with the 5′ phosphate group of the adjoining sugar.

types of bases occur in nucleic acids: pyrimidines and purines (Figure 2.54). **Pyrimidines** are smaller molecules, consisting of a single ring; **purines** are larger, consisting of two rings. RNAs contain two different purines, **adenine** and **guanine**,
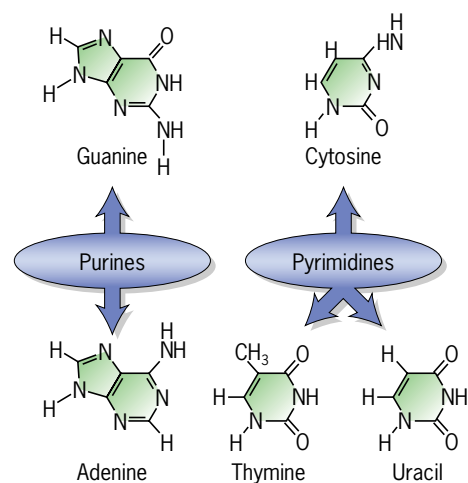
FIGURE 2.54 Nitrogenous bases in nucleic acids. Of the four standard bases found in RNA, adenine and guanine are purines, and uracil and cytosine are pyrimidines. In DNA, the pyrimidines are cytosine and thymine, which differs from uracil by a methyl group attached to the ring.

and two different pyrimidines, **cytosine** and **uracil**. In DNA, uracil is replaced by **thymine**, a pyrimidine with an extra methyl group attached to the ring (Figure 2.54).

Although RNAs consist of a continuous single strand, they often fold back on themselves to produce molecules having extensive double-stranded segments and complex three-dimensional structures. This is illustrated by the two RNAs shown in Figure 2.55. The RNA whose secondary structure is shown in Figure 2.55*a* is a component of the small subunit of the bacterial ribosome (see Figure 2.56). Ribosomal RNAs are not molecules that carry genetic information; rather, they serve as structural scaffolds on which the proteins of the ribosome can be attached and as elements that recognize and bind various soluble components required for protein synthesis. One of the ribosomal RNAs of the large subunit acts as the catalyst for the reaction by which amino acids are covalently joined during protein synthesis. RNAs having a catalytic role are called RNA enzymes, or **ribozymes**. Figure 2.55*b* depicts the tertiary structure of the so-called hammerhead ribozyme, which is able to cleave its own RNA strand. In both examples shown in Figure 2.55, the double-stranded regions are held together by hydrogen bonds between the bases. This same principle is responsible for holding together the two strands of a DNA molecule.

Nucleotides are not only important as building blocks of nucleic acids, they also have important functions in their own right. Most of the energy being put to use at any given moment in any living organism is derived from the nucleotide **adenosine triphosphate (ATP)**. The structure of ATP and its key role in cellular metabolism are discussed in the following chapter. **Guanosine triphosphate (GTP)** is another nucleotide of enormous importance in cellular activities. GTP binds to a variety of proteins (called G proteins) and acts as a switch to turn on their activities (see Figure 11.49 for an example).

(a)

(b)
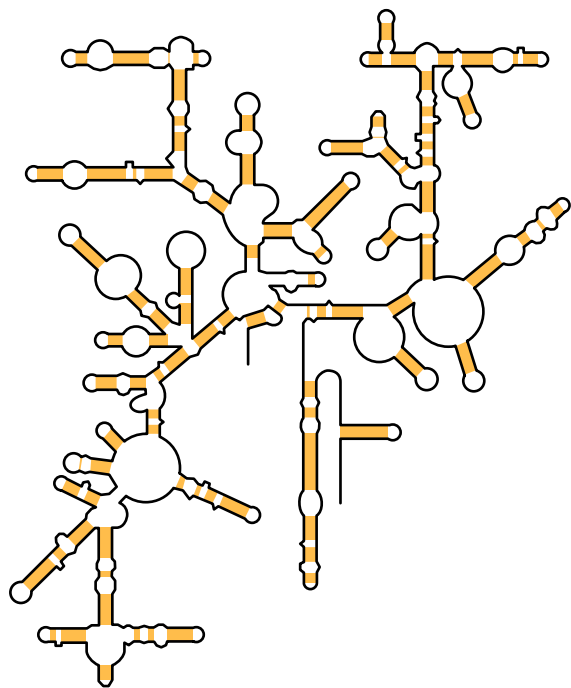
FIGURE 2.55   RNAs can assume complex shapes. (*a*) This ribosomal RNA is an integral component of the small ribosomal subunit of a bacterium. In this two-dimensional profile, the RNA strand is seen to be folded back on itself in a highly ordered pattern so that most of the molecule is double-stranded. (*b*) This hammerhead ribozyme, as it is called, is a small RNA molecule from a viroid (page 24). The helical nature of the double-stranded portions of this RNA can be appreciated in this three-dimensional model of the molecule. (B: FROM WILLIAM G. SCOTT ET AL., CELL 81:993, 1995; BY PERMISSION OF CELL PRESS.)
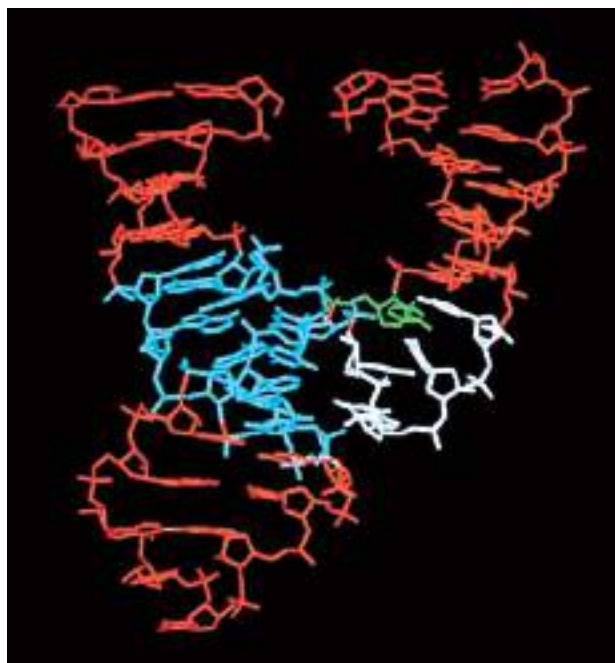
REVIEW ?

1. Which macromolecules are polymers? What is the basic structure of each type of monomer? How do the various monomers of each type of macromolecule vary among themselves?

2. Describe the structure of nucleotides and the manner in which these monomers are joined to form a polynucleotide strand. Why would it be overly simplistic to describe RNA as a single-stranded nucleic acid?

3. Name three polysaccharides composed of polymers of glucose. How do these macromolecules differ from one another?

4. Describe the properties of three different types of lipid molecules. What are their respective biological roles?

5. What are the major properties that distinguish different amino acids from one another? What roles do these differences play in the structure and function of proteins?

6. What are the properties of glycine, proline, and cysteine that distinguish these amino acids?

7. How are the properties of an α helix different from a β strand? How are they similar?

8. Given that proteins act as molecular machines, explain why conformational changes are so important in protein function.

2.6 THE FORMATION OF COMPLEX MACROMOLECULAR STRUCTURES

To what degree can the lessons learned from the study of protein architecture be applied to more complex structures in the cell? Can structures, such as membranes, ribosomes, and cytoskeletal elements, which consist of different types of subunits, also assemble by themselves? How far can subcellular organization be explained simply by having the pieces fit together to form the most stable arrangement? The assembly of cellular organelles is poorly understood, but it is apparent from the following examples that different types of subunits can self-assemble to form higher-order arrangements.

The Assembly of Tobacco Mosaic Virus Particles and Ribosomal Subunits

The most convincing evidence that a particular assembly process is self-directed is the demonstration that the assembly can occur outside the cell (in vitro) under physiologic conditions when the only macromolecules present are those that make up the final structure. In 1955, Heinz Fraenkel-Conrat and Robley Williams of the University of California, Berkeley, demonstrated that TMV particles, which consist of one long RNA molecule (approximately 6600 nucleotides) wound within a helical capsule made of 2130 identical protein subunits (see Figure 1.20), were capable of self-assembly. In their
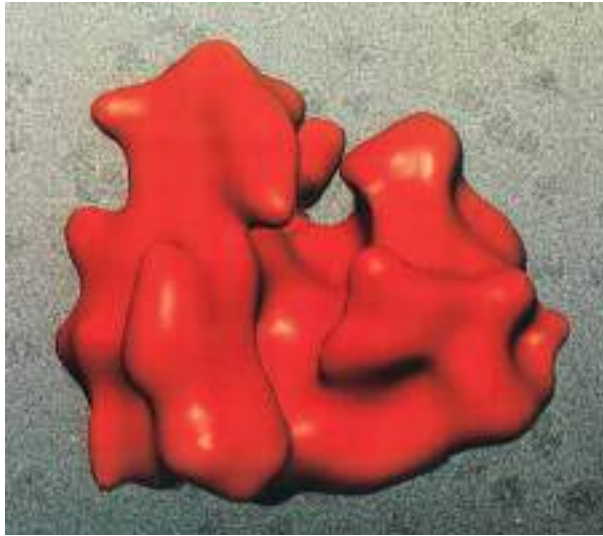
FIGURE 2.56 Reconstruction of a ribosome from the cytoplasm of a wheat germ cell. This reconstruction is based on high-resolution electron micrographs and shows the two subunits of this eukaryotic ribosome, the small (40S) subunit on the left and the large (60S) subunit on the right. The internal structure of a ribosome is discussed in Section 11.8. (FROM ADRIANA VERSCHOOR, ET AL., J. CELL BIOL. VOL. 133 (COVER #3), 1996; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

experiments, they purified TMV RNA and protein separately, mixed them together under suitable conditions, and recovered mature, infective particles after a short period of incubation. Clearly the two components contain all the information necessary for particle formation.

Ribosomes, like TMV particles, are made of RNA and protein. Unlike the simpler TMV, ribosomes contain several different types of RNA and a considerable collection of different proteins. All ribosomes, regardless of their source, are composed of two subunits of different size. Although ribosomal subunits are often depicted in drawings as symmetric structures, in fact they have a highly irregular shape, as indicated in Figure 2.56. The large (or 50S) ribosomal subunit of bacteria contains two molecules of RNA and approximately 32 different proteins. The small (or 30S) ribosomal subunit of bacteria contains one molecule of RNA and 21 different pro-

teins. The structure and function of the ribosome are discussed in detail in Section 11.8.

One of the milestones in the study of ribosomes came in the mid-1960s, when Masayasu Nomura and his co-workers at the University of Wisconsin succeeded in reconstituting complete, fully functional 30S bacterial subunits by mixing the 21 purified proteins of the small subunit with purified small-subunit ribosomal RNA. Apparently, the components of the small subunit contain all the information necessary for the assembly of the entire particle. Analysis of the intermediates that form at different stages during reconstitution in vitro indicates that subunit assembly occurs in a sequential step-by-step manner that closely parallels the process in vivo. At least one of the proteins of the small subunit (S16) appears to function solely in ribosome assembly; deletion of this protein from the reconstitution mixture greatly slowed the assembly process but did not block the formation of fully functional ribosomes. Reconstitution of the large subunit of the bacterial ribosome was accomplished in the following decade. It should be kept in mind that although it takes approximately 2 hours at 50°C to reconstitute the ribosome in vitro, the bacterium can assemble the same structure in a few minutes at temperatures as low as 10°C. It may be that the bacterium uses something that is not available to the investigator who begins with purified components. Assembly of the ribosome within the cell, for example, may include the participation of accessory factors that function in protein folding, such as the chaperones described in the Experimental Pathways. In fact, the formation of ribosomes within a *eukaryotic* cell requires the transient association of many proteins that do not end up in the final particle, as well as the removal of approximately half the nucleotides of the large ribosomal RNA precursor (Section 11.3). As a result, the components of the mature eukaryotic ribosome no longer possess the information to reconstitute themselves in vitro.

REVIEW ?

1. What type of evidence suggests that bacterial ribosomal subunits are capable of self-assembly, but eukaryotic subunits are not?

2. What evidence would indicate that a particular ribosomal protein had a role in ribosome function but not assembly?

EXPERIMENTAL PATHWAYS

Chaperones: Helping Proteins Reach Their Proper Folded State

In 1962, F. M. Ritossa, an Italian biologist studying the development of the fruit-fly *Drosophila*, reported a curious finding.[1] When the temperature at which fruit-fly larvae were developing was raised from the normal 25°C to 32°C, a number of new sites on the giant chromosomes of the larval cells became activated. As we will see in Chapter 10, the giant chromosomes of these insect larvae provide a visual exhibit of gene expression (see Figure 10.8). The results suggested that increased temperature induced the expression of new genes, a finding that was confirmed a decade later with the characterization of several proteins that appeared in larvae following temperature elevation.[2] It was soon found that this response, called the **heat-shock response**, was not confined to fruit flies, but can be initiated in many different cells from virtually every type of organism—from bacteria to plants and mammals. Closer examination revealed that the proteins produced during the response were found not only in heat-shocked cells, but also at lower concentration in cells under normal conditions. What is the function of these so-called *heat-shock proteins* (*hsps*)? The answer to this question was gradually revealed by a series of seemingly unrelated studies.

We saw on page 77 that some complex, multisubunit structures, such as a bacterial ribosome or a tobacco mosaic virus particle, can self-assemble from purified subunits. It was demonstrated in the 1960s that the proteins that make up bacteriophage particles (see Figure 1.21*c*) also possess a remarkable ability to self-assemble, but they are generally unable to form a complete, functional virus particle by themselves in vitro. Experiments on phage assembly in bacterial cells confirmed that phages require bacterial help. It was shown in 1973, for example, that a certain mutant strain of bacteria, called *GroE*, could not support the assembly of normal phages. Depending on the type of phage, the head or the tail of the phage particle was assembled incorrectly.[3,4] These studies suggested that a protein encoded by the bacterial chromosome participated in the assembly of viruses, even though this host protein was not a component of the final virus particles. As it obviously did not evolve as an aid for virus assembly, the bacterial protein required for phage assembly had to play some role in the cell's normal activities, but the precise role remained obscure. Subsequent studies revealed that the *GroE* site on the bacterial chromosome actually contains two separate genes, *GroEL* and *GroES*, that encode two separate proteins GroEL and GroES. Under the electron microscope, the purified GroEL protein appeared as a cylindrical assembly consisting of two disks. Each disk was composed of seven subunits arranged symmetrically around the central axis (Figure 1).[5,6]

Several years later, a study on pea plants hinted at the existence of a similar assembly-promoting protein in the chloroplasts of plants.[7] Rubisco is a large protein in chloroplasts that catalyzes the reaction in which CO_2 molecules taken up from the atmosphere are covalently linked to organic molecules during photosynthesis (Section 6.6). Rubisco comprises 16 subunits: 8 small subunits (molecular mass of 14,000 daltons) and 8 large subunits (55,000 daltons). It was found that large Rubisco subunits, synthesized inside the chloroplast, are not present in an independent state, but are associated with a huge protein assembly consisting of identical subunits of 60,000 daltons (60 kDa) molecular mass. In their paper, the researchers considered the possibility that the complex formed by the large Rubisco subunit and the 60-kDa polypeptide was an intermediate in the assembly of a complete Rubisco molecule.

FIGURE 1   A model of the GroEL complex built according to data from electron microscopy and molecular-weight determination. The complex is seen to consist of two disks, each composed of seven identical subunits arranged symmetrically around a central axis. Subsequent studies showed the complex contains two internal chambers. (FROM T. HOHN ET AL., J. MOL. BIOL. 129:371, 1979. COPYRIGHT 1979, BY PERMISSION OF THE PUBLISHER ACADEMIC PRESS, ELSEVIER SCIENCE.)

A separate line of investigation on mammalian cells also revealed the existence of proteins that appeared to assist the assembly of multisubunit proteins. Like Rubisco, antibody molecules consist of a complex of two different types of subunits, smaller light chains and larger heavy chains. Just as the large subunits of Rubisco become associated with another protein not found in the final complex, so too do the heavy chains of an antibody complex.[8] This protein, which associates with newly synthesized heavy chains, but not with heavy chains that are already bound to light chains, was named *binding protein*, or BiP. BiP was subsequently found to have a molecular mass of 70,000 daltons (70 kDa).

To this point, we have been discussing two lines of investigation: one concerned with the heat-shock response and the other with proteins that promote protein assembly. These two fields came together in 1986, when it was shown that one of the proteins that figures most prominently in the heat-shock response, a protein that had been named *heat-shock protein 70* (*hsp70*) because of its molecular mass, was identical to BiP, the protein implicated in the assembly of antibody molecules.[9]

Even before the discovery of the heat-shock response, the structure of proteins was known to be sensitive to temperature, with a small rise in temperature causing these delicate molecules to begin to unfold. Unfolding exposes hydrophobic residues that were previously buried in the protein's core. Just as fat molecules in a bowl of soup are pushed together into droplets, so too are proteins with hydrophobic patches on their surface. Consequently, when a cell is heat shocked, soluble proteins become denatured and form aggregates. A report in 1985 demonstrated that, following temperature elevation, newly synthesized hsp70 molecules enter cell nuclei and bind to aggregates of nuclear proteins, where they act like molecular crowbars to promote disaggregation.[10] Because of their role in assisting the assembly of proteins by preventing undesirable interactions, hsp70 and related molecules were named **molecular chaperones**.[11]

It was soon demonstrated that the bacterial heat-shock protein GroEL and the Rubisco assembly protein in plants were homologous

proteins. In fact, the two proteins share the same amino acids at nearly half of the more than 500 residues in their respective molecules.[12] The fact that the two proteins—both members of the *Hsp60 chaperone family*—have retained so many of the same amino acids reflects their similar and essential function in the two types of cells. But what was that essential function? At this point it was thought that their primary function was to mediate the assembly of multisubunit complexes, such as Rubisco. This view was changed by experiments studying molecular chaperones in mitochondria. It was known that newly made mitochondrial proteins produced in the cytosol had to cross the outer mitochondrial membranes in an unfolded, extended, monomeric form. A mutant was found that altered the activity of another member of the Hsp60 chaperone family that resided inside mitochondria. In cells containing this mutant chaperone, proteins that were transported into mitochondria failed to fold into their active forms.[13] Even proteins that consisted of a single polypeptide chain failed to fold into their native conformation. This finding changed the perception of chaperone function from a notion that they assist assembly of already-folded subunits into larger complexes, to our current understanding that they assist polypeptide chain folding.

The results of these and other studies indicated the presence in cells of at least two major families of molecular chaperones: the Hsp70 chaperones, such as BiP, and the Hsp60 chaperones (which are also called *chaperonins*), such as Hsp60, GroEL, and the Rubisco assembly protein. We will focus on the Hsp60 chaperonins, such as GroEL, which are best understood.

As first revealed in 1979, GroEL is a huge molecular complex of 14 polypeptide subunits arranged in two stacked rings resembling a double doughnut.[5,6] Fifteen years after these first electron micrographs were taken, the three-dimensional structure of the GroEL complex was determined by X-ray crystallography.[14] The study revealed the presence of a central cavity within the GroEL cylinder. Subsequent studies demonstrated that this cavity was divided into two separate chambers. Each chamber was situated within the cen-

ter of one of the rings of the GroEL complex and was large enough to enclose a polypeptide undergoing folding.

Electron microscopic studies also provided information about the structure and function of a second protein, GroES, which acts in conjunction with GroEL. Like GroEL, GroES is a ring-like protein with seven subunits arrayed symmetrically around a central axis. GroES, however, consists of only one ring, and its subunits are much smaller (10,000 daltons) than those of GroEL (60,000 daltons). GroES is seen as a cap or dome that fits on top of either end of a GroEL cylinder (Figure 2). The attachment of GroES to one end of GroEL causes a dramatic conformational change in the GroEL protein that markedly increases the volume of the enclosed chamber at that end of the complex.[15]

The importance of this conformational change has been revealed in remarkable detail by X-ray crystallographic studies in the laboratories of Arthur Horwich and Paul Sigler at Yale University.[16] As shown in Figure 3, the binding of the GroES cap is accompanied by a 60° rotation of the apical (red) domain of the subunits that make
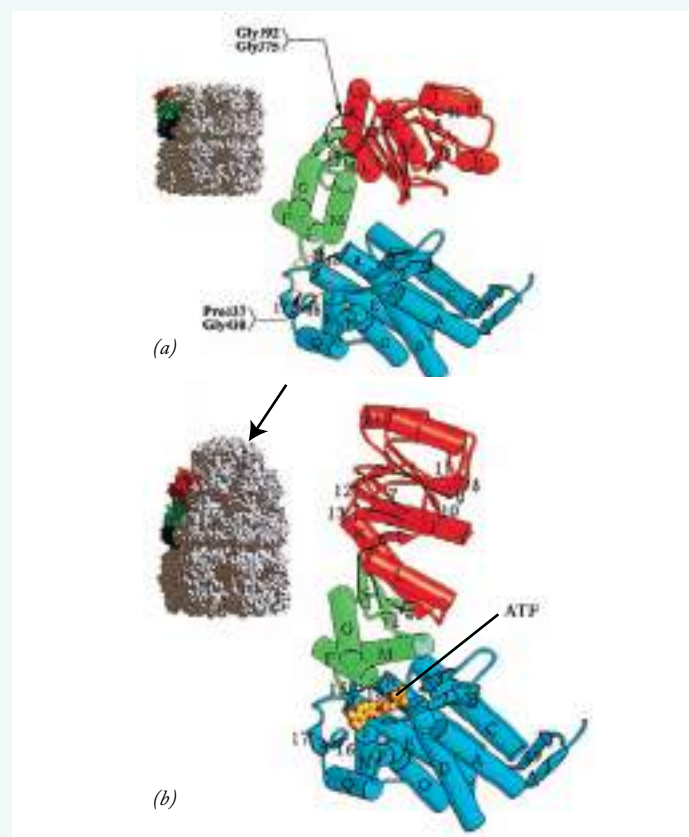
FIGURE 3 Conformational change in GroEL. (*a*) The model on the left shows a surface view of the two rings that make up the GroEL chaperonin. The drawing on the right shows the tertiary structure of one of the subunits of the top GroEL ring. The polypeptide chain can be seen to fold into three domains. (*b*) When a GroES ring (arrow) binds to the GroEL cylinder, the apical domain of each GroEL subunit of the adjacent ring undergoes a dramatic rotation of approximately 60° with the intermediate domain (shown in green) acting like a hinge. The effect of this shift in parts of the polypeptide is a marked elevation of the GroEL wall and enlargement of the enclosed chamber. (Reprinted with permission from Z. Xu, A. L. Horwich, and P. B. Sigler, Nature 388:744, 1997. Copyright © 1997 Macmillan Magazines Limited.)

FIGURE 2 Reconstructions of GroEL based on high-resolution electron micrographs taken of specimens that had been frozen in liquid ethane and examined at −170°C. The image on the left shows the GroEL complex, and that on the right shows the GroEL complex with GroES, which appears as a dome on one end of the cylinder. It is evident that the binding of the GroES is accompanied by a marked change in conformation of the apical end of the proteins that make up the top GroEL ring (arrow), which results in a marked enlargement of the upper chamber. (Reprinted with permission from S. Chen et al., courtesy of Helen R. Saibil, Nature 371:263, 1994. Copyright © 1994 Macmillan Magazines Limited.)

up the GroEL ring at that end of the GroEL cylinder. The attachment of GroES does more than trigger a conformational change that enlarges the GroEL chamber. Before attachment of GroES, the inner wall of the GroEL chamber has exposed hydrophobic residues that give the lining a hydrophobic character. Nonnative polypeptides also have exposed hydrophobic residues that become buried in the interior of the native polypeptide. Because hydrophobic surfaces tend to interact, the hydrophobic lining of the GroEL cavity binds to the surface of nonnative polypeptides. Binding of GroES to GroEL buries the hydrophobic residues of the GroEL wall and exposes a number of polar residues, thereby changing the character of the chamber wall. As a result of this change, a nonnative polypeptide that had been bound to the GroEl wall by hydrophobic interactions is displaced into the space within the chamber. Once freed from its attachment to the chamber wall, the polypeptide is given the opportunity to continue its folding in a protected environment. After about 15 seconds, the GroES cap dissociates from the GroEL ring, and the polypeptide is ejected from the chamber. If the polypeptide has not reached its native conformation by the time it is ejected, it can rebind to the same or another GroEL, and the process is repeated. A model depicting some of the steps thought to occur during GroEL-GroES-assisted folding is shown in Figure 4.

Approximately 250 of the roughly 2400 proteins present in the cytosol of an *E. coli* cell normally interact with GroEL.[17] How is it possible for a chaperone to bind so many different polypeptides? The GroEL binding site consists of a hydrophobic surface formed largely by two α helices of the apical domain that is capable of binding virtually any sequence of hydrophobic residues that might be accessible in a partially folded or misfolded polypeptide.[18] A comparison of the crystal structure of the unbound GroEL molecule with that of GroEL bound to several different peptides revealed that the binding

site on the apical domain of a GroEL subunit can locally adjust its positioning when bound to different partners. This finding indicates that the binding site has structural flexibility that allows it to adjust its shape to fit the shape of the particular polypeptide with which it has to interact.

A number of studies have also suggested that GroEL does more than simply provide a passive chamber in which proteins can fold without outside interference. In one study, site-directed mutagenesis was utilized to modify a key residue, Tyr71 of GroES, whose side chain hangs from the ceiling of the folding chamber.[19] Because of its aromatic ring, tyrosine is a modestly hydrophobic residue (Figure 2.26). When Tyr71 was replaced by a positively or negatively charged amino acid, the resulting GroEL-GroES variant exhibited an increased ability to assist the folding of one specific foreign polypeptide, the green fluorescent protein (GFP). However, substitutions for Tyr71 that improved the ability of GroES-GroEL to increase GFP folding made the chaperonin less competent to help its natural substrates fold. Thus as the chaperonin became more and more specialized to interact with GFP, it lost its general ability to assist folding of proteins having an unrelated structure. This finding suggests that individual amino acids in the wall of the folding chamber may participate somehow in the folding reaction. Data from another study has suggested that binding of a nonnative protein to GroEL is followed by a forced unfolding of the substrate protein.[20] FRET (fluorescence resonance energy transfer) is a technique (discussed in Section 18.1) that allows researchers to determine the distance between different parts of a protein molecule at different times during a given process. In this study, investigators found that the protein undergoing folding, in this case Rubisco, bound to the apical domain of the GroEL ring in a relatively compact state. The compact nature of the bound protein was revealed by the close proximity to one another
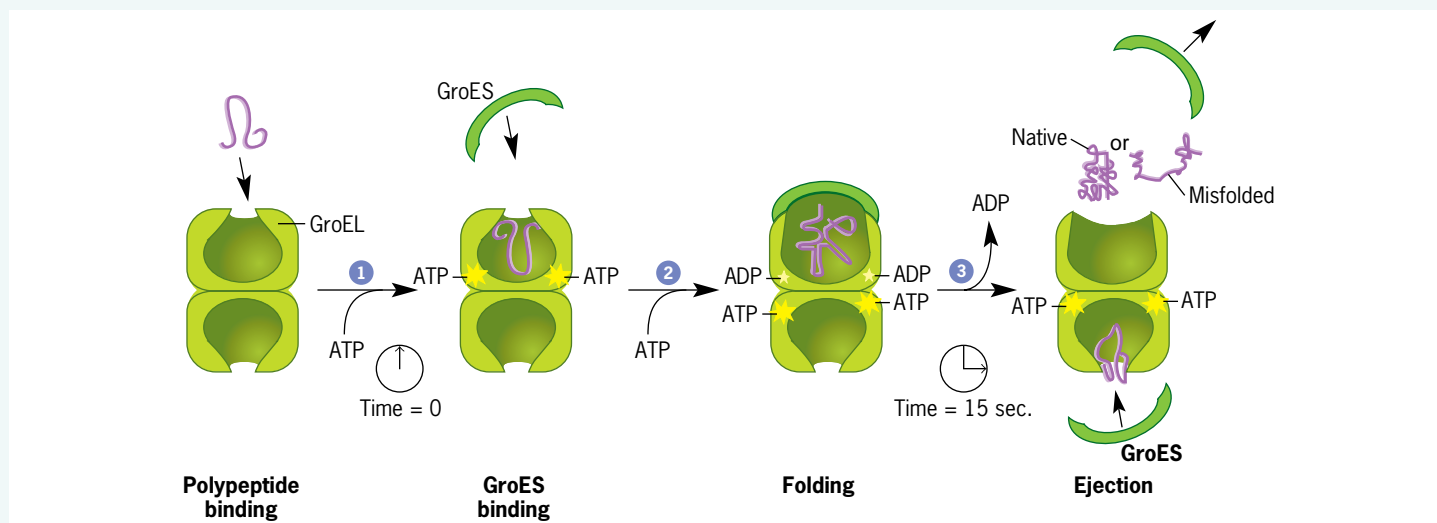
FIGURE 4 A schematic illustration of the proposed steps that occur during the GroEL-GroES-assisted folding of a polypeptide. The GroEL is seen to consist of two chambers that have equivalent structures and functions and that alternate in activity. Each chamber is located within one of the two rings that make up the GroEL complex. The nonnative polypeptide enters one of the chambers (step 1) and binds to hydrophobic sites on the chamber wall. Binding of the GroES cap produces a conformational change in the wall of the top chamber, causing the enlargement of the chamber and release of the nonnative polypeptide from the wall into the encapsulated space (step 2). After about 15 seconds have elapsed, the GroES dissociates from the complex and the polypeptide is ejected from the chamber (step 3). If the polypeptide has achieved its native conformation, as has the molecule on the left, the folding process is complete. If, however, the polypeptide is only partially folded, or is misfolded, it will rebind the GroEL chamber for another round of folding. (*Note:* As indicated, the mechanism of GroEL action is driven by the binding and hydrolysis of ATP, an energy-rich molecule whose function is discussed at length in the following chapter.) (AFTER A. L. HORWICH, ET AL., PROC. NAT'L. ACAD. SCI. U.S.A. 96:11037, 1999.)

of the FRET tags, which were attached to amino acids located at opposite ends of the Rubisco chain. Then, during the conformational change that enlarges the volume of the GroEL cavity (Figure 3), the bound Rubisco protein was forcibly unfolded, as evidenced by the increased distance between the two tagged ends of the molecule. This study suggests that the Rubisco polypeptide is taken completely back to the unfolded state, where it is given the opportunity to refold from scratch. This action should help prevent the nonnative protein from becoming trapped permanently in a misfolded state. In other words, each individual visit to a GroEL-GroES chamber provides an all-or-none attempt to reach the native state, rather than just one stage in a series of steps in which the protein moves closer to the native state with each round of folding.

Keep in mind that molecular chaperones do not convey information for the folding process but instead prevent proteins from veering off their correct folding pathway and finding themselves in misfolded or aggregated states. Just as Anfinsen discovered decades ago, the three-dimensional structure of a protein is determined by its amino acid sequence.

References

1. RITOSSA, F. 1962. A new puffing pattern induced by temperature shock and DNP in *Drosophila*. *Experentia* 18:571–573.
2. TISSIERES, A., MITCHELL, H. K., & TRACY, U. M. 1974. Protein synthesis in salivary glands of *Drosophila melanogaster*: Relation to chromosomal puffs. *J. Mol. Biol.* 84:389–398.
3. STERNBERG, N. 1973. Properties of a mutant of *Escherichia coli* defective in bacteriophage lambda head formation (groE). *J. Mol. Biol.* 76:1–23.
4. GEORGOPOULOS, C. P. ET AL. 1973. Host participation in bacteriophage lambda head assembly. *J. Mol. Biol.* 76:45–60.
5. HOHN, T. ET AL. 1979. Isolation and characterization of the host protein groE involved in bacteriophage lambda assembly. *J. Mol. Biol.* 129:359–373.
6. HENDRIX, R. W. 1979. Purification and properties of groE, a host protein involved in bacteriophage assembly. *J. Mol. Biol.* 129:375–392.
7. BARRACLOUGH, R. & ELLIS, R. J. 1980. Protein synthesis in chloroplasts. IX. *Biochim. Biophys. Acta* 608:19–31.
8. HAAS, I. G. & WABL, M. 1983. Immunoglobulin heavy chain binding protein. *Nature* 306:387–389.
9. MUNRO, S. & PELHAM, H.R.B. 1986. An Hsp70-like protein in the ER: Identity with the 78 kD glucose-regulated protein and immunoglobin heavy chain binding protein. *Cell* 46:291–300.
10. LEWIS, M. J. & PELHAM, H.R.B. 1985. Involvement of ATP in the nuclear and nucleolar functions of the 70kD heat-shock protein. *EMBO J.* 4:3137–3143.
11. ELLIS, J. 1987. Proteins as molecular chaperones. *Nature* 328:378–379.
12. HEMMINGSEN, S. M. ET AL. 1988. Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* 333:330–334.
13. CHENG, M. Y. ET AL. 1989. Mitochondrial heat-shock protein Hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature* 337:620–625.
14. BRAIG, K. ET AL. 1994. The crystal structure of the bacterial chaperonin GroEL at 2.8A. *Nature* 371:578–586.
15. CHEN, S. ET AL. 1994. Location of a folding protein and shape changes in GroEL-GroES complexes. *Nature* 371:261–264.
16. XU, Z., HORWICH, A. L., & SIGLER, P. B. 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)$_7$ chaperonin complex. *Nature* 388:741–750.
17. KERNER, M. J. ET AL. 2005. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* 122:209–220.
18. CHEN, L. & SIGLER, P. 1999. The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. *Cell* 99:757–768.
19. WANG, J. D. ET AL. 2002. Directed evolution of substrate-optimized GroEL/S chaperonins. *Cell* 111:1027–1039.
20. LIN, Z. ET AL. 2008. GroEL stimulates protein folding through forced unfolding. *Nature Struct. Mol. Biol.* 15:303–311.

SYNOPSIS

Covalent bonds hold atoms together to form molecules. Covalent bonds are stable partnerships formed when atoms share their outer-shell electrons, each participant gaining a filled shell. Covalent bonds can be single, double, or triple depending on the number of pairs of shared electrons. If electrons in a bond are shared unequally by the component atoms, the atom with the greater attraction for electrons (the more electronegative atom) bears a partial negative charge, whereas the other atom bears a partial positive charge. Molecules that lack polarized bonds have a nonpolar, or hydrophobic, character, which makes them insoluble in water. Molecules that have polarized bonds have a polar, or hydrophilic, character, which makes them water soluble. Polar molecules of biological importance contain atoms other than just carbon and hydrogen, usually O, N, S, or P. (p. 32)

Noncovalent bonds are formed by weak attractive forces between positively and negatively charged regions within the same molecule or between two nearby molecules. Noncovalent bonds play a key role in maintaining the structure of biological molecules and mediating their dynamic activities. Noncovalent bonds include ionic bonds, hydrogen bonds, and van der Waals forces. Ionic bonds form between fully charged positive and negative groups; hydrogen bonds form between a covalently bonded hydrogen atom (which bears a partial positive charge) and a covalently bonded nitrogen or oxygen atom (which bears a partial negative charge); van der Waals forces form between two atoms exhibiting a transient charge due to a momentary asymmetry in the distribution of electrons around the atoms. Nonpolar molecules or nonpolar portions of larger molecules tend to associate with one another in aqueous environments to form hydrophobic interactions. Examples of these various types of noncovalent interactions include the association of DNA and proteins by ionic bonds, the association of pairs of DNA strands by hydrogen bonds, and the formation of the hydrophobic core of soluble proteins as the result of hydrophobic interactions and van der Waals forces. (p. 33)

Water has unique properties on which life depends. The covalent bonds that make up a water molecule are highly polarized. As a result, water is an excellent solvent capable of forming hydrogen bonds with virtually all polar molecules. Water is also a major determinant of the structure of biological molecules and the types of interactions in which they can engage. The pH of a solution is a measure of the concentration of hydrogen (or hydronium) ions. Most biological processes are acutely sensitive to pH because changes in hydrogen ion concentration alter the ionic state of biological molecules. Cells are protected from pH fluctuations by buffers—compounds that react with hydrogen or hydroxyl ions. (p. 36)

Carbon atoms play a pivotal role in the formation of biological molecules. Each carbon atom is able to bond with up to four other atoms, including other carbon atoms. This property allows the formation of large molecules whose backbone consists of a chain of carbon atoms. Molecules consisting solely of hydrogen and carbon are called hydrocarbons. Most of the molecules of biological importance contain functional groups that include one or more electronegative atoms, making the molecule more polar, more water soluble, and more reactive. (p. 39)

Biological molecules are members of four distinct types: carbohydrates, lipids, proteins, and nucleic acids. Carbohydrates include simple sugars and larger molecules (polysaccharides) constructed of sugar monomers. Carbohydrates function primarily as a storehouse of chemical energy and as durable building materials for biological construction. Simple biological sugars consist of a backbone of three to seven carbon atoms, with each carbon linked to a hydroxyl group except one, which bears a carbonyl. Sugars with five or more carbon atoms self-react to form a ring-shaped molecule. Those carbon atoms along the sugar backbone that are linked to four different groups are sites of stereoisomerism, generating pairs of isomers that cannot be superimposed. The asymmetric carbon farthest from the carbonyl determines whether the sugar is D or L. Sugars are linked to one another by glycosidic bonds to form disaccharides, oligosaccharides, and polysaccharides. In animals, sugar is stored primarily as the branched polysaccharide glycogen, which provides a readily available energy source. In plants, glucose reserves are stored as starch, which is a mixture of unbranched amylose and branched amylopectin. Most of the sugars in both glycogen and starch are joined by $\alpha(1 \rightarrow 4)$ linkages. Cellulose is a structural polysaccharide manufactured by plant cells that serves as a major component of the cell wall. Glucose monomers in cellulose are joined by $\beta(1 \rightarrow 4)$ linkages, which are cleaved by cellulase, an enzyme that is absent in virtually all animals. Chitin is a structural polysaccharide composed of N-acetylglucosamine monomers. (p. 41)

Lipids are a diverse array of hydrophobic molecules having widely divergent structures and functions. Fats consist of a glycerol molecule esterified to three fatty acids. Fatty acids differ in chain length and the number and position of double bonds (sites of unsaturation). Fats are very rich in chemical energy; a gram of fat contains over twice the energy content of a gram of carbohydrate. Steroids are a group of lipids containing a characteristic four-ringed hydrocarbon skeleton. Steroids include cholesterol as well as numerous hormones (e.g., testosterone, estrogen, and progesterone) that are synthesized from cholesterol. Phospholipids are phosphate-containing lipid molecules that contain both a hydrophobic end and a hydrophilic end and play a pivotal role in the structure and function of cell membranes. (p. 46)

Proteins are macromolecules of diverse function consisting of amino acids linked by peptide bonds into polypeptide chains. Included among the diverse array of proteins are enzymes, structural materials, membrane receptors, gene regulatory factors, hormones, transport agents, and antibodies. The order in which the 20 different amino acids are incorporated into a protein is encoded in the sequence of nucleotides in DNA. All 20 amino acids share a common structural organization consisting of an α-carbon bonded to an amino group, a carboxyl group, and a side chain of varying structure. In the present scheme, the side chains are classified into four categories: those that are fully charged at physiologic pH; those that are polar, but uncharged and capable of forming hydrogen bonds; those that are nonpolar and interact by means of van der Waals forces; and three amino acids (proline, cysteine, and glycine) that possess unique properties. (p. 49)

The structure of a protein can be described at four levels of increasing complexity. Primary structure is described by the amino acid sequence of a polypeptide; secondary structure by the three-dimensional structure (conformation) of sections of the polypeptide backbone; tertiary structure by the conformation of the entire polypeptide; and quaternary structure by the arrangement of the subunits if the protein consists of more than one polypeptide chain. The α helix and β-pleated sheet are both stable, maximally hydrogen-bonded secondary structures that are common in many proteins. The tertiary structure of a protein is highly complex and unique to each individual type of protein. Most proteins have an overall globular shape in which the polypeptide is folded to form a compact molecule in which specific residues are strategically situated to allow the protein to carry out its specific function. Most proteins consist of two or more domains that maintain a structural and functional independence from one another. Using the technique of site-directed mutagenesis, researchers can learn about the role of specific amino acid residues by making specific substitutions. In recent years, a new field of proteomics has emerged that uses advanced technologies such as mass spectrometry and high-speed computation to study various properties of proteins on a comprehensive, large scale. For example, the various interactions among thousands of proteins encoded by the fruit-fly genome have been analyzed by large-scale techniques. (p. 53)

The information required for a polypeptide chain to achieve its native conformation is encoded in its primary structure. Some proteins fold into their final conformation by themselves; others require the assistance of nonspecific chaperones, which prevent aggregation of partially folded intermediates. (p. 62)

Nucleic acids are primarily informational molecules that consist of strands of nucleotide monomers. Each nucleotide in a strand consists of a sugar, phosphate, and nitrogenous base. The nucleotides are linked by bonds between the 3′ hydroxyl group of the sugar of one nucleotide and the 5′ phosphate group of the adjoining nucleotide. Both RNA and DNA are assembled from four different nucleotides; nucleotides are distinguished by their bases, which can be a pyrimidine (cytosine or uracil/thymine) or a purine (adenine or guanine). DNA is a double-stranded nucleic acid, and RNA is generally single stranded, though the single strand is often folded back on itself to form double-stranded sections. Information in nucleic acids is encoded in the specific sequence of nucleotides that constitute a strand. (p. 74)

ANALYTIC QUESTIONS

1. Sickle cell anemia results from a substitution of a valine for a glutamic acid. What do you expect the effect might be if the mutation were to have placed a leucine at that site? An aspartic acid?

2. Of the following amino acids, glycine, isoleucine, and lysine, which would you expect to be the most soluble in an acidic aqueous solution? Which the least?

3. How many structural isomers could be formed from a molecule with the formula C_5H_{12}? C_4H_8?

4. Glyceraldehyde is the only three-carbon aldotetrose, and it can exist as two stereoisomers. What is the structure of dihydroxy-acetone, the only ketotriose? How many stereoisomers does it form?

5. Bacteria are known to change the kinds of fatty acids they produce as the temperature of their environment changes. What types of changes in fatty acids would you expect as the temperature drops? Why would this be adaptive?

6. In the polypeptide backbone —C—C—N—C—C—N—C—C—NH$_2$, identify the α-carbons.

7. Which of the following are true? Increasing the pH of a solution would (1) suppress the dissociation of a carboxylic acid, (2) increase the charge on an amino group, (3) increase the dissociation of a carboxylic acid, (4) suppress the charge on an amino group.

8. Which of the four classes of amino acids has side chains with the greatest hydrogen-bond-forming potential? Which has the greatest potential to form ionic bonds? Hydrophobic interactions?

9. If the three enzymes of the pyruvate dehydrogenase complex existed as physically separate proteins rather than as a complex, what effect might this have on the rate of reactions catalyzed by these enzymes?

10. Would you agree that neither ribonuclease nor myoglobin had quaternary structure? Why or why not?

11. How many different tripeptides are possible? How many carboxyl terminals of polypeptide chains are present in a molecule of hemoglobin?

12. You have isolated a pentapeptide composed of four glycine residues and one lysine residue that resides at the C-terminus of the peptide. Using the information provided in the legend of Figure 2.27, if the pK of the side chain of lysine is 10 and the pK of the terminal carboxyl group is 4, what is the structure of the peptide at pH 7? At pH 12?

13. The side chains of glutamic acid (pK 4.3) and arginine (pK 12.5) can form an ionic bond under certain conditions. Draw the relevant portions of the side chains and indicate whether or not an ionic bond could form at the following: (a) pH 4; (b) pH 7; (c) pH 12; (d) pH 13.

14. Would you expect a solution of high salt to be able to denature ribonuclease? Why or why not?

15. You have read in the Human Perspective that (1) mutations in the *PRNP* gene can make a polypeptide more likely to fold into the PrP^{Sc} conformation, thus causing CJD and (2) exposure to the PrP^{Sc} prion can lead to an infection that also causes CJD. How can you explain the occurrence of rare sporadic cases of the disease in persons who have no genetic propensity for it?

16. Persons who are born with Down syndrome have an extra (third) copy of chromosome #21 in their cells. Chromosome #21 contains the gene that encodes the APP protein. Why do you suppose that individuals with Down syndrome develop Alzheimer's disease at an early age?

17. We saw on page 74 how evolution has led to the existence of protein families composed of related molecules with similar functions. A few examples are also known where proteins with very similar functions have primary and tertiary structures that show no evidence of evolutionary relationship. Subtilisin and trypsin, for example, are two protein-digesting enzymes (proteases) that show no evidence they are homologous despite the fact that they utilize the same mechanism for attacking their substrates. How can this coincidence be explained?

18. Would you agree with the statement that many different amino acid sequences can fold into the same basic tertiary structure? What data can you cite as evidence for your position.

19. In the words of one scientist: "The first question any structural biologist asks upon being told that a new [protein] structure has been solved is no longer 'What does it *look* like?; it is now 'What does it look *like*?'" What do you suppose he meant by this statement?

20. It was noted in the Human Perspective that persons with arthritis that had taken certain NSAIDs over a long period of time exhibited a lower incidence of Alzheimer's disease, yet double-blinded clinical trials on these same drugs did not appear to benefit patients with AD. These would appear to be contradictory findings. The first type of study is referred to as a *retrospective* study in that researchers look backwards from a correlation that is made at the present time, in this case a conclusion that taking NSAIDs over a period of time may prevent the development of AD. The second type of study is referred to as a *prospective* study in that it looks forward to future results based on an experimental plan to give some patients a drug and others a placebo. Can you give a reason why these two different approaches might lead to differing conclusions about the use of these drugs?

3
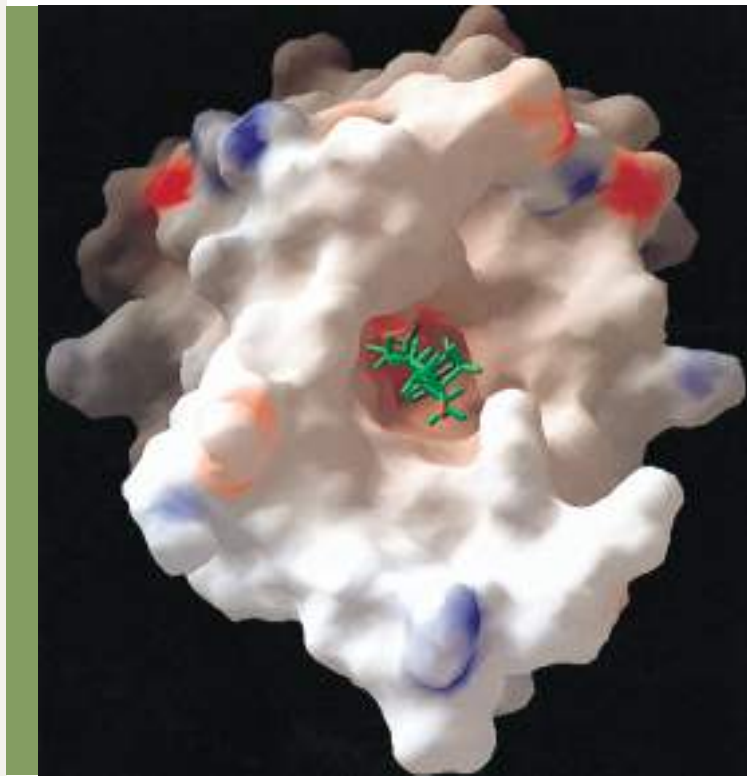
Bioenergetics, Enzymes, and Metabolism

The interrelationship between structure and function is evident at all levels of biological organization from the molecular to the organismal. We saw in the last chapter that proteins have an intricate three-dimensional structure that depends on particular amino acid residues being present in precisely the correct place. In this chapter, we will look more closely at one large group of proteins, the enzymes, and see how their complex architecture endows them with the capability to vastly increase the rate of biological reactions. To understand how enzymes are able to accomplish such feats, it is necessary to consider the flow of energy during a chemical reaction, which brings us to the subject of thermodynamics. A brief survey of the principles of thermodynamics also helps explain many of the cellular processes that will be discussed in this and following chapters, including the movement of ions across membranes, the synthesis of macromolecules, and the assembly of cytoskeletal networks. As we will see, the thermodynamic analysis of a particular system can reveal whether or not the events can occur spontaneously and, if not, provide a measure of the energy a cell must expend for the process to be accomplished. In the final section of this chapter, we will see how individual chemical reactions are linked together to form metabolic pathways and how the flow of energy and raw materials through certain pathways can be controlled. ■

A model showing the surface of the enzyme Δ^5-3-ketosteroid isomerase with a substrate molecule (green) in the active site. The electrostatic character of the surface is indicated by color (red, acidic; blue, basic). (REPRINTED WITH PERMISSION FROM ZHENG RONG WU ET AL., SCIENCE 276:417, 1997, COURTESY OF MICHAEL F. SUMMERS, UNIVERSITY OF MARYLAND, BALTIMORE COUNTY; COPYRIGHT 1997 AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

3.1 BIOENERGETICS

A living cell bustles with activity. Macromolecules of all types are assembled from raw materials, waste products are produced and excreted, genetic instructions flow from the nucleus to the cytoplasm, vesicles are moved along the secretory pathway, ions are pumped across cell membranes, and so forth. To maintain such a high level of activity, a cell must acquire and expend energy. The study of the various types of energy transformations that occur in living organisms is referred to as **bioenergetics**.

The Laws of Thermodynamics and the Concept of Entropy

Energy is defined as the capacity to do work, that is, the capacity to change or move something. **Thermodynamics** is the study of the changes in energy that accompany events in the universe. In the following pages, we will focus on a set of concepts that allow us to predict the direction that events will take and whether or not an input of energy is required to cause the event to happen. However, thermodynamic measurements provide no help in determining how rapidly a specific process will occur or the mechanism used by the cell to carry out the process.

The First Law of Thermodynamics   **The first law of thermodynamics** is the law of conservation of energy. It states that energy can neither be created nor destroyed. Energy can, however, be converted (*transduced*) from one form to another. The **transduction** of electric energy to mechanical energy occurs when we plug in a clock (Figure 3.1*a*), and chemical energy is converted to thermal energy when fuel is burned in an oil heater. Cells are also capable of energy transduction. As discussed in later chapters, the chemical energy stored in certain biological molecules, such as ATP, is converted to mechanical energy when organelles are moved from place to place in a cell, to electrical energy when ions flow across a membrane, or to thermal energy when heat is released during muscle contraction (Figure 3.1*b*). The most important energy transduction in the biological world is the conversion of sunlight into chemical energy—the process of photosynthesis—which provides the fuel that directly or indirectly powers the activities of nearly all forms of life.[1] A number of animals, including fireflies and luminous fish, are able to convert chemical energy back into light (Figure 3.1*c*). Regardless of the transduction process, however, the total amount of energy in the universe remains constant.

To discuss energy transformations involving matter, we need to divide the universe into two parts: the *system* under study and the remainder of the universe, which we will refer to as the *surroundings*. A system can be defined in various ways: it may be a certain space in the universe or a certain amount of matter. For example, the system may be a living cell. The changes in a system's energy that occur during an event are manifested in two ways—as a change in the heat content of the system and in the performance of work. Even though the system may lose or gain energy, the first law of thermodynamics indicates that the loss or gain must be balanced by a corresponding gain or loss in the surroundings, so that the amount in the universe as a whole remains constant. The energy of the system is termed the *internal energy* (E), and its change during a transformation is ΔE (delta E). One way to describe the first law of thermodynamics is that $\Delta E = Q - W$, where Q is the heat energy and W is the work energy.

---

[1] Several communities of organisms are known that are independent of photosynthesis. These include communities that reside in the hydrothermal vents at the bottom of the ocean floor that depend on energy obtained by bacterial chemosynthesis.

(a)   *(b)*   *(c)*
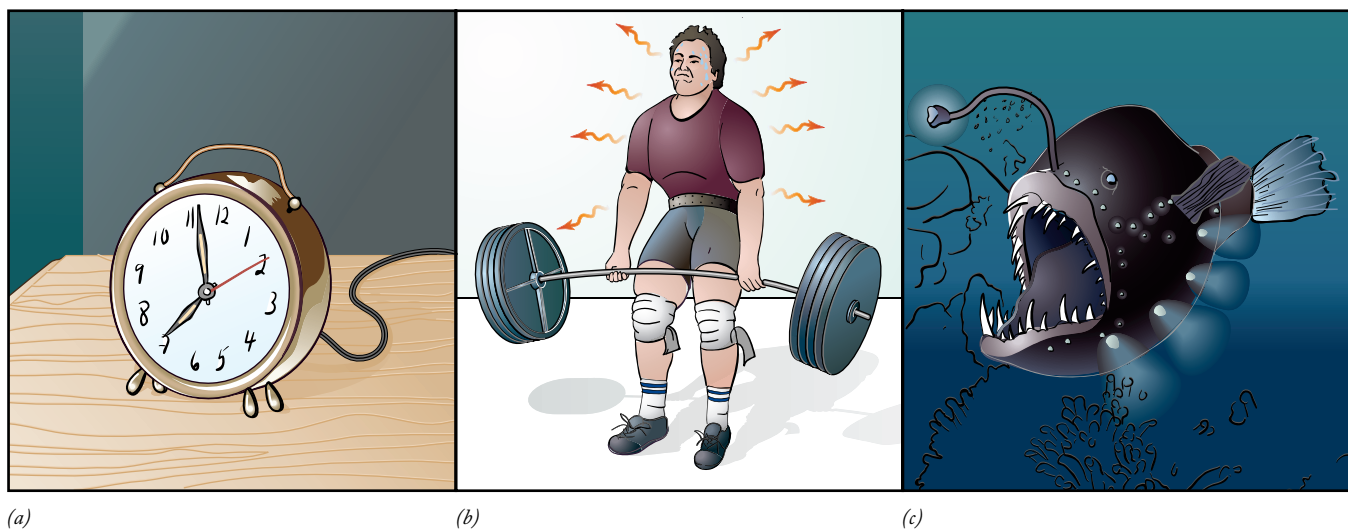
FIGURE 3.1 Examples of energy transduction. (*a*) Conversion of electrical energy to mechanical energy, (*b*) conversion of chemical energy to mechanical and thermal energy, (*c*) conversion of chemical energy to light energy.

Depending on the process, the internal energy of the system at the end can be greater than, equal to, or less than its internal energy at the start, depending on its relationship to its surroundings (Figure 3.2). In other words, ΔE can be positive, zero, or negative. Consider a system to be the contents of a reaction vessel. As long as there is no change in pressure or volume of the contents, there is no work being done by the system on its surroundings, or vice versa. In that case, the internal energy at the end of the transformation will be greater than that at the beginning if heat is absorbed and less if heat is released. Reactions that lose heat are termed **exothermic**, and ones that gain heat are **endothermic**, and there are many reactions of both types. Because ΔE for a particular reaction can be positive or negative, it gives us no information as to the likelihood that a given event will occur. To determine the likelihood of a particular transformation, we need to consider some additional concepts.

The Second Law of Thermodynamics The second law of thermodynamics expresses the concept that events in the universe have direction; they tend to proceed "downhill" from a state of higher energy to a state of lower energy. Thus, in any energy transformation, there is a decreasing availability of energy for doing additional work. Rocks fall off cliffs to the ground below, and once at the bottom, their ability to do additional work is reduced; it is very unlikely that they will lift themselves back to the top of the cliff. Similarly, opposite charges normally move together, not apart, and heat flows from a warmer to a cooler body, not the reverse. Such events are said to be **spontaneous**, a term that indicates they are thermodynamically favorable and can occur *without the input of external energy*.

The concept of the second law of thermodynamics was formulated originally for heat engines, and the law carried with it the idea that it is thermodynamically impossible to construct a perpetual-motion machine. In other words, it is impossible for a machine to be 100 percent efficient, which would be required if it were to continue functioning without the input of external energy. Some of the energy is inevitably lost as the machine carries out its activity. A similar relationship holds true for living organisms. For example, when a giraffe browses on the leaves of a tree or a lion preys on the giraffe, much of the chemical energy in the food never becomes available to the animal having the meal.

The loss of available energy during a process is a result of a tendency for the randomness, or disorder, of the universe to increase every time there is a transfer of energy. This gain in disorder is measured by the term **entropy**, and the loss of available energy is equal to $T\Delta S$, where ΔS is the change in entropy between the initial and final states. Entropy is associated with the *random* movements of particles of matter, which, because they are random, cannot be made to accomplish a *directed* work process. According to the second law of thermodynamics, every event is accompanied by an increase in the entropy of the universe. When a sugar cube is dropped into a cup of hot water, for example, there is a spontaneous shift of the molecules from an ordered state in the crystal to a much more disordered condition when the sugar molecules are spread throughout the solution (Figure 3.3a). As the molecules of the sugar cube dissolve into solution, their freedom of movement increases, as does the entropy of the system. The change from a concentrated to a dispersed state results from the random movements of the molecules. The sugar molecules eventually spread themselves equally through the available volume because the state of uniform distribution is the most probable state.

The release of heat, for example, from the oxidation of glucose within a cell or from the friction generated as blood flows through a vessel, is another example of an increase in entropy. The release of thermal energy by living organisms increases the rate of random movements of atoms and molecules; it cannot be redirected to accomplish additional work.

(a)   (b)
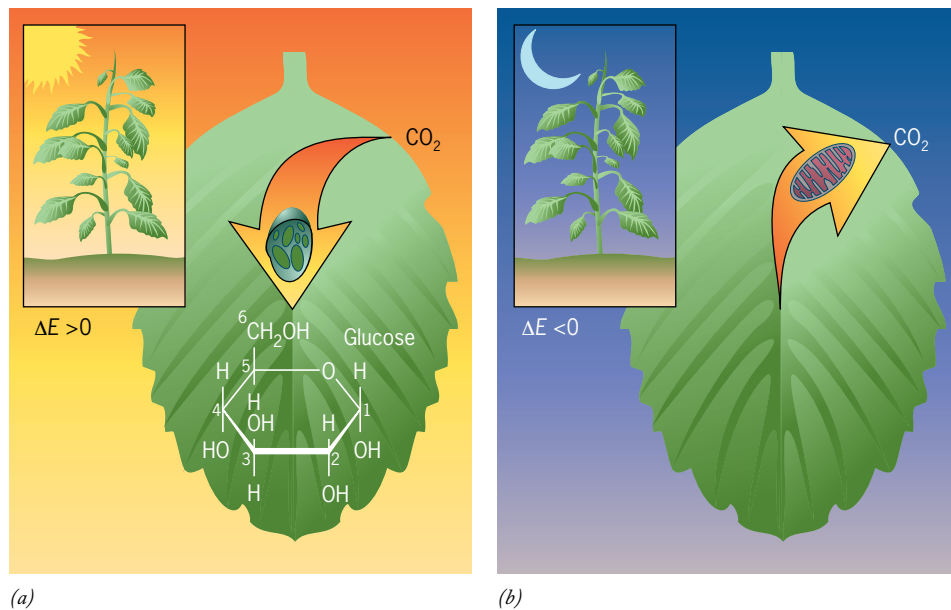
FIGURE 3.2 A change in a system's internal energy. In this example, the system will be defined as a particular leaf of a plant. (*a*) During the day, sunlight is absorbed by photosynthetic pigments in the leaf's chloroplasts and used to convert CO_2 into carbohydrates, such as the glucose molecule shown in the drawing (which is subsequently incorporated into sucrose or starch). As the cell absorbs light, its internal energy increases; the energy present in the remainder of the universe has to decrease. (*b*) At night, the energy relationship between the cell and its surroundings is reversed as the carbohydrates produced during the day are oxidized to CO_2 in the mitochondria and the energy is used to run the cell's nocturnal activities.

(a)

(b)
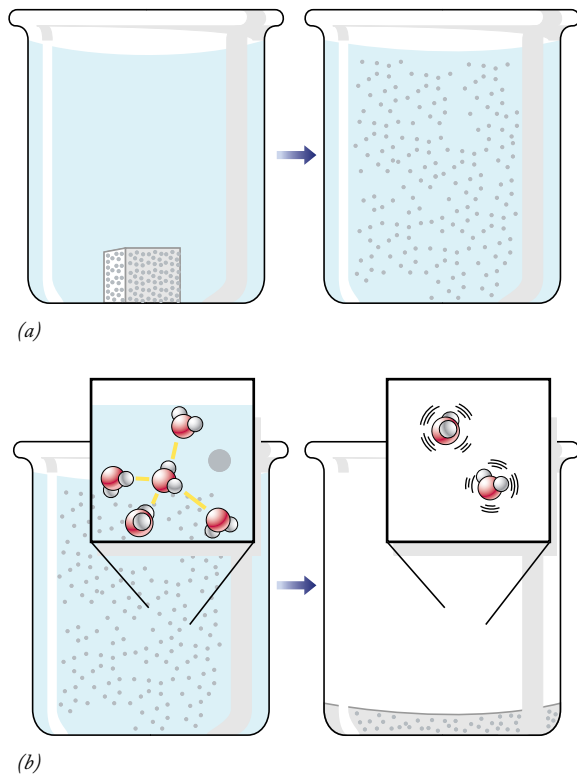
FIGURE 3.3 Events are accompanied by an increase in the entropy of the universe. (*a*) A sugar cube contains sucrose molecules in a highly ordered arrangement in which their freedom of movement is restricted. As the cube dissolves, the freedom of movement of the sucrose molecules is greatly increased, and their random movement causes them to become equally distributed throughout the available space. Once this occurs, there will be no further tendency for redistribution, and the entropy of the system is at a maximum. (*b*) Sugar molecules spread randomly through a solution can be returned to an ordered state, but only if the entropy of the surroundings is increased, as occurs when the more ordered water molecules of the liquid phase become disordered following evaporation.

Since the energy of molecular and atomic movements increases with temperature, so too does the entropy. It is only at absolute zero (0 K), when all movements cease, that the entropy is zero.

As with other spontaneous events, we must distinguish between the system and its surroundings. The second law of thermodynamics indicates only that the total entropy in the universe must increase; the disorder within one part of the universe (the system) can decrease at the greater expense of its surroundings. The dissolved sugar in Figure 3.3*a* can decrease in entropy; it can be recrystallized by evaporating off the water (Figure 3.3*b*). The consequence of this process, however, is an even greater increase in the entropy of the surroundings. The increased freedom of movement of the water molecules in the gaseous phase more than balances the decrease in freedom of the molecules of the sugar crystals.

Life operates on a similar principle. Living organisms are able to decrease their own entropy by increasing the entropy of their environment. Entropy is decreased in an organism when relatively simple molecules, such as amino acids, are ordered

into more complex molecules, such as the protein myoglobin in a muscle cell. For this to occur, however, the entropy of the environment must increase, which is accomplished as complex, ordered molecules such as glycogen stored in liver or muscle tissue are converted into heat and smaller, less ordered compounds (such as CO_2 and H_2O) are released to the environment. It is this feature of metabolism that allows living organisms to maintain such a highly ordered and improbable state—at least for a while.

Another measure of the energy state of a living organism is provided by the information content of its macromolecules. Information is a subject that is difficult to define but easy to recognize. Information can be measured in terms of the ordered arrangement of a structure's subunits. For example, proteins and nucleic acids, in which the specific linear sequence of the subunits is highly ordered, are low in entropy and high in information content. Maintaining a state of high information content (low entropy) requires the input of energy. Consider just one molecule of DNA located in one cell in your liver. That cell has dozens of different proteins whose sole job is to patrol the DNA, looking for damage and repairing it (discussed in Section 13.2). Nucleotide damage in an active cell can be so great that, without this expenditure of energy, the information content of DNA would rapidly deteriorate. Those organisms that are better able to slow the inevitable increase in entropy have longer life spans (page 34).

Free Energy

Together, the first and second laws of thermodynamics indicate that the energy of the universe is constant, but the entropy continues to increase toward a maximum. The concepts inherent in the first two laws were combined by the American chemist, J. Willard Gibbs, in 1878 into the expression $\Delta H = \Delta G + T\Delta S$, where ΔG (delta G) is the change in **free energy**, that is, the change during a process in the energy available to do work; ΔH is the change in *enthalpy* or total energy content of the system (equivalent to ΔE for our purposes), T is the absolute temperature (K = °C + 273), and ΔS is the change in the entropy of the system. The equation states that the total energy change is equal to the sum of the changes in useful energy (ΔG) and energy that is unavailable to do further work ($T\Delta S$).

When rearranged and written as $\Delta G = \Delta H - T\Delta S$, the above equation provides a measure of the spontaneity of a particular process. It allows one to predict the direction in which a process will proceed and the extent to which that process will occur. All *spontaneous* energy transformations must have a negative ΔG; that is, the process must proceed toward a state of lower free energy. The magnitude of ΔG indicates the maximum amount of energy that can be passed on for use in another process but tells us nothing about how rapidly the process will occur. Processes that can occur spontaneously, that is, processes that are thermodynamically favored (have a $-\Delta G$), are described as **exergonic**. In contrast, if the ΔG for a given process is positive, then it cannot occur spontaneously. Such processes are thermodynamically unfavorable and are described as **endergonic**. As we will see,

FIGURE 3.4 When water freezes, its entropy decreases because the water molecules of ice exist in a more ordered state with less freedom of movement than in a liquid state. The decrease in entropy is particularly apparent in the formation of a snowflake. (© NURIDSANY AND PERENNOU/PHOTO RESEARCHERS.)

TABLE 3.1 Thermodynamics of the Ice–Water Transformation

Temp. (°C)	ΔE (cal/mol)	ΔH (cal/mol)	ΔS (cal/mol · °C)	$T\Delta S$ (cal/mol)	ΔG (cal/mol)
−10	−1343	−1343	−4.9	−1292	−51
0	−1436	−1436	−5.2	−1436	0
+10	−1529	−1529	−5.6	−1583	+54

Source: I. M. Klotz, *Energy in Biochemical Reactions,* Academic Press, 1967.

reactions that are normally endergonic can be made to occur by coupling them to energy-releasing processes.

The signs of ΔH and ΔS for a given transformation can be positive or negative, depending on the relation between the system and its surroundings. (Hence, ΔH will be positive if heat is gained by the system and negative if heat is lost; ΔS will be positive if the system becomes more disordered and negative if it becomes more ordered.) The counterplay between ΔH and ΔS is illustrated by the ice–water transformation. The conversion of water from the liquid to the solid state is accompanied by a decrease in entropy (ΔS is negative, as illustrated in Figure 3.4) and a decrease in enthalpy (ΔH is negative). In order for this transformation to occur (i.e., for ΔG to be negative), ΔH must be more negative than $T\Delta S$, a condition that occurs only below 0°C. This relationship can be seen from Table 3.1, which indicates the values for the different terms if one mole of water were to be converted to ice at 10°C, 0°C, or −10°C. In all cases, regardless of the temperature, the energy level of the ice is less than that of the liquid (the ΔH is negative). However, at the higher temperature, the entropy term of the equation ($T\Delta S$) is more negative than the enthalpy term; therefore, the free-energy change is positive, and the process cannot occur spontaneously. At 0°C, the system is in equilibrium; at −10°C, the solidification process is favored, that is, the ΔG is negative.

Free-Energy Changes in Chemical Reactions   Now that we have discussed the concept of free energy in general terms, we can apply the information to chemical reactions within the cell. All chemical reactions within the cell are reversible, and therefore we must consider two reactions occurring simultaneously, one forward and the other in reverse. According to the law of mass action, the rate of a reaction is proportional to

the concentration of the reactants. Consider, for example, the following reaction:

$$A + B \rightleftharpoons C + D$$

The rate of the forward reaction is directly proportional to the product of the molar concentrations of A and B. The rate of the forward reaction can be expressed as $k_1[A][B]$, where k_1 is a rate constant for the forward reaction. The rate of the backward reaction equals $k_2[C][D]$. All chemical reactions proceed, however slowly, toward a state of equilibrium, that is, toward a point at which the rates of the forward and backward reactions are equal. At equilibrium, the same number of A and B molecules are converted into C and D molecules per unit time as are formed from them. At equilibrium, therefore,

$$k_1[A][B] = k_2[C][D]$$

which can be rearranged to

$$\frac{k_1}{k_2} = \frac{[C][D]}{[A][B]}$$

In other words, at equilibrium there is a predictable ratio of the concentration of products to the concentration of reactants. This ratio, which is equal to k_1/k_2, is termed the **equilibrium constant, K_{eq}.**

The equilibrium constant allows one to predict the direction (forward or reverse) in which the reaction is favored under a given set of conditions. Suppose, for example, that we are studying the above reaction, and we have just mixed the four components (A, B, C, D) so that each is present at an initial concentration of 0.5 M.

$$\frac{[C][D]}{[A][B]} = \frac{[0.5][0.5]}{[0.5][0.5]} = 1$$

The direction in which this reaction will proceed depends on the equilibrium constant. If the K_{eq} is greater than 1, the reaction will proceed at a greater rate toward the formation of the products C and D than in the reverse direction. For example, if the K_{eq} happens to be 9.0, then the concentration of the reactants and products at equilibrium in this particular reaction mixture will be 0.25 M and 0.75 M, respectively.

$$\frac{[C][D]}{[A][B]} = \frac{[0.75][0.75]}{[0.25][0.25]} = 9$$

If, on the other hand, the K_{eq} is less than 1, the reverse reaction will proceed at a greater rate than the forward reaction, so that the concentration of A and B will rise at the expense of C and D. It follows from these points that the net direction in

which the reaction is proceeding at any moment depends on the relative concentrations of all participating molecules and can be predicted from the K_{eq}.

Let us return to the subject of energetics. The ratio of reactants to products present at equilibrium is determined by the relative free-energy levels of the reactants and products. As long as the total free energy of the reactants is greater than the total free energy of the products, then ΔG has a negative value and the reaction proceeds in the direction of formation of products. The greater the ΔG, the farther the reaction is from equilibrium and the more work that can be performed by the system. As the reaction proceeds, the difference in free-energy content between the reactants and products decreases (ΔG becomes less negative), until at equilibrium the difference is zero ($\Delta G = 0$) and no further work can be obtained.

Because ΔG for a given reaction depends on the reaction mixture present at a given time, it is not a useful term to compare the energetics of various reactions. To place reactions on a comparable basis and to allow various types of calculations to be made, a convention has been adopted to consider the free-energy change that occurs during a reaction under a set of *standard conditions*. For biochemical reactions, the conditions are arbitrarily set at 25°C (298 K) and 1 atm of pressure, with all the reactants and products present at 1.0 M concentration, except for water, which is present at 55.6 M, and H^+ at 10^{-7} M (pH 7.0).[2] The **standard free-energy change ($\Delta G°'$)** describes the free energy released when reactants are converted to products under these standard conditions. It must be kept in mind that standard conditions do not prevail in the cell, and therefore one must be cautious in the use of values for standard free-energy differences in calculations of cellular energetics.

The relationship between the equilibrium constant and standard free-energy change is given by the equation

$$\Delta G°' = -RT \ln K'_{eq}$$

When the natural log (ln) is converted to $\log_{10}$, the equation becomes

$$\Delta G°' = -2.303RT \log K'_{eq}$$

where R is the gas constant (1.987 cal/mol · K) and T is the absolute temperature (298 K).[3] Recall that the log of 1.0 is zero. Consequently, it follows from the above equation that reactions having equilibrium constants greater than 1.0 have negative $\Delta G°'$ values, which indicates that they can occur spontaneously under *standard conditions*. Reactions having equilibrium constants of less than 1 have positive $\Delta G°'$ values and cannot occur spontaneously under standard conditions. In other words, given the reaction written as follows: $A + B \rightleftharpoons C + D$, if the $\Delta G°'$ is negative, the reaction will go to the right when reactants and products are all present at 1.0 M concentration at pH 7. The greater the negative value, the farther to the right the reaction will proceed before equi-

[2] $\Delta G°'$ indicates standard conditions include pH 7, whereas $\Delta G°$ indicates standard conditions in 1.0 M H^+ (pH 0.0). The designation K'_{eq} also indicates a reaction mixture at pH 7.

[3] The right-hand side of this equation is equivalent to the amount of free energy lost as the reaction proceeds from standard conditions to equilibrium.

TABLE 3.2 Relation between ΔG°' and K'eq at 25°C

K'_{eq}	$\Delta G°'$ (kcal/mol)
10^6	−8.2
10^4	−5.5
10^2	−2.7
10^1	−1.4
10^0	0.0
10^{-1}	1.4
10^{-2}	2.7
10^{-4}	5.5
10^{-6}	8.2

librium is reached. Under the same conditions, if the $\Delta G°'$ is positive, the reaction will proceed to the left; that is, the reverse reaction is favored. The relation between $\Delta G°'$ and K'_{eq} is shown in Table 3.2.

Free-Energy Changes in Metabolic Reactions One of the most important chemical reactions in the cell is the hydrolysis of ATP (Figure 3.5). In the reaction,

$$ATP + H_2O \rightarrow ADP + P_i$$
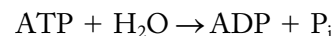
the standard free-energy difference between the products and reactants is −7.3 kcal/mol. Based on this information, it is evident that the hydrolysis of ATP is a highly favorable (exergonic) reaction, that is, one that tends toward a large [ADP]/[ATP] ratio at equilibrium. There are several reasons why this reaction is so favorable, one of which is evident in Figure 3.5. The electrostatic repulsion created by four closely spaced negative charges on ATP^{4-} is partially relieved by formation of ADP^{3-}.

It is important that the difference between ΔG and $\Delta G°'$ be kept clearly in mind. The $\Delta G°'$ is a fixed value for a given reaction and indicates the direction in which that reaction would proceed were the system at standard conditions. Since standard conditions do not prevail within a cell, $\Delta G°'$ values cannot be used to predict the direction in which a particular reaction is proceeding at a given moment within a particular cellular compartment. To do this, one must know the ΔG, which is determined by the concentrations of the reactants and products that are present at the time. At 25°C

$$\Delta G = \Delta G°' + 2.303RT \log \frac{[C][D]}{[A][B]}$$

$$\Delta G = \Delta G°' + 2.303(1.987 \text{ cal/mol} \cdot K)(298 K) \log \frac{[C][D]}{[A][B]}$$

$$\Delta G = \Delta G°' + (1.4 \text{ kcal/mol}) \log \frac{[C][D]}{[A][B]}$$

where [A], [B], [C], and [D] are the actual concentrations at the time. Calculation of the ΔG reveals the direction in which the reaction in the cell is proceeding and how close the
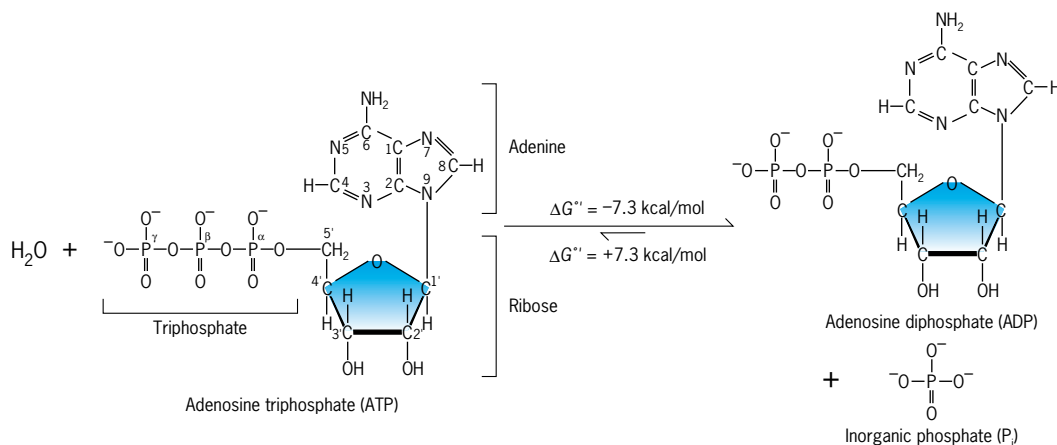
FIGURE 3.5 ATP hydrolysis. Adenosine triphosphate (ATP) is hydrolyzed as part of many biochemical processes. In most reactions, as shown here, ATP is hydrolyzed to ADP and inorganic phosphate (P_i), but in some cases (not shown) it is hydrolyzed to AMP, a compound with only one phosphate group, and pyrophosphate (PP_i). Both of these reactions have essentially the same $\Delta G°'$ of -7.3 kcal/mol (-30.5 kJ/mol).

particular reaction in question is to equilibrium. For example, typical cellular concentrations of the reactants and products in the reaction for ATP hydrolysis might be [ATP] = 10 mM; [ADP] = 1 mM; [P_i] = 10 mM. Substituting these values into the equation,

$$\Delta G = \Delta G°' + 2.303RT \log \frac{[ADP][P_i]}{[ATP]}$$

$$\Delta G = -7.3 \text{ kcal/mol} + (1.4 \text{ kcal/mol}) \log \frac{[10^{-3}][10^{-2}]}{[10^{-2}]}$$

$$\Delta G = -7.3 \text{ kcal/mol} + (1.4 \text{ kcal/mol})(-3)$$
$$\Delta G = -11.5 \text{ kcal/mol (or } -46.2 \text{ kJ/mol)}$$

Thus, even though the $\Delta G°'$ for the hydrolysis of ATP is -7.3 kcal/mol, the typical ΔG in the cell for this reaction is about -12 kcal/mol because a cell maintains a high [ATP]/[ADP] ratio.

Cells carry out many reactions with positive $\Delta G°'$ values because the relative concentrations of reactants and products favor the progress of the reactions. This can happen in two ways. The first illustrates the important difference between ΔG and $\Delta G°'$, and the second reveals the manner in which reactions with positive $\Delta G°'$ values can be driven in the cell by the input of stored chemical energy.

Consider the reaction of glycolysis (see Figure 3.24) in which dihydroxyacetone phosphate is converted to glyceraldehyde 3-phosphate. The $\Delta G°'$ for this reaction is $+1.8$ kcal/mol, yet the formation of the product of this reaction takes place in the cell. The reaction proceeds because other cellular reactions maintain the ratio of the reactant to product above that defined by the equilibrium constant. As long as this condition holds, the ΔG will be negative and the reaction will continue spontaneously in the direction of formation of glyceraldehyde 3-phosphate. This brings up an important characteristic of cellular metabolism, namely, that specific reactions cannot be considered independently as if they were occurring in isolation in a test tube. Hundreds of reactions occur simultaneously in a cell. All these reactions are interrelated because a product of one reaction becomes a reactant for the next reaction in the sequence and so on throughout one metabolic pathway and into the next. To maintain the production of glyceraldehyde 3-phosphate at the expense of dihydroxyacetone phosphate, the reaction is situated within a metabolic pathway in such a way that the product is removed by the next reaction in the sequence at a sufficiently rapid rate that a favorable ratio of the concentrations of these two molecules is retained.

Coupling Endergonic and Exergonic Reactions   Reactions with large positive $\Delta G°'$ values are typically driven by the input of energy. Consider the formation of the amino acid glutamine from glutamic acid by the enzyme glutamine synthetase:

Glutamic acid + NH_3 →
glutamine $\Delta G°' = +3.4$ kcal/mol
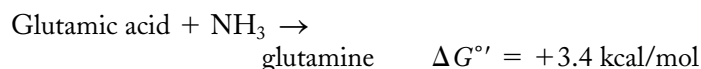
This endergonic reaction takes place in the cell because glutamic acid is actually converted to glutamine in two sequential reactions, both of which are exergonic:

1st reaction: Glutamic acid + ATP →
glutamyl phosphate + ADP

2nd reaction: Glutamyl phosphate + NH_3 →
glutamine + P_i

Overall reaction: Glutamic acid + ATP + NH_3 →
glutamine + ADP + Pi
$\Delta G°' = -3.9$ kcal/mol
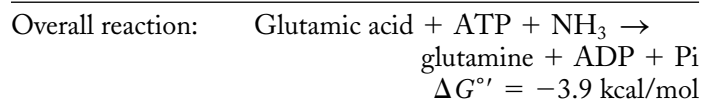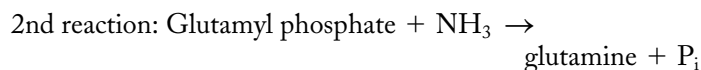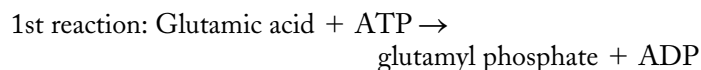
The formation of glutamine is said to be *coupled* to the hydrolysis of ATP. As long as the ΔG for ATP hydrolysis is more negative than the ΔG for glutamine synthesis from glutamic acid is positive, the "downhill" ATP hydrolysis reaction can be used to drive the "uphill" synthesis of glutamine. To couple the two chemical reactions, the product of the first reaction becomes the reactant for the second. The bridge between the two reactions—glutamyl phosphate in this case—is called the *common intermediate*. What happens, in essence, is that the

exergonic hydrolysis of ATP is taking place in two steps. In the first step, glutamic acid acts as an acceptor of the phosphate group, which is displaced by NH_3 in the second step.

The hydrolysis of ATP can be used in the cell to drive reactions leading to the formation of molecules such as glutamine because ATP levels are maintained at much higher levels (relative to those of ADP) than they would be at equilibrium. This can be demonstrated by the following calculation. As noted above, a typical cellular concentration of P_i might be 10 mM. To calculate the equilibrium ratio of [ATP]/[ADP] under these conditions, we can set ΔG at the equilibrium value of 0 and solve the following equation (taken from page 89) for [ADP]/[ATP]:

$$\Delta G = \Delta G^{\circ\prime} + (1.4 \text{ kcal/mol}) \log \frac{[\text{ADP}][\text{P}_i]}{[\text{ATP}]}$$

$$0 = -7.3 \text{ kcal/mol} + (1.4 \text{ kcal/mol}) \log \frac{[\text{ADP}][10^{-2}]}{[\text{ATP}]}$$

$$0 = -7.3 \text{ kcal/mol} + (1.4 \text{ kcal/mol})\left(\log 10^{-2} + \log \frac{[\text{ADP}]}{[\text{ATP}]} \right)$$

$$+ 7.3 \text{ kcal/mol} = (1.4 \text{ kcal/mol})(-2) + (1.4 \text{ kcal/mol}) \log \frac{[\text{ADP}]}{[\text{ATP}]}$$

$$\log \frac{[\text{ADP}]}{[\text{ATP}]} = \frac{10.1 \text{ kcal/mol}}{1.4 \text{ kcal/mol}} = 7.2$$

$$\frac{[\text{ADP}]}{[\text{ATP}]} = 1.6 \times 10^7$$

Thus, at equilibrium, the ADP concentration would be expected to be more than 10^7 times that of ATP, but in fact, the ATP concentrations in most cells are 10 to 100 times that of ADP. This is a crucial point because it's the relative concentrations of ATP and ADP that matter. If a cell were to contain an equilibrium mixture of ATP, ADP, and P_i, it wouldn't matter how much ATP was present, the cell would have no capacity to perform work.

ATP hydrolysis is used to drive most endergonic processes within the cell, including chemical reactions such as the one just described, the separation of charge across a membrane, the concentration of a solute, the movement of filaments in a muscle cell, and the properties of proteins (Figure 3.6). ATP can be used for such diverse processes because its terminal phosphate group can be transferred to a variety of different types of molecules, including amino acids, sugars, lipids, and proteins. In most coupled reactions, the phosphate group is transferred in an initial step from ATP to one of these acceptors and is subsequently removed in a second step (see Figure 4.46 as an example).

Equilibrium versus Steady-State Metabolism As reactions tend toward equilibrium, the free energy available to do work decreases toward a minimum and the entropy increases toward a maximum. Thus, the farther a reaction is kept from its equilibrium state, the less its capacity to do work is lost to the increase in entropy. Cellular metabolism is essentially nonequilibrium metabolism; that is, it is characterized by nonequilibrium ratios of products to reactants. This does not mean that some reactions do not occur at or near equilibrium

(a)

(b)

Glutamic acid + NH_3 $\rightleftharpoons$ Glutamine
Glutamic acid + ATP + NH_3 $\rightleftharpoons$ Glutamine + ADP + P_i

(c)

(d)

(e)
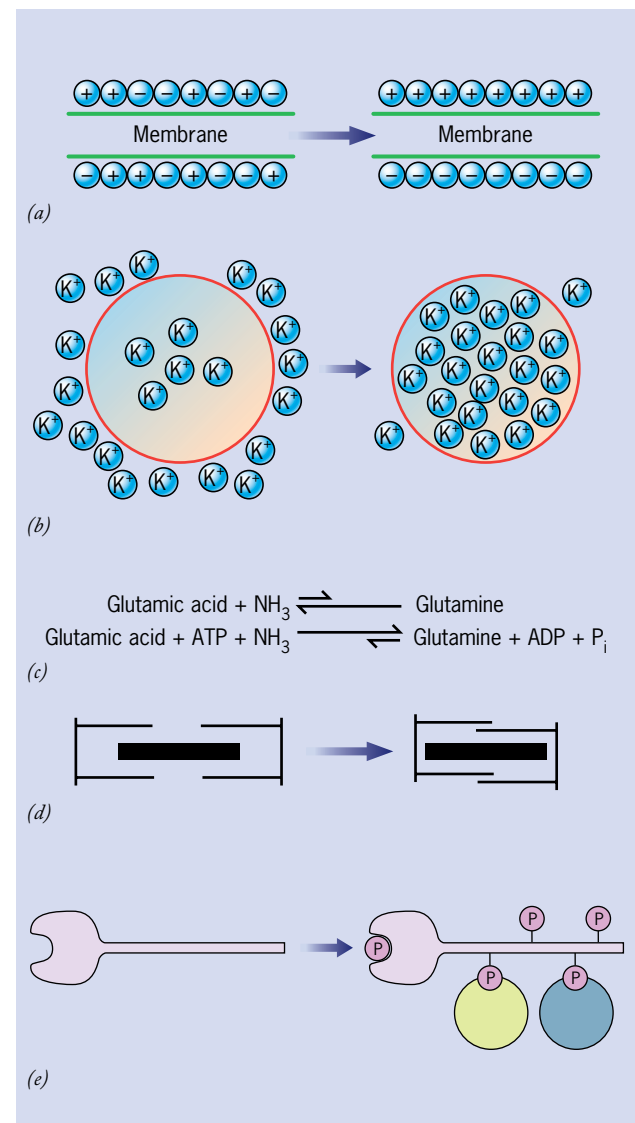
FIGURE 3.6 A few roles for ATP hydrolysis. In the cell, ATP may be used to (*a*) separate charge across a membrane; (*b*) concentrate a particular solute within the cell; (*c*) drive an otherwise unfavorable chemical reaction; (*d*) slide filaments across one another, as occurs during the shortening of a muscle cell; (*e*) donate a phosphate group to a protein, thereby changing its properties and bringing about a desired response. In this case the added phosphate groups serve as binding sites for other proteins.

within cells. In fact, many of the reactions of a metabolic pathway may be near equilibrium (see Figure 3.25). However, at least one and often several reactions in a pathway are poised far from equilibrium, making them essentially irreversible. These are the reactions that keep the pathway going in a single direction. These are also the reactions that are subject to cellular regulation because the flow of materials through the entire pathway can be greatly increased or decreased by stimulating or inhibiting the activity of the enzymes that catalyze these reactions.

The basic principles of thermodynamics were formulated using nonliving, *closed* systems (no exchange of matter between

the system and its surroundings) under reversible equilibrium conditions. The unique features of cellular metabolism require a different perspective. Cellular metabolism can maintain itself at irreversible, nonequilibrium conditions because unlike the environment within a test tube, the cell is an *open* system. Materials and energy are continually flowing into the cell from the bloodstream or culture medium. The extent of this input into cells from the outside becomes apparent if you simply hold your breath. We depend from minute to minute on an external source of oxygen because oxygen is a very important reactant in cellular metabolism. The continual flow of oxygen and other materials into and out of cells allows cellular metabolism to exist in a **steady state** (Figure 3.7). In a steady state, the concentrations of reactants and products remain relatively constant, even though the individual reactions are not necessarily at equilibrium. This is not to suggest that the concentrations of cellular metabolites do not change. Cells are capable of continually adjusting the concentration of key substances in response to changing conditions. A rise or fall in the level of regulatory substances, such as the hormone insulin, for exam-

ple, can cause a dramatic surge or decline in the production of sugars, amino acids, or fats. Cells, in other words, exist in a state of *dynamic* nonequilibrium, in which the rates of forward and reverse reactions can be increased or decreased instantaneously in response to changing conditions.

REVIEW ?

1. Describe the differences between the first and second laws of thermodynamics and how, when taken together, they are able to describe the direction of events that take place in the universe.

2. How is the maintenance of the ordered living state consistent with the second law of thermodynamics?

3. Describe two examples in which the entropy of a system decreases and two examples in which the entropy of a system increases.

4. Discuss the differences between ΔG and $\Delta G^{\circ\prime}$; between the relative rates of the forward and reverse reactions when ΔG is negative, zero, or positive. What is the relationship between $\Delta G^{\circ\prime}$ and K'_{eq}? How can a cell carry out a reaction that has a $+\Delta G^{\circ\prime}$?

5. How is it possible for a cell to maintain an [ATP]/[ADP] ratio greater than one? How does this ratio differ from that expected at equilibrium?

6. How is it that ice formation does not take place at temperatures above 0°C?

3.2 ENZYMES AS BIOLOGICAL CATALYSTS

At the end of the nineteenth century, a debate was raging as to whether or not the process of ethanol formation required the presence of intact yeast cells. On one side of the debate was the organic chemist Justus von Liebig, who argued that the reactions of fermentation that produced the alcohol were no different from the types of organic reactions he had been studying in the test tube. On the other side was the biologist Louis Pasteur, who argued that the process of fermentation could only occur within the confines of an intact, highly organized, living cell.

In 1897, two years after Pasteur's death, Hans Büchner, a bacteriologist, and his brother Eduard, a chemist, were preparing "yeast juice"—an extract made by grinding yeast cells with sand grains and then filtering the mixture through filter paper. They wanted to preserve the yeast juice for later use. After trying to preserve the extract with antiseptics and failing, they attempted to protect the preparation from spoilage by adding sugar, the same procedure used to preserve jams and jellies. Instead of preserving the solution, the yeast juice produced gas from the sugar, and it bubbled continuously for days. After further analysis, Eduard discovered that fermentation was producing ethanol and bubbles of carbon dioxide. Büchner had shown that fermentation did not require the presence of intact cells.

It was soon found, however, that fermentation was very different from the types of reactions carried out by organic

FIGURE 3.7 **Steady state versus equilibrium.** (*a*) As long as this amoeba can continue to take in nutrients from the outside world, it can harvest the energy necessary to maintain concentrations of compounds at a steady state, which may be far from equilibrium. The steady-state concentrations of ATP and ADP are indicated by the colored dots and the histogram. (*b*) When the amoeba dies, the concentrations of ATP and ADP (as well as other biochemicals) shift toward their equilibrium ratios.

chemists. Fermentation required the presence of a unique set of catalysts that had no counterpart in the nonliving world. These catalysts were called **enzymes** (after the Greek for "in yeast"). Enzymes are the mediators of metabolism, responsible for virtually every reaction that occurs in a cell. Without enzymes, metabolic reactions would proceed so slowly as to be imperceptible.

The first evidence that enzymes are proteins was obtained in 1926 by James Sumner of Cornell University, when he crystallized the enzyme urease from jack beans and determined its composition. Though this finding was not greeted with much positive acclaim at the time, several other enzymes were soon shown to be proteins, and it became accepted for the next few decades that all biological catalysts were proteins. Eventually, it became evident that certain biological reactions are catalyzed by RNA molecules. For the sake of clarity, the term *enzyme* is still generally reserved for protein catalysts, while the term *ribozyme* is used for RNA catalysts. We will restrict the discussion in the present chapter to protein catalysts and discuss the properties of RNA catalysts in Chapter 11.

Even though enzymes are proteins, many of them are conjugated proteins; that is, they contain nonprotein components, called **cofactors**, which may be inorganic (metals) or organic (**coenzymes**). When present, cofactors are important participants in the functioning of the enzyme, often carrying out activities for which amino acids are not suited. In myoglobin, for example, as discussed in the previous chapter, the iron atom of a heme group is the site where oxygen is bound and stored until required by cellular metabolism.

The Properties of Enzymes

As is true of all catalysts, enzymes exhibit the following properties: (1) they are required only in small amounts; (2) they are not altered irreversibly during the course of the reaction, and therefore each enzyme molecule can participate repeatedly in individual reactions; and (3) they have no effect on the thermodynamics of the reaction. This last point is particularly important. Enzymes do not supply energy for a chemical reaction and therefore do not determine whether a reaction is thermodynamically favorable or unfavorable. Similarly, enzymes do not determine the ratio of products to reactants at equilibrium. These are inherent properties of the reacting chemicals. As catalysts, enzymes can only accelerate the rate at which a favorable chemical reaction proceeds.

There is no necessary relationship between the magnitude of the ΔG for a particular reaction and the rate at which that reaction takes place. The magnitude of ΔG informs us only of the difference in free energy between the beginning state and equilibrium. It is totally independent of the pathway or the time it takes to reach equilibrium. The oxidation of glucose, for example, is a very favorable process, as can be determined by the amount of energy released during its combustion. However, glucose crystals can be left out in room air virtually indefinitely without a noticeable conversion into less energetic materials. Glucose, in other words, is *kinetically* stable, even if it is *thermodynamically* unstable. Even if the sugar were to be dissolved, as long as the solution were kept sterile, it would not rapidly deteriorate. However, if one were to add a few bacteria, within a short period of time the sugar would be taken up by the cells and enzymatically degraded.

Enzymes are remarkably adept catalysts. Catalysts employed by organic chemists in the lab, such as acid, metallic platinum, and magnesium, generally accelerate reactions a hundred to a thousand times over the noncatalyzed rate. In contrast, enzymes typically increase the velocity of a reaction 10^8 to 10^{13} fold (Table 3.3). On the basis of these numbers, enzymes can accomplish in one second what would require from 3 years to 300,000 years if the enzyme were absent. Even more remarkably, they accomplish this feat at the mild temperature and pH found within a cell. In addition, unlike the inorganic catalysts used by chemists, enzymes are highly specific in the

TABLE 3.3 Catalytic Activity of a Variety of Enzymes

Enzyme	Nonenzymatic $t_{1/2}$[1]		Turnover number[2]	Rate enhancement[3]
OMP decarboxylase	78,000,000	yr	39	1.4×10^{17}
Staphylococcal nuclease	130,000	yr	95	5.6×10^{14}
Adenosine deaminase	120	yr	370	2.1×10^{12}
AMP nucleosidase	69,000	yr	60	6.0×10^{12}
Cytidine deaminase	69	yr	299	1.2×10^{12}
Phosphotriesterase	2.9	yr	2,100	2.8×10^{11}
Carboxypeptidase A	7.3	yr	578	1.9×10^{11}
Ketosteroid isomerase	7	wk	66,000	3.9×10^{11}
Triosephosphate isomerase	1.9	d	4,300	1.0×10^{9}
Chorismate mutase	7.4	hr	50	1.9×10^{6}
Carbonic anhydrase	5	sec	1×10^6	7.7×10^{6}
Cyclophilin, human	23	sec	13,000	4.6×10^{5}

Source: A. Radzicka and R. Wolfenden, *Science* 267:91, 1995. Copyright 1995 American Association for the Advancement of Science.
[1]The time that would elapse for half the reactants to be converted to product in the absence of enzyme.
[2]The number of reactions catalyzed by a single enzyme molecule per second when operating at a saturating substrate concentration.
[3]The increase in reaction rate achieved by the enzyme-catalyzed reaction over the noncatalyzed reaction.

reactants they can bind and the reaction they can catalyze. The reactants bound by an enzyme are called **substrates**. If, for example, the enzyme hexokinase is present in solution with a hundred low-molecular-weight compounds in addition to its substrate, glucose, only the glucose molecules will be recognized by the enzyme and undergo reaction. For all practical purposes, the other compounds may as well be absent. This type of specificity, whether between enzymes and substrates or other types of proteins and the substances that they bind, is crucial for maintaining the order required to sustain life.

In addition to their high level of activity and specificity, enzymes act as metabolic traffic directors in the sense that enzyme-catalyzed reactions are very orderly—the only products formed are the appropriate ones. This is very important because the formation of chemical by-products would rapidly take its toll on the life of a fragile cell. Finally, unlike other catalysts, the activity of enzymes can be regulated to meet the particular needs of the cell at a particular time. As will be evident in this chapter and the remainder of the text, the enzymes of a cell are truly a wondrous collection of miniature molecular machines.

Overcoming the Activation Energy Barrier

How are enzymes able to accomplish such effective catalysis? The first question to consider is why thermodynamically favorable reactions do not proceed on their own at relatively rapid rates in the absence of enzymes. Even ATP, whose hydrolysis is so favorable, is stable in a cell until broken down in a controlled enzymatic reaction. If this weren't the case, ATP would have little biological use.

Chemical transformations require that certain covalent bonds be broken within the reactants. For this to occur, the reactants must contain sufficient kinetic energy (energy of mo-

tion) that they overcome a barrier, called the **activation energy** **(E_A)**. This is expressed diagrammatically in Figure 3.8, where the activation energy is represented by the height of the curves. The reactants in a chemical reaction are often compared to an object resting on top of a cliff ready to plunge to the bottom. If left to itself, the object in all likelihood would remain there indefinitely. However, if someone were to come along and provide the object with sufficient energy to overcome the friction or other small barrier in its way and cause it to reach the edge of the cliff, it would spontaneously drop to the bottom. The object has the potential to fall to a lower energy state once the kinetic barriers have been removed.

In a solution at room temperature, molecules exist in a state of random movement, each possessing a certain amount of kinetic energy at a given instant. Among a population of molecules, their energy is distributed in a bell-shaped curve (Figure 3.9), some having very little energy and others much more. Molecules of high energy (activated molecules) remain as such only for a brief time, losing their excess energy to other molecules by collision. Let us consider a reaction in which one reactant molecule is split into two product molecules. If a given reactant molecule acquires sufficient energy to overcome the activation barrier, then the possibility exists that it will split into the two product molecules. The rate of the reaction depends on the number of reactant molecules that contain the necessary kinetic energy at any given time. One way to increase the reaction rate is to increase the energy of the reactants. This is most readily done in the laboratory by heating the reaction mixture (Figure 3.9). In contrast, applying heat to an enzyme-mediated reaction leads to the rapid inactivation of the enzyme due to its denaturation.

Reactants, when they are at the crest of the energy hump and ready to be converted to products, are said to be at the **transition state** (Figure 3.8). At this point, the reactants have
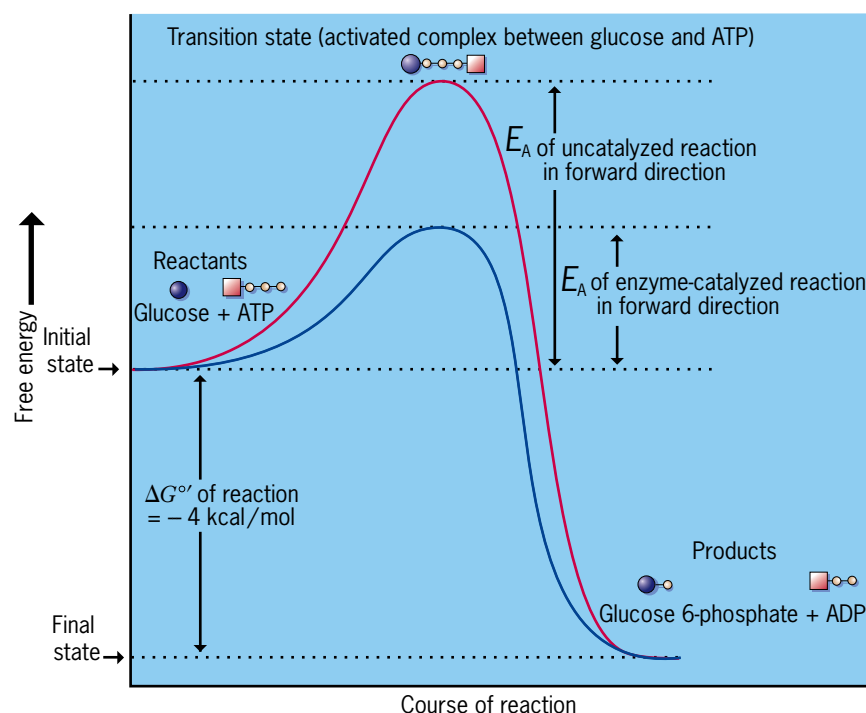
FIGURE 3.8   Activation energy and enzymatic reactions. Even though the formation of glucose 6-phosphate is a thermodynamically favored reaction ($\Delta G^{\circ\prime} = -4$ kcal/mol), the reactants must possess sufficient energy to achieve an activated state in which the atomic rearrangements necessary for the reaction can occur. The amount of energy required is called the activation energy (E_A) and is represented by the height of the curve. The activation energy is not a fixed value, but varies with the particular reaction pathway. E_A is greatly reduced when the reactants combine with an enzyme catalyst. (This diagram depicts a simple, one-step reaction mechanism. Many enzymatic reactions take place in two or more steps leading to the formation of intermediates (as in Figure 3.13). Each step in the reaction has a distinct E_A and a separate transition state.)
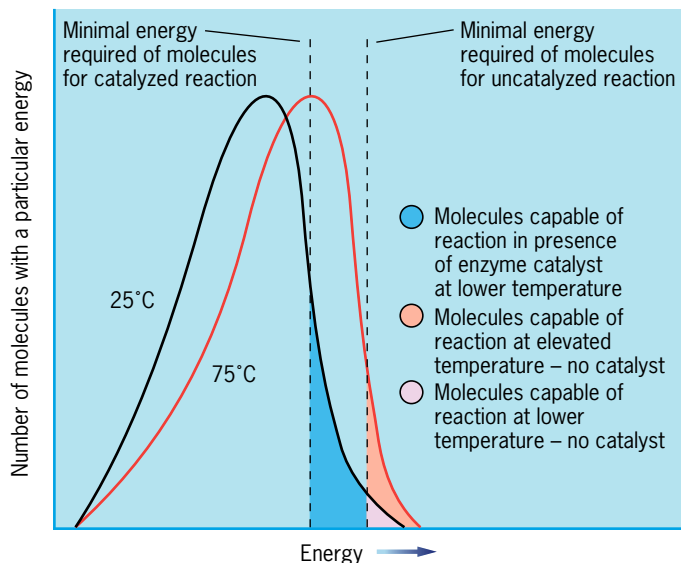
FIGURE 3.9 **The effect of lowering activation energy on the rate of a reaction.** The bell-shaped curves indicate the energy content of a population of molecules present in a reaction mixture at two different temperatures. The number of reactant molecules containing sufficient energy to undergo reaction is increased by either heating the mixture or adding an enzyme catalyst. Heat increases the rate of reaction by increasing the energy content of the molecules, while the enzyme does so by lowering the activation energy required for the reaction to occur.

formed a fleeting, activated complex in which bonds are being formed and broken. We can illustrate the nature of a transition state structure by examining the interconversion of D and L stereoisomers of proline, a reaction catalyzed by the bacterial enzyme proline racemase.
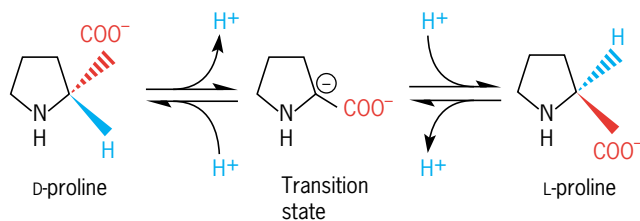
This reaction proceeds in either direction through the loss of a proton from the α-carbon of the proline molecule. As a result, the transition state structure contains a negatively charged carbanion in which the three bonds formed by the carbon atom are all in the same plane.

Unlike the difference in standard free energy for a reaction, the activation energy required to reach the transition state is not a fixed value but varies with the particular reaction mechanism utilized. Enzymes catalyze reactions by decreasing the magnitude of the activation energy barrier. Consequently, unlike catalysis by heat, enzymes cause their substrates to be very reactive without having to be raised to particularly high energy levels. A comparison between the percentage of molecules able to react in an enzyme-catalyzed reaction versus a noncatalyzed reaction is shown in Figure 3.9. Enzymes are able to lower activation energies by binding more tightly to the transition state than to the reactants, which stabilizes this acti-

vated complex, thereby decreasing its energy. The importance of the transition state can be demonstrated in numerous ways. For example:

- Compounds that resemble the transition state of a reaction tend to be very effective inhibitors of that reaction because they are able to bind so tightly to the catalytic region of the enzyme.
- Antibodies normally do not behave like enzymes but, rather, simply bind to molecules with high affinity. However, antibodies that bind to a compound that resembles a transition state for a reaction are often able to act like enzymes and catalyze the breakdown of that compound.

As the transition state is converted to products, the affinity of the enzyme for the bound molecule(s) decreases and the products are expelled.

The Active Site

As catalysts, enzymes accelerate bond-breaking and bond-forming processes. To accomplish this task, enzymes become intimately involved in the activities that are taking place among the reactants. Enzymes do this by forming a complex with reactants, called an **enzyme–substrate (ES) complex**. An ES complex is illustrated schematically in Figure 3.10 and the image in Figure 3.14. In most cases, the association between enzyme and substrate is noncovalent, though many examples are known in which a transient covalent bond is formed.
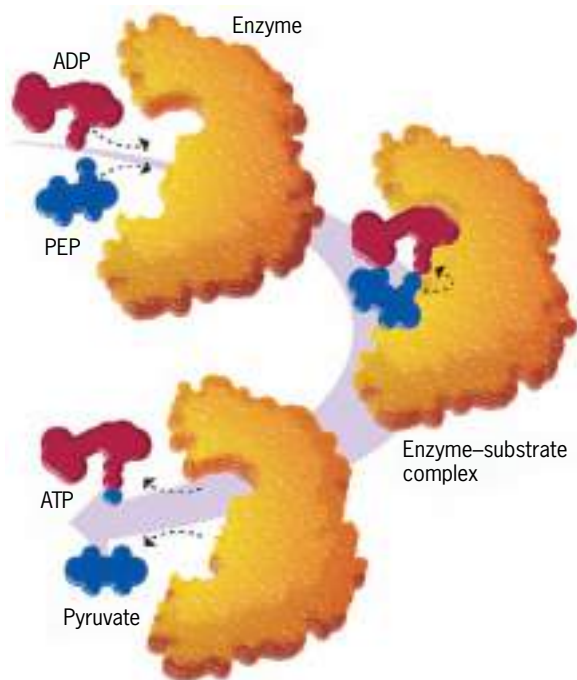
FIGURE 3.10 **Formation of an enzyme–substrate complex.** Schematic drawing of the reaction catalyzed by pyruvate kinase (see Figure 3.24) in which the two substrates, phosphoenolpyruvate (PEP) and ADP, bind to the enzyme to form an enzyme–substrate (ES) complex, which leads to the formation of the products, pyruvate and ATP.

That part of the enzyme molecule that is directly involved in binding the substrate is termed the **active site**. The active site and the substrate(s) have complementary shapes, enabling them to bind together with a high degree of precision. The binding of substrate to enzyme is accomplished by the same types of noncovalent interactions (ionic bonds, hydrogen bonds, hydrophobic interactions) that determine the structure of the protein itself. For example, the enzyme depicted in Figure 3.11*a* contains a number of positively charged residues strategically located to bind negatively charged atoms of the substrate. In addition to binding the substrate, the active site contains a particular array of amino acid side chains that influence the substrate and lower the activation energy of the reaction. The importance of individual side chains of the active site can be evaluated by site-directed mutagenesis (page 72), a technique in which one particular amino acid is replaced by another amino acid having different properties.

The active site is typically buried in a cleft or crevice that leads from the aqueous surroundings into the depths of the protein. When a substrate enters the active site cleft, it typically gives up its bound water molecules (desolvation) and enters a hydrophobic environment within the enzyme. The reactivity of the active-site side chains may be much greater in this protected environment as compared to the aqueous solvent of the cell. The amino acids that make up the active site are usually situated at distant points along the length of the extended polypeptide chain, but they are brought into close proximity to one another as the polypeptide folds into its final tertiary structure (Figure 3.11*b,c*). The structure of the active site accounts not only for the catalytic activity of the enzyme, but also for its *specificity*. As noted above, most enzymes are capable of binding only one or a small number of closely related biological molecules.

(a)

(b)

FIGURE 3.11 The active site of an enzyme. (*a*) Diagrammatic representation of the active site of the enzyme ribulose bisphosphate carboxylase showing the various sites of interaction between the bound substrates (RuBP and CO_2) and certain amino acid side chains of the enzyme. In addition to determining the substrate-binding properties of the active site, these noncovalent interactions alter the properties of the substrate in ways that accelerate its conversion to products. (*b*) An electron density map of the active site of a viral thymidine kinase with the substrate, deoxythymidine, shown in the center of the map (arrow). The blue mesh provides an indication of the outer reaches of the electron orbitals of the atoms that make up the substrate and the side chains of the enzyme, thus portraying a visual representation of the space occupied by the atoms of the active site. (*c,d*) Examples of the precise fit that occurs between parts of an enzyme and a substrate during catalysis. These two examples show the tight spatial relationship between (*c*) a glutamic acid (yellow) and (*d*) a histidine (green) of the enzyme triosephosphate isomerase and the substrate (red). (A: AFTER D. A. HARRIS, BIOENERGETICS AT A GLANCE, P. 88, BLACKWELL, 1995; B: REPRINTED WITH PERMISSION FROM D. G. BROWN, M. R. SANDERSON ET AL., NATURE STR. BIOL. 2:878, 1995; C–D: REPRINTED WITH PERMISSION FROM J. R. KNOWLES, NATURE 350:122, 1991; B–D: COPYRIGHT 1995, 1991, MACMILLAN MAGAZINES LIMITED.)
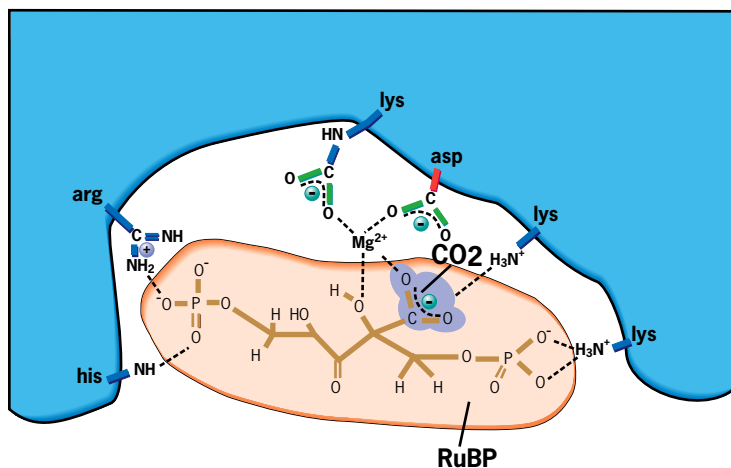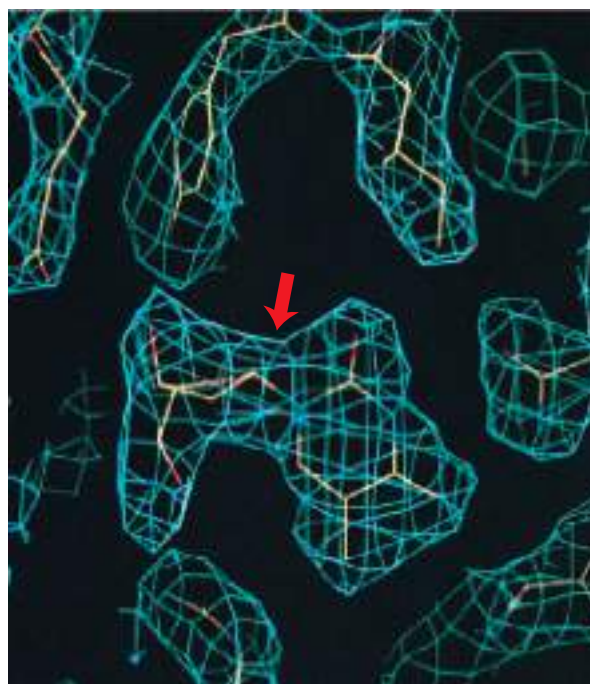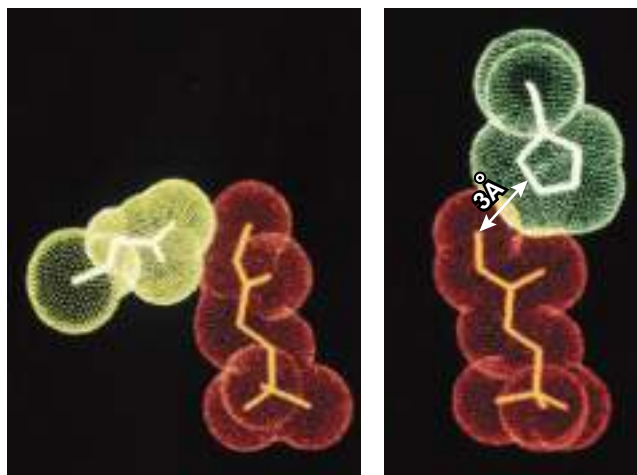
(c)   (d)

Mechanisms of Enzyme Catalysis

How is it that an enzyme can cause a reaction to occur hundreds of times a second when that same reaction might occur at undetectable rates in the enzyme's absence? The answer lies in the formation of the enzyme–substrate complex, which allows the substrate(s) to be taken out of solution and held onto the surface of the large catalyst molecule. Once there, the physical and chemical properties of the substrate can be affected in a number of ways, several of which are described in the following sections.

Substrate Orientation   Suppose you were to place a handful of nuts and bolts into a bag and shake the bag for 5 minutes. It is very unlikely that any of the bolts would have a nut firmly attached to its end when you were finished. In contrast, if you were to pick up a bolt in one hand and a nut in the other, you could rapidly guide the bolt into the nut. By holding the nut and bolt in the proper orientation, you have greatly decreased the entropy of the system. Enzymes lower the entropy of their substrates in a similar manner.

Substrates bound to the surface of an enzyme are brought very close together in precisely the correct orientation to undergo reaction (Figure 3.12*a*). In contrast, when reactants are present in solution, they are free to undergo translational and rotational movements, and even those possessing sufficient energy do not necessarily undergo a collision that results in the formation of a transition-state complex.

Changing Substrate Reactivity   Enzymes are composed of amino acids having a variety of different types of side chains, from fully charged to highly nonpolar. When a substrate is bound to the surface of an enzyme, the distribution of electrons within that substrate molecule is influenced by the neighboring side chains of the enzyme (Figure 3.12*b*). This influence increases the reactivity of the substrate and stabilizes the transition-state complex formed during the reaction. These effects are accomplished without the input of external energy, such as heat.

There are several general mechanisms whereby the reactivity of substrates is increased by association with enzymes. Basically, these mechanisms are similar to those characterized by organic chemists studying the mechanisms of organic reactions in a test tube. For example, the rate of reactions can be greatly affected by changes in pH. Although enzymes cannot change the pH of their medium, they do contain numerous amino acids with acidic or basic side chains. These groups are capable of donating or accepting protons to and from the substrate, thereby altering the electrostatic character of the substrate, making it more reactive.

The active sites of many enzymes contain side chains with a partial positive or negative charge. Such groups are capable of interacting with a substrate to alter its electrostatic properties (Figure 3.12*b*) and hence its reactivity. Such groups are also capable of reacting with a substrate to produce a temporary, covalent enzyme–substrate linkage. Chymotrypsin, an enzyme that digests food proteins within the small intestine,
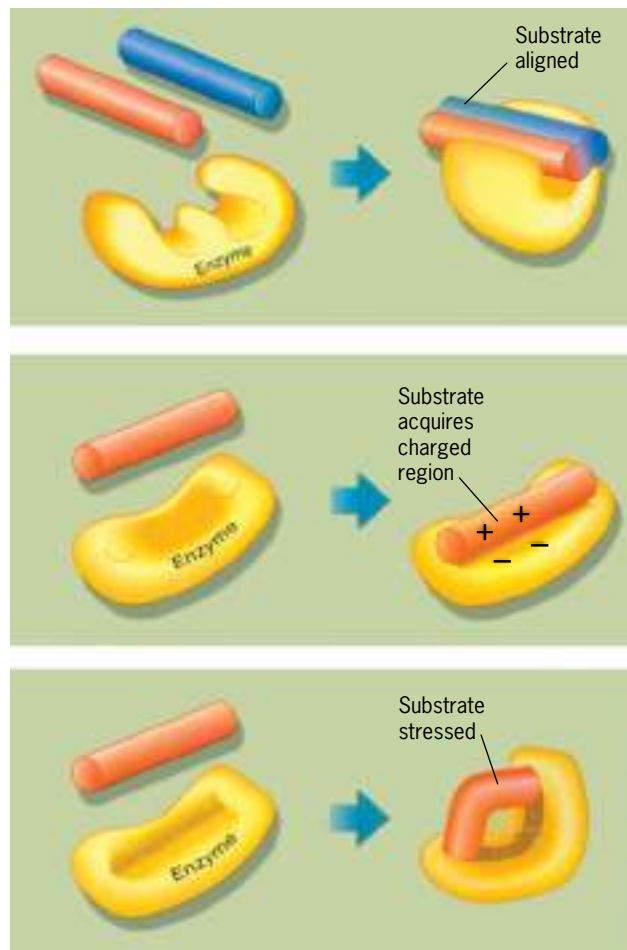
FIGURE 3.12 **Three mechanisms by which enzymes accelerate reactions:** (*a*) maintaining precise substrate orientation, (*b*) changing substrate reactivity by altering its electrostatic structure, (*c*) exerting physical stress on bonds in the substrate to be broken.

acts in this way. The series of reactions that take place as chymotrypsin hydrolyzes a peptide bond in a substrate protein is shown in Figure 3.13. The reaction in Figure 3.13 is divided into two steps. In the first step, the electronegative oxygen atom of the side chain of a serine of the enzyme "attacks" a carbon atom of the substrate. As a result, the peptide bond of the substrate is hydrolyzed, and a covalent bond is formed between the serine and the substrate, displacing the remainder of the substrate as one of the products. As discussed in the figure legend, the ability of serine to carry out this reaction depends on a nearby histidine residue, which attracts the proton of serine's hydroxyl group, increasing the nucleophilic power of the group's oxygen atom. Enzymes are also adept at utilizing water molecules in the reactions they catalyze. In the second step shown in Figure 3.13*b*, the covalent bond between enzyme and substrate is cleaved by a water molecule, returning the enzyme to its original unbonded state and releasing the remainder of the substrate as the second product.

Although amino acid side chains can engage in a variety of reactions, they are not well suited to donate or accept electrons.

(a)   *(b)*
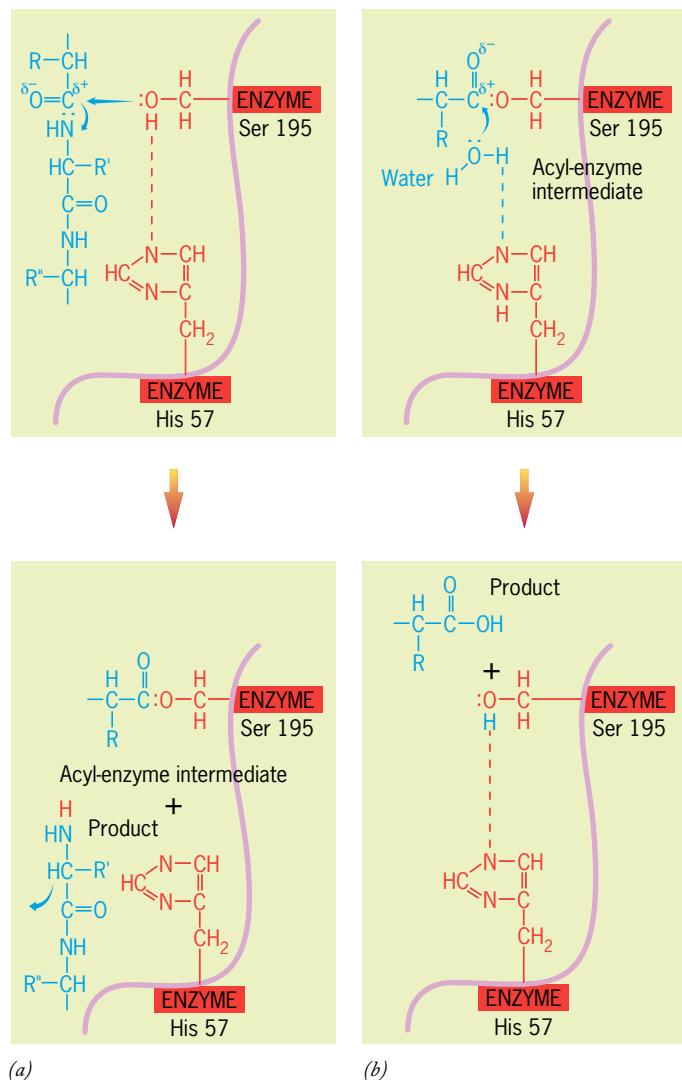
FIGURE 3.13 Diagrammatic representation of the catalytic mechanism of chymotrypsin. The reaction is divided into two steps. (*a*) The electronegative oxygen atom of a serine residue (Ser 195) in the enzyme, which carries a partial negative charge, carries out a nucleophilic attack on the carbonyl carbon atom of the substrate, which carries a partial positive charge, splitting the peptide bond. The polypeptide substrate is shown in blue. The serine is made more reactive by a closely applied histidine residue (His 57) that draws the proton from the serine and subsequently donates the proton to the nitrogen atom of the cleaved peptide bond. Histidine is able to do this because its side chain is a weak base that is capable of gaining and losing a proton at physiologic pH. (A stronger base, such as lysine, would remain fully protonated at this pH.) Part of the substrate forms a transient covalent bond with the enzyme by means of the serine side chain, while the remainder of the substrate is released. (It can be noted that the serine and histidine residues are situated 138 amino acids away from each other in the primary sequence but are brought together within the enzyme by the folding of the polypeptide. An aspartic acid, residue 102, which is not shown, also plays a role in catalysis by influencing the ionic state of the histidine.) (*b*) In the second step, the electronegative oxygen atom of a water molecule displaces the covalently linked substrate from the enzyme, regenerating the unbound enzyme molecule. As in the first step, the histidine plays a role in proton transfer; in this step, the proton is removed from water, making it a much stronger nucleophile. The proton is subsequently donated to the serine residue of the enzyme.

As we will see in the following section (and more fully in Chapters 5 and 6), the transfer of electrons is the central event in the oxidation–reduction reactions that play such a vital role in cellular metabolism. To catalyze these reactions, enzymes contain cofactors (metal ions or coenzymes) that increase the reactivity of substrates by removing or donating electrons.

Inducing Strain in the Substrate   Although the active site of an enzyme may be complementary to its substrate(s), various studies reveal a shift in the relative positions of certain of the atoms of the enzyme once the substrate has bound. In many cases, the conformation shifts so that the complementary fit between the enzyme and reactants is improved (an **induced fit**), and the proper reactive groups of the enzyme move into place. An example of induced fit is shown in Figure 3.14. These types of movements within an enzyme molecule provide a good example of a protein acting as a molecular machine. As these conformational changes occur, mechanical work is performed, allowing the enzyme to exert a physical force on certain bonds within a substrate molecule. This has the effect of destabilizing the substrate, causing it to adopt the transition state in which the strain is relieved (Figure 3.12*c*).

Studying Conformational Changes and Catalytic Intermediates To fully understand the mechanism by which an enzyme catalyzes a particular reaction, it is necessary to describe the various changes in atomic and electronic structure in both the enzyme and the substrate(s) as the reaction proceeds. We saw in the last chapter how X-ray crystallographic techniques can reveal details of the structure of a large enzyme molecule. Since 40 to 60 percent of the volume of a typical protein crystal consists of trapped solvent, most crystallized enzymes retain a high level of enzymatic activity. It should be possible, therefore, to use X-ray diffraction techniques to study reaction mechanisms. There is one major limitation, namely, time. In a standard crystallographic study, enzyme crystals must be subjected to an X-ray beam for a period of hours or days while the necessary data are collected. The portrait that emerges from such studies captures the structure of the molecule averaged out over time. Recent innovations, however, have made it possible to use X-ray crystallographic techniques to observe the fleeting structural changes that take place in the active site while an enzyme is catalyzing a single reaction cycle. This approach, which is called *time-resolved crystallography*, can include:

- Use of ultra-high-intensity X-ray beams generated by a synchrotron, an instrument used by nuclear physicists to study subatomic particles. This can cut the X-ray exposure period to a matter of picoseconds, which is the same time scale required for an enzyme to catalyze a single chemical transformation.

- Cooling the enzyme crystals to temperatures within 20 to 40 degrees of absolute zero, which slows the reaction by a factor as high as 10 billion, greatly increasing the lifetime of transient intermediates.

- Use of techniques to simultaneously trigger a reaction throughout an entire crystal so that all of the enzyme molecules in the crystal are in the same stage of the

reaction at the same time. For example, in the case of a reaction in which ATP is a substrate, the enzyme crystals can be infiltrated with ATP molecules that have been made nonreactive by linking them to an inert group (e.g., a nitrophenyl group) by a photosensitive bond. When the crystals are exposed to a brief pulse of light, all of the "caged" ATP molecules are released, initiating the reaction simultaneously in active sites throughout the crystal.

■ Use of enzymes bearing site-directed mutations (page 72) that impose kinetic barriers at specific stages of the reaction, leading to an increased lifetime for particular intermediates.

■ Determination of the structure at ultra-high (atomic) resolution (e.g., 0.8 Å), which allows visualization of individual hydrogens that may be present in hydrogen bonds or associated with acidic groups in the protein; the presence or absence of bound water molecules; the precise conformation of catalytic side chains; and the subtle strain that appears in

parts of the substrate during catalysis. The remarkable detail that can be extracted from these high-resolution images is illustrated by the hydrogen bond in Figure 3.15.

By combining these techniques, researchers have been able to determine the three-dimensional structure of an enzyme at successive stages during a single catalyzed reaction.

FIGURE 3.15 **Electron density map of a single hydrogen bond** (green-dotted line). This map shows a very small part of the proteolytic enzyme subtilisin at atomic (0.78 Å) resolution. The hydrogen atom (yellow) is seen to be shared between a nitrogen atom on the ring of a histidine residue and an oxygen atom of an aspartic acid residue. (REPRINTED WITH PERMISSION FROM PETER KUHN ET AL., BIOCHEMISTRY 37:13450, 1998, COURTESY OF RICHARD BOTT; COPYRIGHT 1998 AMERICAN CHEMICAL SOCIETY.)

(a)

(b)

FIGURE 3.14 **An example of induced fit.** When a glucose molecule binds to the enzyme hexokinase, the protein undergoes a conformational change that encloses the substrate within the active site pocket and aligns the reactive groups of the enzyme and substrate. (COURTESY OF THOMAS A. STEITZ.)

When these individual "snapshots" are put together in sequence, they produce a "movie" showing the various catalytic intermediates that appear and disappear as the reaction proceeds. An example of the data that can be collected using several of these approaches is shown in Figure 3.16.

While the spotlight has been focused on the conformational changes that occur as an enzyme transforms its substrates into products, considerable attention has also been paid in recent years to the conformational changes that occur within the protein itself. As shown in Figure 2.37, proteins are dynamic molecules that are capable of built-in (intrinsic) motion that can have important functional relevance. In the case depicted in Figure 2.37, this motion allows entry of substrate to the enzyme's active site. Analysis of this type of intrinsic motion by various experimental techniques has suggested that proteins—even in the absence of substrate—are capable of many of the same movements that can be detected during the protein's catalytic cycle. If true, this suggests that evolution has selected for protein structures that are capable of intrinsic motions that can be useful for potential functions that a protein might serve. Rather than inducing a specific conformational change, the substrate(s) might simply be "waiting" for a protein to assume a conformation to which it can bind effectively.

Enzyme Kinetics

Enzymes vary greatly in their ability to catalyze reactions. The catalytic activity of an enzyme is revealed by studying its **kinetics**, that is, the rate at which it catalyzes a reaction under various experimental conditions. In 1913, Leonor Michaelis and Maud Menten reported on the mathematical relationship between substrate concentration and the velocity of enzyme reactions as measured by the amount of product formed (or substrate consumed) in a given amount of time. This relationship can be expressed by an equation (given on p. 101) that generates a hyperbola, as shown in Figure 3.17. Rather than considering the theoretical aspects of enzyme kinetics, we can obtain the same curve in a practical manner, as it is done for each enzyme studied. To determine the velocity of a reaction, an incubation mixture is set up at the desired temperature, containing all the ingredients required except one, which when added initiates the reaction. If, at the time the reaction begins, no product is present in the mix-
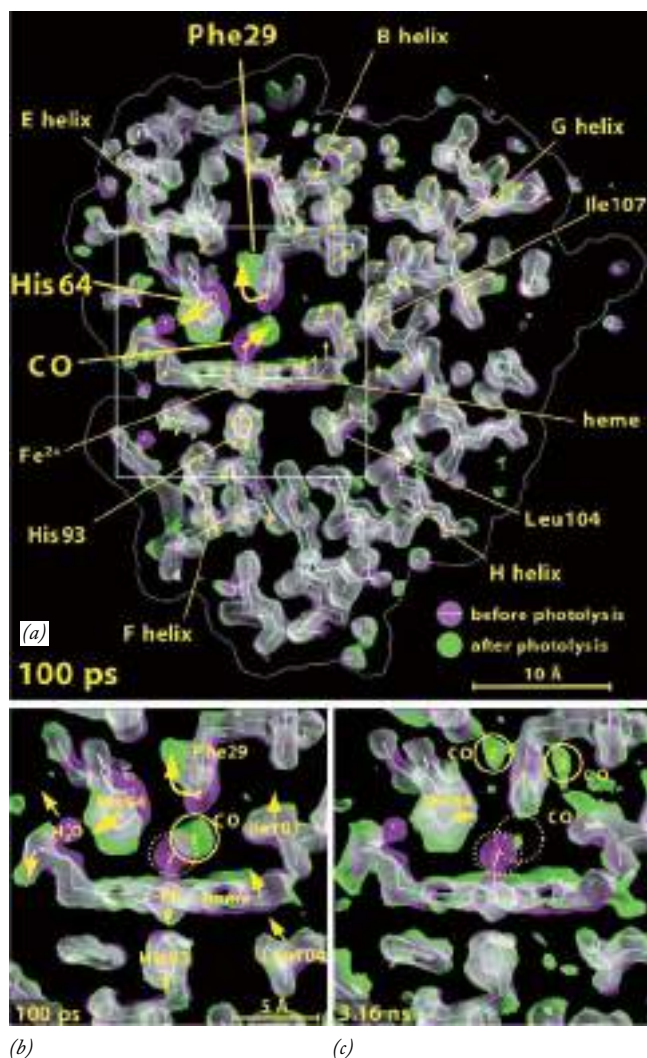
FIGURE 3.16 Myoglobin: the movie. In this example of time-resolved X-ray crystallography, the structure of myoglobin (Mb) was determined with a molecule of CO bound to the heme group and at various times after the CO molecule was released. (CO binds to the same site on Mb as O_2 but is better suited for these types of studies.) Release of CO was induced simultaneously throughout the crystal by exposure to a flash of laser light (photolysis). Each of the structures was determined following a single, intense X-ray pulse from a synchrotron. The myoglobin molecule being studied had a single amino acid substitution that made it a better subject for analysis. (*a*) A 6.5 Å-thick slice through a Mb molecule showing the changes that occur by 100 picoseconds (1 ps = one-trillionth of a second) following the release of CO from its binding site. The structure of Mb prior to the laser flash is shown in magenta, and the structure of the protein 100 ps after the laser flash is shown in green. Those parts of the molecule that did not change in structure over this time period appear white. Three large-scale displacements near the CO-binding site can be seen (indicated by the yellow arrows). (*b*) An enlarged view of the boxed region in part *a*. The released CO (solid circle) is situated about 2 Å from its original binding site (dashed circle). Movement of CO is accommodated by the rotation of Phe29, which pushes His64 outward, which in turn dislodges a bound water molecule. (*c*) By 3.16 nanoseconds after the laser flash, the CO molecule has migrated to either of the two positions shown (labeled 2 and 3), Phe29 and His64 have relaxed toward their initial states, and the heme group has undergone considerable displacement as indicated by the increased amount of green shading in the region of the heme. (REPRINTED WITH PERMISSION FROM FRIEDRICH SCHOTTE ET AL., SCIENCE 300:1946, 2003; COPYRIGHT © 2003 AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)
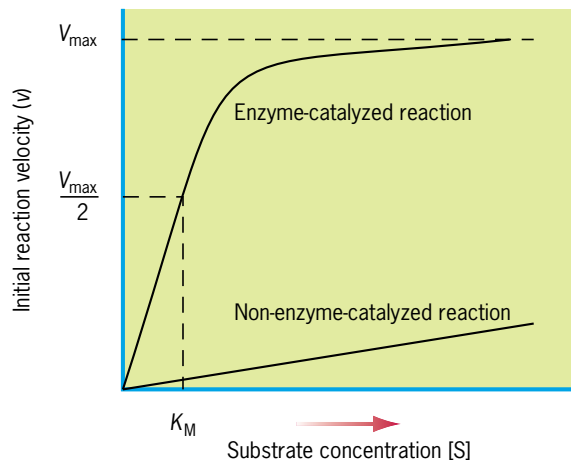
FIGURE 3.17 The relationship between the rate (velocity) of an enzyme-catalyzed reaction and the substrate concentration. Since each enzyme molecule is only able to catalyze a certain number of reactions in a given amount of time, the velocity of the reaction (typically expressed as moles of product formed per second) approaches a maximal rate as the substrate concentration increases. The substrate concentration at which the reaction is at half-maximal velocity ($V_{max}/2$) is called the Michaelis constant, or K_M.

ture, then the amount of product that appears over time provides a measure of the velocity of the reaction. There are complicating factors in this procedure. If the incubation time is too great, the substrate concentration becomes measurably reduced. In addition, as product appears, it can be converted back to substrate by the reverse reaction, which is also catalyzed by the enzyme. Ideally, what we want to determine is the *initial* velocity, that is, the velocity at the instant when no product has yet been formed. To accurately measure the initial reaction velocity, brief incubation times and sensitive measuring techniques are employed.

To generate a curve such as that shown in Figure 3.17, the initial velocity is determined for a series of incubation mixtures that contain the same amount of enzyme but an increasing concentration of substrate. It is evident from this curve that the initial reaction velocity varies markedly with substrate concentration. At low substrate concentrations, enzyme molecules are subjected to relatively few collisions with substrate in a given amount of time. Consequently, the enzyme has "idle time"; that is, substrate molecules are rate limiting. At high substrate concentrations, enzymes are colliding with substrate molecules more rapidly than they can be converted to product. Thus, at high substrate concentrations, individual enzyme molecules are working at their maximal capacity; that is, enzyme molecules are rate limiting. Thus, as a greater and greater concentration of substrate is present in the reaction mixture, the enzyme approaches a state of *saturation*. The initial velocity at this theoretical saturation point is termed the **maximal velocity (V_{max})**.

The simplest measure of the catalytic activity of an enzyme is provided by its **turnover number**, which can be calculated from the V_{max}. The turnover number (or *catalytic constant, k_{cat}*, as it is also called) is the maximum number of molecules of substrate that can be converted to product by one enzyme

molecule per unit time. A turnover number (per second) of 1 to 10^3 is typical for enzymes, although values as great as 10^6 are known (see Table 3.3). It is apparent from these values that a few molecules of enzyme can rapidly convert a large number of substrate molecules into product.

The value of V_{max} is only one useful term obtained from a plot such as that in Figure 3.17; another is the **Michaelis constant (K_M)**, which is equal to the substrate concentration when the reaction velocity is one-half of V_{max}. As its name implies, the K_M is constant for a given enzyme and, thus, independent of substrate or enzyme concentration. The relationship between V_{max} and K_M is best appreciated by considering the Michaelis-Menten equation, which can be used to generate the plot shown in Figure 3.17.

$$V = V_{max} \frac{[S]}{[S] + K_M}$$

According to the equation, when the substrate concentration [S] is set at a value equivalent to K_M, then the velocity of the reaction (V) becomes equal to $V_{max}/2$, or one-half the maximal velocity. Thus, $K_M = [S]$, when $V = V_{max}/2$.

To generate a hyperbolic curve such as that in Figure 3.17 and make an accurate determination of the values for V_{max} and K_M, a considerable number of points must be plotted. An easier and more accurate description is gained by plotting the reciprocals of the velocity and substrate concentration against one another, as formulated by Hans Lineweaver and Dean Burk. When this is done, the hyperbola becomes a straight line (Figure 3.18) whose *x* intercept is equal to $-1/K_M$, *y* intercept is equal to $1/V_{max}$, and slope is equal to K_M/V_{max}. The values of K_M and V_{max} are, therefore, readily determined by extrapolation of the line drawn from a relatively few points.

In most cases, the value for K_M provides a measure of the affinity of the enzyme for the substrate. The higher the K_M, the greater the substrate concentration that is required to reach one-half V_{max} and, thus, the lower the affinity of the
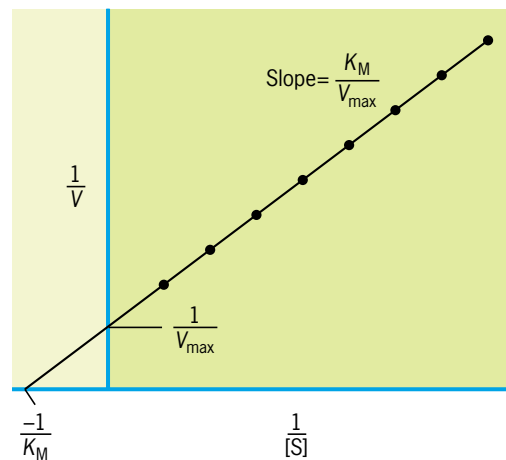
FIGURE 3.18 A Lineweaver-Burk plot of the reciprocals of velocity and substrate concentration from which the values for the V_{max} and K_M are readily calculated.
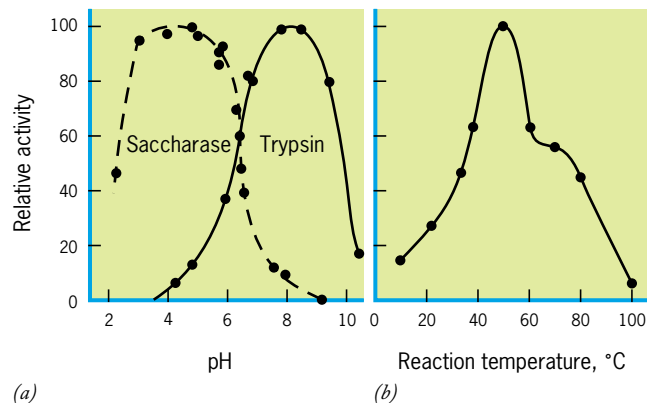
FIGURE 3.19 Dependence of the rate of an enzyme-catalyzed reaction on (a) pH, and (b) temperature. The shape of the curves and the optimal pH and temperature vary with the particular reaction. (a) Changes in pH affect the ionic properties of the substrate and the enzyme as well as the enzyme's conformation. (b) At lower temperatures, reaction rates rise with increasing temperature due to the increased energy of the reactants. At higher temperatures, this positive effect is offset by enzyme denaturation. (A: FROM E. A. MOELWYN-HUGHES, IN THE ENZYMES, J. B. SUMNER AND K. MYRBACK, EDS., VOL. 1, ACADEMIC PRESS, 1950; B: FROM K. HAYASHI ET AL., J. BIOCHEM. 64:93, 1968.)

enzyme for that substrate. The K_M for most enzymes ranges between 10^{-1} and 10^{-7}, with a typical value about 10^{-4} M. Other factors that strongly influence enzyme kinetics are the pH and temperature of the incubation medium. Each enzyme has an optimal pH and temperature at which it operates at maximal activity (Figure 3.19). The influence of temperature on enzyme activity is illustrated by the striking effect of refrigeration in slowing the growth of microorganisms.

Enzyme Inhibitors   **Enzyme inhibitors** are molecules that are able to bind to an enzyme and decrease its activity. The cell depends on inhibitors to regulate the activity of many of its enzymes; biochemists use inhibitors to study the properties of enzymes; and many biochemical companies produce enzyme inhibitors that act as drugs, antibiotics, or pesticides. Enzyme inhibitors can be divided into two types: reversible or irreversible. Reversible inhibitors, in turn, can be considered as competitive or noncompetitive.

   Irreversible inhibitors are those that bind very tightly to an enzyme, often by forming a covalent bond to one of its amino acid residues. A number of nerve gases, such as diisopropylphosphofluoridate and the organophosphate pesticides, act as irreversible inhibitors of acetylcholinesterase, an enzyme that plays a crucial role in destroying acetylcholine, the neurotransmitter responsible for causing muscle contraction. With the enzyme inhibited, the muscle is stimulated continuously and remains in a state of permanent contraction. As discussed in the accompanying Human Perspective, the antibiotic penicillin acts as an irreversible inhibitor of a key enzyme in the formation of the bacterial cell wall.

   Reversible inhibitors, on the other hand, bind only loosely to an enzyme, and thus are readily displaced. **Competitive inhibitors** are reversible inhibitors that compete with a substrate for access to the active site of an enzyme. Since substrates have a complementary structure to the active site to which they bind, competitive inhibitors must resemble the substrate to compete for the same binding site, but differ in a way that prevents them from being transformed into product (Fig-

ure 3.20). Analysis of the types of molecules that can compete with the substrate for a binding site on the enzyme provides insight into the structure of the active site and the nature of the interaction between a substrate and its enzyme.

   Competitive enzyme inhibition forms the basis of action of many common drugs, as illustrated in the following exam-
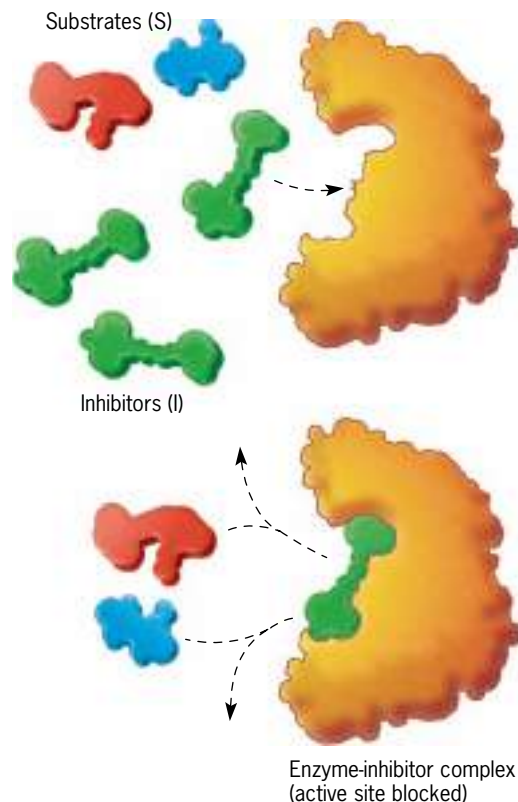
FIGURE 3.20 Competitive inhibition. Because of their molecular similarity, competitive inhibitors are able to compete with the substrate for a binding site on the enzyme. The effect of a competitive inhibitor depends on the relative concentrations of the inhibitor and substrate.

ple. Angiotensin converting enzyme (ACE) is a proteolytic enzyme that acts on a 10-residue peptide (angiotensin I) to produce an 8-residue peptide (angiotensin II). Elevated levels of angiotensin II are a major risk factor in the development of high blood pressure (*hypertension*). In the 1960s John Vane and his colleagues at the Eli Lilly Company began a search for compounds that could inhibit ACE. Previous studies had found that the venom of a Brazilian pit viper contained inhibitors of proteolytic enzymes, and it was found that one of the components of this venom, a peptide called teprotide
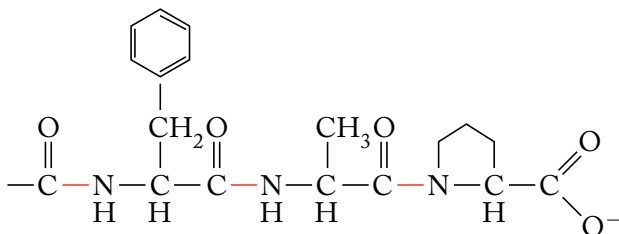
Teprotide (peptide bonds marked in red)

was a potent competitive inhibitor of ACE. Although teprotide was shown to lower blood pressure in hypertensive patients, it was not a very useful drug because it had a peptide structure and thus was rapidly degraded if taken orally. Subsequent efforts to develop nonpeptide inhibitors of the enzyme led researchers to synthesize a compound called captopril,
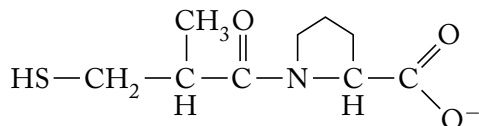
Captopril

which became the first useful antihypertensive drug that acted by binding to ACE.

The effectiveness of a competitive inhibitor depends on its relative affinity for the enzyme. Regardless, competitive in-hibition can be overcome if the substrate/inhibitor ratio is great enough. In other words, if the number of collisions between the enzyme and inhibitor becomes insignificant relative to those between the enzyme and its substrate, then the effect of the inhibitor becomes minimal. Given a sufficient substrate concentration, it remains theoretically possible to achieve the enzyme's maximal velocity even in the presence of the competitive inhibitor.

In **noncompetitive inhibition**, the substrate and inhibitor do not compete for the same binding site; generally, the inhibitor acts at a site other than the enzyme's active site. The level of inhibition depends only on the concentration of the inhibitor, and increasing the concentration of the substrate cannot overcome it. Since, in the presence of a noncompetitive inhibitor, a certain fraction of the enzyme molecules are necessarily inactive at a given instant, the maximal velocity of the population of enzyme molecules cannot be reached. The effects on the kinetics of enzymes in the presence of noncompetitive and competitive inhibitors are shown in Figure 3.21. In one case, the V_{max} is lowered, and in the other, the K_M is increased. In both types, the slope (K_M/V_{max}) is increased relative to the uninhibited reaction. As we will see on page 113, cells utilize a version of noncompetitive inhibition to regulate the activity of key enzymes of metabolic pathways.

REVIEW ?

1. How is it possible to have a reaction characterized by a large ΔG and a small E_A? a large E_A and a small ΔG?

2. Explain how enzymes can be so specific toward the substrates they bind.

3. Examine Figure 3.13 illustrating the steps in an enzymatically catalyzed reaction and describe what is occurring without reading the legend.

4. Distinguish between turnover number and V_{max}.

(a)

(b)
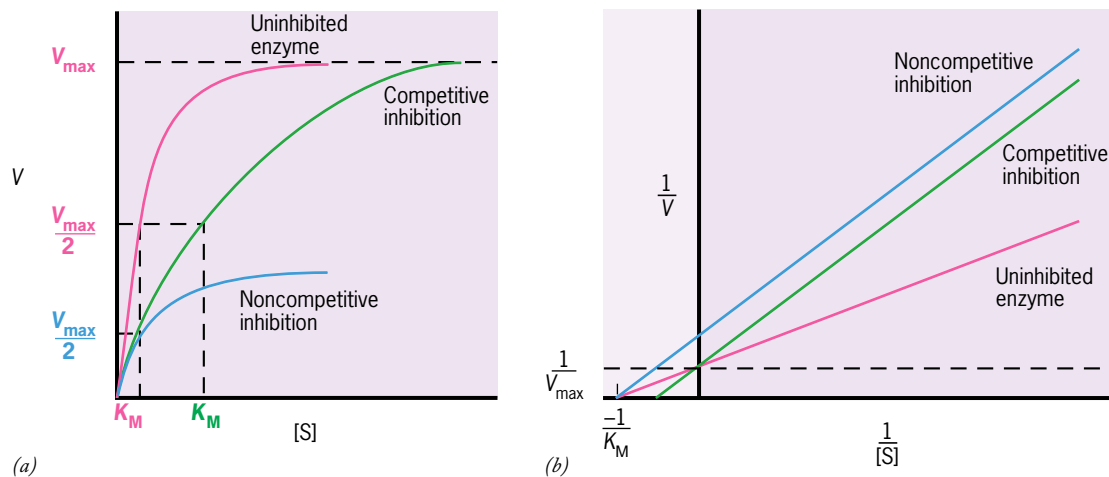
FIGURE 3.21 The effects of inhibitors on enzyme kinetics. The effect of both competitive and noncompetitive inhibitors is shown when the kinetics of the reaction are plotted as velocity of reaction versus substrate concentration (*a*) or its reciprocal (*b*). The noncompetitive inhibitor reduces V_{max} without affecting K_M, whereas the competitive inhibitor increases K_M without affecting V_{max}.

THE HUMAN PERSPECTIVE

The Growing Problem of Antibiotic Resistance

Not too long ago, it was widely believed that human health would no longer be threatened by serious bacterial infections. Bacterial diseases such as leprosy, pneumonia, gonorrhea, and dozens of others were being cured by administration of any one of a number of antibiotics—compounds that selectively kill bacteria without harming the human host in which they grow. This state of affairs has changed dramatically in the past couple of decades as pathogenic bacteria have become increasingly resistant to these "wonder drugs," leading to the deaths of many people who would once have been successfully treated. Even tuberculosis, which had virtually disappeared from developed countries, has reemerged around the world in a form referred to as XDR (extremely drug resistant) that is virtually untreatable. It is feared that XDR-TB is on the verge of becoming a major global health crisis. To make matters worse, the pharmaceutical industry has drastically cut resources devoted to the development of new antibiotics. This change of action by the drug industry is generally attributed to (1) a lack of financial incentives: antibiotics are taken only for a short period of time (as opposed to drugs prescribed for chronic conditions such as diabetes or depression); (2) new antibiotics running the risk of having a relatively short lifetime in the marketplace as bacteria become resistant to each successive product; and (3) the most effective antibiotics being held back from widespread use and being kept instead as weapons of last resort when other drugs have failed. Ideas for the formation of nonprofit institutions for the development of new antibiotics have been widely discussed.

Here, we will look briefly at the mechanism of action of antibiotics—particularly those that target enzymes, which are the subjects of this chapter—and the development of bacterial resistance. Most antibiotics are derived from natural products, produced by microorganisms to kill other microorganisms. Several types of targets in bacterial cells have proven most vulnerable. These include the following:

1. *Enzymes involved in the synthesis of the bacterial cell wall.* Penicillin and its derivatives (e.g., methicillin) are structural analogues of the substrates of a family of transpeptidases that catalyze the final cross-linking reactions that give the cell wall its rigidity. If these reactions do not occur, a rigid cell wall fails to develop. Penicillin is an irreversible inhibitor of the transpeptidases; the antibiotic fits into the active site of these enzymes, forming a covalently bound complex that cannot be dislodged. The antibiotic vancomycin (which was originally derived from a microorganism living in soil samples taken from Borneo) inhibits transpeptidation by binding to the peptide substrate of the transpeptidase, rather than to the enzyme itself. Normally, the transpeptidase substrate terminates in a D-alanine–D-alanine dipeptide. To become resistant to vancomycin, a bacterial cell must synthesize an alternate terminus that doesn't bind the drug, which is a roundabout process that requires the acquisition of several new enzymatic activities. As a result, vancomycin is the antibiotic to which bacteria have proven least able to develop resistance and thus is usually given as a last resort when other antibiotics have failed. Unfortunately, vancomycin-resistant strains of several pathogenic bacteria, including *Staphylococcus aureus*, have emerged in recent years. *S. aureus* is a common inhabitant of the skin and nasal passages. Although it is usually relatively harmless, this organism is the most frequent cause of life-threatening infections that develop in hospitalized patients who have open wounds or invasive

tubes. For years, methicillin-resistant *S. aureus* (MRSA) strains have created havoc in hospital wards, killing tens of thousands of patients. MRSA infections have also begun to appear in community settings, such as high school gyms or children's daycare centers. In many cases, vancomycin is the only drug that can stop these infections. Consequently, it was alarming to discover that MRSA can become resistant to vancomycin by acquiring a cluster of vancomycin-resistance genes from another bacterium (*E. faecium*), which is also a common cause of hospital-based infections. So far, no known vancomycin-resistant MRSA have been able to gain a foothold in either a hospital or community setting, but it may only be a matter of time. For this reason, infectious disease specialists have urged hospitals to institute better hygiene programs and to isolate patients at the first sign of infection. These practices have been proven to reduce the incidence of lethal infections where they have been put into place.

2. *Components of the system by which bacteria duplicate, transcribe, and translate their genetic information.* Although bacterial and eukaryotic cells have a similar system for storing and utilizing genetic information, there are many differences between the two types of cells that pharmacologists can take advantage of. Streptomycin and the tetracyclines, for example, bind to bacterial ribosomes, but not eukaryotic ribosomes. Quinolones, such as ciprofloxacin (brand name Cipro), are a rare example of wholly synthetic antibiotics (i.e., they are not based on natural products). Quinolones inhibit the enzyme DNA gyrase, which is required for bacterial DNA replication.

Nearly all new antibiotics are derivatives of existing compounds that have been modified chemically in the laboratory. New compounds are typically screened in one of two ways. The compound is either tested for its ability to kill bacterial cells that are growing in a culture dish or a laboratory animal, or it is tested for its ability to bind and inhibit a particular target protein that has been purified from bacterial cells. It was hoped that the sequencing of the genomes of pathogenic bacteria beginning in 1995 would lead to the identification of a host of new drug targets, but this has not been the case. Remarkably, there have only been two new classes of antibiotics developed and approved since 1963. One of these classes includes the antibiotic linezolid (brand name Zyvox), introduced in 2000, which acts specifically on bacterial ribosomes and interferes with protein synthesis. The other new class of antibiotics, introduced in 2003 and represented by daptomycin (brand name Cubicin), are cyclic lipopeptides, which disrupt bacterial membrane function. Many researchers are hoping that Zyvox and Cubicin will be used sparingly so that resistance can be kept to a minimum.

3. *Enzymes that catalyze metabolic reactions specifically in bacteria.* Sulfa drugs, for example, are effective antibiotics because they closely resemble the compound *p*-aminobenzoic acid (PABA),

$$H_2N\!-\!\!\bigcirc\!\!-\!COOH \qquad H_2N\!-\!\!\bigcirc\!\!-\!SO_2\!-\!NH\!-\!R$$
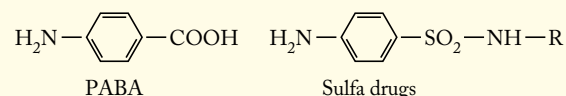
$$\text{PABA} \qquad\qquad\qquad \text{Sulfa drugs}$$

which bacteria convert enzymatically to the essential coenzyme folic acid. Since humans lack a folic acid-synthesizing enzyme, they must obtain this essential coenzyme in their diet and, consequently, sulfa drugs have no effect on human metabolism.

Bacteria become resistant to antibiotics through a number of distinct mechanisms, many of which can be illustrated using penicillin as an example. Penicillin is a β-lactam; that is, it contains a four-membered β-lactam ring (shown in color).
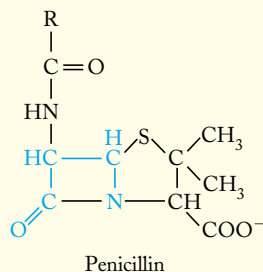
Penicillin

By the 1940s, researchers had discovered that certain bacteria possess an enzyme called β-lactamase (or penicillinase) that can open the lactam ring, rendering the compound harmless to the bacterium. At the time when penicillin was first introduced as an antibiotic during World War II, none of the major disease-causing bacteria possessed a gene for β-lactamase. This can be verified by examining the DNA of bacteria descended from laboratory cultures that were started in the preantibiotic era. Today, the β-lactamase gene is present in a wide variety of infectious bacteria, and the production of β-lactamase by these cells is the primary cause of penicillin resistance. How did these species acquire the gene?

The widespread occurrence of the β-lactamase gene illustrates how readily genes can spread from one bacterium to another, not only among the cells of a given species, but between species. There are several ways this can happen, including conjugation (shown in Figure 1.13), in which DNA is passed from one bacterial cell to another; transduction, in which a bacterial gene is carried from cell to cell by a virus; and transformation, in which a bacterial cell is able to pick up naked DNA from its surrounding medium. Pharmacologists have attempted to counter the spread of β-lactamase by synthesizing penicillin derivatives (e.g., methicillin) that are more resistant to the

hydrolytic enzyme. As might be expected, natural selection leads rapidly to the evolution of bacteria whose β-lactamase can split the new forms of the antibiotic.

Not all penicillin-resistant bacteria have acquired a β-lactamase gene. Some are resistant because they possess modifications in their cell walls that block entry of the antibiotic; others are resistant because they are able to selectively export the antibiotic once it has entered the cell; still others are resistant because they possess modified transpeptidases that fail to bind the antibiotic. Bacterial meningitis, for example, is caused by the bacterium *Neisseria meningitidis*, which has yet to display evidence it has acquired β-lactamase. Yet these bacteria are becoming resistant to penicillin because their transpeptidases have lost an affinity for the antibiotics as the result of mutations in the gene that encodes the enzyme.

The problem of drug resistance is not restricted to bacterial diseases, but has also become a major issue in the treatment of AIDS. Unlike bacteria, whose DNA-replicating enzymes operate at a very high level of accuracy, the replicating enzyme of the AIDS virus (HIV), called *reverse transcriptase*, makes a large number of mistakes, which leads to a high rate of mutation. This high error rate (about 1 mistake for every 10,000 bases duplicated), combined with the high rate of virus production ($>10^8$ virus particles produced in a person per day), makes it very likely that drug-resistant variants will emerge within an individual as the infection progresses. This problem is being combatted:

- By having patients take several different drugs targeted at different viral enzymes. This greatly reduces the likelihood that a variant will emerge that is resistant to all of the drugs.
- By designing drugs that interact with the most highly conserved portions of each targeted enzyme, that is, those portions where mutations are most likely to produce a defective enzyme. This point underscores the importance of knowing the structure and function of the target enzyme and the manner in which potential drugs interact with that target.

3.3 METABOLISM

Metabolism is the collection of biochemical reactions that occur within a cell, which includes a tremendous diversity of molecular conversions. Most of these reactions can be grouped into **metabolic pathways** containing a sequence of chemical reactions in which each reaction is catalyzed by a specific enzyme, and the product of one reaction is the substrate for the next. The enzymes constituting a metabolic pathway are usually confined to a specific region of the cell, such as the mitochondria or the cytosol. Increasing evidence suggests that the enzymes of a metabolic pathway are often physically linked to one another, a feature that allows the product of one enzyme to be delivered directly as a substrate to the active site of the next enzyme in the reaction sequence.

The compounds formed in each step along the pathway are **metabolic intermediates** (or *metabolites*) that lead ultimately to the formation of an end product. End products are molecules with a particular role in the cell, such as an amino acid that can be incorporated into a polypeptide, or a sugar

that can be consumed for its energy content. The metabolic pathways of a cell are interconnected at various points so that a compound generated by one pathway may be shuttled in a number of directions depending on the requirements of the cell at the time. We will concentrate in this section on aspects of metabolism that lead to the transfer and utilization of chemical energy, for this topic is one we will draw on throughout the book.

An Overview of Metabolism

Metabolic pathways can be divided into two broad types. **Catabolic pathways** lead to the disassembly of complex molecules to form simpler products. Catabolic pathways serve two functions: they make available the raw materials from which other molecules can be synthesized, and they provide chemical energy required for the many activities of a cell. As will be discussed at length, energy released by catabolic pathways is stored temporarily in two forms: as high-energy phosphates (primarily ATP) and as high-energy electrons (primarily in

NADPH). **Anabolic pathways** lead to the synthesis of more complex compounds from simpler starting materials. Anabolic pathways are energy-requiring and utilize chemical energy released by the exergonic catabolic pathways.

Figure 3.22 shows a greatly simplified profile of the ways in which the major anabolic and catabolic pathways are interconnected. Macromolecules are first disassembled (hydrolyzed) into the building blocks of which they are made (stage I, Figure 3.22). Once macromolecules have been hydrolyzed into their components—amino acids, sugars, and fatty acids—the cell can reutilize the building blocks directly to (1) form other macromolecules of the same class (stage I), (2) convert them into different compounds to make other products, or (3) degrade them further (stages II and III, Figure 3.22) and extract a measure of their free-energy content.

The pathways for degradation of the diverse building blocks of macromolecules vary according to the particular compound being catabolized. Ultimately, however, all of these molecules are converted into a small variety of compounds that can be metabolized similarly. Thus, even though substances begin as macromolecules having a very different structure, they are converted by catabolic pathways to the same low-molecular-weight metabolites. For this reason, catabolic pathways are said to be convergent.

Remarkably, the chemical reactions and metabolic pathways described in this chapter are found in virtually every living cell, from the simplest bacterium to the most complex plant or animal. It is evident that these pathways appeared very early in, and have been retained throughout, the course of biological evolution.

Oxidation and Reduction: A Matter of Electrons

Both catabolic and anabolic pathways include key reactions in which electrons are transferred from one reactant to another. Reactions that involve a change in the electronic state of the reactants are called **oxidation–reduction** (or **redox**) **reactions**. Changes of this type involve the gain or loss of electrons. Consider the conversion of metallic iron (Fe^0) to the ferrous state (Fe^{2+}), in which the iron atom loses a pair of electrons, thereby attaining a more positive state. When an atom loses one or more electrons, it is said to be *oxidized*. The reaction is reversible, meaning ferrous ions can be converted to metallic iron, a more negative state, by the acquisition of a pair of electrons. When an atom gains one or more electrons, it is said to be *reduced*.

For metallic iron to be oxidized, there must be some substance to accept the electrons that are released. Conversely, for ferrous ions to be reduced, there must be some substance to donate the necessary electrons. In other words, the oxidation of one reactant must be accompanied by the simultaneous reduction of some other reactant, and vice versa. One possible reaction involving iron might be

$$Fe^0 + Cu^{2+} \rightleftharpoons Fe^{2+} + Cu^0$$
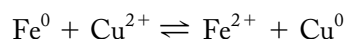
The substance that is oxidized during a redox reaction, that is, the one that loses electrons, is called a **reducing agent**, and the
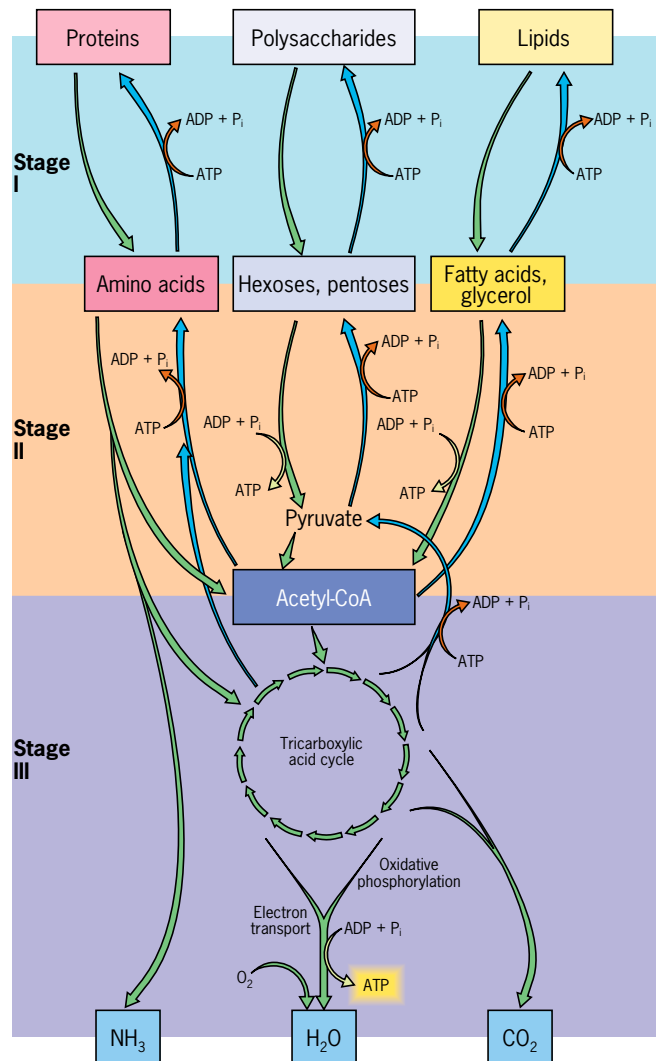
FIGURE 3.22   Three stages of metabolism. The catabolic pathways (green arrows downward) converge to form common metabolites and lead to ATP synthesis in stage III. The anabolic pathways (blue arrows upward) start from a few precursors in stage III and utilize ATP to synthesize a large variety of cellular materials. Metabolic pathways for nucleic acids are more complex and are not shown here. (FROM A. L. LEHNINGER, BIOCHEMISTRY, 2D ED., 1975. WORTH PUBLISHERS, NEW YORK.)

one that is reduced, that is, the one that gains electrons, is called an **oxidizing agent**.

The oxidation or reduction of metals, such as iron or copper, involves the complete loss or gain of electrons. The same cannot occur with most organic compounds for the following reason: The oxidation and reduction of organic substrates during cellular metabolism involve carbon atoms that are covalently bonded to other atoms. As discussed in Chapter 2, when a pair of electrons is shared by two different atoms, the electrons are attracted more strongly to one of the two atoms of the polarized bond. In a C—H bond, the carbon atom has the strongest pull on the electrons; thus it can be said that the carbon atom is in a reduced state. In contrast, if a carbon atom is bonded to a more electronegative atom, as in a C—O or

— Covalent bond in which carbon atom has greater share of electron pair
— Covalent bond in which oxygen atom has greater share of electron pair
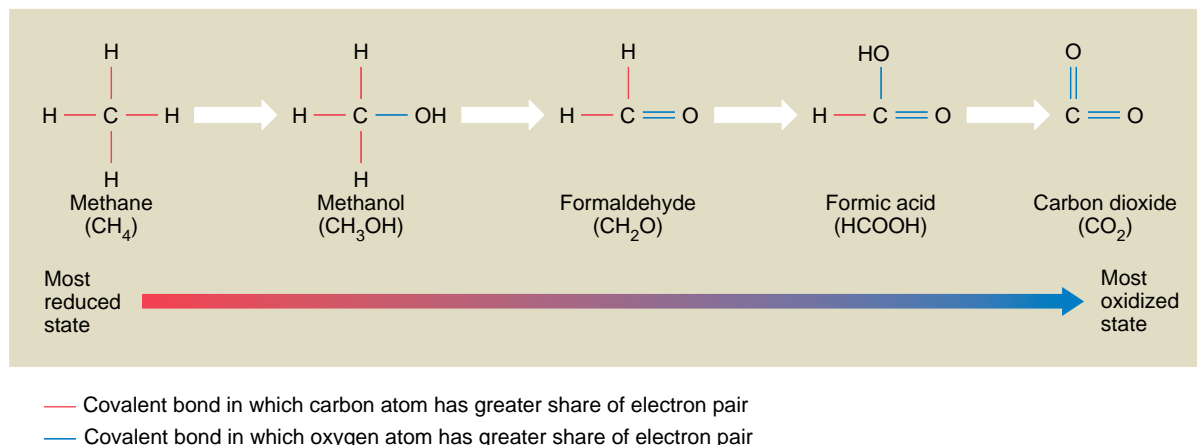
FIGURE 3.23 **The oxidation state of a carbon atom depends on the other atoms to which it is bonded.** Each carbon atom can form a maximum of four bonds with other atoms. This series of simple, one-carbon molecules illustrates the various oxidation states in which the carbon atom can exist. In its most reduced state, the carbon is bonded to four hydrogens (forming methane); in its most oxidized state, the carbon atom is bonded to two oxygens (forming carbon dioxide).

C—N bond, the electrons are pulled away from the carbon atom, which is thus in an oxidized state. Since carbon has four outer-shell electrons it can share with other atoms, it can exist in a variety of oxidation states. This is illustrated by the carbon atom in a series of one-carbon molecules (Figure 3.23) ranging from the fully reduced state in methane (CH_4) to the fully oxidized state in carbon dioxide (CO_2). The relative oxidation state of an organic molecule can be roughly determined by counting the number of hydrogen versus oxygen and nitrogen atoms per carbon atom. As we will see shortly, the oxidation state of the carbon atoms in an organic molecule provides a measure of the molecule's free-energy content.

The Capture and Utilization of Energy

The compounds that we use as chemical fuels to run our furnaces and automobiles are highly reduced organic compounds, such as natural gas (CH_4) and petroleum derivatives. Energy is released when these molecules are burned in the presence of oxygen, converting the carbons to more oxidized states, as in carbon dioxide and carbon monoxide gases. The degree of reduction of a compound is also a measure of its ability to perform chemical work within the cell. The more hydrogen atoms that can be stripped from a "fuel" molecule, the more ATP that ultimately can be produced. Carbohydrates are rich in chemical energy because they contain strings of (H—C—OH) units. Fats contain even greater energy per unit weight because they contain strings of more reduced (H—C—H) units. In the following discussion we will concentrate on carbohydrates.

As the sole building block of both starch and glycogen, glucose is a key molecule in the energy metabolism of both plants and animals. The free energy released by the complete oxidation of glucose is very large:

$$C_6H_{12}O_6 + 6\,O_2 \rightarrow 6\,CO_2 + 6\,H_2O$$
$$\Delta G^{\circ\prime} = -686 \text{ kcal/mol}$$
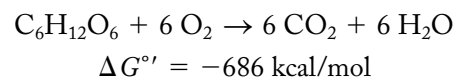
By comparison, the free energy required to convert ADP to ATP is relatively small:

$$ADP + P_i \rightarrow ATP + H_2O \qquad \Delta G^{\circ\prime} = +7.3 \text{ kcal/mol}$$

It is evident from these numbers that the complete oxidation of a molecule of glucose to CO_2 and H_2O can release enough energy to produce a large number of ATPs. As we will see in Chapter 5, up to about 36 molecules of ATP are formed per molecule of glucose oxidized under conditions that exist within most cells. For this many ATP molecules to be produced, the sugar molecule is disassembled in many small steps. Those steps in which the free-energy difference between the reactants and products is relatively large can be coupled to reactions that lead to the formation of ATP.

There are basically two stages in the catabolism of glucose, and they are virtually identical in all aerobic organisms. The first stage, **glycolysis**, occurs in the soluble phase of the cytoplasm (the cytosol) and leads to the formation of pyruvate. The second stage is the **tricarboxylic acid** (or **TCA**) **cycle**, which occurs within the mitochondria of eukaryotic cells and the cytosol of prokaryotes and leads to the final oxidation of the carbon atoms to carbon dioxide. Most of the chemical energy of glucose is stored in the form of high-energy electrons, which are removed as substrate molecules are oxidized during both glycolysis and the TCA cycle. It is the energy of these electrons that is ultimately used to synthesize ATP. We will concentrate in the following pages on the steps during glycolysis, the first stage in the oxidation of glucose, which occurs without the involvement of oxygen. This is presumably the pathway of energy capture utilized by our
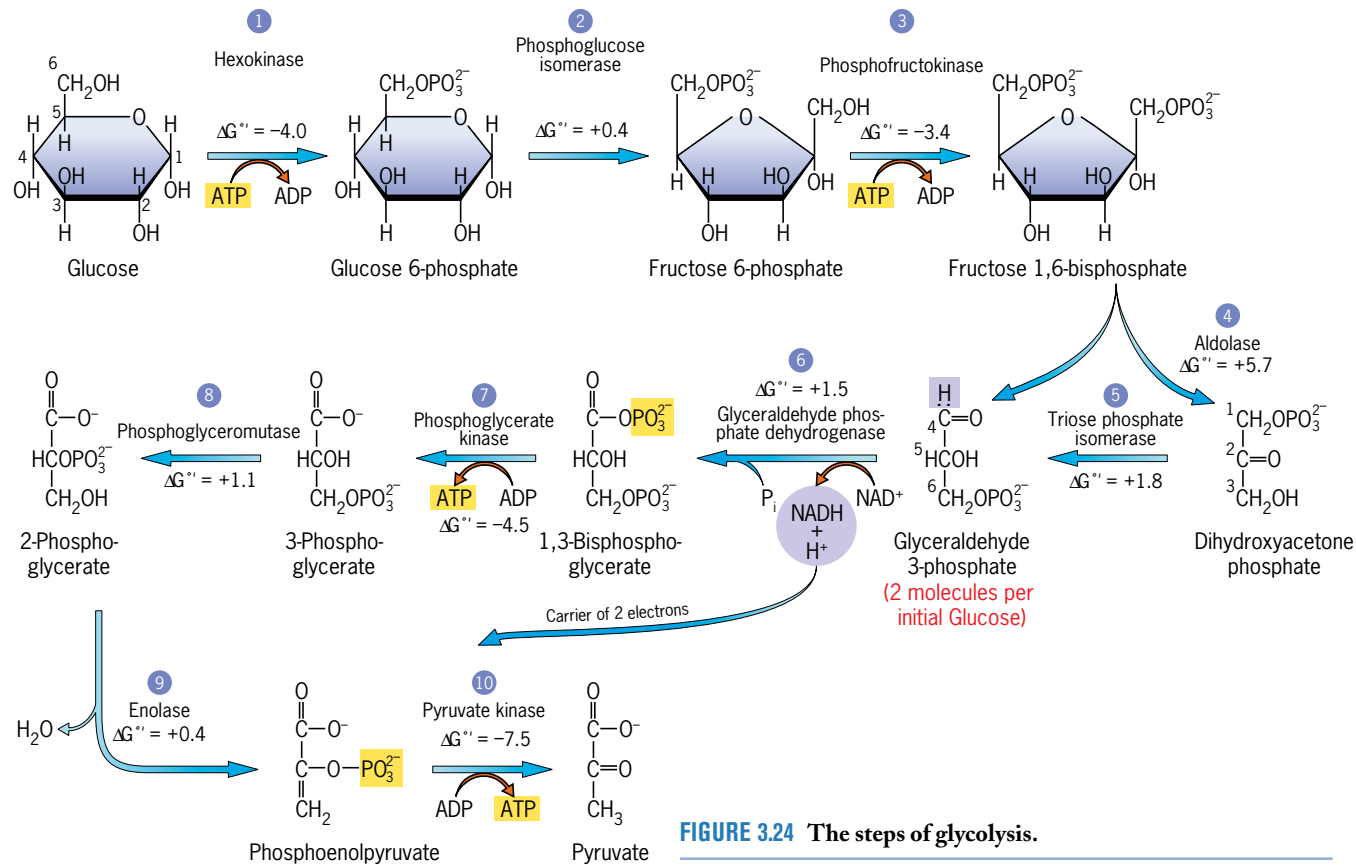
FIGURE 3.24 The steps of glycolysis.

early anaerobic ancestors and remains the major anabolic pathway utilized by anaerobic organisms living today. We will complete the story of glucose oxidation in Chapter 5 when we discuss the mitochondrion and its role in aerobic respiration.

Glycolysis and ATP Formation   The reactions of glycolysis and the enzymes that catalyze them are shown in Figure 3.24. Before discussing the specific reactions, an important point concerning the thermodynamics of metabolism can be made. In an earlier discussion, the difference between ΔG and $\Delta G^{\circ\prime}$ was stressed; it is the ΔG for a particular reaction that determines its direction in the cell. Actual measurements of the concentrations of metabolites in the cell can reveal the value for ΔG of a reaction at any given time. Figure 3.25 shows the typical ΔG values measured for the reactions of glycolysis. In contrast to the $\Delta G^{\circ\prime}$ values of Figure 3.24, all but three reactions have ΔG values of nearly zero; that is, they are near equilibrium. The three reactions that are far from equilibrium, which makes them essentially irreversible in the cell, provide the driving force that moves the metabolites through the glycolytic pathway in a directed manner.

In 1905, two British chemists, Arthur Harden and William Young, were studying glucose breakdown by yeast cells, a process that generates bubbles of CO_2 gas. Harden and Young noted that the bubbling eventually slowed and stopped, even though there was plenty of glucose left to metabolize. Apparently, some other essential component of the broth was being exhausted. After experimenting with a number of substances, the chemists found that adding inorganic
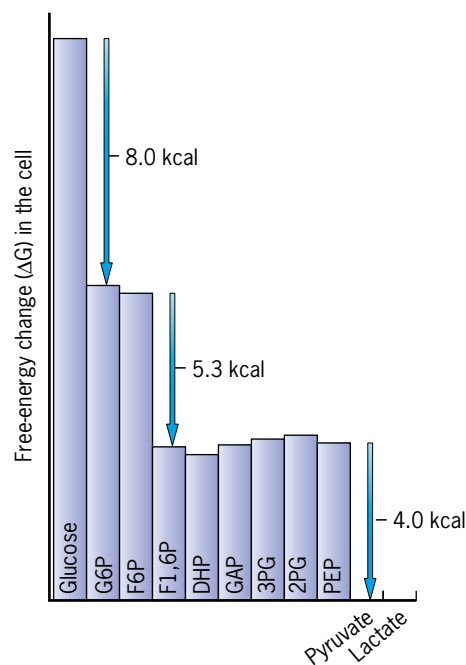
FIGURE 3.25 **Free-energy profile of glycolysis in the human erythrocyte.** All reactions are at or near equilibrium except those catalyzed by hexokinase, phosphofructokinase, and pyruvate kinase, which exhibit large differences in free energy. In the cell, all reactions must proceed with a decline in free energy; the slight increases in free energy depicted here for several steps must be regarded as deriving from errors in experimental measurements of metabolite concentrations. (FROM A. L. LEHNINGER, BIOCHEMISTRY, 2D ED., 1975. WORTH PUBLISHERS, NEW YORK.)

phosphates started the reaction going again. They concluded that the reaction was exhausting phosphate, the first clue that phosphate groups played a role in metabolic pathways. The importance of the phosphate group is illustrated by the initial reactions of glycolysis.

Glycolysis begins with the linkage of the sugar to a phosphate group (step 1, Figure 3.24) at the expense of one molecule of ATP. The use of ATP at this stage can be considered an energy investment—the cost of getting into the glucose oxidation business. Phosphorylation activates the sugar, making it capable of participating in subsequent reactions whereby phosphate groups are moved around and transferred to other acceptors. Phosphorylation also lowers the cytoplasmic concentration of glucose, which promotes continued diffusion of the sugar from the blood into the cell. Glucose 6-phosphate is converted to fructose 6-phosphate and then to fructose 1,6-bisphosphate at the expense of a second molecule of ATP (steps 2, 3). The six-carbon bisphosphate is split into two three-carbon monophosphates (step 4), which sets the stage for the first exergonic reaction to which the formation of ATP can be coupled. Let us turn now to the formation of ATP.

ATP is formed in two basically different ways, both illustrated by one chemical reaction of glycolysis: the conversion of glyceraldehyde 3-phosphate to 3-phosphoglycerate (steps 6, 7). The overall reaction is an oxidation of an aldehyde to a carboxylic acid (as in Figure 3.23), and it occurs in two steps catalyzed by two different enzymes (Figure 3.24). The first of these enzymes requires a nonprotein cofactor (a coenzyme), called nicotinamide adenine dinucleotide (NAD), to catalyze the reaction. As will be evident in this and following chapters, NAD plays a key role in energy metabolism by accepting and donating electrons. The first reaction (Figure 3.26a,b) is an oxidation–reduction in which two electrons and a proton (equivalent to a hydride ion, $:H^-$) are transferred from glyceraldehyde 3-phosphate (which becomes oxidized) to NAD^+ (which becomes reduced). The reduced form of the coenzyme
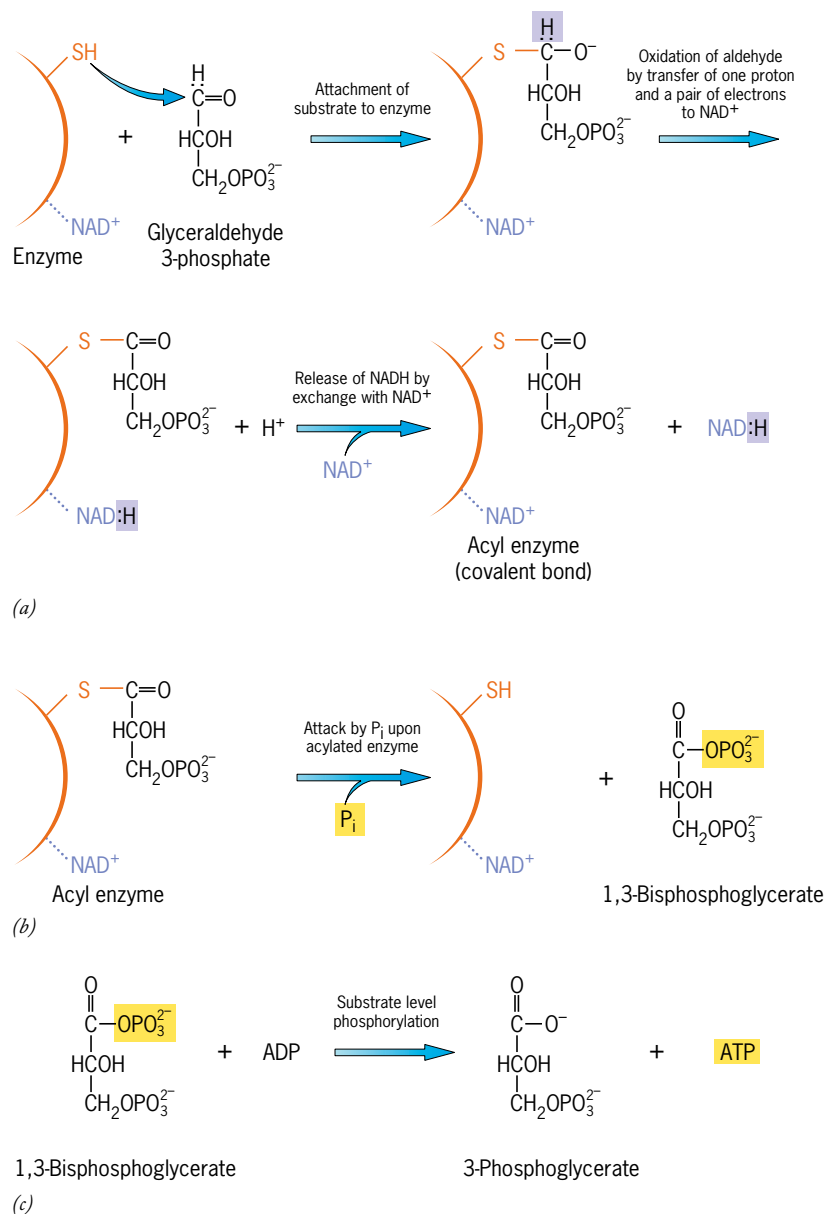
FIGURE 3.26 **The transfer of energy during a chemical oxidation.** The oxidation of glyceraldehyde 3-phosphate to 3-phosphoglycerate, which is an example of the oxidation of an aldehyde to a carboxylic acid, occurs in two steps catalyzed by two enzymes. The first reaction (parts a and b) is catalyzed by the enzyme glyceraldehyde 3-phosphate dehydrogenase, which transfers a hydride ion (two electrons and a proton) from the substrate to NAD^+. Once reduced, the NADH molecules are displaced by NAD^+ molecules from the cytosol. (c) The second reaction, which is catalyzed by the enzyme phosphoglycerate kinase, is an example of a substrate-level phosphorylation in which a phosphate group is transferred from a substrate molecule, in this case 1,3-bisphosphoglycerate, to ADP to form ATP.
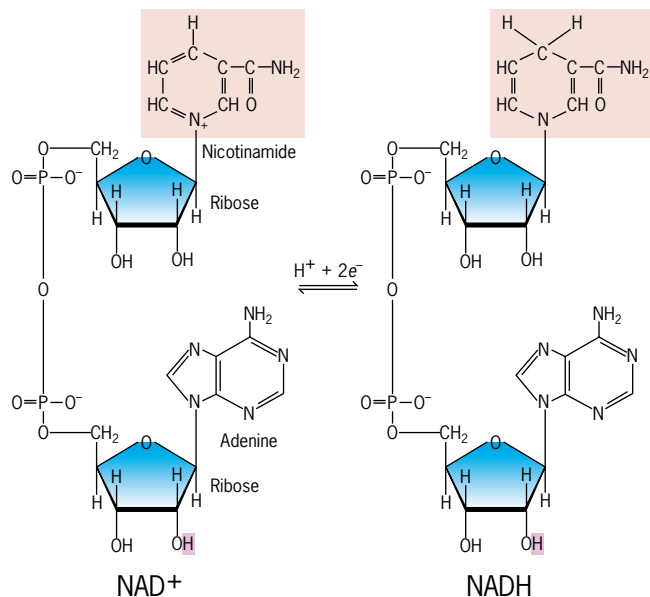
FIGURE 3.27 **The structure of NAD⁺ and its reduction to NADH.** When the 2′ OH of the ribose moiety (indicated by the purple colored screen) is covalently bonded to a phosphate group, the molecule is NADP⁺/NADPH, whose function is discussed later in the chapter.

is NADH (Figure 3.27). An enzyme that catalyzes this type of reaction is termed a **dehydrogenase**; the enzyme catalyzing the above reaction is *glyceraldehyde 3-phosphate dehydrogenase*. The NAD⁺ molecule, which is derived from the vitamin niacin, acts as a loosely associated coenzyme to the dehydrogenase in position to accept the hydride ion (i.e., both the electrons and the proton). The NADH formed in the reaction is then released from the enzyme in exchange for a fresh molecule of NAD⁺.

We will return to this reaction in a moment, but first we will continue with the consequences of the formation of NADH. NADH is considered a high-energy compound because of the ease with which it is able to transfer electrons to other electron-attracting molecules. (NADH is said to have a high electron-transfer potential relative to other electron acceptors in the cell; see Table 5.1). Electrons are usually transferred from NADH through a series of membrane-embedded electron carriers that constitute an *electron-transport chain*. As the electrons move along this chain, they move to a lower and lower free-energy state and are ultimately passed to molecular oxygen, reducing it to water. The energy released during electron transport is utilized to form ATP by a process called *oxidative phosphorylation*. Electron transport and oxidative phosphorylation will be explored in detail in Chapter 5.

In addition to the indirect route for ATP formation involving NADH and an electron-transport chain, the conversion of glyceraldehyde 3-phosphate to 3-phosphoglycerate includes a direct route for ATP formation. In the second step of this overall reaction (step 7, Figure 3.24 and 3.26*c*), a phosphate group is transferred from 1,3-bisphosphoglycerate to ADP to form a molecule of ATP. The reaction is catalyzed by the enzyme *phosphoglycerate kinase*. This direct route of ATP

formation is referred to as **substrate-level phosphorylation** because it occurs by transfer of a phosphate group from one of the substrates (in this case, 1,3-bisphosphoglycerate) to ADP. The remaining reactions of glycolysis (steps 8–10), which includes a second substrate-level phosphorylation of ADP (step 10), are indicated in Figure 3.24.

The substrate-level phosphorylation of ADP illustrates an important point about ATP. Its formation is not *that* endergonic. In other words, ATP can be readily formed by metabolic reactions. There are numerous phosphorylated molecules whose hydrolysis has a more negative $\Delta G^{\circ\prime}$ than that of ATP. Figure 3.28 illustrates the relative $\Delta G^{\circ\prime}$'s for the hydrolysis of several phosphorylated compounds. Any donor higher on the scale can be used to phosphorylate any molecule lower on the scale, and the $\Delta G^{\circ\prime}$ of this reaction will be equal to the difference between the two values given in the figure. For example, the $\Delta G^{\circ\prime}$ for the transfer of a phosphate group from 1,3-bisphosphoglycerate to ADP to form ATP is equal to -4.5 kcal/mol (-11.8 kcal/mol $+7.3$ kcal/mol). This concept of **transfer potential** is useful for comparing any series of donors and acceptors regardless of the group being transferred, whether protons, electrons, oxygen, or phosphate groups. Those molecules higher on the scale, that is, ones with greater free energy (larger $-\Delta G^{\circ\prime}$'s), are molecules with less affinity for the group being transferred than are ones lower on the scale. The less the affinity, the better the donor; the greater the affinity, the better the acceptor.
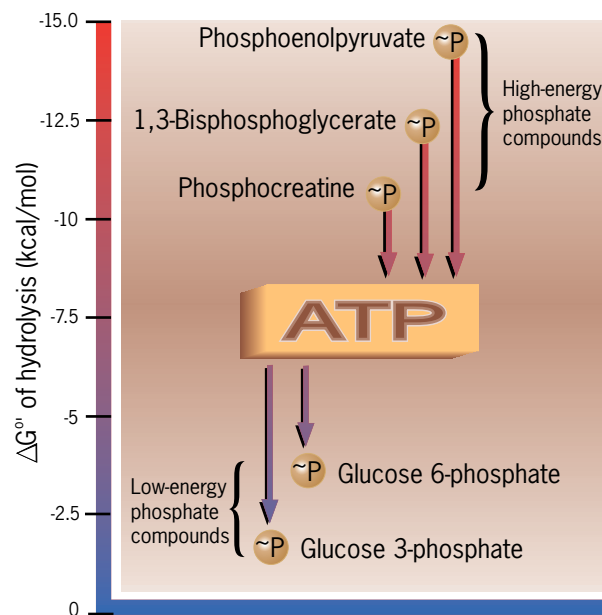
FIGURE 3.28 **Ranking compounds by phosphate transfer potential.** Those phosphorylated compounds higher on the scale (ones with a more negative $\Delta G^{\circ\prime}$ of hydrolysis) have a lower affinity for their phosphate group than those compounds lower on the scale. As a result, compounds higher on the scale readily transfer their phosphate group to form compounds that are lower on the scale. Thus, phosphate groups can be transferred from 1,3-bisphosphate or phosphoenolpyruvate to ADP during glycolysis.

An important feature of glycolysis is that it can generate a limited number of ATP molecules in the absence of oxygen. Neither the substrate-level phosphorylation of ADP by 1,3-bisphosphoglycerate nor a later one by phosphoenolpyruvate (step 10, Figure 3.24) requires molecular oxygen. Thus, glycolysis can be considered an **anaerobic pathway** to ATP production, indicating that it can proceed in the absence of molecular oxygen to continue to provide ATP. Two molecules of ATP are produced by substrate-level phosphorylation during glycolysis from each molecule of glyceraldehyde 3-phosphate oxidized to pyruvate. Since each molecule of glucose produces two molecules of glyceraldehyde 3-phosphate, four ATP molecules are generated per molecule of glucose oxidized to pyruvate. On the other hand, two molecules of ATP must be hydrolyzed to initiate glycolysis, leaving a net profit for the cell of two molecules of ATP per glucose oxidized. The net equation for the reactions of glycolysis can be written

$$\text{Glucose} + 2\,\text{ADP} + 2\,\text{P}_i + 2\,\text{NAD}^+ \rightarrow$$
$$2\,\text{pyruvate} + 2\,\text{ATP} + 2\,\text{NADH} + 2\,\text{H}^+ + 2\,\text{H}_2\text{O}$$
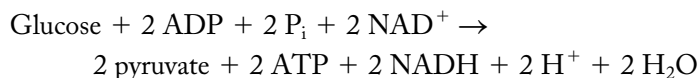
Pyruvate, the end product of glycolysis, is a key compound because it stands at the junction between anaerobic (oxygen-independent) and aerobic (oxygen-dependent) pathways. In the absence of molecular oxygen, pyruvate is subjected to fermentation, as discussed in the following section. When oxygen is available, pyruvate is further catabolized by aerobic respiration, as discussed in Chapter 5.

Anaerobic Oxidation of Pyruvate: The Process of Fermentation We have seen that glycolysis is able to provide a cell with a small net amount of ATP for each molecule of glucose converted to pyruvate. However, glycolytic reactions occur at rapid rates so that a cell is able to produce a significant amount of ATP using this pathway. In fact, a number of cells, including yeast cells, tumor cells, and muscle cells, rely heavily on glycolysis as a means of ATP formation. There is, however, a problem that must be confronted by such cells. One of the products of the oxidation of glyceraldehyde 3-phosphate is NADH. The formation of NADH occurs at the expense of one of the reactants, NAD^+, which is in short supply in cells. Since NAD^+ is a required reactant in this important step of glycolysis, it must be regenerated from NADH. If not, the oxidation of glyceraldehyde 3-phosphate can no longer take place, nor can any of the succeeding reactions of glycolysis. However, without oxygen, NADH cannot be oxidized to NAD^+ by means of the electron-transport chain because oxygen is the final electron acceptor of the chain. Cells are able, however, to regenerate NAD^+ by **fermentation**, the transfer of electrons from NADH to pyruvate, the end product of glycolysis, or to a compound derived from pyruvate (Figure 3.29). Like glycolysis, fermentation takes place in the cytosol. For most organisms, which depend on O_2, fermentation is a stopgap measure to regenerate NAD^+ when O_2 levels are low, so that glycolysis can continue and ATP production can be maintained.

The product of fermentation varies from one type of cell or organism to another. When muscle cells are required to
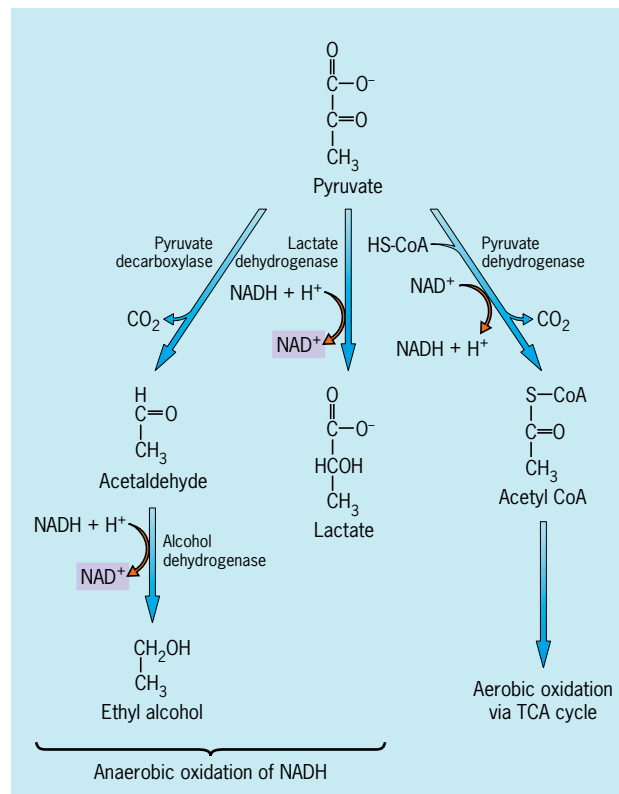
FIGURE 3.29 Fermentation. Most cells carry out aerobic respiration, which depends on molecular oxygen. If oxygen supplies should diminish, as occurs in a skeletal muscle cell undergoing strenuous contraction or a yeast cell living under anaerobic conditions, these cells are able to regenerate NAD^+ by fermentation. Muscle cells accomplish fermentation by formation of lactate, whereas yeast cells do so by formation of ethanol. Aerobic oxidation of pyruvate by means of the TCA cycle is discussed at length in Chapter 5.

contract repeatedly, the oxygen level becomes too low to keep pace with the cell's metabolic demands. Under these conditions, skeletal muscle cells regenerate NAD^+ by converting pyruvate to lactate. When oxygen once again becomes available in sufficient amounts, the lactate can be converted back to pyruvate for continued oxidation. Yeast cells have faced the challenges of anaerobic life with a different metabolic solution: they convert pyruvate to ethanol, as illustrated in Figure 3.29.

Although fermentation is a necessary adjunct to metabolism in many organisms, and the sole source of metabolic energy in some anaerobes, the energy gained by glycolysis alone is meager when compared with the complete oxidation of glucose to carbon dioxide and water. When a mole of glucose is completely oxidized, 686 kilocalories are released. In contrast, only 57 kilocalories are released when this same amount of glucose is converted to ethanol under standard conditions, and only 47 kilocalories are released when it is converted to lactate. In any case, only two molecules of ATP are formed per glucose oxidized by glycolysis and fermentation; over 90 percent of the energy is simply discarded in the fermentation product (as evidenced by the flammability of ethyl alcohol).

During the early stages of life on Earth, when oxygen had not yet appeared, glycolysis and fermentation were probably the primary metabolic pathways by which energy was extracted from sugars by primitive prokaryotic cells. Following the appearance of cyanobacteria, the oxygen of the atmosphere rose dramatically, and the evolution of an aerobic metabolic strategy became possible. As a result, the products of glycolysis could be completely oxidized, and much more ATP could be harvested. We will see how this is accomplished in Chapter 5 when we discuss the structure and function of the mitochondrion.

Reducing Power   The energy required to synthesize complex biological molecules, such as proteins, fats, and nucleic acids, is derived largely from the ATP generated by glycolysis and electron transport. But many of these materials, particularly fats and other lipids, are more reduced than the metabolites from which they are built. Synthesis of fats requires the reduction of metabolites, which is accomplished by the transfer of high-energy electrons from NADPH, a compound similar in structure to NADH, but containing an additional phosphate group (described in the legend to Figure 3.27). A cell's reservoir of NADPH represents its **reducing power**, which is an important measure of the cell's usable energy content. The use of NADPH can be illustrated by one of the key reactions of photosynthesis:

$$\underset{\text{1,3-Bisphosphoglycerate}}{\overset{\displaystyle \overset{O}{\underset{\displaystyle \|}{}}\!\overset{\displaystyle C-OPO_3^{2-}}{\underset{\displaystyle \big|}{}}\overset{\displaystyle HC-OH}{\underset{\displaystyle \big|}{}}\overset{\displaystyle CH_2OPO_3^{2-}}{}}{}\ +\ NADPH\ \rightarrow\ \underset{\text{Glyceraldehyde 3-phosphate}}{\overset{\displaystyle \overset{O}{\underset{\displaystyle \|}{}}\!\overset{\displaystyle CH}{\underset{\displaystyle \big|}{}}\overset{\displaystyle HC-OH}{\underset{\displaystyle \big|}{}}\overset{\displaystyle CH_2OPO_3^{2-}}{}}{}\ +\ NADP^+\ +\ P_i}$$
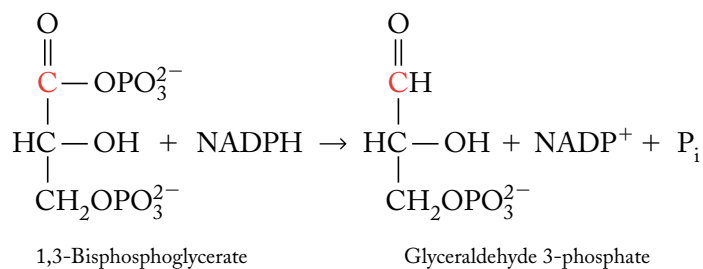
In this reaction, a pair of electrons (together with a proton) are transferred from NADPH to the substrate 1,3-bisphosphoglycerate, reducing a carbon atom (indicated in red).

The oxidized form of NADPH, $NADP^+$, is formed from NAD^+ in the following reaction:

$$NAD^+ + ATP \rightleftharpoons NADP^+ + ADP$$
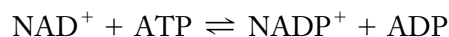
NADPH can then be formed by the reduction of $NADP^+$. Like NADH, NADPH is a high-energy compound because of its high electron-transfer potential. The transfer of free energy in the form of these electrons raises the acceptor to a more reduced, more energetic state.

The separation of reducing power into two distinct but related molecules, NADH and NADPH, reflects a separation of their primary metabolic roles. NADH and NADPH are recognized as coenzymes by different enzymes. Enzymes having a reductive role in anabolic pathways utilize NADPH as their coenzyme, whereas enzymes acting as dehydrogenases in catabolic pathways utilize NAD^+. Even though they are used differently, one coenzyme can reduce the other in the following reaction, which is catalyzed by the enzyme *transhydrogenase*:

$$NADH + NADP^+ \rightleftharpoons NAD^+ + NADPH$$
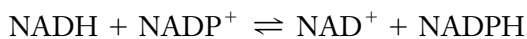
When energy is abundant, the production of NADPH is favored, providing a supply of electrons needed for biosynthesis of new macromolecules that are essential for growth. When energy resources are scarce, however, most of the high-energy electrons of NADH are "cashed in" for ATP, and only enough NADPH is generated to meet the cell's minimal biosynthetic requirements.

Metabolic Regulation

The amount of ATP present in a cell at a given moment is surprisingly small. A bacterial cell, for example, contains approximately one million molecules of ATP, whose half-life is very brief (on the order of a second or two). With so limited a supply, it is evident that ATP is not a molecule in which a large total amount of free energy is stored. The energy reserves of a cell are stored instead as polysaccharides and fats. When the levels of ATP start to fall, reactions are set in motion to increase ATP formation at the expense of the energy-rich storage forms. Similarly, when ATP levels are high, reactions that would normally lead to ATP production are inhibited. Cells regulate these important energy-releasing reactions by controlling certain key enzymes in a number of metabolic pathways.

The function of a protein is closely related to its structure (i.e., conformation). It is not surprising, therefore, that cells regulate protein activity by modifying the conformation of key protein molecules. In the case of enzymes, regulation of catalytic activity is centered on the modification of the structure of the active site. Two common mechanisms for altering the shape of an enzyme's active site are covalent modification and allosteric modulation, both of which play key roles in regulating glucose oxidation.[4]

Altering Enzyme Activity by Covalent Modification   During the mid-1950s, Edmond Fischer and Edwin Krebs of the University of Washington were studying glycogen phosphorylase, an enzyme found in muscle cells that disassembles glycogen into its glucose subunits. The enzyme could exist in either an inactive or an active state. Fischer and Krebs prepared a crude extract of muscle cells and found that inactive enzyme molecules in the extract could be converted to active ones simply by adding ATP to the test tube. Further analysis revealed a second enzyme in the extract—a "converting enzyme," as they called it—that transferred a phosphate group from ATP to one of the 841 amino acids that make up the glycogen phosphorylase molecule. The presence of the phosphate group altered the shape of the active site of the enzyme molecule and increased its catalytic activity.

Subsequent research has shown that covalent modification of enzymes, as illustrated by the addition (or removal) of phosphates, is a general mechanism for changing the activity of enzymes. Enzymes that transfer phosphate groups to other proteins are called **protein kinases** and regulate such diverse activities as hormone action, cell division, and gene expression. The "converting enzyme" discovered by Krebs and Fischer was later named phosphorylase kinase, and its regulation is discussed

---

[4]Metabolism is also regulated by controlling the concentrations of enzymes. The relative rates at which enzymes are synthesized and degraded are considered in subsequent chapters.

in Section 15.3. There are two basically different types of protein kinases: one type adds phosphate groups to specific tyrosine residues in a substrate protein; the other type adds phosphates to specific serine or threonine residues in the substrate. The importance of protein kinases is reflected in the fact that approximately 2 percent of all the genes of a yeast cell (113 out of about 6200) encode members of this class of enzymes.

Altering Enzyme Activity by Allosteric Modulation
Allosteric modulation is a mechanism whereby the activity of an enzyme is either inhibited or stimulated by a compound that binds to a site, called the **allosteric site**, that is spatially distinct from the enzyme's active site. Like the sequential collapse of a row of dominoes, the binding of a compound to the allosteric site sends a "ripple" through the protein, producing a defined change in the shape of the active site, which may be located on the opposite side of the enzyme or even on a different polypeptide within the protein molecule. Depending on the particular enzyme and the particular allosteric modulator, the change in shape of the active site may either stimulate or inhibit its ability to catalyze the reaction. Allosteric modulation illustrates the intimate relationship between molecular structure and function. Very small changes in the structure of the enzyme induced by the allosteric modulator can cause marked changes in enzyme activity.

Cells are highly efficient manufacturing plants that do not waste energy and materials producing compounds that are not utilized. One of the primary mechanisms cells use to shut down anabolic assembly lines is a type of allosteric modulation called **feedback inhibition**, in which the enzyme catalyzing the first committed step in a metabolic pathway is temporarily inactivated when the concentration of the end product of that pathway—an amino acid, for example—reaches a certain level. This is illustrated by the simple pathway depicted in Figure 3.30 in which two substrates, A and B, are converted to the end product E. As the concentration of the product E rises, it binds to the allosteric site of enzyme BC, causing a conformational change in the active site that decreases the enzyme's activity. Feedback inhibition exerts immediate, sensitive control over a cell's anabolic activity.

Separating Catabolic and Anabolic Pathways A brief consideration of the anabolic pathway leading to the synthesis of glucose (**gluconeogenesis**) will illustrate a few important aspects about synthetic pathways. Most cells are able to synthesize glucose from pyruvate at the same time they are oxidizing glucose as their major source of chemical energy. How are cells able to utilize these two opposing pathways? The first important point is that, even though enzymes can catalyze a reaction in both directions, the reactions of gluconeogenesis cannot proceed simply by the reversal of the reactions of glycolysis. The glycolytic pathway contains three thermodynamically irreversible reactions (Figure 3.25), and these steps must somehow be bypassed. Even if all the reactions of glycolysis could be reversed, it would be a very undesirable way for the cell to handle its metabolic activities, since the two pathways could not be controlled independently of one another. Thus, a cell could not shut down glucose synthesis and crank up glucose breakdown because the same enzymes would be active in both directions.
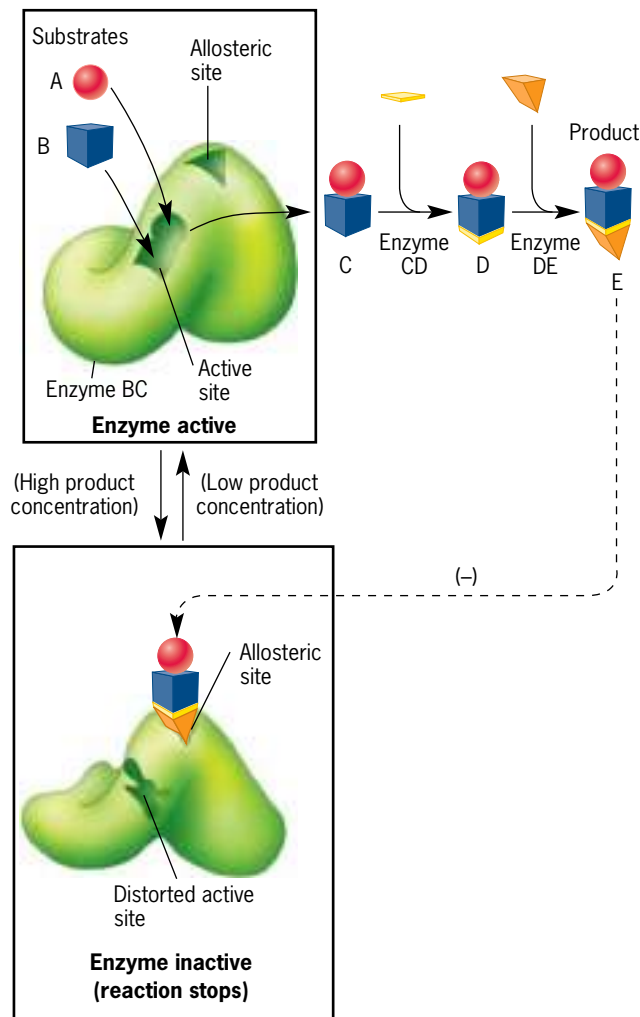
FIGURE 3.30 Feedback inhibition. The flow of metabolites through a metabolic pathway stops when the first enzyme of the pathway (enzyme BC) is inhibited by the end product of that pathway (compound E), which binds to an allosteric site on the enzyme. Feedback inhibition prevents a cell from wasting resources by continuing to produce compounds that are no longer required.

If one compares the pathway by which glucose is degraded to the pathway by which glucose is synthesized, it is apparent that some of the reactions are identical, although they are running in opposite directions, while others are very different (steps 1 to 3, Figure 3.31). By using different enzymes to catalyze different key reactions in the two opposing pathways, a cell can solve both the thermodynamic and the regulatory problems inherent in being able to manufacture and degrade the same molecules.

We can see this better by looking more closely at key enzymes of both glycolysis and gluconeogenesis. As indicated in step 2 of Figure 3.31, *phosphofructokinase*, an enzyme of glycolysis, catalyzes the reaction

Fructose 6-phosphate + ATP $\rightleftharpoons$

fructose 1,6-bisphosphate + ADP
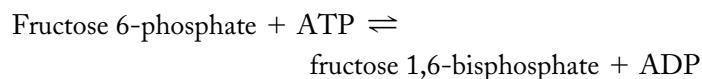
which has a $\Delta G^{\circ\prime}$ of -3.4 kcal/mol, making it essentially irreversible. The reaction has such a large $-\Delta G^{\circ\prime}$ because it is

GLYCOLYSIS   **GLUCONEOGENESIS**

Glucose

1

ATP

$\Delta G^{\circ\prime} = -4.0$ kcal/mol

ADP

Hexokinase

P_i $\Delta G^{\circ\prime} = -2.9$ kcal/mol

Glucose 6-phosphatase

Glucose 6-phosphate

Fructose 6-phosphate

2

ATP

$\Delta G^{\circ\prime} = -3.4$ kcal/mol

ADP

Phosphofructokinase

P_i $\Delta G^{\circ\prime} = -3.9$ kcal/mol

Fructose bisphosphatase

Fructose 1,6-bisphosphate

Phosphoenolpyruvate

3

$\Delta G^{\circ\prime} = -7.5$ kcal/mol

GDP

$\Delta G^{\circ\prime} = +0.2$ kcal/mol

GTP

CO_2

Phosphoenolpyruvate carboxykinase

ADP

Oxaloacetate

ATP

Pyruvate kinase

ADP

ATP Pyruvate carboxylase

CO_2

Pyruvate

FIGURE 3.31 Glycolysis versus gluconeogenesis. Whereas most of the reactions are the same in the two pathways, even though they run in opposite directions, the three irreversible reactions of glycolysis (steps 1 to 3 here) are replaced in the gluconeogenic pathway by different, thermodynamically favored reactions.

coupled to the hydrolysis of ATP. In gluconeogenesis, the formation of fructose 6-phosphate is catalyzed by the enzyme *fructose 1,6-bisphosphatase* by a simple hydrolysis:

Fructose 1,6-bisphosphate + $H_2O \rightleftharpoons$

fructose 6-phosphate + P_i $\Delta G^{\circ\prime} = -3.9$ kcal/mol
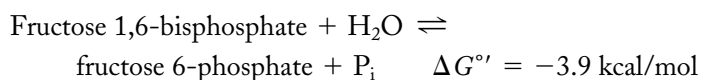
The particular enzymes of glycolysis and gluconeogenesis described above are key enzymes in their respective pathways. These enzymes are regulated, in part, by AMP and ATP. Keep in mind that the concentration of AMP within cells is inversely related to the concentration of ATP; when ATP levels are low, AMP levels are high, and vice versa. Consequently, elevated AMP concentrations signal to a cell that its ATP fuel reserves are becoming depleted. Even though ATP happens to be a substrate of phosphofructokinase, ATP is also an allosteric inhibitor, while AMP is an allosteric activator. When ATP levels are high, the activity of the enzyme is decreased

so that no additional ATP is formed by glycolysis. Conversely, when ADP and AMP levels are high relative to ATP, the activity of the enzyme is increased, which promotes the synthesis of ATP. In contrast, the activity of fructose 1,6-bisphosphatase, a key enzyme of gluconeogenesis, is inhibited by elevated levels of AMP.

The importance of allosteric modulation in metabolic control is illustrated by recent studies on the enzyme AMP-activated protein kinase or simply AMPK. Like phosphorylase kinase discussed above, AMPK controls the activity of other enzymes by adding phosphate groups to specific serine or threonine residues within their structure. AMPK is regulated allosterically by AMP. As AMP concentrations rise, AMPK is activated, causing it to phosphorylate and inhibit key enzymes involved in anabolic pathways while at the same time phosphorylating and activating key enzymes in catabolic pathways. The end result of AMPK activation is a decrease in the activity of pathways that consume ATP and an increase in the activity of pathways that produce ATP, leading to an elevation in the concentration of ATP within the cell. AMPK is involved not only in cellular energy regulation but, at least in mammals, in regulation of energy balance within the whole body. In mammals, appetite is controlled by centers within the hypothalamus of the brain that respond to the levels of certain nutrients and hormones within the blood. A drop in blood glucose levels, for example, acts on the hypothalamus to stimulate appetite, which appears to be mediated through activation of AMPK in hypothalamic nerve cells. Conversely, the feeling of being "full" that we experience after eating a meal is thought to be triggered by certain hormones (e.g., leptin secreted by fat cells) that inhibit the activity of AMPK in these same brain cells. Because of its role as an "energy thermostat," AMPK has become a prime target in the development of drugs to treat both obesity and diabetes.

REVIEW ?

1. Why are catabolic pathways described as convergent, whereas anabolic pathways are described as divergent?

2. Compare the energy obtained by a cell that oxidizes glucose anaerobically and aerobically. What is the difference in the end products of these two types of metabolism?

3. Explain what is meant by phosphate transfer potential. How does the phosphate transfer potential of phosphoenolpyruvate compare with that of ATP? What does this mean thermodynamically (i.e., in terms of relative $\Delta G^{\circ\prime}$s of hydrolysis)? What does it mean in terms of affinity for phosphate groups?

4. Why is reducing power considered a form of energy?

5. In the reaction in which glyceraldehyde 3-phosphate is oxidized to 3-phosphoglycerate, which compound is acting as the reducing agent? Which as the oxidizing agent? Which compound is being oxidized? Which is being reduced?

6. Which reactions of glycolysis are coupled to ATP hydrolysis? Which reactions involve substrate-level phosphorylation? Which reactions depend on either fermentation or aerobic respiration to continue?

SYNOPSIS

Energy is the capacity to do work. Energy can occur in various forms, including chemical, mechanical, light, electrical, and thermal energy, which are interconvertible. Whenever energy is exchanged, the total amount of energy in the universe remains constant, but there is a loss of free energy, that is, energy available to do additional work. The usable energy lost to entropy results from an increase in the randomness and disorder of the universe. Living organisms are systems of low entropy, maintained by the constant input of external energy that is ultimately derived from the sun. (p. 85)

All spontaneous (exergonic) energy transformations proceed from a state of higher free energy to a state of lower free energy; the ΔG must be negative. In a chemical reaction, ΔG is equivalent to the difference in the free-energy content between the reactants and products. The greater the ΔG, the farther the reaction is from equilibrium. As the reaction proceeds, ΔG decreases and reaches 0 at equilibrium. To compare the energy changes during different chemical reactions, free-energy differences between reactants and products are determined for a set of standard conditions and denoted as $\Delta G^{\circ\prime}$, in which $\Delta G^{\circ\prime} = -2.303\ RT \log K'_{eq}$. Reactions having equilibrium constants greater than one have negative $\Delta G^{\circ\prime}$ values. It must be kept in mind that $\Delta G^{\circ\prime}$ is a fixed value that describes the direction in which a reaction proceeds when the reaction mixture is at standard conditions. It is of no value in determining the direction of a reaction at a particular moment in the cell; this is governed by ΔG, which is dependent on the concentrations of reactants and products at the time. Reactions that have positive $\Delta G^{\circ\prime}$ values (such as the reaction in which dihydroxyacetone phosphate is converted to glyceraldehye 3-phosphate) can occur in the cell because the ratio of reactants to products is kept at a value greater than that predicted by K'_{eq}. (p. 87)

The hydrolysis of ATP is a highly favorable reaction ($\Delta G^{\circ\prime} = -7.3$ kcal/mol) and can be used to drive reactions that would otherwise be unfavorable. The use of ATP hydrolysis to drive unfavorable reactions is illustrated by glutamine synthesis from glutamic acid and NH_3 ($\Delta G^{\circ\prime} = +3.4$ kcal/mol). The reaction is driven through the formation of a common intermediate, glutamyl phosphate. ATP hydrolysis can play a role in such processes because the cell maintains an elevated [ATP]/[ADP] ratio that is far above that at equilibrium, which illustrates that cellular metabolism operates under nonequilibrium conditions. This does not mean that every reaction is kept from equilibrium. Rather, certain key reactions in a metabolic pathway occur with large negative ΔG values, which makes them essentially irreversible within the cell and allows them to drive the entire pathway. The concentrations of reactants and products can be maintained at relatively constant nonequilibrium (steady state) values within a cell because materials are continually flowing into the cell from the external medium and waste products are continually being removed. (p. 89)

Enzymes are proteins that vastly accelerate the rate of specific chemical reactions by binding to the reactant(s) and increasing the likelihood that they will be converted to products. Like all true catalysts, enzymes are present in small amounts; they are not altered irreversibly during the course of the reaction; and they have no effect on the thermodynamics of the reaction. Enzymes, therefore, cannot cause an unfavorable reaction (one with a $+\Delta G$) to proceed in the forward direction, nor can they change the ratio of products to reactants at equilibrium. As catalysts, enzymes can only accelerate the rate of favorable reactions, which they accomplish under conditions of mild temperature and pH found in the cell. Enzymes are also characterized by specificity toward their substrates, highly efficient

catalysis with virtually no undesirable by-products, and the opportunity for their catalytic activity to be regulated. (p. 92)

Enzymes act by lowering the activation energy (E_A)—the kinetic energy required by reactants to undergo reaction. As a result, a much greater percentage of reactant molecules possess the necessary energy to be converted to products in the presence of an enzyme. Enzymes lower E_A through the formation of an enzyme–substrate complex. The part of the enzyme that binds the substrate(s) is called the active site, which also contains the necessary amino acid side chains and/or cofactors to influence the substrates so as to facilitate the chemical transformation. Among the mechanisms that facilitate catalysis, enzymes are able to hold reactants in the proper orientation; they are able to make the substrates more reactive by influencing their electronic character; and they are able to exert physical stress that weakens certain bonds within the substrate. (p. 94)

Metabolism is the collection of biochemical reactions that occur within the cell. These reactions can be grouped into metabolic pathways containing a sequence of chemical reactions in which each reaction is catalyzed by a specific enzyme. Metabolic pathways are divided into two broad types: catabolic pathways, in which compounds are disassembled and energy is released, and anabolic pathways that lead to the synthesis of more complex compounds using energy stored in the cell. Macromolecules of diverse structure are degraded by catabolic pathways into a relatively small number of low-molecular-weight metabolites that provide the raw materials from which anabolic pathways diverge. Both types of pathways include oxidation–reduction reactions in which electrons are transferred from one substrate to another, increasing the state of reduction of the recipient and increasing the state of oxidation of the donor. (p. 105)

The state of reduction of an organic molecule, as measured by the number of hydrogens per carbon atom, provides a rough measure of the molecule's energy content. A mole of glucose releases 686 kcal when completely oxidized to CO_2 and H_2O, whereas the conversion of a mole of ADP to ATP requires only 7.3 kcal. Thus, the oxidation of a molecule of glucose can produce enough energy to generate a large number of ATP molecules. The first stage in the catabolism of glucose is glycolysis, during which glucose is converted to pyruvate with a net gain of two molecules of ATP and two molecules of NADH. The ATP molecules are produced by substrate-level phosphorylation in which a phosphate group is transferred from a substrate to ADP. The NADHs are produced by oxidation of an aldehyde to a carboxylic acid with the accompanying transfer of a hydride ion (a proton and two electrons) from the substrate to NAD^+. In the presence of O_2, most cells oxidize the NADH by means of an electron-transport chain, forming ATP by aerobic respiration. In the absence of O_2, NAD^+ is regenerated by fermentation in which the high-energy electrons of NADH are used to reduce pyruvate. NAD^+ regeneration is required for glycolysis to continue. (p. 106)

An enzyme's activity is commonly regulated by two mechanisms: covalent modification and allosteric modulation. Covalent modification is most often accomplished by transfer of a phosphate group from ATP to one or more serine, threonine, or tyrosine residues of the enzyme in a reaction catalyzed by a protein kinase. Allosteric modulators act by binding noncovalently to a site on the enzyme that is removed from the active site. Binding of the modulator alters the conformation of the active site, increasing or decreasing the enzyme's catalytic activity. A common example of allosteric modulation is feedback inhibition, in which the end product of a pathway allosterically inhibits the first enzyme that is unique to that pathway. The same compound degraded

by a catabolic pathway may serve as the end product of an anabolic pathway. Glucose, for example, is degraded by glycolysis and synthesized by gluconeogenesis. While most of the enzymes are shared by the two pathways, three key enzymes are unique to each pathway, which allows the cell to regulate the two pathways independently and to overcome what would otherwise be irreversible reactions. (p. 112)

ANALYTIC QUESTIONS

1. How would you expect a drop in pH to affect a reaction catalyzed by chymotrypsin? How might an increase in pH affect this reaction?

2. Feedback inhibition typically alters the activity of the first enzyme of a metabolic pathway rather than one of the latter enzymes of the pathway. Why is this adaptive?

3. After reviewing the reactions of glutamine formation on page 90, explain why each of the following statements concerning the third (or overall) reaction is either true or false.
   a. If the reaction were written in reverse, its $\Delta G^{\circ\prime}$ would be $+3.9$ kcal/mol.
   b. If all reactants and products were at standard conditions at the beginning of an experiment, after a period of time, the $[NH_3]/[ADP]$ ratio would decrease.
   c. As the reaction proceeds, the $\Delta G^{\circ\prime}$ moves closer to zero.
   d. At equilibrium, the forward and reverse reactions are equal and the $[ATP]/[ADP]$ ratio becomes one.
   e. In the cell, it is possible for glutamine to be formed when the $[glutamine]/[glutamic\ acid]$ ratio is greater than one.

4. You have just isolated a new enzyme and have determined the velocity of reaction at three different substrate concentrations. You find that the slope of the product versus time curve is the same for all three concentrations. What can you conclude about the conditions in the reaction mixture?

5. Lysozyme is a slow-acting enzyme, requiring approximately two seconds to catalyze a single reaction. What is the turnover number of lysozyme?

6. In the reaction $R \rightleftharpoons P$, if one mole of product (P) has the same free energy as one mole of reactant (R), what is the value for the K'_{eq} of this reaction? What is the value of the $\Delta G^{\circ\prime}$?

7. What is meant in terms of concentration ratios when it is said that the ΔG of ATP hydrolysis in the cell is approximately -12 kcal/mol, whereas the $\Delta G^{\circ\prime}$ is -7.3 kcal/mol?

8. The enzymes under cellular regulation are those whose reactions typically proceed under nonequilibrium conditions. What would be the effect of allosteric inhibition of an enzyme that operated close to equilibrium?

9. In the reaction $A \rightleftharpoons B$, if the K'_{eq} is 10^3, what is the $\Delta G^{\circ\prime}$? What is the $\Delta G^{\circ\prime}$ if the K'_{eq} had been determined to be 10^{-3}? What is the K'_{eq} of the hexokinase reaction indicated in Figure 3.24 (step 1)?

10. In 1926, James Sumner concluded that urease was a protein based on the fact that crystals of the enzyme tested positive for reagents that reacted with proteins and negative for reagents that reacted with fats, carbohydrates, and other substances. His conclusion was attacked by other enzymologists, who found that their highly active enzyme solutions failed to contain evidence of protein. How can these seemingly opposing findings be reconciled?

11. If the reaction $XA + Y \rightleftharpoons XY + A$ has a $\Delta G^{\circ\prime}$ of $+7.3$ kcal/mol, could this reaction be driven in the cell by coupling it to ATP hydrolysis? Why or why not?

12. In a series of reactions, $A \rightarrow B \rightarrow C \rightarrow D$, it was determined that the equilibrium constant for the second reaction ($B \rightarrow C$) is 0.1. You would expect the concentration of C in a living cell to be: (1) equal to B, (2) one-tenth of B, (3) less than one-tenth of B, (4) 10 times that of B, (5) more than 10 times that of B. (Circle any correct answer.)

13. The reaction of compound X with compound Y to produce compound Z is an unfavorable reaction ($\Delta G^{\circ\prime} = +5$ kcal/mol). Draw the chemical reactions that would occur if ATP was utilized to drive the reaction.

14. ATP has evolved as the central molecule in energy metabolism. Could 1,3-bisphosphoglycerate serve the same function? Why or why not?

15. Calculate the ΔG for ATP hydrolysis in a cell in which the $[ATP]/[ADP]$ ratio had climbed to $100:1$, while the Pi concentration remained at 10 mM. How does this compare to the ratio of $[ATP]/[ADP]$ when the reaction is at equilibrium and the P_i concentration remains at 10 mM? What would be the value for ΔG when the reactants and products were all at standard conditions (1 M)?

16. Consider the reaction:

    Glucose $+ P_i \rightleftharpoons$ glucose 6-phosphate $+ H_2O$
    $$\Delta G^{\circ\prime} = +3\ \text{kcal/mol}$$

    What is the equilibrium constant, K'_{eq}, for this reaction? (*Note:* the concentration of water is ignored.) Does the positive $\Delta G^{\circ\prime}$ in the above reaction mean that the reaction can never go spontaneously from left to right?

17. Under physiologic conditions, $[Glucose] = 5$ mM, $[Glucose\ 6\text{-phosphate}] = 83$ mM, and $[P_i] = 1$ mM. Under these conditions, will the reaction of problem 16 go spontaneously from left to right? If not, what would the concentration of glucose need to be for the reaction to go from left to right, if the concentrations of the other reactants and products are as stated above?

18. Consider the reaction:

    Glutamate $+$ ammonia $\rightleftharpoons$ glutamine $+ H_2O$
    $$\Delta G^{\circ\prime} = +3.4\ \text{kcal/mol}$$

    If the concentration of ammonia (NH_3) is 10 mM, what is the ratio of glutamate/glutamine required for the reaction to proceed spontaneously from left to right at 25°C?

19. It should be clear that the synthesis of glutamine cannot occur in a cell by the reaction described in problem 18. The actual reaction couples glutamine synthesis to ATP hydrolysis:

    Glutamate $+$ ammonia $+$ ATP $\rightleftharpoons$ glutamine $+$ ADP $+ P_i$

    What is the $\Delta G^{\circ\prime}$ for this reaction? Suppose that all reactants and products, except ammonia, are present at 10 mM. What concentration of ammonia would be required to drive the reaction to the right, that is, to drive the net synthesis of glutamine?

20. A noncompetitive inhibitor does not prevent the enzyme from binding its substrate. What will be the effect of increasing the substrate concentration in the presence of a noncompetitive inhibitor? Do you expect a noncompetitive inhibitor to affect the enzyme V_{max}? K_M? Explain briefly.

4

The Structure and Function of the Plasma Membrane

T he outer walls of a house or car provide a strong, inflexible barrier that protects its human inhabitants from an unpredictable and harsh external world. You might expect the outer boundary of a living cell to be constructed of an equally tough and impenetrable barrier because it must also protect its delicate internal contents from a nonliving, and often inhospitable, environment. Yet cells are separated from the external world by a thin, fragile structure called the **plasma membrane** that is only 5 to 10 nm wide. It would require about five thousand plasma membranes stacked one on top of the other to equal the thickness of a single page of this book.

Because it is so thin, no hint of the plasma membrane is detected when a section of a cell is examined under a light microscope. In fact, it wasn't until the late 1950s that techniques for preparing and staining tissue had progressed to the point where the plasma membrane could be resolved in the electron microscope. These early electron micrographs, such as those taken by J. D. Robertson of Duke University (Figure 4.1*a*), portrayed the plasma membrane as a three-layered structure, consisting of darkly staining inner and outer layers and lightly staining middle layer. All membranes that were examined closely—whether they were plasma, nuclear, or cytoplasmic membranes (Figure 4.1*b*), or taken from plants, animals, or microorganisms—showed this same ultrastructure. In addition to providing a visual image of this critically important cellular structure, these electron micrographs touched off a vigorous debate as to the molecular composition of the various layers of a membrane, an argument that went to the very

A model of a fully hydrated lipid bilayer composed of phosphatidylcholine molecules (each containing two myristoyl fatty acids) penetrated by a single transmembrane helix composed of 32 alanine residues. (FROM LIYANG SHEN, DONNA BASSOLINO, AND TERRY STOUCH, BRISTOL-MYERS SQUIBB RESEARCH INSTITUTE, FROM BIOPHYSICAL JOURNAL, VOL. 73, P. 6, 1997.)

(a)

50 nm
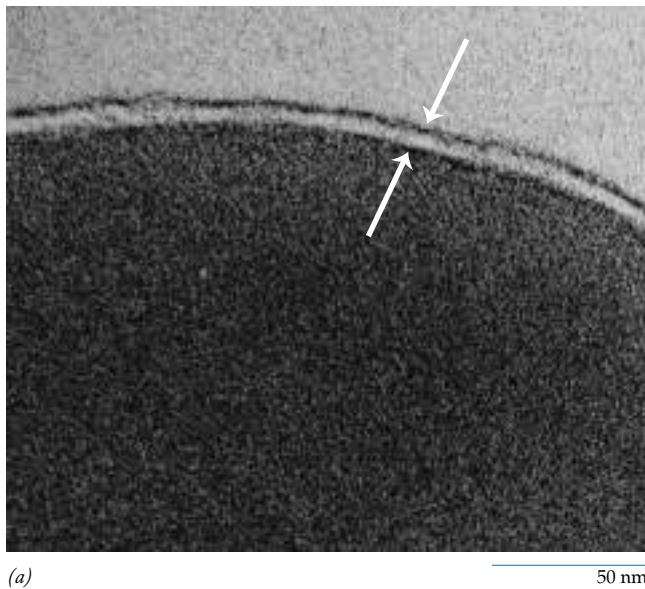
FIGURE 4.1   The trilaminar appearance of membranes. (*a*) Electron micrograph showing the three-layered (trilaminar) structure of the plasma membrane of an erythrocyte after staining the tissue with the heavy metal osmium. Osmium binds preferentially to the polar head groups of the lipid bilayer, producing the trilaminar pattern. The arrows denote the inner and outer edges of the membrane. (*b*) The outer edge of a differentiated muscle cell grown in culture showing the similar trilaminar structure of both the plasma membrane (PM) and the membrane of the sarcoplasmic reticulum (SR), a calcium-storing compartment of the cytoplasm.   (A: COURTESY OF J. D. ROBERTSON; B: FROM ANDREW R. MARKS ET AL., J. CELL BIOL. 114:305, 1991; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

(b)

0.1 μm
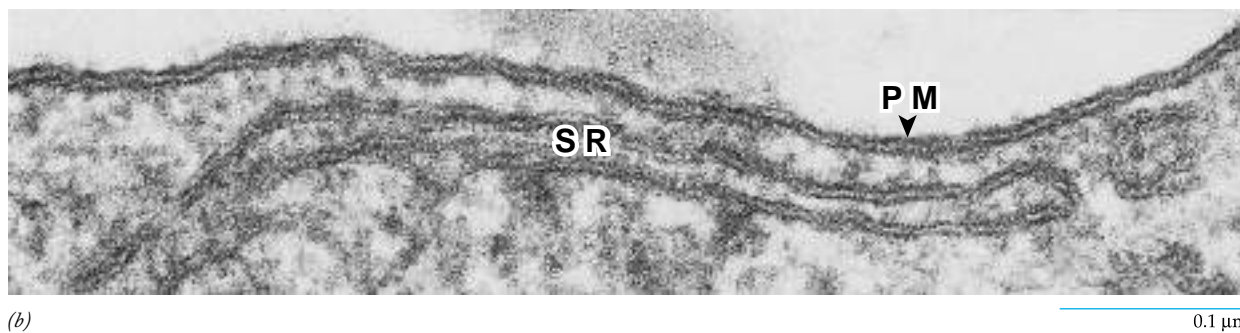
heart of the subject of membrane structure and function. As we will see shortly, cell membranes contain a lipid bilayer, and the two dark-staining layers in the electron micrographs of Figure 4.1 correspond primarily to the inner and outer polar surfaces of the bilayer (equivalent to the yellow atoms of the chapter-opening image). We will return to the structure of membranes below, but first we will survey some of the major functions of membranes in the life of a cell (Figure 4.2). ■

4.1 AN OVERVIEW
OF MEMBRANE FUNCTIONS

1. ***Compartmentalization.*** Membranes are continuous, unbroken sheets and, as such, inevitably enclose compartments. The plasma membrane encloses the contents of the entire cell, whereas the nuclear and cytoplasmic membranes enclose diverse intracellular spaces. The various membrane-bounded compartments of a cell possess markedly different contents. Membrane compartmentalization allows specialized activities to proceed without external interference and enables cellular activities to be regulated independently of one another.

2. ***Scaffold for biochemical activities.*** Membranes not only enclose compartments but are also a distinct compartment themselves. As long as reactants are present in solu-

tion, their relative positions cannot be stabilized and their interactions are dependent on random collisions. Because of their construction, membranes provide the cell with an extensive framework or scaffolding within which components can be ordered for effective interaction.

3. ***Providing a selectively permeable barrier.*** Membranes prevent the unrestricted exchange of molecules from one side to the other. At the same time, membranes provide the means of communication between the compartments they separate. The plasma membrane, which encircles a cell, can be compared to a moat around a castle: both serve as a general barrier, yet both have gated "bridges" that promote the movement of select elements into and out of the enclosed living space.

4. ***Transporting solutes.*** The plasma membrane contains the machinery for physically transporting substances from one side of the membrane to another, often from a region where the solute is present at low concentration into a region where that solute is present at much higher concentration. The membrane's transport machinery allows a cell to accumulate substances, such as sugars and amino acids, that are necessary to fuel its metabolism and build its macromolecules. The plasma membrane is also able to transport specific ions, thereby establishing ionic gradients across itself. This capability is especially critical for nerve and muscle cells.
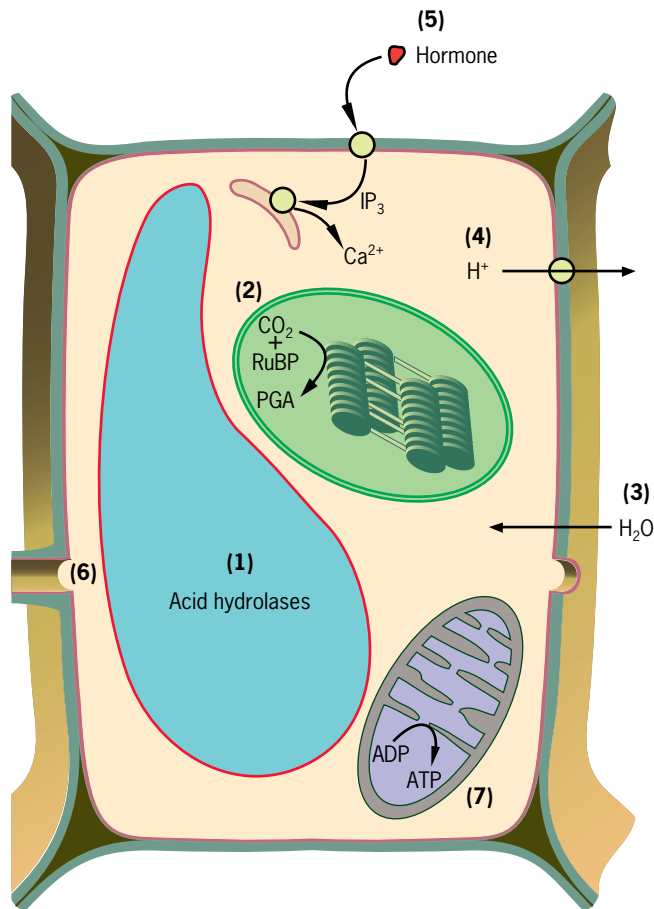
FIGURE 4.2 A summary of membrane functions in a plant cell. (1) An example of membrane compartmentalization in which hydrolytic enzymes (acid hydrolases) are sequestered within the membrane-bounded vacuole. (2) An example of the role of cytoplasmic membranes as a site of enzyme localization. The fixation of CO_2 by the plant cell is catalyzed by an enzyme that is associated with the outer surface of the thylakoid membranes of the chloroplasts. (3) An example of the role of membranes as a selectively permeable barrier. Water molecules are able to penetrate rapidly through the plasma membrane, causing the plant cell to fill out the available space and exert pressure against its cell wall. (4) An example of solute transport. Hydrogen ions, which are produced by various metabolic processes in the cytoplasm, are pumped out of plant cells into the extracellular space by a transport protein located in the plasma membrane. (5) An example of the involvement of a membrane in the transfer of information from one side to another (signal transduction). In this case, a hormone (e.g., abscisic acid) binds to the outer surface of the plasma membrane and triggers the release of a chemical message (such as IP_3) into the cytoplasm. In this case, IP_3 causes release of Ca^{2+} ions from a cytoplasmic warehouse. (6) An example of the role of membranes in cell–cell communication. Openings between adjoining plant cells, called plasmodesmata, allow materials to move directly from the cytoplasm of one cell into its neighbors. (7) An example of the role of membranes in energy transduction. The conversion of ADP to ATP occurs in close association with the inner membrane of the mitochondrion.

5. *Responding to external signals.* The plasma membrane plays a critical role in the response of a cell to external stimuli, a process known as **signal transduction**. Membranes possess **receptors** that combine with specific molecules (or **ligands**) having a complementary structure. Different types of cells have membranes with different receptors and are, therefore, capable of recognizing and responding to different ligands in their environment. The interaction of a plasma membrane receptor with an external ligand may cause the membrane to generate a signal that stimulates or inhibits internal activities. For example, signals generated at the plasma membrane may tell a cell to manufacture more glycogen, to prepare for cell division, to move toward a higher concentration of a particular compound, to release calcium from internal stores, or possibly to commit suicide.

6. *Intercellular interaction.* Situated at the outer edge of every living cell, the plasma membrane of multicellular organisms mediates the interactions between a cell and its neighbors. The plasma membrane allows cells to recognize and signal one another, to adhere when appropriate, and to exchange materials and information.

7. *Energy transduction.* Membranes are intimately involved in the processes by which one type of energy is converted to another type (energy transduction). The most fundamental energy transduction occurs during photosynthesis when energy in sunlight is absorbed by membrane-bound pigments, converted into chemical energy, and stored in carbohydrates. Membranes are also involved in the transfer of chemical energy from carbohydrates and fats to ATP. In eukaryotes, the machinery for these energy conversions is contained within membranes of chloroplasts and mitochondria.

We will concentrate in this chapter on the structure and functions of the plasma membrane, but remember that the principles discussed here are common to all cell membranes. Specialized aspects of the structure and functions of mitochondrial, chloroplast, cytoplasmic, and nuclear membranes will be discussed in Chapters 5, 6, 8, and 12, respectively.

4.2 A BRIEF HISTORY OF STUDIES ON PLASMA MEMBRANE STRUCTURE

The first insights into the chemical nature of the outer boundary layer of a cell were obtained by Ernst Overton of the University of Zürich during the 1890s. Overton knew that nonpolar solutes dissolved more readily in nonpolar solvents than in polar solvents, and that polar solutes had the opposite solubility. Overton reasoned that a substance entering a cell from the medium would first have to dissolve in the outer boundary layer of that cell. To test the permeability of the

outer boundary layer, Overton placed plant root hairs into hundreds of different solutions containing a diverse array of solutes. He discovered that the more lipid-soluble the solute, the more rapidly it would enter the root hair cells (see p. 144). He concluded that the dissolving power of the outer boundary layer of the cell matched that of a fatty oil.

The first proposal that cellular membranes might contain a lipid bilayer was made in 1925 by two Dutch scientists, E. Gorter and F. Grendel. These researchers extracted the lipid from human red blood cells and measured the amount of surface area the lipid would cover when spread over the surface of water (Figure 4.3*a*). Since mature mammalian red blood cells lack both nuclei and cytoplasmic organelles, the plasma membrane is the only lipid-containing structure in the cell. Consequently, all of the lipids extracted from the cells can be assumed to have resided in the cells' plasma membranes. The ratio of the surface area of water covered by the extracted lipid to the surface area calculated for the red blood cells from which the lipid was extracted varied between 1.8 to 1 and 2.2 to 1. Gorter and Grendel speculated that the actual ratio was 2:1 and concluded that the plasma membrane contained a bimolecular layer of lipids, that is, a **lipid bilayer** (Figure 4.3*b*). They also suggested that the polar groups of each molecular layer (or *leaflet*) were directed outward toward the aqueous environment, as shown in Figure 4.3*b,c*. This would be the thermodynamically favored arrangement, because the polar head groups of the lipids could interact with surrounding water molecules, just as the hydrophobic fatty acyl chains would be

protected from contact with the aqueous environment (Figure 4.3*c*). Thus, the polar head groups would face the cytoplasm on one edge and the blood plasma on the other. Even though Gorter and Grendel made several experimental errors (which fortuitously canceled one another out), they still arrived at the correct conclusion that membranes contain a lipid bilayer.

In the 1920s and 1930s, cell physiologists obtained evidence that there must be more to the structure of membranes than simply a lipid bilayer. It was found, for example, that lipid solubility was not the sole determining factor as to whether or not a substance could penetrate the plasma membrane. Similarly, the surface tensions of membranes were calculated to be much lower than those of pure lipid structures. This decrease in surface tension could be explained by the presence of protein in the membrane. In 1935, Hugh Davson and James Danielli proposed that the plasma membrane was composed of a lipid bilayer that was lined on both its inner and outer surface by a layer of globular proteins. They revised their model in the early 1950s to account for the selective permeability of the membranes they had studied. In the revised version (Figure 4.4*a*), Davson and Danielli suggested that, in addition to the outer and inner protein layers, the lipid bilayer was also

(*b*)

(*c*)
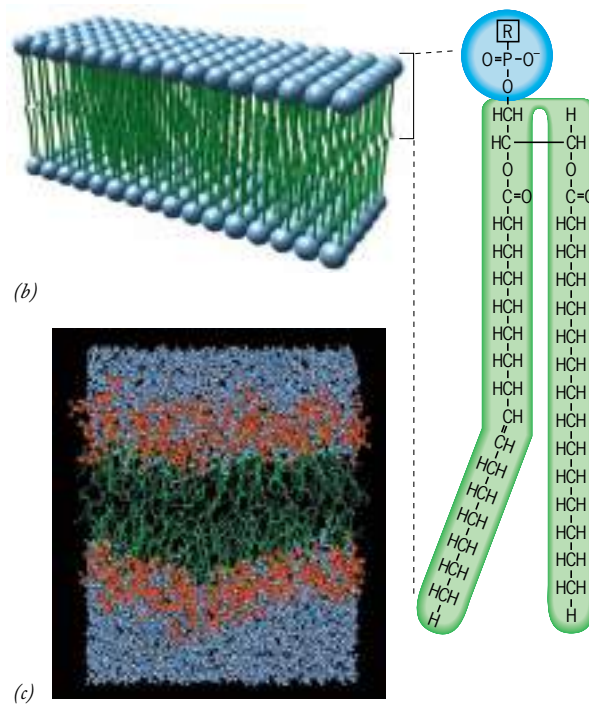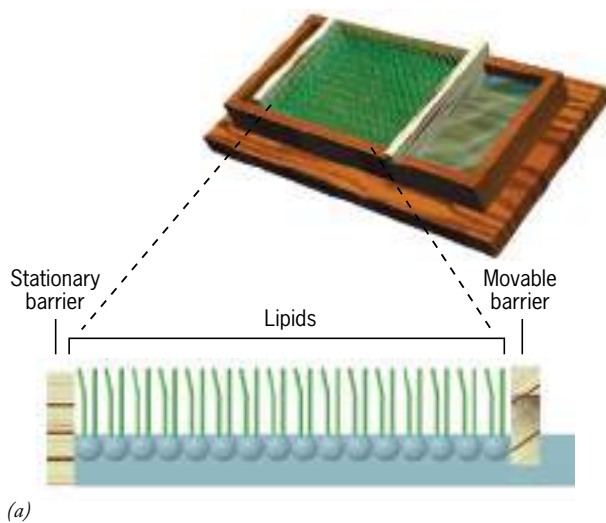
FIGURE 4.3 **The plasma membrane contains a lipid bilayer.** (*a*) Calculating the surface area of a lipid preparation. When a sample of phospholipids is dissolved in an organic solvent, such as hexane, and spread over an aqueous surface, the phospholipid molecules form a layer over the water that is a single molecule thick: a monomolecular layer. The molecules in the layer are oriented with their hydrophilic groups bonded to the surface of the water and their hydrophobic chains directed into the air. To estimate the surface area the lipids would cover if they were part of a membrane, the lipid molecules can be compressed into the smallest possible area by means of movable barriers. Using this type of apparatus, which is called a Langmuir trough after its inventor, Gorter

and Grendel concluded that red blood cells contained enough lipid to form a layer over their surface that was two molecules thick: a bilayer. (*b*) As Gorter and Grendel first proposed, the core of a membrane contains a bimolecular layer of phospholipids oriented with their water-soluble head groups facing the outer surfaces and their hydrophobic fatty acid tails facing the interior. The structures of the head groups are given in Figure 4.6*a*. (*c*) Simulation of a fully hydrated lipid bilayer composed of the phospholipid phosphatidylcholine. Phospholipid head groups are orange, water molecules are blue and white, fatty acid chains are green. (C: FROM S.-W. CHIU, TRENDS IN BIOCHEM. SCI. 22:341, 1997, COPYRIGHT 1997, WITH PERMISSION FROM ELSEVIER SCIENCE.)

penetrated by protein-lined pores, which could provide conduits for polar solutes and ions to enter and exit the cell.

Experiments conducted in the late 1960s led to a new concept of membrane structure, as detailed in the fluid-mosaic model proposed in 1972 by S. Jonathan Singer and Garth Nicolson of the University of California, San Diego (Figure 4.4*b*). In the **fluid-mosaic model**, which has served as the "central dogma" of membrane biology for more than three decades, the lipid bilayer remains the core of the membrane, but attention is focused on the physical state of the lipid. Unlike previous models, the bilayer of a fluid-mosaic membrane is present in a fluid state, and individual lipid molecules can move laterally within the plane of the membrane.

The structure and arrangement of membrane proteins in the fluid-mosaic model differ from those of previous models in that they occur as a "mosaic" of discontinuous particles that penetrate the lipid sheet (Figure 4.4*b*). Most importantly, the fluid-mosaic model presents cellular membranes as dynamic structures in which the components are mobile and capable of coming together to engage in various types of transient or semipermanent interactions. In the following sections, we will examine some of the evidence used to formulate and support this dynamic portrait of membrane structure and look at some of the recent data that bring the model up to date (Figure 4.4*c*).
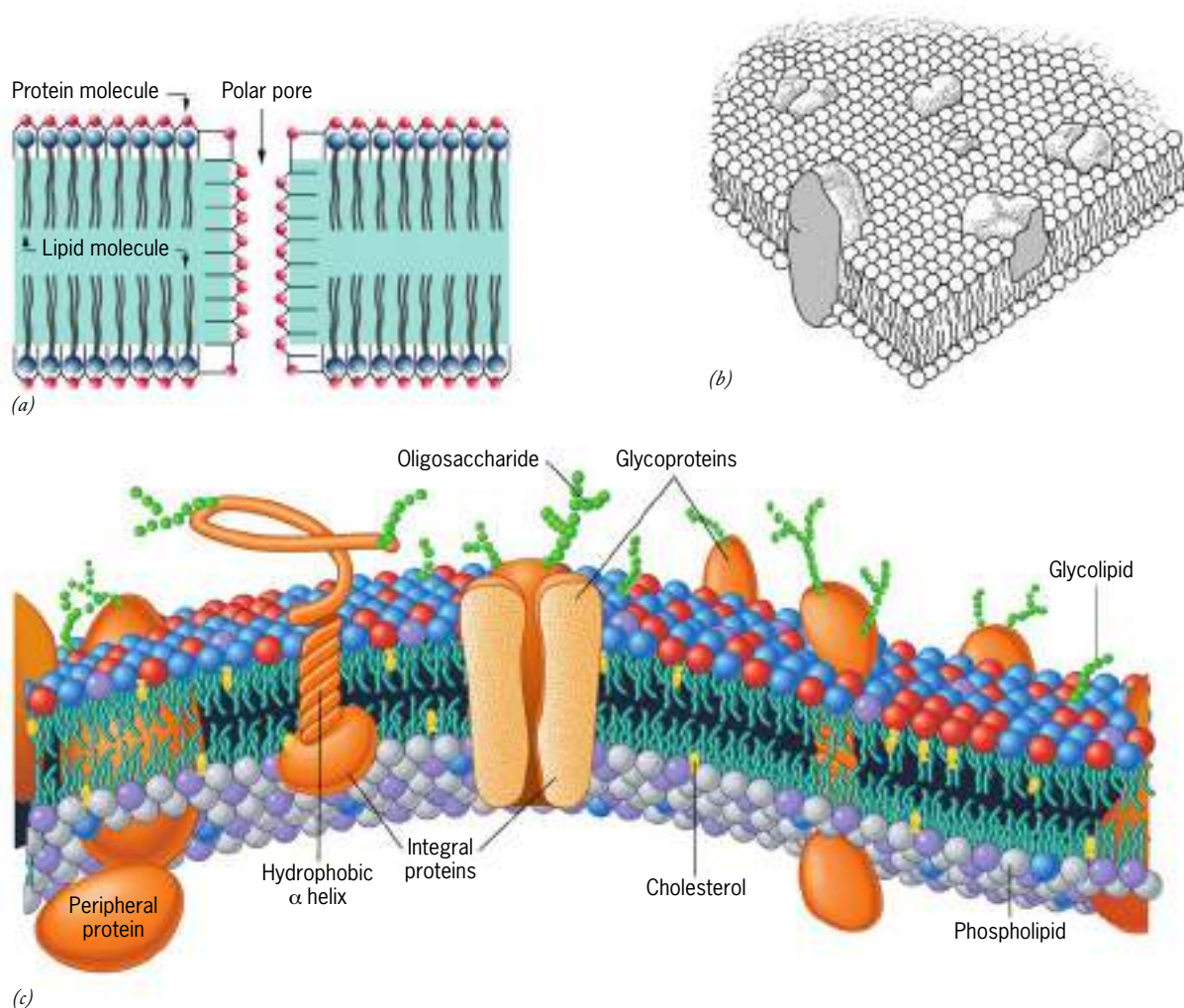
FIGURE 4.4 A brief history of the structure of the plasma membrane. (*a*) A revised 1954 version of the Davson-Danielli model showing the lipid bilayer, which is lined on both surfaces by a monomolecular layer of proteins that extends through the membrane to form protein-lined pores. (*b*) The fluid-mosaic model of membrane structure as initially proposed by Singer and Nicolson in 1972. Unlike previous models, the proteins penetrate the lipid bilayer. Although the original model shown here depicted a protein that was only partially embedded in the bilayer, lipid-penetrating proteins that have been studied span the entire bilayer. (*c*) A current representation of the plasma membrane showing the same basic organization as that proposed by Singer and Nicolson. The external surface of most membrane proteins, as well as a small percentage of the phospholipids, contain short chains of sugars, making them glycoproteins and glycolipids. Those portions of the polypeptide chains that extend through the lipid bilayer typically occur as α helices composed of hydrophobic amino acids. The two leaflets of the bilayer contain different types of lipids as indicated by the differently colored head groups. The outer leaflet may contain microdomains ("rafts") consisting of clusters of specific lipid species. (A: FROM J. F. DANIELLI, COLLSTON PAPERS 7:8, 1954; B: REPRINTED WITH PERMISSION FROM S. J. SINGER AND G. L. NICOLSON, SCIENCE 175:720, 1972; COPYRIGHT 1972, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

REVIEW **?**

1. Describe some of the important roles of membranes in the life of a eukaryotic cell. What do you think might be the effect of a membrane that was incapable of performing one or another of these roles?

2. Summarize some of the major steps leading to the current model of membrane structure. How does each new model retain certain basic principles of earlier models?

4.3 THE CHEMICAL COMPOSITION OF MEMBRANES

Membranes are lipid–protein assemblies in which the components are held together in a thin sheet by noncovalent bonds. As noted above, the core of the membrane consists of a sheet of lipids arranged in a bimolecular layer (Figure 4.3*b,c*). The lipid bilayer serves primarily as a structural backbone of the membrane and provides the barrier that prevents random movements of water-soluble materials into and out of the cell. The proteins of the membrane, on the other hand, carry out most of the specific functions summarized in Figure 4.2. Each type of differentiated cell contains a unique complement of membrane proteins, which contributes to the specialized activities of that cell type (see Figure 4.32*d* for an example).

The ratio of lipid to protein in a membrane varies, depending on the type of cellular membrane (plasma vs. endoplasmic reticulum vs. Golgi), the type of organism (bacterium vs. plant vs. animal), and the type of cell (cartilage vs. muscle vs. liver). For example, the inner mitochondrial membrane has a very high ratio of protein/lipid in comparison to the red blood cell plasma membrane, which is high in comparison to the membranes of the myelin sheath that form a multilayered wrapping around a nerve cell (Figure 4.5). To a large degree, these differences can be correlated with the basic functions of these membranes. The inner mitochondrial membrane contains the protein carriers of the electron-transport chain, and relative to other membranes, lipid is diminished. In contrast, the myelin sheath acts primarily as electrical insulation for the nerve cell it encloses, a function that is best carried out by a thick lipid layer of high electrical resistance with a minimal content of protein. Membranes also contain carbohydrates, which are attached to the lipids and proteins as indicated in Figure 4.4*c*.

Membrane Lipids

Membranes contain a wide diversity of lipids, all of which are **amphipathic**; that is, they contain both hydrophilic and hydrophobic regions. There are three main types of membrane lipids: phosphoglycerides, sphingolipids, and cholesterol.

Phosphoglycerides Most membrane lipids contain a phosphate group, which makes them **phospholipids**. Because most membrane phospholipids are built on a glycerol backbone, they are called **phosphoglycerides** (Figure 4.6*a*). Unlike
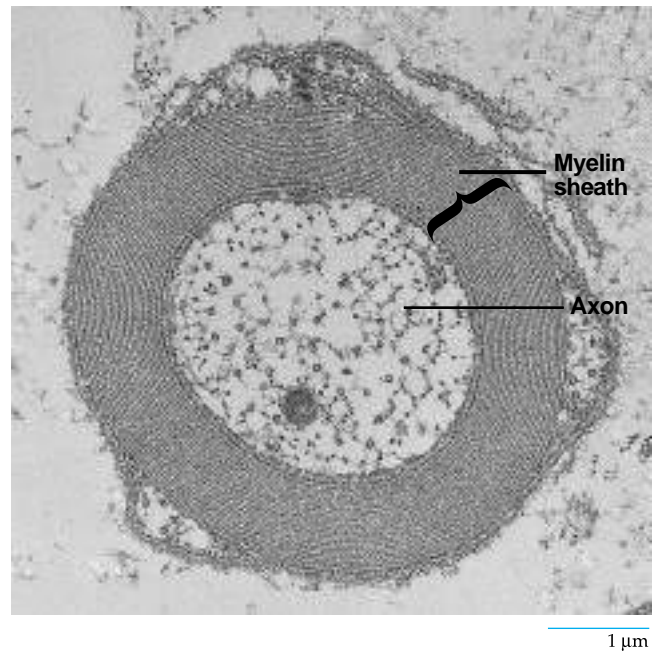
FIGURE 4.5 The myelin sheath. Electron micrograph of a nerve cell axon surrounded by a myelin sheath consisting of concentric layers of plasma membrane that have an extremely low protein/lipid ratio. The myelin sheath insulates the nerve cell from the surrounding environment, which increases the velocity at which impulses can travel along the axon (discussed on page 162). The perfect spacing between the layers is maintained by interlocking protein molecules (called P_0) that project from each membrane. (FROM LEONARD NAPOLITANO, FRANCIS LEBARON, AND JOSEPH SCALETTI, J. CELL BIOL. 34:820, 1967; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

triglycerides, which have three fatty acids (page 47) and are not amphipathic, membrane glycerides are *diglycerides*—only two of the hydroxyl groups of the glycerol are esterified to fatty acids; the third is esterified to a hydrophilic phosphate group. Without any additional substitutions beyond the phosphate and the two fatty acyl chains, the molecule is called *phosphatidic acid*, which is virtually absent in most membranes. Instead, membrane phosphoglycerides have an additional group linked to the phosphate, most commonly either choline (forming *phosphatidylcholine*, PC), ethanolamine (forming *phosphatidylethanolamine*, PE), serine (forming *phosphatidylserine*, PS), or inositol (forming *phosphatidylinositol*, PI). Each of these groups is small and hydrophilic and, together with the negatively charged phosphate to which it is attached, forms a highly water-soluble domain at one end of the molecule, called the **head group**. At physiologic pH, the head groups of PS and PI have an overall negative charge, whereas those of PC and PE are neutral. In contrast, the fatty acyl chains are hydrophobic, unbranched hydrocarbons approximately 16 to 22 carbons in length (Figure 4.6). A membrane fatty acid may be fully saturated (i.e., lack double bonds), monounsaturated (i.e., possess one double bond), or polyunsaturated (i.e., possess more than one double bond). Phosphoglycerides often contain one unsaturated and one saturated fatty acyl chain.

Recent interest has focused on the apparent health benefits of two highly unsaturated fatty acids (EPA and DHA) found at high concentration in fish oil. EPA and DHA contain five and six double bonds, respectively, and are incorporated primarily into PE and PC molecules of certain membranes, most notably in the brain and retina. EPA and DHA are described as omega-3 fatty acids because their last double bond is situated 3 carbons from the omega (CH_3) end of the fatty acyl chain. With fatty acid chains at one end of the molecule and a polar head group at the other end, all of the phosphoglycerides exhibit a distinct amphipathic character.

Sphingolipids   A less abundant class of membrane lipids, called **sphingolipids**, are derivatives of sphingosine, an amino alcohol that contains a long hydrocarbon chain (Figure 4.6b). Sphingolipids consist of sphingosine linked to a fatty acid (R of Figure 4.6b) by its amino group. This molecule is a *ceramide*. The various sphingosine-based lipids have additional groups esterified to the terminal alcohol of the sphingosine moiety. If the substitution is phosphorylcholine, the molecule is *sphingomyelin*, which is the only phospholipid of the membrane that is not built with a glycerol backbone. If the substitution is a carbohydrate, the molecule is a **glycolipid**. If the carbohydrate is a simple sugar, the glycolipid is called a *cerebroside*; if it is a small cluster of sugars, the glycolipid is called a *ganglioside*. Since all sphingolipids have two long, hydrophobic hydrocarbon chains at one end and a hydrophilic region at the other, they are also amphipathic and basically similar in overall structure to the phosphoglycerides.

Glycolipids are interesting membrane components. Relatively little is known about them, yet tantalizing hints have emerged to suggest they play crucial roles in cell function. The nervous system is particularly rich in glycolipids. The myelin sheath pictured in Figure 4.5 contains a high content of a particular glycolipid, called galactocerebroside (shown in Figure 4.6b), which is formed when a galactose is added to ceramide. Mice lacking the enzyme that carries out this reaction exhibit severe muscular tremors and eventual paralysis. Similarly, humans who are unable to synthesize a particular ganglioside (G_{M3}) suffer from a serious neurological disease characterized by severe seizures and blindness. Glycolipids also play a role in certain infectious diseases; the toxins that cause cholera and botulism both enter their target cell by first binding to cell-surface gangliosides, as does the influenza virus.

Cholesterol   Another lipid component of certain membranes is the sterol **cholesterol** (see Figure 2.21), which in certain animal cells may constitute up to 50 percent of the

FIGURE 4.6 The chemical structure of membrane lipids. (a) The structures of phosphoglycerides (see also Figure 2.22). (b) The structures of sphingolipids. Sphingomyelin is a phospholipid; cerebrosides and gangliosides are glycolipids. A third membrane lipid is cholesterol, which is shown in the next figure. (R = fatty acyl chain). [The green portion of each lipid, which represents the hydrophobic tail(s) of the molecule, is actually much longer than the hydrophilic head group (see Figure 4.23).]
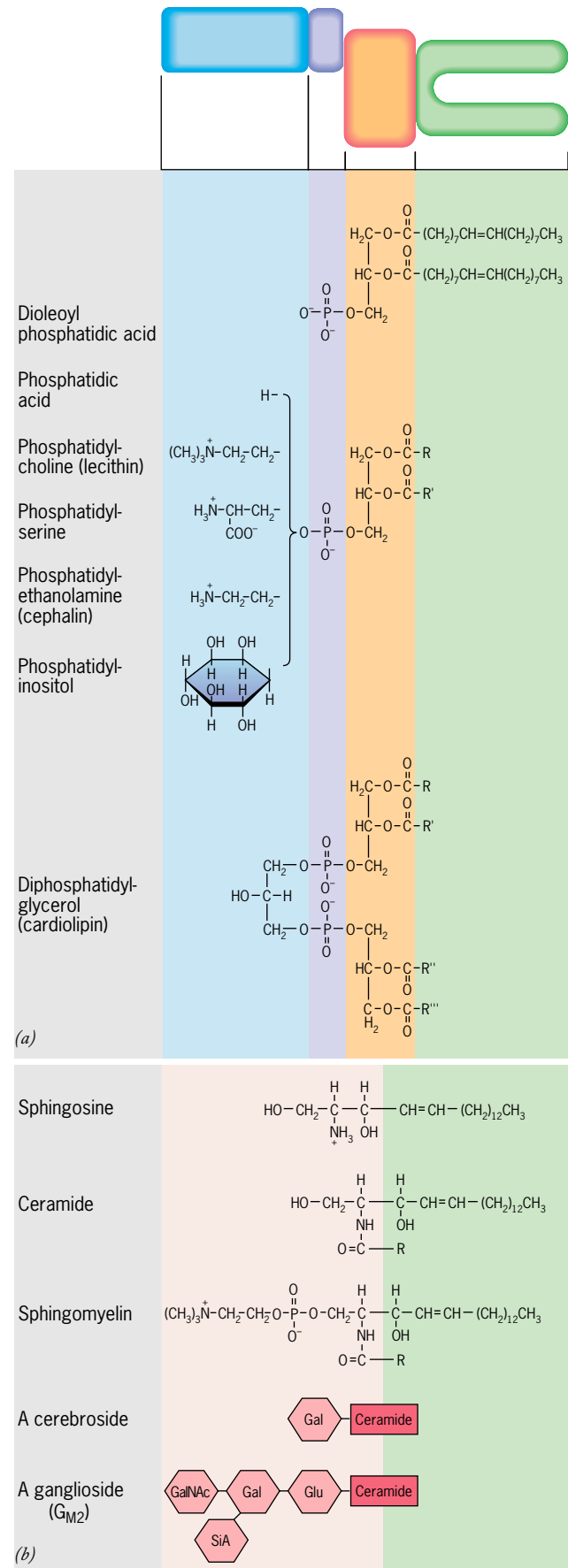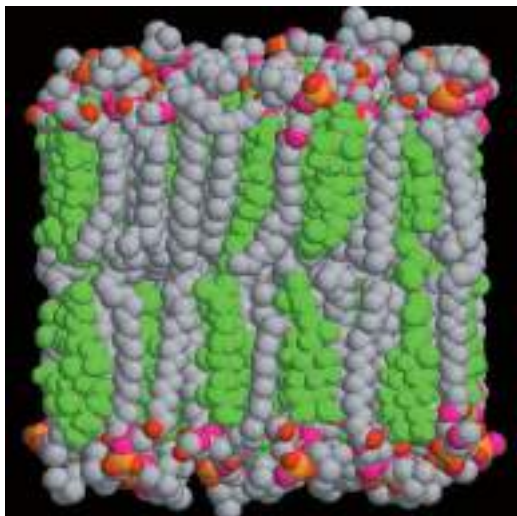
FIGURE 4.7 **The cholesterol molecules** (shown in green) of a lipid bilayer are oriented with their small hydrophilic end facing the external surface of the bilayer and the bulk of their structure packed in among the fatty acid tails of the phospholipids. The placement of cholesterol molecules interferes with the flexibility of the lipid hydrocarbon chains, which tends to stiffen the bilayer while maintaining its overall fluidity. Unlike other lipids of the membrane, cholesterol is often rather evenly distributed between the two layers (leaflets). (REPRINTED FROM H. L. SCOTT, CURR. OPIN. STRUCT. BIOL. 12:499, 2002, COPYRIGHT 2002, WITH PERMISSION FROM ELSEVIER SCIENCE.)

lipid molecules in the plasma membrane. Cholesterol is absent from the plasma membranes of most plant and all bacterial cells. Cholesterol molecules are oriented with their small hydrophilic hydroxyl group toward the membrane surface and the remainder of the molecule embedded in the lipid bilayer (Figure 4.7). The hydrophobic rings of a cholesterol molecule are flat and rigid, and they interfere with the movements of the fatty acid tails of the phospholipids (page 134).

The Nature and Importance of the Lipid Bilayer   Each type of cellular membrane has its own characteristic lipid composition, differing from one another in the types of lipids, the nature of the head groups, and the particular species of fatty acyl chain(s). Because of this structural variability, it is estimated that some biological membranes contain hundreds of chemically distinct species of phospholipid. The role of this remarkable diversity of lipid species remains the subject of interest and speculation.

   The percentages of some of the major types of lipids of a variety of membranes are given in Table 4.1. The lipids of a membrane are more than simple structural elements; they can have important effects on the biological properties of a membrane. Lipid composition can determine the physical state of the membrane (page 134) and influence the activity of particular membrane proteins. Membrane lipids also provide the precursors for highly active chemical messengers that regulate cellular function (Section 15.3).

   Various types of measurements indicate that the combined fatty acyl chains of both leaflets of the lipid bilayer span a

width of about 30 Å and that each row of head groups (with its adjacent shell of water molecules) adds another 15 Å (refer to the chapter-opening illustration on page 117). Thus, the entire lipid bilayer is only about 60 Å (6 nm) thick. The presence in membranes of this thin film of amphipathic lipid molecules has remarkable consequences for cell structure and function. Because of thermodynamic considerations, the hydrocarbon chains of the lipid bilayer are never exposed to the surrounding aqueous solution. Consequently, membranes are never seen to have a free edge; they are always continuous, unbroken structures. As a result, membranes form extensive interconnected networks within the cell. Because of the flexibility of the lipid bilayer, membranes are deformable and their overall shape can change, as occurs during locomotion (Figure 4.8*a*) or cell division (Figure 4.8*b*). The lipid bilayer is thought to facilitate the regulated fusion or budding of membranes. For example, the events of secretion, in which cytoplasmic vesicles fuse to the plasma membrane, or of fertilization, where two cells fuse to form a single cell (Figure 4.8*c*), involve processes in which two separate membranes come together to become one continuous sheet (see Figure 8.32). The importance of the lipid bilayer in maintaining the proper internal composition of a cell, in separating electric charges across the plasma membrane, and in many other cellular activities will be apparent throughout this and subsequent chapters.

   Another important feature of the lipid bilayer is its ability to self-assemble, which can be demonstrated more easily within a test tube than a living cell. If, for example, a small amount of phosphatidylcholine is dispersed in an aqueous solution, the phospholipid molecules assemble spontaneously to form the walls of fluid-filled spherical vesicles, called **liposomes**. The walls of these liposomes consist of a continuous lipid bilayer that is organized in the same manner as that of the lipid bilayer of a natural membrane. Liposomes have proven invaluable in membrane research. Membrane proteins can be inserted into liposomes and their function studied in a much simpler environment than that of a natural membrane. Liposomes have also been developed as vehicles to deliver drugs or DNA molecules within the body.

TABLE 4.1 **Lipid Compositions of Some Biological Membranes***

Lipid	Human erythrocyte	Human myelin	Beef heart mitochondria	E. coli
Phosphatidic acid	1.5	0.5	0	0
Phosphatidylcholine	19	10	39	0
Phosphatidyl- ethanolamine	18	20	27	65
Phosphatidylgycerol	0	0	0	18
Phosphatidylserine	8.5	8.5	0.5	0
Cardiolipin	0	0	22.5	12
Sphingomyelin	17.5	8.5	0	0
Glycolipids	10	26	0	0
Cholesterol	25	26	3	0

*The values given are weight percent of total lipid.
Source: C. Tanford, *The Hydrophobic Effect*, p. 109, copyright 1980, John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.

(a)

(b)

(c)
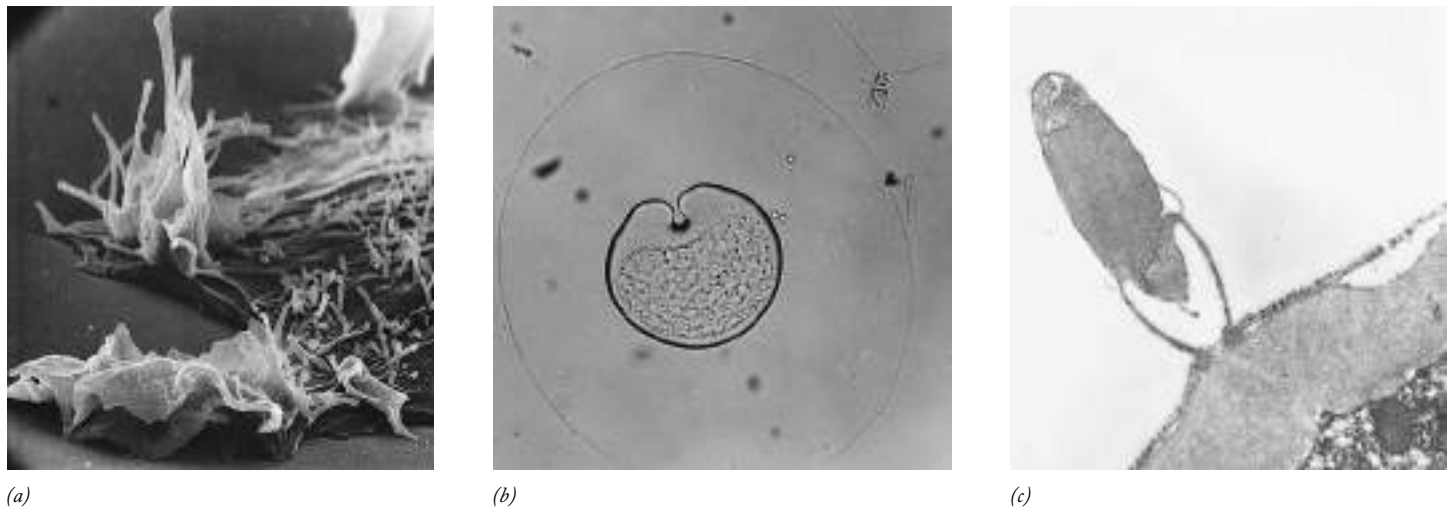
FIGURE 4.8 The dynamic properties of the plasma membrane. (*a*) The leading edge of a moving cell often contains sites where the plasma membrane displays undulating ruffles. (*b*) Division of a cell is accompanied by the deformation of the plasma membrane as it is pulled toward the center of the cell. Unlike most dividing cells, the cleavage furrow of this dividing ctenophore egg begins at one pole and moves unidirectionally through the egg. (*c*) Membranes are capable of fusing with other membranes. This sperm and egg are in a stage leading to the fusion of their plasma membranes. (A: COURTESY OF JEAN PAUL REVEL; B: COURTESY OF GARY FREEMAN; C: COURTESY OF A. L. COLWIN AND L. H. COLWIN.)

The drugs or DNA can be linked to the wall of the liposome or contained at high concentration within its lumen (Figure 4.9). In these studies, the walls of the liposomes are constructed to contain specific proteins (such as antibodies or hormones) that allow the liposomes to bind selectively to the surfaces of particular target cells where the drug or DNA is intended to go. Most of the early clinical studies with liposomes met with failure because the injected vesicles were rapidly removed by phagocytic cells of the immune system. This obstacle has been overcome with the development of so-called stealth liposomes (e.g., Caelyx) that contain an outer coating of a synthetic polymer that protects the liposomes from immune destruction (Figure 4.9).

The Asymmetry of Membrane Lipids

The lipid bilayer consists of two distinct leaflets that have a distinctly different lipid composition. One line of experiments that has led to this conclusion takes advantage of the fact that lipid-digesting enzymes cannot penetrate the plasma membrane and, consequently, are only able to digest lipids that reside in the outer leaflet of the bilayer. If intact human red blood cells are treated with a lipid-digesting phospholipase, approximately 80 percent of the phosphatidylcholine (PC) of the membrane is hydrolyzed, but only about 20 percent of the membrane's phosphatidylethanolamine (PE) and less than 10 percent of its phosphatidylserine (PS) are attacked. These data indicate that, compared to the inner leaflet, the outer leaflet has a relatively high concentration of PC (and sphingomyelin) and a low concentration of PE and PS (Figure 4.10). It follows that the lipid bilayer can be thought of as composed of two more-or-less stable, independent monolayers having different physical and chemical properties.

The different classes of lipids in Figure 4.10 exhibit different properties. All the glycolipids of the plasma membrane are in the outer leaflet where they often serve as receptors for extracellular ligands. Phosphatidylethanolamine, which is concentrated in the inner leaflet, tends to promote the curvature of the membrane, which is important in membrane budding and fusion. Phosphatidylserine, which is concentrated in the inner leaflet, has a net negative charge at physiologic pH, which makes it a candidate for binding positively charged lysine and arginine residues, such as those adjacent to the membrane-spanning α helix of glycophorin A
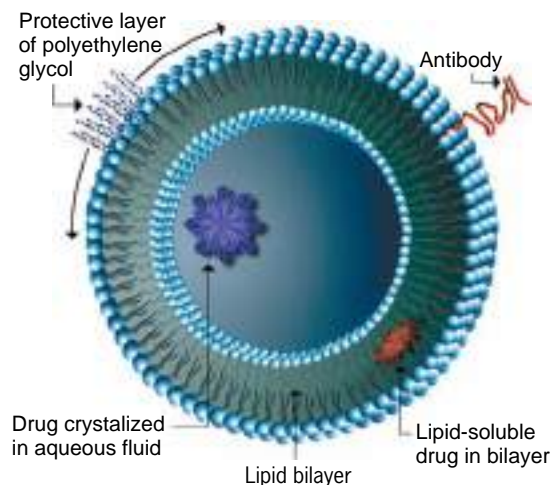
FIGURE 4.9 Liposomes. A schematic diagram of a stealth liposome containing a hydrophilic polymer (such as polyethylene glycol) to protect it from destruction by immune cells, antibody molecules that target it to specific body tissues, a water-soluble drug enclosed in the fluid-filled interior chamber, and a lipid-soluble drug in the bilayer.
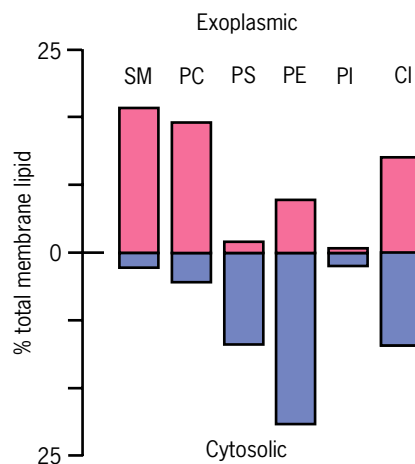
FIGURE 4.10 The asymmetric distribution of phospholipids (and cholesterol) in the plasma membrane of human erythrocytes. (SM, sphingomyelin; PC, phosphatidylcholine; PS, phosphatidylserine; PE, phosphatidylethanolamine; PI, phosphatidylinositol; Cl, cholesterol.)

FIGURE 4.11 Two types of linkages that join sugars to a polypeptide chain. The *N*-glycosidic linkage between asparagine and *N*-acetylglucosamine is more common than the *O*-glycosidic linkage between serine or threonine and *N*-acetylgalactosamine.

in Figure 4.18. The appearance of PS on the outer surface of aging lymphocytes marks the cells for destruction by macrophages, whereas its appearance on the outer surface of platelets leads to blood coagulation. Phosphatidylinositol, which is concentrated in the inner leaflet, plays a key role in the transfer of stimuli from the plasma membrane to the cytoplasm (Section 15.3).

Membrane Carbohydrates

The plasma membranes of eukaryotic cells also contain carbohydrate. Depending on the species and cell type, the carbohydrate content of the plasma membrane ranges between 2 and 10 percent by weight. More than 90 percent of the membrane's carbohydrate is covalently linked to proteins to form glycoproteins; the remaining carbohydrate is covalently linked to lipids to form glycolipids, which were discussed on page 123. As indicated in Figure 4.4c, all of the carbohydrate of the plasma membrane faces outward into the extracellular space.[1] The carbohydrate of internal cellular membranes also faces away from the cytosol (the basis for this orientation is illustrated in Figure 8.14).

The modification of proteins was discussed briefly on page 53. The addition of carbohydrate, or **glycosylation**, is the most complex of these modifications. The carbohydrate of glycoproteins is present as short, branched hydrophilic **oligosaccharides**, typically having fewer than about 15 sugars per chain. In contrast to most high-molecular-weight carbohydrates (such as glycogen, starch, or cellulose), which are polymers of a single sugar, the oligosaccharides attached to membrane proteins and lipids can display considerable variability in composition and structure. Oligosaccharides may be

attached to several different amino acids by two major types of linkages (Figure 4.11). These carbohydrate projections play an important role in mediating the interactions of a cell with its environment (Chapter 7) and sorting of membrane proteins to different cellular compartments (Chapter 8). The carbohydrates of the glycolipids of the red blood cell plasma membrane determine whether a person's blood type is A, B, AB, or O (Figure 4.12). A person having blood type A has an enzyme that adds an *N*-acetylgalactosamine to the end of the chain, whereas a person with type B blood has an enzyme that adds galactose to the chain terminus. These two enzymes are encoded by alternate versions of the same gene, yet they recognize different substrates. People with AB blood type possess both enzymes, whereas people with O blood type lack enzymes capable of attaching either terminal sugar. The function of the ABO blood-group antigens remains a mystery.

REVIEW ?

1. Draw the basic structure of the major types of lipids found in cellular membranes. How do sphingolipids differ from glycerolipids? Which lipids are phospholipids? Which are glycolipids? How are these lipids organized into a bilayer? How is the bilayer important for membrane activities?

2. What is a liposome? How are liposomes used in medical therapies?

3. What is an oligosaccharide? How are they linked to membrane proteins? How are they related to human blood types?

[1]It can be noted that even though phosphatidylinositol contains a sugar group (Figure 4.6), it is not considered to be part of the carbohydrate portion of the membrane in this discussion.
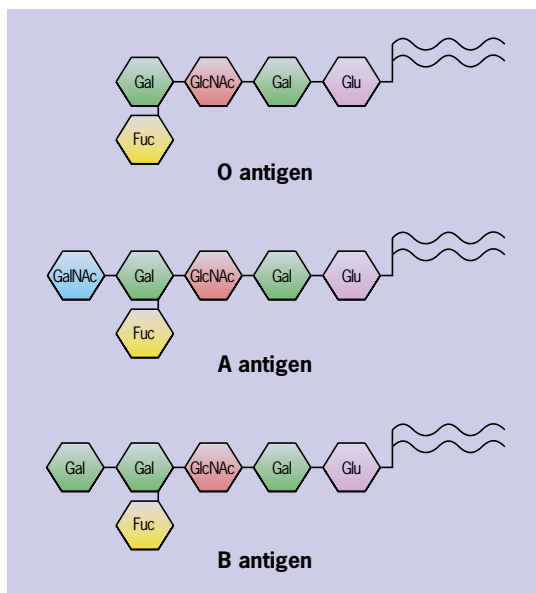
FIGURE 4.12 Blood-group antigens. Whether a person has type A, B, AB, or O blood is determined by a short chain of sugars covalently attached to membrane lipids and proteins of the red blood cell membrane. The oligosaccharides attached to membrane lipids (forming a ganglioside) that produce the A, B, and O blood types are shown here. A person with type AB blood has gangliosides with both the A and B structure. (Gal, galactose; GlcNAc, *N*-acetylglucosamine; Glu, glucose; Fuc, fucose; GalNAc, *N*-acetylgalactosamine.)

4.4 THE STRUCTURE AND FUNCTIONS OF MEMBRANE PROTEINS

Depending on the cell type and the particular organelle within that cell, a membrane may contain hundreds of different proteins. Each membrane protein has a defined orientation relative to the cytoplasm, so that the properties of one surface of a membrane are very different from those of the other surface. This asymmetry is referred to as membrane "sidedness." In the plasma membrane, for example, those parts of membrane proteins that interact with other cells or with extracellular substances project outward into the extracellular space, whereas those parts of membrane proteins that interact with cytoplasmic molecules project into the cytosol. Membrane proteins can be grouped into three distinct classes distinguished by the intimacy of their relationship to the lipid bilayer (Figure 4.13). These are

1. **Integral proteins** that penetrate the lipid bilayer. Integral proteins are **transmembrane proteins**; that is, they pass entirely through the lipid bilayer and thus have domains that protrude from both the extracellular and cytoplasmic sides of the membrane. Some integral proteins have only one membrane-spanning segment, whereas others are multispanning. Genome-sequencing studies suggest that integral proteins constitute 20–30 percent of all encoded proteins.

2. **Peripheral proteins** that are located entirely outside of
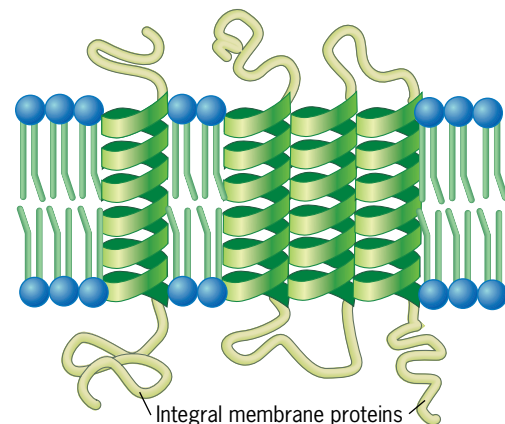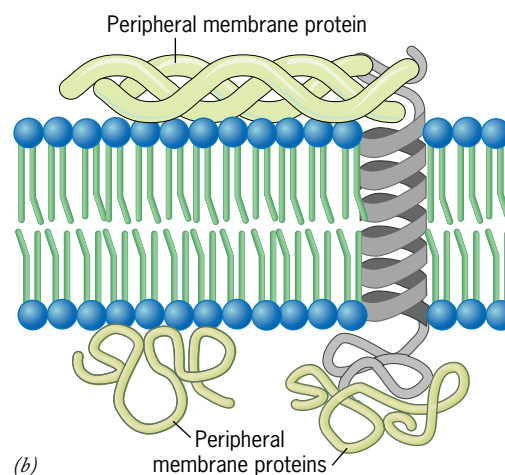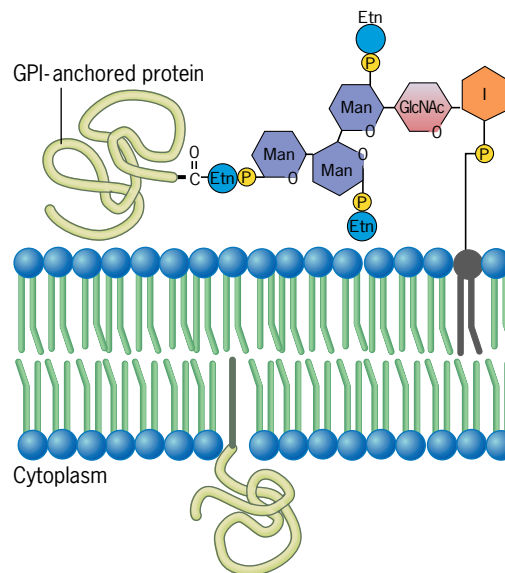
FIGURE 4.13 Three classes of membrane protein. (*a*) Integral proteins typically contain one or more transmembrane helices (see Figure 5.4 for an exception). (*b*) Peripheral proteins are noncovalently bonded to the polar head groups of the lipid bilayer and/or to an integral membrane protein. (*c*) Lipid-anchored proteins are covalently bonded to a lipid group that resides within the membrane. The lipid can be phosphatidylinositol, a fatty acid, or a prenyl group (a long-chain hydrocarbon built from five-carbon isoprenoid units). I, inositol; GlcNAc, *N*-acetylglucosamine; Man, mannose; Etn, ethanolamine; GPI, glycosylphosphatidylinositol.

the lipid bilayer, on either the cytoplasmic or extracellular side, yet are associated with the surface of the membrane by noncovalent bonds.

3. **Lipid-anchored proteins** that are located outside the lipid bilayer, on either the extracellular or cytoplasmic surface, but are covalently linked to a lipid molecule that is situated within the bilayer.

Integral Membrane Proteins

Most integral membrane proteins function in the following capacities: as receptors that bind specific substances at the membrane surface, as channels or transporters involved in the movement of ions and solutes across the membrane, or as agents that transfer electrons during the processes of photosynthesis and respiration. Like the phospholipids of the bilayer, integral membrane proteins are also amphipathic, having both hydrophilic and hydrophobic portions. As discussed below, those portions of an integral membrane protein that reside within the lipid bilayer tend to have a hydrophobic character. Amino acid residues in these transmembrane domains form van der Waals interactions with the fatty acyl chains of the bilayer, which seals the protein into the lipid "wall" of the membrane. As a result, the permeability barrier of the membrane is preserved and the protein is brought into
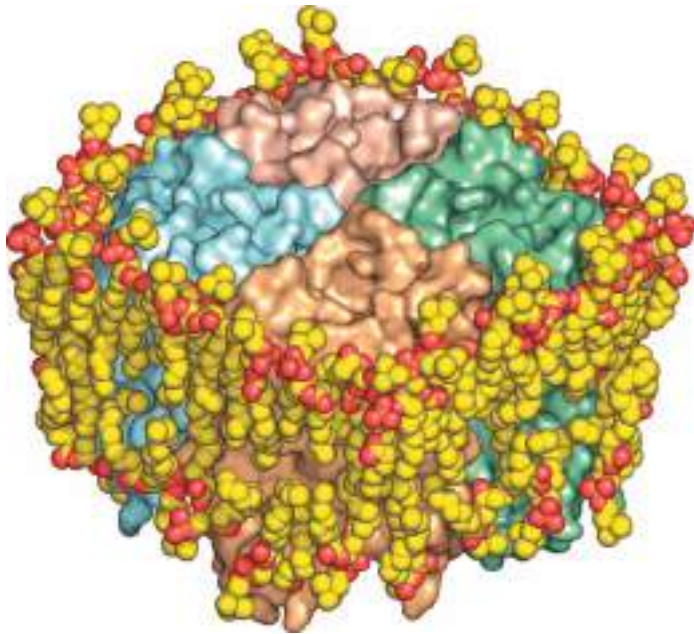
FIGURE 4.14 **Proteins can be surrounded by a closely applied shell of lipid molecules.** Aquaporin is a membrane protein containing four subunits (colored differently in the illustration) surrounding an aqueous channel. Analysis of the protein's structure revealed the presence of a surrounding layer of bound lipid molecules. Whether or not these lipid molecules affect the function of the aquaporin molecule is unclear, but it is likely that they are held in close proximity to the protein and thus unable to move freely within the bilayer. (FROM CAROLA HUNTE AND SEBASTIAN RICHERS, CURR. OPIN. STRUCT. BIOL. 18:407, 2008, COPYRIGHT 2008, WITH PERMISSION FROM ELSEVIER SCIENCE.)

direct contact with surrounding lipid molecules (Figure 4.14). Lipid molecules that are closely associated with a membrane protein can play an important role in the activity of the protein, although the degree to which a particular protein requires specific interactions with particular lipid molecules remains unclear.

Those portions of an integral membrane protein that project into either the cytoplasm or extracellular space tend to be more like the globular proteins discussed in Section 2.5. These nonembedded domains tend to have hydrophilic surfaces that interact with water-soluble substances (low-molecular-weight substrates, hormones, and other proteins) at the edge of the membrane. Several large families of membrane proteins contain an interior channel that provides an aqueous passageway through the lipid bilayer. The linings of these channels typically contain key hydrophilic residues at strategic locations. As will be discussed later, integral proteins need not be fixed structures but may be able to move laterally within the membrane.

Distribution of Integral Proteins: Freeze-Fracture Analysis

The concept that proteins penetrate through membranes, rather than simply remaining external to the bilayer, was derived primarily from the results of a technique called **freeze-fracture replication** (see Section 18.2). In this procedure, tissue is frozen solid and then struck with a knife blade, which fractures the block into two pieces. As this occurs, the fracture plane often takes a path between the two leaflets of the lipid bilayer (Figure 4.15*a*). Once the membranes are split in this manner, metals are deposited on their exposed surfaces to form a shadowed *replica*, which is viewed in the electron microscope (see Figure 18.17). As shown in Figure 4.15*b*, the replica resembles a road strewn with pebbles, which are called *membrane-associated particles*. Since the fracture plane passes through the center of the bilayer, most of these particles correspond to integral membrane proteins that extend at least halfway through the lipid core. When the fracture plane reaches a given particle, it goes around it rather than cracking it in half. Consequently, each protein (particle) separates with one half of the plasma membrane (Figure 4.15*c*), leaving a corresponding pit in the other half (see Figure 7.30*c*). One of the great values of the freeze-fracturing technique is that it allows an investigation of the microheterogeneity of the membrane. Localized differences in parts of the membrane stand out in these replicas and can be identified (as illustrated by the replica of a gap junction shown in Figure 7.32*d*). Biochemical analyses, in contrast, average out such differences.

Studying the Structure and Properties of Integral Membrane Proteins

Because of their hydrophobic transmembrane domains, integral membrane proteins are difficult to isolate in a soluble form. Removal of these proteins from the membrane normally requires the use of a detergent, such as the ionic (charged) detergent SDS (which denatures proteins) or the nonionic

FIGURE 4.16 Solubilization of membrane proteins with detergents. The nonpolar ends of the detergent molecules associate with the nonpolar residues of the protein that had previously been in contact with the fatty acyl chains of the lipid bilayer. In contrast, the polar ends of the detergent molecules interact with the surrounding water molecules, keeping the protein in solution. Nonionic detergents, as shown here, solubilize membrane proteins without disrupting their structure.

(a)

(b)   0.2 µm

(c)   0.3 µm

FIGURE 4.15 Freeze fracture: a technique for investigating cell membrane structure. (*a*) When a block of frozen tissue is struck by a knife blade, a fracture plane runs through the tissue, often following a path that leads it through the middle of the lipid bilayer. The fracture plane goes around the proteins rather than cracking them in half, and they segregate with one of the two halves of the bilayer. The exposed faces within the center of the bilayer can then be covered with a metal deposit to form a metallic replica. These exposed faces are referred to as the E, or ectoplasmic face, and the P, or protoplasmic face. (*b*) Replica of a freeze-fractured human erythrocyte. The P fracture face is seen to be covered with particles approximately 8 nm in diameter. A small ridge (arrow) marks the junction of the particulate face with the surrounding ice. (*c*) This micrograph shows the surface of an erythrocyte that was frozen and then fractured, but rather than preparing a replica, the cell was thawed, fixed, and labeled with a marker for the carbohydrate groups that project from the external surface of the integral protein glycophorin (Figure 4.18). Thin sections of the labeled, fractured cell reveal that glycophorin molecules (black particles) have preferentially segregated with the outer half of the membrane. The red line shows the path of the fracture plane. (B: FROM THOMAS W. TILLACK AND VINCENT T. MARCHESI, J. CELL BIOL. 45:649, 1970; C: FROM PEDRO PINTO DA SILVA AND MARIA R. TORRISI, J. CELL BIOL. 93:467, 1982; B,C: BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

(uncharged) detergent Triton X-100 (which generally does not alter a protein's tertiary structure).

$$CH_3{-}(CH_2)_{11}{-}OSO_3^-Na^+$$

Sodium dodecyl sulfate (SDS)

$$CH_3{-}\overset{\displaystyle CH_3}{\underset{\displaystyle CH_3}{C}}{-}CH_2{-}\overset{\displaystyle CH_3}{\underset{\displaystyle CH_3}{C}}{-}\langle \bigcirc \rangle{-}(O{-}CH_2{-}CH_2)_{10}{-}OH$$

Triton X-100
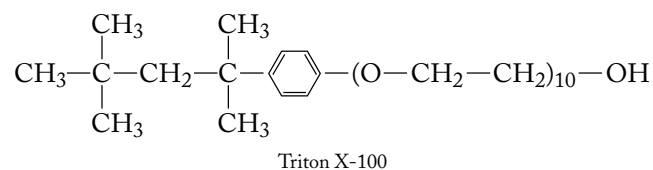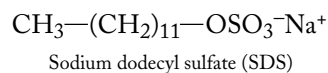
Like membrane lipids, detergents are amphipathic, being composed of a polar end and a nonpolar hydrocarbon chain (see Figure 2.20). As a consequence of their structure, detergents can substitute for phospholipids in stabilizing integral proteins while rendering them soluble in aqueous solution (Figure 4.16). Once the proteins have been solubilized by the detergent, various analyses can be carried out to determine the protein's amino acid composition, molecular mass, amino acid sequence, and so forth.

Researchers have had great difficulty obtaining crystals of most integral membrane proteins for use in X-ray crystallography. In fact, fewer than 1 percent of the known high-resolution protein structures represent integral membrane proteins (see http://blanco.biomol.uci.edu/MemPro_resources.html for a discussion of principles of protein structure and http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html for an updated gallery).[2] Furthermore, most of these structures represent prokaryotic versions of a particular protein, which are often smaller than their eukaryotic homologues and easier to obtain in large quantity. Once the structure of

---

[2]Many integral membrane proteins have a substantial portion that is present in the cytoplasm or extracellular space. In many cases, this soluble portion has been cleaved from its transmembrane domain, crystallized, and its tertiary structure determined. While this approach provides valuable data about the protein, it fails to provide information about the protein's orientation within the membrane. Another crystallographic approach to the study of membrane proteins, which uses the electron microscope rather than X-ray diffraction, is discussed in the Experimental Pathways on page 168.
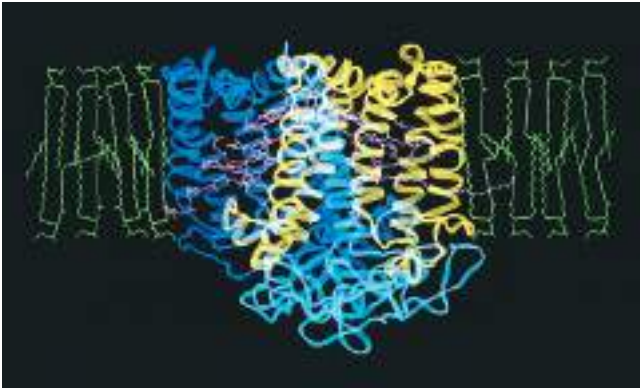
FIGURE 4.17 An integral protein as it resides within the plasma membrane. Tertiary structure of the photosynthetic reaction center of a bacterium as determined by X-ray crystallography. The protein contains three different membrane-spanning polypeptides, shown in yellow, light blue, and dark blue. The helical nature of each of the transmembrane segments is evident. (FROM G. FEHER, J. P. ALLEN, M. Y. OKAMURA, AND D. C. REES, REPRINTED WITH PERMISSION FROM NATURE 339:113, 1989; COPYRIGHT 1989, MACMILLAN MAGAZINES LIMITED.)

one member of a membrane protein family is determined, researchers can usually apply a strategy called *homology modeling* to learn about the structure and activity of other members of the family. For example, solution of the structure of the bacterial K$^+$ channel KcsA (shown in Figure 4.39) provided a wealth of data that could be applied to the structure and mechanism of action of the much more complex eukaryotic K$^+$ channels (Figure 4.42).

One of the first membrane proteins whose entire three-dimensional structure was determined by X-ray crystallography is shown in Figure 4.17. This protein—the bacterial photosynthetic reaction center—consists of three subunits

containing 11 membrane-spanning α helices. Some of the technical difficulties in preparing membrane protein crystals have been overcome using new methodologies and laborious efforts. In one study, for example, researchers were able to obtain high-quality crystals of a bacterial transporter after testing and refining more than 95,000 different conditions for crystallization. Despite increasing success in protein crystallization, researchers still rely largely on indirect approaches for determining the three-dimensional organization of most membrane proteins. We will examine some of these approaches in the following paragraphs.

Identifying Transmembrane Domains A great deal can be learned about the structure of a membrane protein and its orientation within the lipid bilayer from a computer-based (computational) analysis of its amino acid sequence, which is readily deduced from the nucleotide sequence of an isolated gene. The first question one might ask is: Which segments of the polypeptide chain are actually embedded in the lipid bilayer? Those segments of a protein embedded within the membrane, which are described as the **transmembrane domains**, have a simple structure; they consist of a string of about 20 predominantly nonpolar amino acids that span the core of the lipid bilayer as an α helix.[3] The chemical structure of a single transmembrane helix is shown in Figure 4.18, which depicts the two-dimensional structure of

---

[3]It was noted on page 54 that the α helix is a favored conformation because it allows for a maximum number of hydrogen bonds to be formed between neighboring amino acid residues, thereby creating a highly stable (low-energy) configuration. This is particularly important for a membrane-spanning polypeptide that is surrounded by fatty acyl chains and, thus, cannot form hydrogen bonds with an aqueous solvent. Transmembrane helices are at least 20 amino acids in length, because this is the minimum stretch of polypeptide capable of spanning the hydrocarbon core of a lipid bilayer of 30 Å width. A few integral membrane proteins have been found to contain loops or helices that penetrate but do not span the bilayer. An example is the P helix in Figure 4.39.
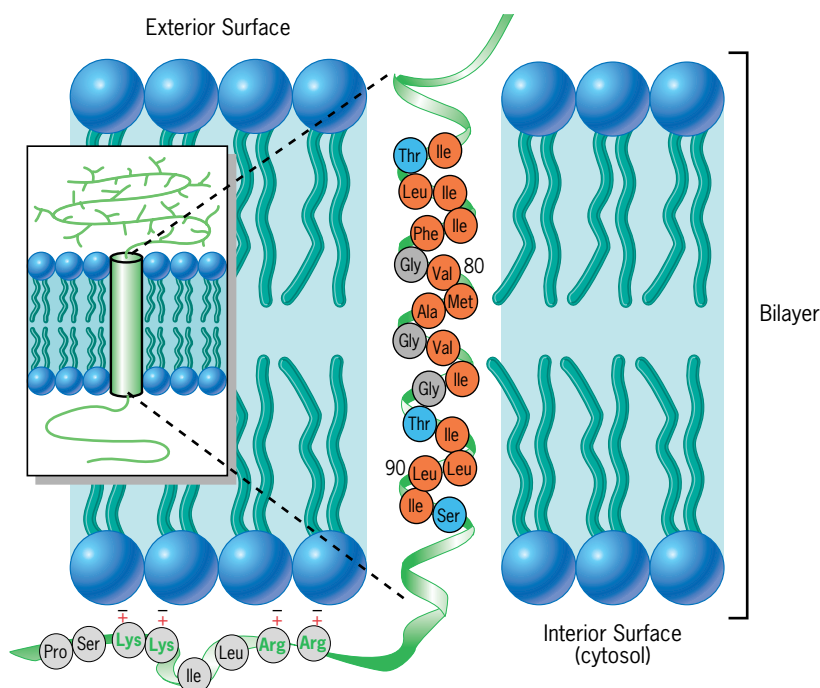
FIGURE 4.18 Glycophorin A, an integral protein with a single transmembrane domain. The single α helix that passes through the membrane consists predominantly of hydrophobic residues (orange-colored circles). The four positively charged amino acid residues of the cytoplasmic domain of the membrane form ionic bonds with negatively charged lipid head groups. Carbohydrates are attached to a number of amino acid residues on the outer surface of the protein (shown in the inset). All but one of the 16 oligosaccharides are small *O*-linked chains (the exception is a larger oligosaccharide linked to the asparagine residue at position 26). Glycophorin molecules are present as homodimers within the erythrocyte membrane (Figure 4.32*d*). The two helices cross over one another in the region between residues 79 and 83. This Gly-X-X-X-Gly sequence is commonly found where transmembrane helices come into close proximity.

(a)   *(b)*   *(c)*   *(d)*
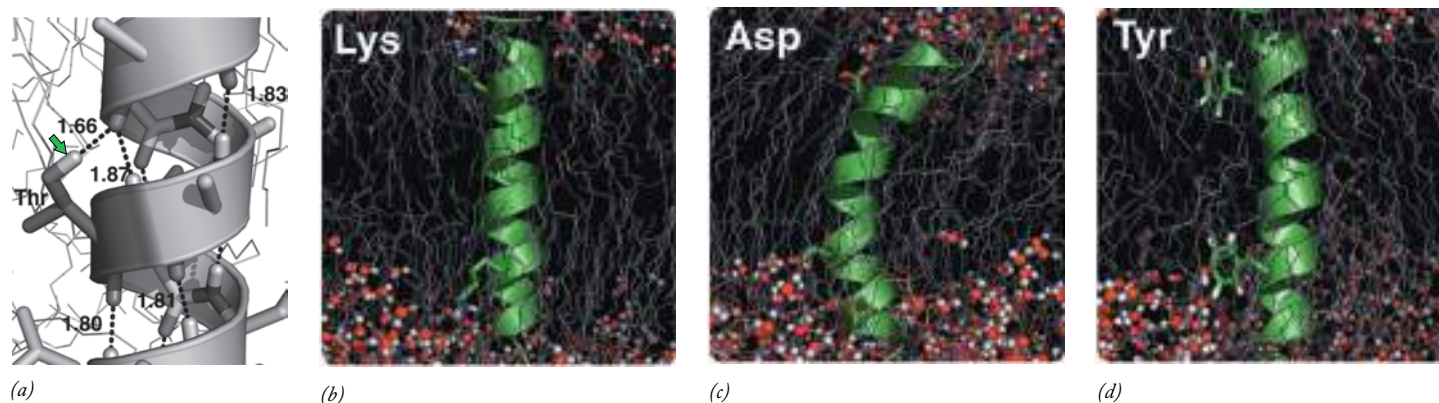
FIGURE 4.19 Accommodating various amino acid residues within transmembrane helices. (*a*) In this portrait of a small portion of a transmembrane helix, the hydroxyl group of the threonine side chain (arrow) is able to form a (shared) hydrogen bond with a backbone oxygen within the lipid bilayer. Hydrogen bonds are indicated by the dashed lines and their distances are shown in angstroms. (*b*) The side chains of the two lysine residues of this transmembrane helix are sufficiently long and flexible to form bonds with the head groups and water molecules of the polar face of the lipid bilayer. (*c*) The side chains of the two aspartic acid residues of this transmembrane helix can also reach the polar face of the bilayer but introduce distortion in the helix to do so. (*d*) The aromatic side chains of the two tyrosine residues of this transmembrane helix are oriented perpendicular to the axis of the membrane and parallel to the fatty acyl chains with which they interact. (FROM ANNA C. V. JOHANSSON AND ERIK LINDAHL, BIOPHYS. J. 91:4459, 4453, 2006.)

glycophorin A, the major integral protein of the erythrocyte plasma membrane. Of the 20 amino acids that make up the lone α helix of a glycophorin monomer (amino acids 73 to 92 of Figure 4.18), all but three have hydrophobic side chains (or an H atom in the case of the glycine residues). The exceptions are serine and threonine, which are noncharged, polar residues (Figure 2.26). Figure 4.19*a* shows a portion of a transmembrane helix with a threonine residue, not unlike those of glycophorin A. The hydroxyl group of the residue's side chain can form a hydrogen bond with one of the oxygen atoms of the peptide backbone. Fully charged residues may also appear in transmembrane helices, but they tend to be accommodated in ways that allow them to fit into their hydrophobic environment. This is illustrated in the model transmembrane helices shown in Figures 4.19*b* and *c*. Each of the helices in these figures contains a pair of charged residues whose side chains are able to reach out and interact with the innermost polar regions of the membrane, even if it requires distorting the helix to do so. Figure 4.19*d* shows two tyrosine residues with their hydrophobic aromatic side chains; each aromatic ring is oriented parallel with the hydrocarbon chains of the bilayer with which it has become integrated.

Knowing the amino acid sequence of an integral membrane protein, we can usually identify the transmembrane segments using a *hydropathy plot*, in which each site along a polypeptide is assigned a value that provides a measure of the *hydrophobicity* of the amino acid at that site as well as that of its neighbors. This approach provides a "running average" of the hydrophobicity of short sections of the polypeptide, and guarantees that one or a few polar amino acids in a sequence do not alter the profile of the entire stretch. Hydrophobicity of amino acids can be determined using various criteria, such as their lipid solubility or the energy that would be required to transfer them from an aqueous into a lipid medium. A hydropathy plot for glycophorin A is shown in Figure 4.20. Transmembrane

segments are usually identified as a jagged peak that extends well into the hydrophobic side of the spectrum. A reliable prediction concerning the orientation of the transmembrane segment within the bilayer can usually be made by examining the flanking amino acid residues. In most cases, as illustrated by glycophorin in Figure 4.18, those parts of the polypeptide at the cytoplasmic flank of a transmembrane segment tend to be more positively charged than those at the extracellular flank.

Not all integral membrane proteins contain transmembrane α helices. A number of membrane proteins contain a relatively large channel positioned within a circle of membrane-spanning
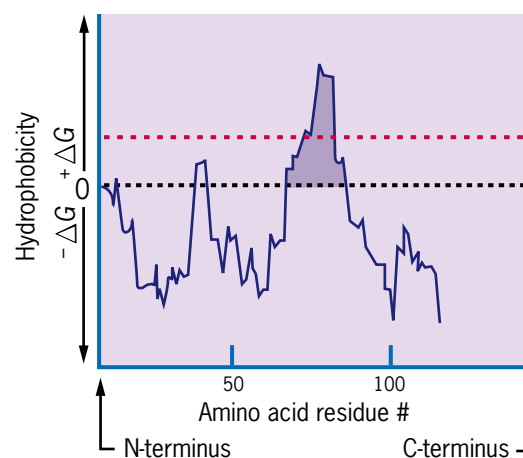
FIGURE 4.20 Hydropathy plot for glycophorin A, a single membrane-spanning protein. Hydrophobicity is measured by the free energy required to transfer each segment of the polypeptide from a nonpolar solvent to an aqueous medium. Values above the 0 line are energy-requiring ($+\Delta G$s), indicating they consist of stretches of amino acids that have predominantly nonpolar side chains. Peaks that project above the red-colored line are interpreted as a transmembrane domain.

β strands organized into a barrel as illustrated in Figure 5.4. To date, aqueous channels constructed of β barrels have only been found in the outer membranes of bacteria, mitochondria, and chloroplasts.

Determining Spatial Relationships within an Integral Membrane Protein   Suppose you have isolated a gene for an integral membrane protein and, based on its nucleotide sequence, determined that it contains four apparent membrane-spanning α helices. You might want to know how these helices are oriented relative to one another and which amino acid side chains of each helix face outward toward the lipid environment. Although these determinations are difficult to make without detailed structural models, considerable insight can be gained by site-directed mutagenesis, that is, by introducing specific changes into the gene that codes for the protein (Section 18.18). For example, site-directed mutagenesis can be employed to replace amino acid residues in neighboring helices with cysteine residues. As discussed on page 52, two cysteine residues can form a covalent disulfide bridge. If two transmembrane helices of a polypeptide each contain a cysteine residue, and the two cysteine residues are able to form a disulfide bridge with one another, then these helices must reside in very close proximity. The results of one site-directed cross-linking study on lactose permease, a sugar-transporting protein in bacterial cell membranes, are shown in Figure 4.21. It was found in this case that helix VII lies in close proximity to both helices I and II.

Determining spatial relationships between amino acids in a membrane protein can provide more than structural information; it can tell a researcher about some of the dynamic events that occur as a protein carries out its function. One way to learn about the distance between selected residues in a protein is to introduce chemical groups whose properties are sensitive to the distance that separates them. *Nitroxides* are chemical groups that contain an unpaired electron, which produces a characteristic spectrum when monitored by a technique called *electron paramagnetic resonance* (*EPR*) *spectroscopy.* A nitroxide group can be introduced at any site in a protein by first mutating that site to a cysteine and then attaching the nitroxide to the —SH group of the cysteine residue. Figure 4.22 shows how this technique was used to discover the conformational changes that occur in a membrane protein as its channel is opened in response to changes in the pH of the medium. The protein in question, a bacterial K$^+$ channel, is a tetramer composed of four identical subunits. The cytoplasmic opening to the channel is bounded by four transmembrane helices, one from each subunit of the protein. Figure 4.22a shows the EPR spectra that were obtained when a nitroxide was introduced near the cytoplasmic end of each transmembrane helix. The red line shows the spectrum obtained at pH 6.5 when the channel is in the closed state, and the blue line shows the spectrum at pH 3.5 when the channel is open. The shape of each line depends on the proximity of the nitroxides to one another. The spectrum is broader at pH 6.5 because the nitroxide groups on the four subunits are closer together at this pH, which decreases the intensity of their EPR signals. These results indicate that the activation of the channel is accompa-
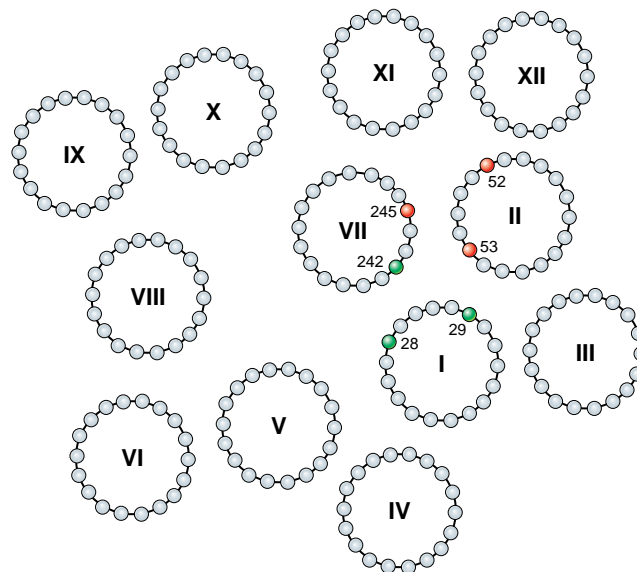
FIGURE 4.21 Determination of helix packing in a membrane protein by site-directed cross-linking. In these experiments, pairs of cysteine residues are introduced into the protein by site-directed mutagenesis, and the ability of the cysteines to form disulfide bridges is determined. Hydropathy plots and other data had indicated that lactose permease has 12 transmembrane helices. It was found that a cysteine introduced at position 242 of helix VII can cross-link to a cysteine introduced at either position 28 or 29 of helix I. Similarly, a cysteine at position 245 of helix VII can cross-link to cysteines at either 52 or 53 of helix II. The proximity of these three helices is thus established. (The X-ray structure of this protein was reported in 2003.) (REPRINTED FROM H. R. KABACK, J. VOSS, AND J. WU, CURR. OPIN. STRUCT. BIOL. 7:539, 1997; COPYRIGHT 1997, WITH PERMISSION FROM ELSEVIER SCIENCE.)

nied by an increased separation between the labeled residues of the four subunits (Figure 4.22b). An increase in diameter of the channel opening allows ions in the cytoplasm to reach the actual permeation pathway (shown in red) within the channel, which allows only the passage of K$^+$ ions (discussed on page 149). An alternate technique, called FRET, that can also be used to measure changes in the distance between labeled groups within a protein, is illustrated in Figure 18.8.

Peripheral Membrane Proteins

Peripheral proteins are associated with the membrane by weak electrostatic bonds (refer to Figure 4.13b). Peripheral proteins can usually be solubilized by extraction with high-concentration salt solutions that weaken the electrostatic bonds holding peripheral proteins to a membrane. In actual fact, the distinction between integral and peripheral proteins is blurred because many integral membrane proteins consist of several polypeptides, some that penetrate the lipid bilayer and others that remain on the periphery.

The best studied peripheral proteins are located on the internal (cytosolic) surface of the plasma membrane, where they form a fibrillar network that acts as a membrane "skeleton" (see Figure 4.32d). These proteins provide mechanical support for

(a)
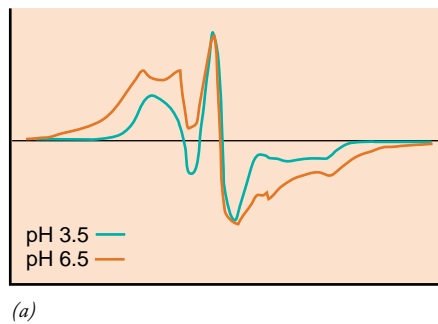
FIGURE 4.22 Use of EPR spectroscopy to monitor changes in conformation of a bacterial K⁺ ion channel as it opens and closes. (*a*) EPR spectra from nitroxides that have been attached to cysteine residues near the cytoplasmic end of the four transmembrane helices that line the channel. The cysteine residue in each helix replaces a glycine residue that is normally at that position. The shapes of the spectra depend on the distances between unpaired electrons in the nitroxides on different subunits. (Nitroxides are described as "spin-labels," and this technique is known as site-directed spin labeling.) (*b*) A highly schematic model of the ion channel in the open and closed states based on the data from part *a*. Opening of the channel is accompanied by the movement of the four nitroxide groups apart from one another. (A: FROM E. PEROZO ET AL., NATURE STRUCT. BIOL. 5:468, 1998.)
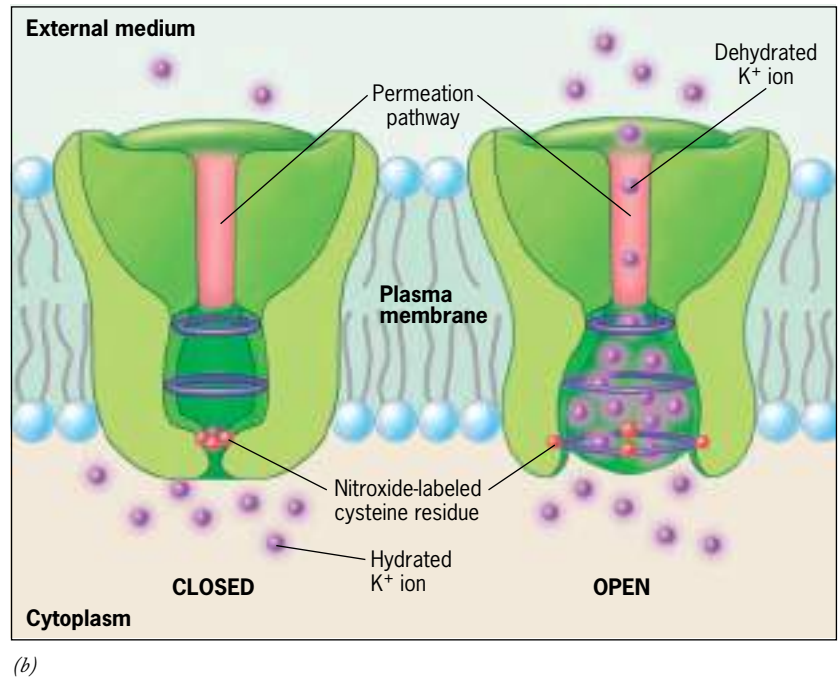
the membrane and function as an anchor for integral membrane proteins. Other peripheral proteins on the internal plasma membrane surface function as enzymes, specialized coats (see Figure 8.24), or factors that transmit transmembrane signals (see Figure 15.17). Peripheral proteins typically have a dynamic relationship with the membrane, being recruited to the membrane or released from the membrane depending on prevailing conditions.

Lipid-Anchored Membrane Proteins

Several types of lipid-anchored membrane proteins can be distinguished. Numerous proteins present on the external face of the plasma membrane are bound to the membrane by a small, complex oligosaccharide linked to a molecule of phosphatidylinositol that is embedded in the outer leaflet of the lipid bilayer (refer to Figure 4.13c). Peripheral membrane proteins containing this type of glycosyl-phosphatidylinositol linkage are called **GPI-anchored proteins**. They were discovered when it was shown that certain membrane proteins could be released by a phospholipase that specifically recognized and cleaved inositol-containing phospholipids. The normal cellular scrapie protein PrP^C (page 64) is a GPI-linked molecule, as are various receptors, enzymes, and cell-adhesion proteins. A rare type of anemia, paroxysmal nocturnal hemoglobinuria, results from a deficiency in GPI synthesis that makes red blood cells susceptible to lysis.

Another group of proteins present on the *cytoplasmic* side of the plasma membrane is anchored to the membrane by one or more long hydrocarbon chains embedded in the inner leaflet of the lipid bilayer (refer to Figure 4.13c and accompanying legend). At least two proteins associated with the plasma membrane in this way (Src and Ras) have been implicated in the transformation of a normal cell to a malignant state.

> **REVIEW** ?
>
> 1. Why are detergents necessary to solubilize membrane proteins? How might one determine the diversity of integral proteins that reside in a purified membrane fraction?
>
> 2. How can one determine: (1) the location of transmembrane segments in the amino acid sequence or (2) the relative locations of transmembrane helices?
>
> 3. What is meant by the statement that the proteins of a membrane are distributed asymmetrically? Is this also true of the membrane's carbohydrate?
>
> 4. Describe the properties of the three classes of membrane proteins (integral, peripheral, and lipid-anchored), how they differ from one another, and how they vary among themselves.

4.5 MEMBRANE LIPIDS AND MEMBRANE FLUIDITY

The physical state of the lipid of a membrane is described by its fluidity (or viscosity).[4] Consider a simple artificial bilayer composed of phosphatidylcholine and phosphatidylethanolamine, whose fatty acids are largely unsaturated. If the temperature of

---

[4]Fluidity and viscosity are inversely related; fluidity is a measure of the ease of flow, and viscosity is a measure of the resistance to flow.

(a)

(b)
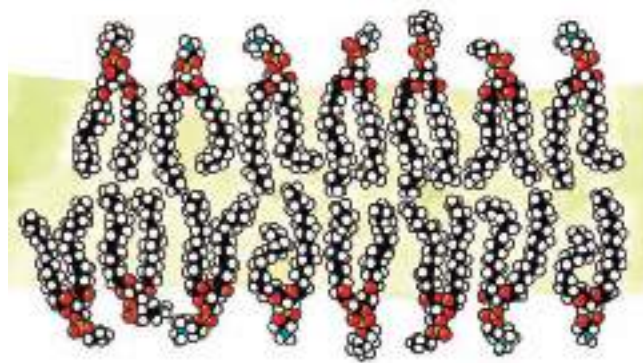
FIGURE 4.23 The structure of the lipid bilayer depends on the temperature. The bilayer shown here is composed of two phospholipids: phosphatidylcholine and phosphatidylethanolamine. (*a*) Above the transition temperature, the lipid molecules and their hydrophobic tails are free to move in certain directions, even though they retain a considerable degree of order. (*b*) Below the transition temperature, the movement of the molecules is greatly restricted, and the entire bilayer can be described as a crystalline gel. (AFTER R. N. ROBERTSON, THE LIVELY MEMBRANES, CAMBRIDGE UNIVERSITY PRESS; 1983, REPRINTED WITH PERMISSION OF CAMBRIDGE UNIVERSITY PRESS.)

the bilayer is kept relatively warm (e.g., 37°C), the lipid exists in a relatively fluid state (Figure 4.23*a*). At this temperature, the lipid bilayer is best described as a two-dimensional liquid crystal. As in a crystal, the molecules still retain a specified orientation; in this case, the long axes of the molecules tend toward a parallel arrangement, yet individual phospholipids can rotate around their axis or move laterally within the plane of the bilayer. If the temperature is slowly lowered, a point is reached where the bilayer distinctly changes (Figure 4.23*b*). The lipid is converted from a liquid crystalline phase to a frozen crystalline gel in which the movement of the phospholipid fatty acid chains is greatly restricted. The temperature at which this change occurs is called the **transition temperature**.

The transition temperature of a particular bilayer depends on the ability of the lipid molecules to be packed together, which depends in turn on the particular lipids of which it is constructed. Saturated fatty acids have the shape of a straight, flexible rod. *Cis*-unsaturated fatty acids, on the other hand, have crooks in the chain at the sites of a double bond (Figures 2.19 and 4.23). Consequently, phospholipids with saturated chains pack together more tightly than those containing unsaturated chains. The greater the degree of unsaturation of the fatty acids of the bilayer, the *lower* the temperature before the bilayer gels. The introduction of one double bond in a molecule of stearic acid lowers the melting temperature almost 60°C (Table 4.2).[5] Another factor that influences bilayer fluidity is fatty acid chain length. The shorter the fatty acyl chains of a phospholipid, the lower its melting temperature. The physical state of the membrane is also affected by cholesterol. Because of their orientation within the bilayer (Figure 4.7), cholesterol molecules disrupt the close packing of fatty acyl chains and interfere with their mobility. The pres-

TABLE 4.2 Melting Points of the Common 18-Carbon Fatty Acids

Fatty acid	*cis* Double bonds	M.p.(°C)
Stearic acid	0	70
Oleic acid	1	13
Linoleic acid	2	−9
Linolenic acid	3	−17
Eicosapentanoic acid (EPA)*	5	−54

*EPA has 20 carbons.

ence of cholesterol tends to abolish sharp transition temperatures and creates a condition of intermediate fluidity. In physiologic terms, cholesterol tends to increase the durability while decreasing the permeability of a membrane.

The Importance of Membrane Fluidity

What effect does the physical state of the lipid bilayer have on the biological properties of the membrane? Membrane fluidity provides a perfect compromise between a rigid, ordered structure in which mobility would be absent and a completely fluid, nonviscous liquid in which the components of the membrane could not be oriented and structural organization and mechanical support would 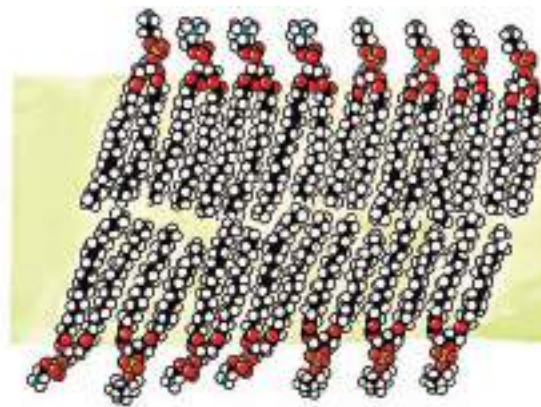be lacking. In addition, fluidity allows for interactions to take place within the membrane. For example, membrane fluidity makes it possible for clusters of membrane proteins to assemble at particular sites within the membrane and form specialized structures, such as intercellular junctions, light-capturing photosynthetic complexes, and synapses. Because of membrane fluidity, molecules that interact can come together, carry out the necessary reaction, and move apart.

Fluidity also plays a key role in membrane assembly, a subject discussed in Chapter 8. Membranes arise only from preexisting membranes, and their growth is accomplished by the insertion of lipids and proteins into the fluid matrix of the membranous sheet. Many of the most basic cellular processes,

---

[5]The effect of fatty acid saturation on melting temperature is illustrated by familiar food products. Vegetable oils remain a liquid in the refrigerator, whereas margarine is a solid. Vegetable oils contain polyunsaturated fatty acids, whereas the fatty acids of margarine have been saturated by a chemical process that hydrogenates the double bonds of the vegetable oils used as the starting material.

including cell movement, cell growth, cell division, formation of intercellular junctions, secretion, and endocytosis, depend on the movement of membrane components and would probably not be possible if membranes were rigid, nonfluid structures.

Maintaining Membrane Fluidity

The internal temperature of most organisms (other than birds and mammals) fluctuates with the temperature of the external environment. Since it is essential for many activities that the membranes of a cell remain in a fluid state, cells respond to changing conditions by altering the types of phospholipids of which they are made. Maintenance of membrane fluidity is an example of homeostasis at the cellular level and can be demonstrated in various ways. For example, if the temperature of a culture of cells is lowered, the cells respond metabolically. The initial "emergency" response is mediated by enzymes that remodel membranes, making the cell more cold resistant. Remodeling is accomplished by (1) desaturating single bonds in fatty acyl chains to form double bonds, and (2) reshuffling the chains between different phospholipid molecules to produce ones that contain two unsaturated fatty acids, which greatly lowers the melting temperature of the bilayer. Desaturation of single bonds to form double bonds is catalyzed by enzymes called *desaturases*. Reshuffling is accomplished by *phospholipases*, which split the fatty acid from the glycerol backbone, and *acyltransferases*, which transfer fatty acids be-

tween phospholipids. In addition, the cell changes the types of phospholipids being synthesized in favor of ones containing more unsaturated fatty acids. As a result of the activities of these various enzymes, the physical properties of a cell's membranes are matched to the prevailing environmental conditions. Maintenance of fluid membranes by adjustments in fatty acyl composition has been demonstrated in a variety of organisms, including hibernating mammals, pond-dwelling fish whose body temperature changes markedly from day to night, cold-resistant plants, and bacteria living in hot springs.

Lipid Rafts

Every so often an issue emerges that splits the community of cell biologists into believers and nonbelievers. The issue of lipid rafts falls into this category. When membrane lipids are extracted from cells and used to prepare *artificial* lipid bilayers, cholesterol and sphingolipids tend to self-assemble into microdomains that are more gelated and highly ordered than surrounding regions consisting primarily of phosphoglycerides. Because of their distinctive physical properties, such microdomains tend to float within the more fluid and disordered environment of the artificial bilayer (Figure 4.24*a*). As a result, these patches of cholesterol and sphingolipid are referred to as **lipid rafts**. When added to these artificial bilayers, certain proteins tend to become concentrated in the lipid rafts, whereas others tend to remain outside their boundaries. GPI-anchored proteins show a particular fondness for the ordered regions of the bilayer (Figure 4.24*a*).

(a)

(b)
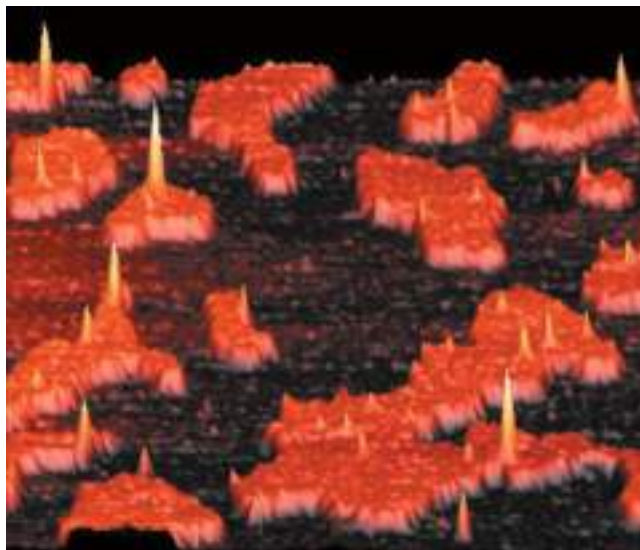
FIGURE 4.24 Lipid rafts. (*a*) Image of the upper surface of an artificial lipid bilayer containing phosphatidylcholine, which appears as the black background, and sphingomyelin molecules, which organize themselves spontaneously into the orange-colored rafts. The yellow peaks show the positions of a GPI-anchored protein, which is almost exclusively raft-associated. This image is provided by an atomic force microscope, which measures the height of various parts of the specimen at the molecular level. (*b*) Schematic model of a lipid raft within a cell. The outer leaflet of the raft consists primarily of cholesterol (yellow) and sphingolipids (red head groups). Phosphatidylcholine molecules (blue head groups)
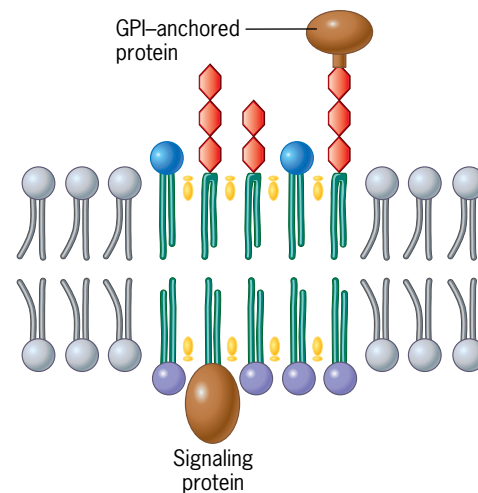
with long saturated fatty acids also tend to concentrate in this region. GPI-anchored proteins are thought to become concentrated in lipid rafts. The lipids in the outer leaflet of the raft have an organizing effect on the lipids of the inner leaflet. As a result, the inner-leaflet raft lipids consist primarily of cholesterol and glycerophospholipids with long saturated fatty acyl tails. The inner leaflet tends to concentrate lipid-anchored proteins, such as Src kinase, that are involved in cell signaling. (The controversy over the existence of lipid rafts is discussed in *Nature Cell Biol*. 9:7, 2007.) (A: FROM D. E. SASLOWSKY, ET AL., J. BIOL. CHEM. 277, COVER OF #30, 2002; COURTESY OF J. MICHAEL EDWARDSON.)

The controversy arises over whether similar types of cholesterol-rich lipid rafts, such as that illustrated in Figure 4.24b, exist within living cells. Most of the evidence in favor of lipid rafts is derived from studies that employ unnatural treatments, such as detergent extraction or cholesterol depletion, which makes the results difficult to interpret. Attempts to demonstrate the presence of lipid rafts in living cells have generally been unsuccessful, which can either mean that such rafts do not exist or they are so small (5 to 25 nm diameter) and short-lived as to be difficult to detect with current techniques. The concept of lipid rafts is very appealing because it provides a means to introduce order into a seemingly random sea of lipid molecules. Lipid rafts are postulated to serve as floating platforms that concentrate particular proteins, thereby organizing the membrane into functional compartments (Figure 4.24b). For example, lipid rafts are thought to provide a favorable local environment for cell-surface receptors to interact with other membrane proteins that transmit signals from the extracellular space to the cell interior.

FIGURE 4.25 The possible movements of phospholipids in a membrane. The types of movements in which membrane phospholipids can engage and the approximate time scales over which they occur. Whereas phospholipids move from one leaflet to another at a very slow rate, they diffuse laterally within a leaflet rapidly. Lipids lacking polar groups, such as cholesterol, can move across the bilayer quite rapidly.

REVIEW ?

1. What is the importance of fatty acid unsaturation for membrane fluidity? of enzymes that are able to desaturate fatty acids?

2. What is meant by a membrane's transition temperature, and how is it affected by the degree of saturation or length of fatty acyl chains? How are these properties important in the formation of lipid rafts?

3. Why is membrane fluidity important to a cell?

4. How can the two sides of a lipid bilayer have different ionic charges?

4.6 THE DYNAMIC NATURE OF THE PLASMA MEMBRANE

It is apparent from the previous discussion that the lipid bilayer can exist in a relatively fluid state. As a result, a phospholipid can move laterally wi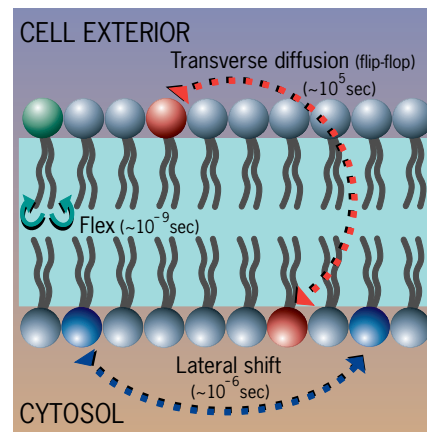thin the same leaflet with considerable ease. The mobility of individual lipid molecules within the bilayer of the plasma membrane can be directly observed under the microscope by linking the polar heads of the lipids to gold particles or fluorescent compounds (see Figure 4.29). It is estimated that a phospholipid can diffuse from one end of a bacterium to the other end in a second or two. In contrast, it takes a phospholipid molecule a matter of hours to days to move across to the other leaflet. Thus, of all the possible motions that a phospholipid can make, its flip-flop to the other side of the membrane is the most restricted (Figure 4.25). This finding is not surprising. For flip-flop to occur, the hydrophilic head group of the lipid must pass through the internal hydrophobic sheet of the membrane, which is thermodynamically unfavorable. However, cells contain enzymes that actively move certain phospholipids from

one leaflet to the other. These enzymes play a role in establishing lipid asymmetry and may also reverse the slow rate of passive transmembrane movement.

Because lipids provide the matrix in which integral proteins of a membrane are embedded, the physical state of the lipid is an important determinant of the mobility of integral proteins. The demonstration that integral proteins can move within the plane of the membrane was a cornerstone in the formulation of the fluid-mosaic model. The dynamic properties of membrane proteins have been revealed in several ways.

The Diffusion of Membrane Proteins after Cell Fusion

Cell fusion is a technique whereby two different types of cells, or cells from two different species, can be fused to produce one cell with a common cytoplasm and a single, continuous plasma membrane. Cells are induced to fuse with one another by making the outer surface of the cells "sticky" so that their plasma membranes adhere to one another. Cells can be induced to fuse by addition of certain inactivated viruses that attach to the surface membrane, by adding the compound polyethylene glycol, or by a mild electric shock. Cell fusion has played an important role in cell biology and is currently used in an invaluable technique to prepare specific antibodies (Section 18.19).

The first experiments to demonstrate that membrane proteins could move within the plane of the membrane utilized cell fusion, and they were reported in 1970 by Larry Frye and Michael Edidin of Johns Hopkins University. In their experiments, mouse and human cells were fused, and the locations of specific proteins of the plasma membrane were followed once the two membranes had become continuous.
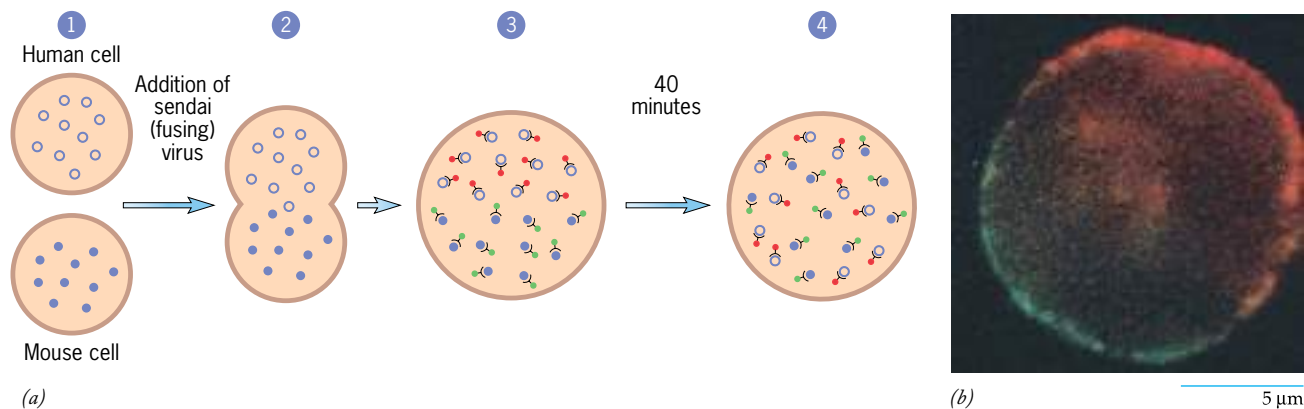
FIGURE 4.26 The use of cell fusion to reveal mobility of membrane proteins. (*a*) Outline of the experiment in which human and mouse cells were fused (steps 1–2) and the distribution of the proteins on the surface of each cell were followed in the hybrids over time (steps 3–4). Mouse membrane proteins are indicated by solid circles, human membrane proteins by open circles. Locations of human and mouse proteins in the plasma membranes of the hybrid cells were monitored by interaction with fluorescent red and fluorescent green antibodies, respectively. (*b*) Micrograph showing a fused cell in which mouse and human proteins are still in their respective hemispheres (equivalent to the hybrid in step 3 of part *a*). (B: FROM L. D. FRYE AND MICHAEL EDIDIN, J. CELL SCI. 7:328, 334, 1970; BY PERMISSION OF THE COMPANY OF BIOLOGISTS LTD.)

To follow the distribution of either the mouse membrane proteins or the human membrane proteins at various times after fusion, antibodies against one or the other type of protein were prepared and covalently linked to fluorescent dyes. The antibodies against the mouse proteins were complexed with a dye that fluoresces green and the antibodies against human proteins with one that fluoresces red. When the antibodies were added to fused cells, they bound to the human or mouse proteins and could be located under a fluorescence light microscope (Figure 4.26*a*). At the time of fusion, the plasma membrane appeared half human and half mouse; that is, the two protein types remained segregated in their own hemisphere (step 3, Figure 4.26*a,b*). As the time after fusion increased, the membrane proteins were seen to move laterally within the membrane into the opposite hemisphere. By about 40 minutes, each species' proteins were uniformly distributed around the entire hybrid cell membrane (step 4, Figure 4.26*a*). If the same experiment was performed at lower temperature, the viscosity of the lipid bilayer increased, and the mobility of the membrane proteins decreased. These early cell fusion experiments gave the impression that integral membrane proteins were capable of virtually unrestricted movements. As we will see shortly, subsequent studies made it apparent that membrane dynamics was a much more complex subject than first envisioned.

Restrictions on Protein and Lipid Mobility

Several techniques allow researchers to follow the movements of molecules in the membranes of living cells using the light microscope. In a technique called **fluorescence recovery after photobleaching (FRAP)**, which is illustrated in Figure 4.27*a*, integral membrane components in cultured cells are first labeled by linkage to a fluorescent dye. A particular membrane protein can be labeled using a specific probe, such as a fluorescent antibody. Once labeled, cells are placed under the microscope and irradiated by a sharply focused laser beam that bleaches the fluorescent molecules in its path, leaving a circular spot (typically about 1 μm diameter) on the surface of the cell that is largely devoid of fluorescence. If the labeled proteins in the membrane are mobile, then the random movements of these molecules should produce a gradual reappearance of fluorescence in the irradiated circle. The rate of fluorescence recovery (Figure 4.27*b*) provides a direct measure of the rate of diffusion (expressed as a diffusion coefficient, *D*) of the mobile molecules. The extent of fluorescence recovery (expressed as a percentage of the original intensity) provides a measure of the percentage of the labeled molecules that are free to diffuse.

Early studies utilizing FRAP suggested that (1) membrane proteins moved much more slowly in a plasma membrane than they would in a pure lipid bilayer and (2) a significant fraction of membrane proteins (30 to 70 percent) were not free to diffuse back into the irradiated circle. But the FRAP technique has its drawbacks. FRAP can only follow the average movement of a relatively large number of labeled molecules (hundreds to thousands) as they diffuse over a relatively large distance (e.g., 1 μm). As a result, researchers using FRAP cannot distinguish between proteins that are truly immobile and ones that can only diffuse over a limited distance in the time allowed. To get around these limitations, alternate techniques have been developed that allow researchers to observe the movements of individual protein molecules over very short distances and to determine how they might be restrained.

In **single-particle tracking (SPT)**, individual membrane protein molecules are labeled, usually with antibody-coated gold particles (approximately 40 nm in diameter), and the movements of the labeled molecules are followed by computer-enhanced video microscopy (Section 18.1). The results of these

(a)

(b)
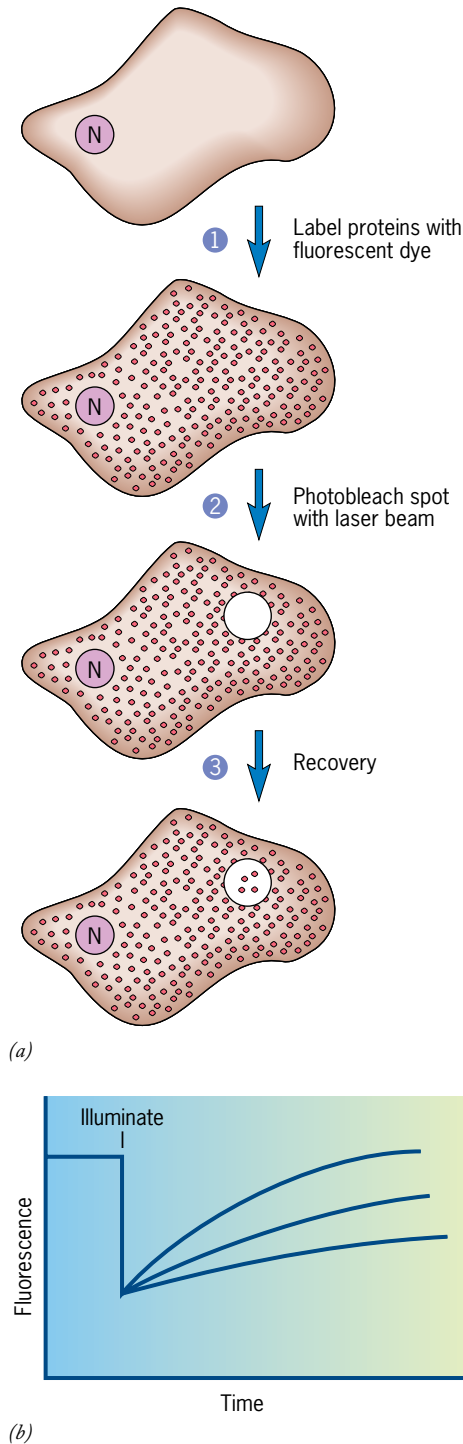
FIGURE 4.27 Measuring the diffusion rates of membrane proteins by fluorescence recovery after photobleaching (FRAP). (*a*) In this technique, a particular component of the membrane is first labeled with a fluorescent dye (step 1). A small region of the surface is then irradiated to bleach the dye molecules (step 2), and the recovery of fluorescence in the bleached region is followed over time (step 3). (N represents the cell nucleus.) (*b*) The rate of fluorescence recovery within the illuminated spot can vary depending on the protein(s) being followed. The rate of recovery is related to the diffusion coefficient of the fluorescently labeled protein.

studies often depend on the particular protein being investigated. For example,

■ Some membrane proteins move randomly throughout the membrane (Figure 4.28, protein A), though generally at rates considerably less than would be measured in an artificial lipid bilayer. (If protein mobility were based strictly on physical parameters such as lipid viscosity and protein size, one would expect proteins to migrate with diffusion coefficients of approximately 10^{-8} to 10^{-9} cm^2/sec rather than 10^{-10} to 10^{-12} cm^2/sec, as is observed for molecules of this group.) The reasons for the reduced diffusion coefficient have been debated.

■ Some membrane proteins fail to move and are considered to be immobilized (Figure 4.28, protein B).

■ In some cases, a protein is found to move in a highly directed (i.e., nonrandom) manner toward one part of the cell or another (Figure 4.28, protein C). For example, a particular membrane protein may tend to move toward the leading or the trailing edge of a moving cell.

■ In most studies, the largest fraction of protein species exhibit random (Brownian) movement within the membrane at rates consistent with free diffusion (diffusion coefficients about 5×10^{-9} cm^2/sec), but the molecules are unable to migrate freely more than a few tenths of a micrometer. Long-range diffusion occurs but at slower rates,
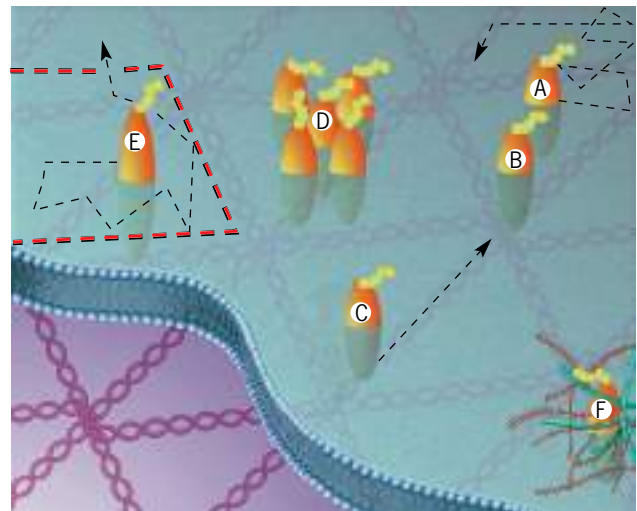
FIGURE 4.28 Patterns of movement of integral membrane proteins. Depending on the cell type and the conditions, integral membrane proteins can exhibit several different types of mobility. Protein A is capable of diffusing randomly throughout the membrane, though its rate of movement may be limited; protein B is immobilized as the result of its interaction with the underlying membrane skeleton; protein C is being moved in a particular direction as the result of its interaction with a motor protein at the cytoplasmic surface of the membrane; movement of protein D is restricted by other integral proteins of the membrane; movement of protein E is restricted by fences formed by proteins of the membrane skeleton, but it can hop into adjacent compartments through transient openings in a fence; movement of protein F is restrained by extracellular materials.

apparently because of the presence of a system of barriers. These barriers are discussed in the following sections.

Control of Membrane Protein Mobility   It is apparent that plasma membrane proteins are not totally free to drift around randomly on the lipid "sea," but instead they are subjected to various influences that affect their mobility. Some membranes are crowded with proteins, so that the random movements of one molecule can be impeded by its neighbors (Figure 4.28, protein D). The strongest influences on an integral membrane protein are thought to be exerted from just beneath the membrane on its cytoplasmic face. The plasma membranes of many cells possess a fibrillar network, or "membrane skeleton," consisting of peripheral proteins situated on the cytoplasmic surface of the membrane. A certain proportion of a membrane's integral protein molecules are either tethered to the membrane skeleton (Figure 4.28, protein B) or otherwise restricted by it.

Information concerning the presence of membrane barriers has been obtained using an innovative technique that allows investigators to trap integral proteins and drag them through the plasma membrane with a known force. This technique, which uses an apparatus referred to as *optical tweezers*, takes advantage of the tiny optical forces that are generated by a focused laser beam. The integral proteins to be studied are tagged with antibody-coated beads, which serve as handles that can be gripped by the laser beam. It is generally found that optical tweezers can drag an integral protein for a limited distance before the protein encounters a barrier that causes it to be released from the laser's grip. As it is released, the protein typically springs backward, suggesting that the barriers are elastic structures.

One approach to studying factors that affect membrane protein mobility is to genetically modify cells so that they produce altered membrane proteins. Integral proteins whose cytoplasmic portions have been genetically deleted often move much greater *distances* than their intact counterparts, indicating that barriers reside on the cytoplasmic side of the membrane. These findings suggest that the membrane's underlying skeleton forms a network of "fences" around portions of the membrane, creating compartments that restrict the distance an integral protein can travel (Figure 4.28, protein E). Proteins move across the boundaries from one compartment to another through breaks in the fences. Such openings are thought to appear and disappear along with the dynamic disassembly and reassembly of parts of the meshwork. Membrane compartments may keep specific combinations of proteins in close enough proximity to facilitate their interaction.

Integral proteins lacking that portion that would normally project into the extracellular space typically move at a much faster *rate* than the wild-type version of the protein. This finding suggests that the movement of a transmembrane protein through the bilayer is slowed by extracellular materials that can entangle the external portion of the protein molecule (Figure 4.28, protein F).

Membrane Lipid Mobility   Proteins are huge molecules, so it isn't surprising that their movement within the lipid bilayer might be restricted. Phospholipids, by comparison, are small molecules that make up the very fabric of the lipid bilayer. One might expect their movements to be completely unfettered, yet a number of studies have suggested that phospholipid diffusion is also restricted. When individual phospholipid molecules of a plasma membrane are tagged and followed under the microscope using ultra high-speed cameras, they are seen to be confined for very brief periods and then hop from one confined area to another. Figure 4.29*a* shows the path taken by an individual phospholipid within the plasma membrane over a period of 56 milliseconds. Computer analysis indicates that the phospholipid diffuses freely within one compartment (shaded in purple) before it jumps the "fence" into a neighboring compartment (shaded in blue) and then over another fence into an adjacent compartment (shaded in green), and so forth. Treatment of the membrane with agents that disrupt the underlying membrane skeleton removes some of the fences that restrict phospholipid diffusion. But if the membrane skeleton lies beneath the lipid bilayer, how can it interfere with phospholipid movement? The authors of these studies speculate that the fences are constructed of rows of integral membrane proteins whose cytoplasmic domains are attached to the membrane skeleton. This
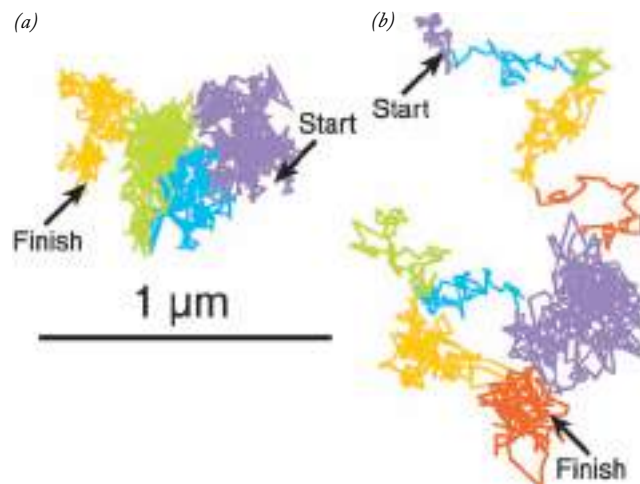
FIGURE 4.29 Experimental demonstration that diffusion of phospholipids within the plasma membrane is confined. (*a*) The track of a single labeled unsaturated phospholipid is followed for 56 ms as it diffuses within the plasma membrane of a rat fibroblast. Phospholipids diffuse freely within a confined compartment before hopping into a neighboring compartment. The rate of diffusion within a compartment is as rapid as that expected by unhindered Brownian movement. However, the overall rate of diffusion of the phospholipid appears slowed because the molecule must hop a barrier to continue its movement. The movement of the phospholipid within each compartment is represented by a single color. (*b*) The same experiment shown in *a* is carried out for 33 ms in an artificial bilayer, which lacks the "picket fences" present in a cellular membrane. The much more open, extended trajectory of the phospholipid can now be explained by simple, unconfined Brownian movement. For the sake of comparison, fake compartments were assigned in *b* and indicated by different colors. (FROM TAKAHIRO FUJIWARA ET AL., COURTESY OF AKIHIRO KUSUMI, J. CELL BIOL. 157:1073, 2002; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)
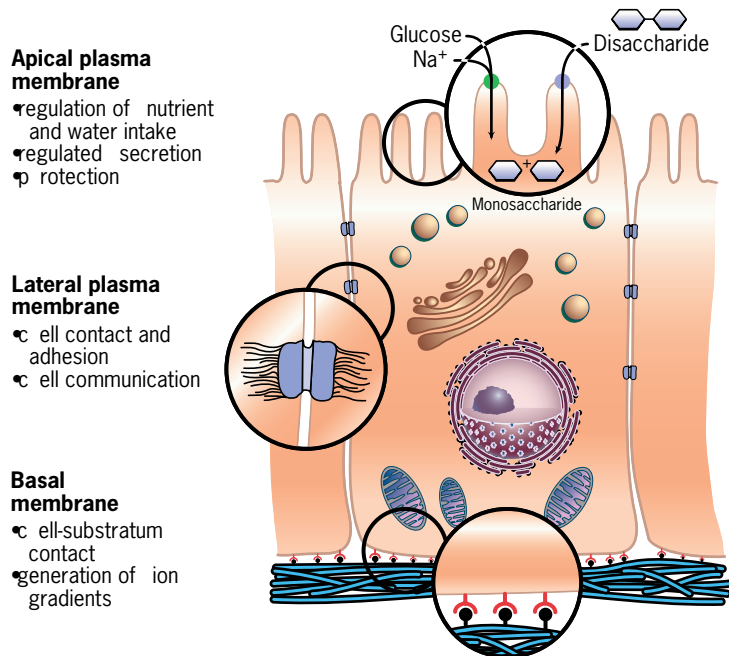
is not unlike the confinement of horses or cows by a picket fence whose posts are embedded in the underlying soil.

Membrane Domains and Cell Polarity For the most part, studies of membrane dynamics, such as those discussed above, are carried out on the relatively homogeneous plasma membrane situated at the upper or lower surface of a cell residing on a culture dish. Most membranes, however, exhibit distinct variations in protein composition and mobility, especially in cells whose various surfaces display distinct functions. For example, the epithelial cells that line the intestinal wall or make up the microscopic tubules of the kidney are highly polarized cells whose different surfaces carry out different functions (Figure 4.30). Studies indicate that the apical plasma membrane, which selectively absorbs substances from the lumen, possesses different enzymes than the lateral plasma membrane, which interacts with neighboring epithelial cells, or the basal membrane, which adheres to an underlying extracellular substrate (a *basement membrane*). In other examples, the receptors for neurotransmitter substances are concentrated into regions of the plasma membrane located within synapses (see Figure 4.56), and receptors for low-density lipoproteins are concentrated into patches of the plasma membrane specialized to facilitate their internalization (see Figure 8.38).

Of all the various types of mammalian cells, sperm may have the most highly differentiated structure. A mature sperm can be divided into head, midpiece, and tail, each having its own specialized functions. Although divided into a number of distinct parts, a sperm is covered by a *continuous* plasma membrane which, as revealed by numerous techniques, consists of a mosaic of different types of localized domains. For example, when sperm are treated with a variety of antibodies, each antibody combines with the surface of the cell in a unique topographic pattern that reflects the distribution within the plasma membrane of the particular protein antigen recognized by that antibody (Figure 4.31).

The Red Blood Cell: An Example of Plasma Membrane Structure

Of all the diverse types of membranes, the plasma membrane of the human erythrocyte (red blood cell) is the most studied and best understood (Figure 4.32). There are several reasons for the popularity of this membrane. The cells are inexpensive to obtain and readily available in huge numbers from whole blood. They are already present as single cells and need not be dissociated from a complex tissue. The cells are simple by comparison with other cell types, lacking nuclear and cytoplasmic membranes that inevitably contaminate plasma membrane preparations from other cells. In addition, purified, *intact* erythrocyte plasma membranes can be obtained simply by placing the cells in a dilute (hypotonic) salt solution. The cells respond to this osmotic shock by taking up water and swelling, a phenomenon termed *hemolysis*. As the surface area of each cell increases, the cell becomes leaky, and the contents, composed almost totally of dissolved hemoglobin, flow out of the cell leaving behind a plasma membrane "ghost" (Figure 4.32*b*).

Apical plasma membrane
- regulation of nutrient and water intake
- regulated secretion
- protection

Lateral plasma membrane
- cell contact and adhesion
- cell communication

Basal membrane
- cell-substratum contact
- generation of ion gradients

FIGURE 4.30 Differentiated functions of the plasma membrane of an epithelial cell. The apical surface of this intestinal epithelial cell contains integral proteins that function in ion transport and hydrolysis of disaccharides, such as sucrose and lactose; the lateral surface contains integral proteins that function in intercellular interaction; and the basal surface contains integral proteins that function in the association of the cell with the underlying basement membrane.

(a)   (b)

(c)   30 μm   (d)

Anterior head
Posterior head

Posterior tail
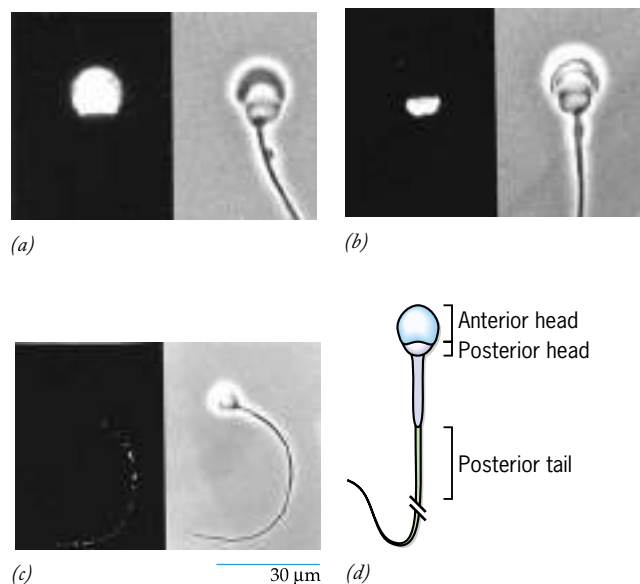
FIGURE 4.31 Differentiation of the mammalian sperm plasma membrane as revealed by fluorescent antibodies. (*a–c*) Three pairs of micrographs, each showing the distribution of a particular protein at the cell surface as revealed by a bound fluorescent antibody. The three proteins are localized in different parts of the continuous sperm membrane. Each pair of photographs shows the fluoresence pattern of the bound antibody and a phase contrast micrograph of the same cell. (*d*) Diagram summarizing the distribution of the proteins. (A–C: FROM DIANA G. MYLES, PAUL PRIMAKOFF, AND ANTHONY R. BELLVÉ, CELL 23:434, 1981, BY PERMISSION OF CELL PRESS.)

(a)

7 μm

(b)

PERIPHERAL   INTEGRAL

Major   Minor

Spectrin

Anykrin

ATPase
Band 3   AChe
Band 4.1   Glycophorin A
Band 4.2

Actin
G3PD

Band 7

Band 4.1

(c)

Actin

Tropomyosin

Ankyrin

Band 4.1

Spectrin
— α
— β

Band 3

Glycophorin A

(d)
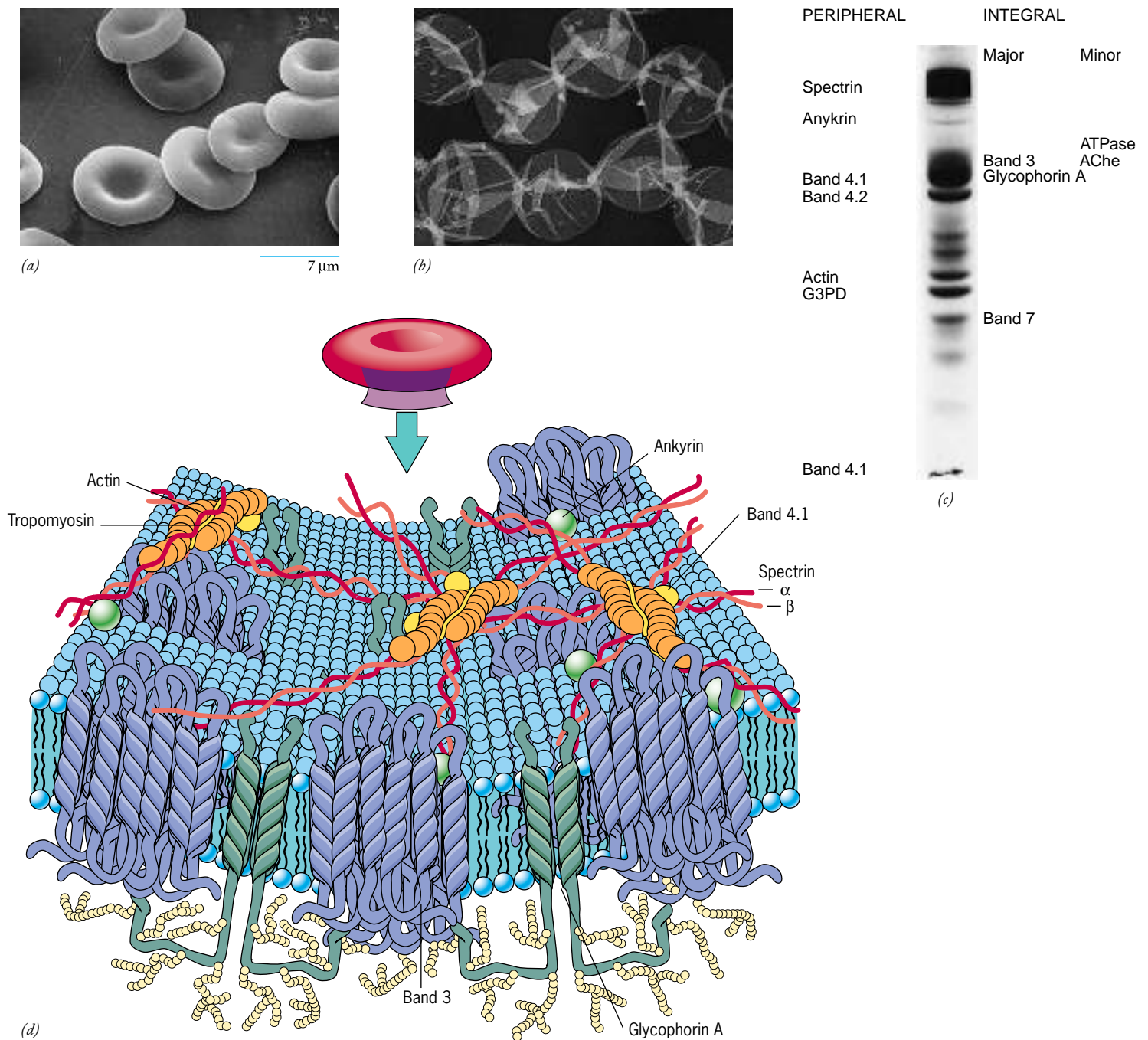
FIGURE 4.32 **The plasma membrane of the human erythrocyte.** (*a*) Scanning electron micrograph of human erythrocytes. (*b*) Micrograph showing plasma membrane ghosts, which were isolated by allowing erythrocytes to swell and hemolyze as described in the text. (*c*) The results of SDS–polyacrylamide gel electrophoresis (SDS–PAGE) used to fractionate the proteins of the erythrocyte membrane, which are identified at the side of the gel. (*d*) A model of the erythrocyte plasma membrane as viewed from the internal surface, showing the integral proteins embedded in the lipid bilayer and the arrangement of peripheral proteins that make up the membrane's internal skeleton. The band 3 dimer shown here is simplified. The band 4.1 protein stabilizes actin–spectrin complexes. (*e*) Electron micrograph showing the arrangement of the proteins of the inner membrane skeleton. (A: COURTESY FRANÇOIS M. M. MOREL, RICHARD F. BAKER, AND HAROLD WAYLAND, J. CELL BIOL. 48:91, 1971; B: COURTESY OF JOSEPH F. HOFFMAN; C: REPRODUCED, WITH PERMISSION, FROM V. T. MARCHESI, H. FURTHMAYR, AND M. TOMITA, ANNU. REV. BIOCHEM. VOL. 45; © 1976 BY ANNUAL REVIEWS INC.; D: AFTER D. VOET AND J. G. VOET, BIOCHEMISTRY, 2D ED.; COPYRIGHT © 1995, JOHN WILEY & SONS, INC.; E: FROM SHIH-CHUN LIU, LAURA H. DERICK, AND JIRI PALEK, J. CELL BIOL. 104:527, 1987; A, E: BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

(e)

100nm

Once erythrocyte plasma membranes are isolated, the proteins can be solubilized and separated from one another (fractionated), providing a better idea of the diversity of proteins within the membrane. Fractionation of membrane proteins can be accomplished using polyacrylamide gel electrophoresis (PAGE) in the presence of the ionic detergent sodium dodecyl sulfate (SDS). (The technique of SDS–PAGE is discussed in Section 18.7.) The SDS keeps the integral proteins soluble and, in addition, adds a large number of negative charges to the proteins with which it associates. Because the number of charged SDS molecules per unit weight of protein tends to be relatively constant, the molecules separate from one another according to their molecular weight. The largest proteins move most slowly through the molecular sieve of the gel. The major proteins of the erythrocyte membrane are separated into about a dozen conspicuous bands by SDS–PAGE (Figure 4.32c). Among the proteins are a variety of enzymes (including glyceraldehyde 3-phosphate dehydrogenase, one of the enzymes of glycolysis), transport proteins (for ions and sugars), and skeletal proteins (e.g., spectrin).

Integral Proteins of the Erythrocyte Membrane   A model of the erythrocyte plasma membrane showing its major proteins is seen in Figure 4.32d. The most abundant integral proteins of this membrane are a pair of carbohydrate-containing, membrane-spanning proteins, called band 3 and glycophorin A. The high density of these proteins within the membrane is evident in the freeze-fracture micrographs of Figure 4.15. Band 3, which gets its name from its position in an electrophoretic gel (Figure 4.32c), is present as a dimer composed of two identical subunits (a *homodimer*). Each subunit spans the membrane at least a dozen times and contains a relatively small amount of carbohydrate (6–8 percent of the molecule's weight). Band 3 protein serves as a channel for the passive exchange of anions across the membrane. As blood circulates through the tissues, carbon dioxide becomes dissolved in the fluid of the bloodstream (the plasma) and undergoes the following reaction:

$$H_2O + CO_2 \rightarrow H_2CO_3 \rightarrow HCO_3^- + H^+$$

The bicarbonate ions (HCO_3^-) enter the erythrocyte in exchange for chloride ions, which leave the cell. In the lungs, where carbon dioxide is released, the reaction is reversed and bicarbonate ions leave the erythrocyte in exchange for chloride ions. The reciprocal movement of HCO_3^- and Cl^- occurs through a channel in the center of each band 3 dimer.

Glycophorin A was the first membrane protein to have its amino acid sequence determined. The arrangement of the polypeptide chain of glycophorin A within the plasma membrane is shown in Figure 4.18. (Other related glycophorins, B, C, D, and E, are also present in the membrane at much lower concentrations.) Like band 3, glycophorin A is also present in the membrane as a dimer. Unlike band 3, each glycophorin A subunit spans the membrane only once, and it contains a bushy carbohydrate cover consisting of 16 oligosaccharide chains that together make up about 60 percent of the molecule's weight. It is thought that the primary function of glycophorin derives from the large number of negative charges

borne on sialic acid, the sugar residue at the end of each carbohydrate chain. Because of these charges, red blood cells repel each other, which prevents the cells from clumping as they circulate through the body's tiny vessels. It is noteworthy that persons who lack both glycophorin A and B in their red blood cells show no ill-effects from their absence. At the same time, the band 3 proteins in these individuals are more heavily glycosylated, which apparently compensates for the otherwise missing negative charges required to prevent cell–cell interaction. Glycophorin also happens to be the receptor utilized by the protozoan that causes malaria, providing a path for entry into the blood cell. Consequently, individuals whose erythrocytes lack glycophorin A and B are thought to be protected from acquiring malaria. Differences in glycophorin amino acid sequence determine whether a person has an MM, MN, or NN blood type.

The Erythrocyte Membrane Skeleton   The peripheral proteins of the erythrocyte plasma membrane are located on its internal surface and constitute a fibrillar membrane skeleton (Figure 4.32d,e) that plays a major role in determining the biconcave shape of the erythrocyte. As discussed on page 139, the membrane skeleton can establish domains within the membrane that enclose particular groups of membrane proteins and may greatly restrict the movement of these proteins. The major component of the skeleton is an elongated fibrous protein, called *spectrin*. Spectrin is a heterodimer approximately 100 nm long, consisting of an α and β subunit that curl around one another. Two such dimeric molecules are linked at their head ends to form a 200-nm-long filament that is both flexible and elastic. Spectrin is attached to the internal surface of the membrane by means of noncovalent bonds to another peripheral protein, *ankyrin* (the green spheres of Figure 4.32d), which in turn is linked noncovalently to the cytoplasmic domain of a band 3 molecule. As evident in Figures 4.32d and e, spectrin filaments are organized into hexagonal or pentagonal arrays. This two-dimensional network is constructed by linking both ends of each spectrin filament to a cluster of proteins that include a short filament of *actin* and *tropomyosin*, proteins typically involved in contractile activities. A number of genetic diseases (*hemolytic anemias*) characterized by fragile, abnormally shaped erythrocytes have been traced to mutations in ankyrin or spectrin.

If the peripheral proteins are removed from erythrocyte ghosts, the membrane becomes fragmented into small vesicles, indicating that the inner protein network is required to maintain the integrity of the membrane. Erythrocytes are circulating cells that are squeezed under pressure through microscopic capillaries whose diameter is considerably less than that of the erythrocytes themselves. To traverse these narrow passageways, and to do so day after day, the red blood cell must be highly deformable, durable, and capable of withstanding shearing forces that tend to pull it apart. The spectrin–actin network gives the cell the strength, elasticity, and pliability necessary to carry out its demanding function.

When first discovered, the membrane skeleton of the erythrocyte was thought to be a unique structure suited to the unique shape and mechanical needs of this cell type. However,

as other cells were examined, similar types of membrane skeletons containing members of the spectrin and ankyrin families have been revealed, indicating that inner membrane skeletons are widespread. Dystrophin, for example, is a member of the spectrin family that is found in the membrane skeleton of muscle cells. Mutations in dystrophin are responsible for causing muscular dystrophy, a devastating disease that cripples and kills children. As in the case of cystic fibrosis (page 157), the most debilitating mutations are ones that lead to a complete absence of the protein in the cell. The plasma membranes of muscle cells lacking dystrophin are apparently destroyed as a consequence of the mechanical stress exerted on them as the muscle contracts. As a result, the muscle cells die and eventually are no longer replaced.

REVIEW

1. Describe two techniques to measure the rates of diffusion of a specific membrane protein.
2. Compare and contrast the types of protein mobility depicted in Figure 4.28.
3. Discuss two major functions of the integral and peripheral proteins of the erythrocyte membrane.
4. Compare the rate of lateral diffusion of a lipid with that of flip-flop. What is the reason for the difference?

4.7 THE MOVEMENT OF SUBSTANCES ACROSS CELL MEMBRANES

Because the contents of a cell are completely surrounded by its plasma membrane, all communication between the cell and the extracellular medium must be mediated by this structure. In a sense, the plasma membrane has a dual function. On one hand, it must retain the dissolved materials of the cell so that they do not simply leak out into the environment, while on the other hand, it must allow the necessary exchange of materials into and out of the cell. The lipid bilayer of the membrane is ideally suited to prevent the loss of charged and polar solutes from a cell. Consequently, some special provision must be made to allow the movement of nutrients, ions, waste products, and other compounds, in and out of the cell. There are basically two means for the movement of substances through a membrane: passively by diffusion or actively by an energy-coupled transport process. Both types of movements lead to the net flux of a particular ion or compound. The term *net flux* indicates that the movement of the substance into the cell (*influx*) and out of the cell (*efflux*) is not balanced, but that one exceeds the other.

Several different processes are known by which substances move across membranes: simple diffusion through the lipid bilayer; simple diffusion through an aqueous, protein-lined channel; diffusion that is facilitated by a protein transporter; and active transport, which requires an energy-driven protein "pump" capable of moving substances against a concentration gradient (Figure 4.33). We will consider each in turn, but first we will discuss the energetics of solute movement.
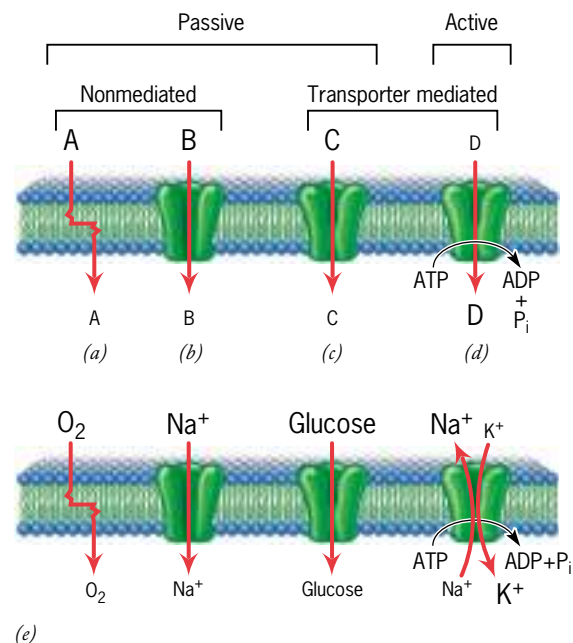
FIGURE 4.33 Four basic mechanisms by which solute molecules move across membranes. The relative sizes of the letters indicate the directions of the concentration gradients. (*a*) Simple diffusion through the bilayer, which always proceeds from high to low concentration. (*b*) Simple diffusion through an aqueous channel formed within an integral membrane protein or a cluster of such proteins. As in *a*, movement is always down a concentration gradient. (*c*) Facilitated diffusion in which solute molecules bind specifically to a membrane protein carrier (a facilitative transporter). As in *a* and *b*, movement is always from high to low concentration. (*d*) Active transport by means of a protein transporter with a specific binding site that undergoes a change in affinity driven with energy released by an exergonic process, such as ATP hydrolysis. Movement occurs against a concentration gradient. (*e*) Examples of each type of mechanism as it occurs in the membrane of an erythrocyte.

The Energetics of Solute Movement

Diffusion is a spontaneous process in which a substance moves from a region of high concentration to a region of low concentration, eventually eliminating the concentration difference between the two regions. As discussed on page 87, diffusion depends on the random thermal motion of solutes and is an exergonic process driven by an increase in entropy. We will restrict the following discussion to diffusion of substances across membranes.

The free-energy change when an uncharged solute (a nonelectrolyte) diffuses across a membrane depends on the magnitude of the concentration gradient, that is, the difference in concentration on each side of the membrane. The following relationship describes the movement of a nonelectrolyte *into* the cell:

$$\Delta G = RT \ln \frac{[C_i]}{[C_o]}$$

$$\Delta G = 2.303 \, RT \log_{10} \frac{[C_i]}{[C_o]}$$

where ΔG is the free-energy change (Section 3.1), R is the gas

constant, T is the absolute temperature, and $[C_i]/[C_o]$ is the ratio of the concentration of the solute on the inside (i) and outside (o) surfaces of the membrane. At 25°C,

$$\Delta G = 1.4 \text{ kcal/mol} \cdot \log_{10} \frac{[C_i]}{[C_o]}$$

If the ratio of $[C_i]/[C_o]$ is less than 1.0, then the log of the ratio is negative, ΔG is negative, and the net influx of solute is thermodynamically favored (exergonic). If, for example, the external concentration of solute is 10 times the internal concentration, $\Delta G = -1.4$ kcal/mole. Thus, the maintenance of a tenfold concentration gradient represents a storage of 1.4 kcal/mol. As solute moves into the cell, the concentration gradient decreases, the stored energy is dissipated, and ΔG decreases until, at equilibrium, ΔG is zero. (To calculate ΔG for movement of a solute out of the cell, the term for concentration ratios becomes $[C_o]/[C_i]$).

If the solute is an electrolyte (a charged species), the overall charge difference between the two compartments must also be considered. As a result of the mutual repulsion of ions of like charges, it is thermodynamically unfavorable for an electrolyte to move across a membrane from one compartment into another compartment having a net charge of the same sign. Conversely, if the charge of the electrolyte is opposite in sign to the compartment into which it is moving, the process is thermodynamically favored. The greater the difference in charge (the potential difference or voltage) between the two compartments, the greater the difference in free energy. Thus, the tendency of an electrolyte to diffuse between two compartments depends on two gradients: a chemical gradient, determined by the concentration difference of the substance between the two compartments, and the electric potential gradient, determined by the difference in charge. Together these differences are combined to form an **electrochemical gradient**. The free-energy change for the diffusion of an electrolyte into the cell is

$$\Delta G = RT \ln \frac{[C_i]}{[C_o]} + zF\Delta E_m$$

where z is the charge of the solute, F is the Faraday constant (23.06 kcal/V · equivalent, where an equivalent is the amount of the electrolyte having one mole of charge), and ΔE_m is the potential difference (in volts) between the two compartments. We saw in the previous example that a tenfold difference in concentration of a nonelectrolyte across a membrane at 25°C generates a ΔG of -1.4 kcal/mol. Suppose the concentration gradient consisted of Na^+ ions, which were present at tenfold higher concentration outside the cell than in the cytoplasm. Because the voltage across the membrane of a cell is typically about -70 mV (page 160), the free-energy change for the movement of a mole of Na^+ ions into the cell under these conditions would be

$$\Delta G = -1.4 \text{ kcal/mol} + zF\Delta E_m$$
$$\Delta G = -1.4 \text{ kcal/mol} + (1)(23.06 \text{ kcal/V} \cdot \text{mol})(-0.07 \text{ V})$$
$$= -3.1 \text{ kcal/mol}$$

Thus under these conditions, the concentration difference and the electric potential make similar contributions to the storage of free energy across the membrane.

The interplay between concentration and potential differences is seen in the diffusion of potassium ions (K^+) out of a cell. The efflux of the ion is favored by the K^+ concentration gradient, which has a higher K^+ concentration inside the cell, but hindered by the electrical gradient that its diffusion creates, which leaves a higher negative charge inside the cell. We will discuss this subject further when we consider the topic of membrane potentials and nerve impulses in Section 4.8.
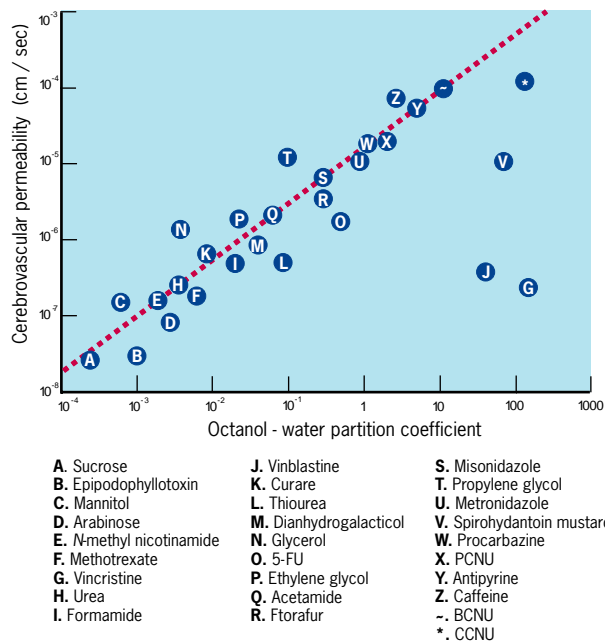
Diffusion of Substances through Membranes

Two qualifications must be met before a nonelectrolyte can diffuse passively across a plasma membrane. The substance must be present at higher concentration on one side of the membrane than the other, and the membrane must be permeable to the substance. A membrane may be permeable to a given solute either (1) because that solute can pass directly through the lipid bilayer, or (2) because that solute can traverse an aqueous pore that spans the membrane. Let us begin by considering the former route in which a substance must dissolve in the lipid bilayer on its way through the membrane.

Discussion of simple diffusion leads us to consider the polarity of a solute. One simple measure of the polarity (or nonpolarity) of a substance is its **partition coefficient**, which is the ratio of its solubility in a nonpolar solvent, such as octanol or a vegetable oil, to that in water under conditions where the nonpolar solvent and water are mixed together. Figure 4.34 shows the relationship between partition coefficient and membrane permeability of a variety of chemicals and drugs. It is evident that the greater the lipid solubility, the faster the penetration.

Another factor determining the rate of penetration of a compound through a membrane is its size. If two molecules have approximately equivalent partition coefficients, the smaller molecule tends to penetrate the lipid bilayer of a membrane more rapidly than the larger one. Very small, uncharged molecules penetrate very rapidly through cellular membranes. Consequently, membranes are highly permeable to small inorganic molecules, such as O_2, CO_2, NO, and H_2O, which are thought to slip between adjacent phospholipids. In contrast, larger polar molecules, such as sugars, amino acids, and phosphorylated intermediates, exhibit poor membrane penetrability. As a result, the lipid bilayer of the plasma membrane provides an effective barrier that keeps these essential metabolites from diffusing out of the cell. Some of these molecules (e.g., sugars and amino acids) must enter cells from the bloodstream, but they cannot do so by simple diffusion. Instead, special mechanisms must be available to mediate their penetration through the plasma membrane. The use of such mechanisms allows a cell to regulate the movement of substances across its surface barrier. We will return to this feature of membranes later.

The Diffusion of Water through Membranes Water molecules move much more rapidly through a cell membrane than do dissolved ions or small polar organic solutes, which are essentially nonpenetrating. Because of this difference in the penetrability of water versus solutes, membranes are said to be

A. Sucrose	J. Vinblastine	S. Misonidazole
B. Epipodophyllotoxin	K. Curare	T. Propylene glycol
C. Mannitol	L. Thiourea	U. Metronidazole
D. Arabinose	M. Dianhydrogalacticol	V. Spirohydantoin mustard
E. N-methyl nicotinamide	N. Glycerol	W. Procarbazine
F. Methotrexate	O. 5-FU	X. PCNU
G. Vincristine	P. Ethylene glycol	Y. Antipyrine
H. Urea	Q. Acetamide	Z. Caffeine
I. Formamide	R. Ftorafur	−. BCNU
		*. CCNU

FIGURE 4.34 The relationship between partition coefficient and membrane permeability. In this case, measurements were made of the penetration of a variety of chemicals and drugs across the plasma membranes of the cells that line the capillaries of the brain. Substances penetrate by passage through the lipid bilayer of these cells. The partition coefficient is expressed as the ratio of solubility of a solute in octanol to its solubility in water. Permeability is expressed as penetrance (P) in cm/sec. For all but a few compounds, such as vinblastine and vincristine, penetrance is directly proportional to lipid solubility. (FROM N. J. ABBOTT AND I. A. ROMERO, MOLEC. MED. TODAY, 2:110, 1996; COPYRIGHT 1996, WITH PERMISSION FROM ELSEVIER SCIENCE.)

semipermeable. Water moves readily through a semipermeable membrane from a region of lower *solute* concentration to a region of higher *solute* concentration. This process is called **osmosis**, and it is readily demonstrated by placing a cell into a solution containing a nonpenetrating solute at a concentration different than that present within the cell itself.

When two compartments of different solute concentration are separated by a semipermeable membrane, the compartment of higher solute concentration is said to be **hypertonic** (or **hyperosmotic**) relative to the compartment of lower solute concentration, which is described as being **hypotonic** (or **hypoosmotic**). When a cell is placed into a hypotonic solution, the cell rapidly gains water by osmosis and swells (Figure 4.35*a*). Conversely, a cell placed into a hypertonic solution rapidly loses water by osmosis and shrinks (Figure 4.35*b*). These simple observations show that a cell's volume is controlled by the difference between the solute concentration inside the cell and that in the extracellular medium. The swelling and shrinking of cells in slightly hypotonic and hypertonic media are usually only temporary events. Within a few minutes, the cells recover and return to their original volume. In a hypotonic medium, recovery occurs as the cells lose ions, thereby reducing their internal osmotic pressure. In a hypertonic medium, recovery occurs as the cells gain ions from the medium. Once the internal solute concentration (which includes a high concentration of dissolved proteins) equals the external solute concentration, the internal and external fluids are **isotonic** (or **isosmotic**), and no net movement of water into or out of the cells occurs (Figure 4.35*c*).

Osmosis is an important factor in a multitude of bodily functions. Your digestive tract, for example, secretes several liters of fluid daily, which is reabsorbed osmotically by the cells that line your intestine. If this fluid weren't reabsorbed, as happens in cases of extreme diarrhea, you would face the prospect of rapid dehydration. Plants utilize osmosis in differ-
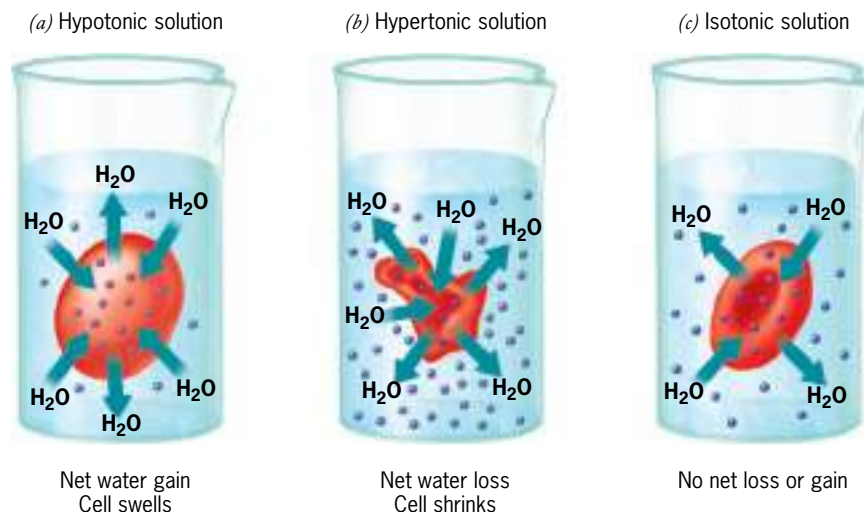
FIGURE 4.35 The effects of differences in the concentration of solutes on opposite sides of the plasma membrane. (*a*) A cell placed in a hypotonic solution (one having a lower solute concentration than the cell) swells because of a net gain of water by osmosis. (*b*) A cell in a hypertonic solution shrinks because of a net loss of water by osmosis. (*c*) A cell placed in an isotonic solution maintains a constant volume because the inward flux of water is equal to the outward flux.

ent ways. Unlike animal cells, which are generally isotonic with the medium in which they are bathed, plant cells are generally hypertonic compared to their fluid environment. As a result, there is a tendency for water to enter the cell, causing it to develop an internal (*turgor*) pressure that pushes against its surrounding wall (Figure 4.36*a*). Turgor pressure provides support for nonwoody plants and for the nonwoody parts of trees, such as the leaves. If a plant cell is placed into a hypertonic medium, its volume shrinks as the plasma membrane pulls away from the surrounding cell wall, a process called **plasmolysis** (Figure 4.36*b*). The loss of water due to plasmolysis causes plants to lose their support and wilt.

Many cells are much more permeable to water than can be explained by simple diffusion through the lipid bilayer. In the early 1990s, Peter Agre and colleagues at Johns Hopkins University were attempting to isolate and purify the mem-

brane proteins responsible for the Rh antigen on the surface of red blood cells. During this pursuit, they identified a protein they thought might be the long-sought water channel of the erythrocyte membrane. To test their hypothesis, they engineered frog oocytes to incorporate the newly discovered protein into their plasma membranes and then placed the oocytes in a hypotonic medium. Just as predicted, the oocytes swelled due to the influx of water and eventually burst. The team had discovered a family of small integral proteins, called *aquaporins*, that allow the passive movement of water from one side of the plasma membrane to the other. Each aquaporin subunit (in the four-subunit protein) contains a central channel that is lined primarily by hydrophobic amino acid residues and is highly specific for water molecules. A billion or so water molecules can pass—in single file—through each channel every second. At the same time, H^+ ions, which normally hop along a chain of water molecules, are not able to penetrate these open pores. The apparent mechanism by which these channels are able to exclude protons has been suggested by a combination of X-ray crystallographic studies, which has revealed the structure of the protein, and computer-based simulations (page 59), which has put this protein structure into operation. A model based on computer simulations is shown in Figure 4.37*a*. Very near its narrowest point, the wall of an aquaporin channel contains a pair of precisely positioned positive charges (residues N203 and N68 in Figure 4.37*b*) that attract the oxygen atom of each water molecule as it speeds through the constriction in the protein. This interaction reorients the central water molecule in a position that prevents it from maintaining the hydrogen bonds that normally link it to its neighboring water molecules. This removes the bridge that would normally allow protons to move from one water molecule to the next.

Aquaporins are particularly prominent in cells, such as those of a kidney tubule or plant root, where the passage of water plays a crucial role in the tissue's physiologic activities. The hormone vasopressin, which stimulates water retention by the collecting ducts of the kidney, acts by way of one of these proteins (AQP2). Some cases of the inherited disorder *congenital nephrogenic diabetes insipidus* arise from mutations in this aquaporin channel. Persons suffering from this disease excrete large quantities of urine because their kidneys do not respond to vasopressin.
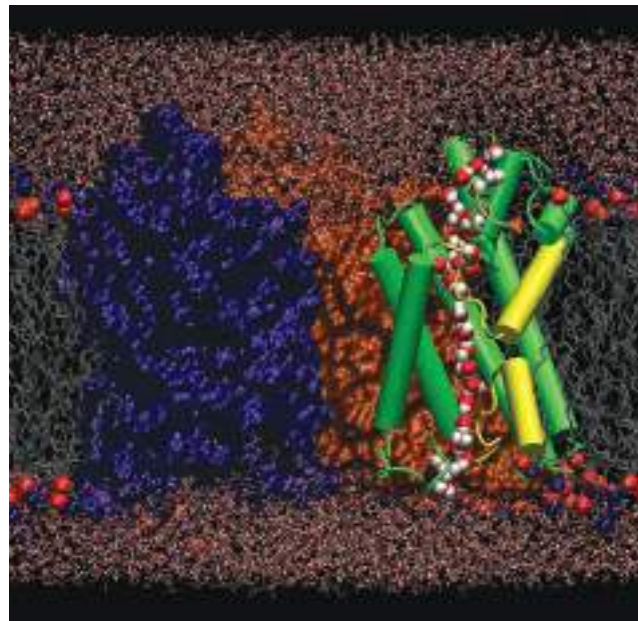
The Diffusion of Ions through Membranes   The lipid bilayer that constitutes the core of biological membranes is highly impermeable to charged substances, including small ions such as Na^+, K^+, Ca^{2+}, and Cl^-. Yet the rapid movement (**conductance**) of these ions across membranes plays a critical role in a multitude of cellular activities, including formation and propagation of a nerve impulse, secretion of substances into the extracellular space, muscle contraction, regulation of cell volume, and the opening of stomatal pores on plant leaves.

In 1955, Alan Hodgkin and Richard Keynes of Cambridge University first proposed that cell membranes contain **ion channels**, that is, openings in the membrane that are permeable to specific ions. During the late 1960s and 1970s,

HYPOTONIC:
Normal turgor pressure

(a)

Plasmolysis

H_2O

HYPERTONIC:
No turgor pressure

(b)

FIGURE 4.36 **The effects of osmosis on a plant cell.** (*a*) Aquatic plants living in freshwater are surrounded by a hypotonic environment. Water therefore tends to flow into the cells, creating turgor pressure. (*b*) If the plant is placed in a hypertonic solution, such as seawater, the cell loses water, and the plasma membrane pulls away from the cell wall. (© ED RESCHKE.)
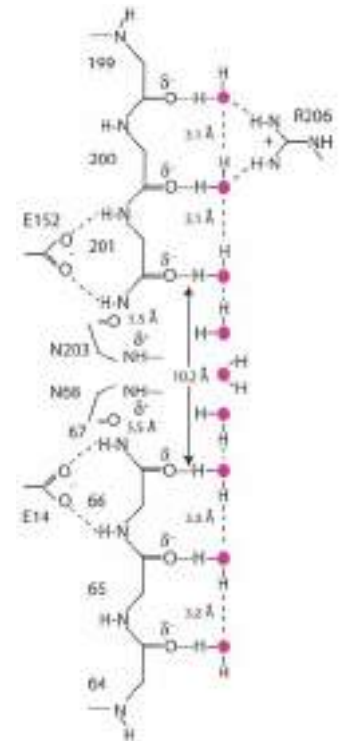
FIGURE 4.37 **Passage of water molecules through an aquaporin channel.** (*a*) Snapshot from a molecular dynamics simulation of a stream of water molecules (red and white spheres) passing in single file through the channel in one of the subunits of an aquaporin molecule residing within a membrane. (*b*) A model describing the mechanism by which water molecules pass through an aquaporin channel with the simultaneous exclusion of protons. Nine water molecules are shown to be lined up in single file along the wall of the channel. Each water molecule is depicted as a red circular O atom with two associated Hs. In this model, the four water molecules at the top and bottom of the channel are oriented, as the result of their interaction with the carbonyl (C=O) groups of the protein backbone (page 50), with their H atoms pointed away from the center of the channel. These water molecules are able to form hydrogen bonds (dashed lines) with their neighbors. In contrast, the single water molecule in the center of the channel is oriented in a position that prevents it from forming hydrogen bonds with other water molecules, which has the effect of interrupting the flow of protons through the channel. Animations of aquaporin channels can be found at www.nobelprize.org/nobel_prizes/chemistry/laureates/200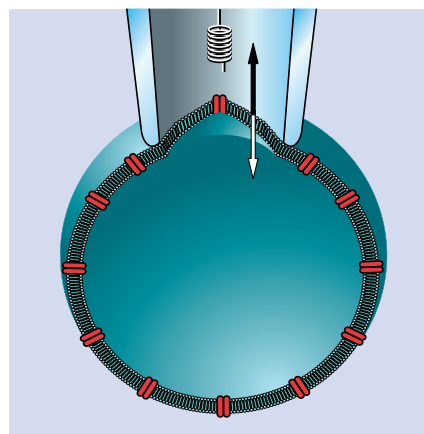3/animations.html. (A: REPRINTED FROM BENOIT ROUX AND KLAUS SCHULTEN, STRUCTURE 12:1344, 2004, BY PERMISSION OF CELL PRESS; B: REPRINTED FROM R. M. STROUD ET AL., CURR. OPIN. STRUCT. BIOL. 13:428, 2003. COPYRIGHT 2003, WITH PERMISSION FROM ELSEVIER SCIENCE.)

(a)

(b)

Bertil Hille of the University of Washington and Clay Armstrong of the University of Pennsylvania began to obtain evidence for the existence of such channels. The final "proof" emerged through the work of Bert Sakmann and Erwin Neher at the Max-Planck Institute in Germany in the late 1970s and early 1980s who developed techniques to monitor the ionic current passing through a single ion channel. This is accomplished using very fine micropipette-electrodes made of polished glass that are placed on the outer cell surface and sealed to the membrane by suction. The voltage across the membrane can be maintained (*clamped*) at any particular value, and the current originating in the small patch of membrane surrounded by the pipette can be measured (Figure 4.38). These landmark studies marked the first successful investigations into the activities of individual protein molecules. Today, biologists have identified a bewilder-

FIGURE 4.38 **Measuring ion channel conductance by patch-clamp recording.** (*a*) In this technique, a highly polished glass micropipette is placed against a portion of the outer surface of a cell, and suction is applied to seal the rim of the pipette against the plasma membrane. Because the pipette is wired as an electrode (a *microelectrode*), a voltage can be applied across the patch of membrane enclosed by the pipette, and the responding flow of ions through the membrane channels can be measured. As indicated in the figure, the micropipette can enclose a patch of membrane containing a single ion channel, which allows investigators to monitor the opening and closing of a single gated channel, as well as its conductance at different applied voltages. (*b*) The micrograph shows patch-clamp recordings being made from a single photoreceptor cell of the retina of a salamander. One portion of the cell is drawn into a glass micropipette by suction, while a second micropipette-electrode (lower right) is sealed against a small patch of the plasma membrane on another portion of the cell. (B: FROM T. D. LAMB, H. R. MATTHEWS, AND V. TORRE, J. PHYSIOLOGY 372:319, 1986. REPRODUCED WITH PERMISSION.)

(a)

(b)   35 μm

ing variety of ion channels, each formed by integral membrane proteins that enclose a central aqueous pore. As might be predicted, mutations in the genes encoding ion channels can lead to many serious diseases (see Table 1 of the Human Perspective, page 156).

Most ion channels are highly selective in allowing only one particular type of ion to pass through the pore. As with the passive diffusion of other types of solutes across membranes, the diffusion of ions through a channel is always downhill, that is, from a state of higher energy to a state of lower energy. Most of the ion channels that have been identified can exist in either an open or a closed conformation; such channels are said to be **gated**. The opening and closing of the gates are subject to complex physiologic regulation and can be induced by a variety of factors depending on the particular channel. Three major categories of gated channels are distinguished:

1. **Voltage-gated channels** whose conformational state depends on the difference in ionic charge on the two sides of the membrane.

2. **Ligand-gated channels** whose conformational state depends on the binding of a specific molecule (the ligand), which is usually not the solute that passes through the channel. Some ligand-gated channels are opened (or closed) following the binding of a molecule to the outer surface of the channel; others are opened (or closed) following the binding of a ligand to the inner surface of the channel. For example, neurotransmitters, such as acetylcholine, act on the outer surface of certain cation channels, while cyclic nucleotides, such as cAMP, act on the inner surface of certain calcium ion channels.

3. **Mechano-gated channels** whose conformational state depends on mechanical forces (e.g., stretch tension) that are applied to the membrane. Members of one family of cation channels, for example, are opened by the movements of stereocilia (see Figure 9.54) on the hair cells of the inner ear in response to sound or motions of the head.

We will focus in the following discussion on the structure and function of voltage-gated potassium ion channels because these are the best understood.

In 1998, Roderick MacKinnon and his colleagues at Rockefeller University provided the first atomic-resolution image of an ion channel protein, in this case, a bacterial K^+ ion channel called KcsA. The relationship between structure and function is evident everywhere in the biological world, but it would be difficult to find a better example than that of the K^+ ion channel depicted in Figure 4.39. As we will see shortly, the formulation of this structure led directly to an understanding of the mechanism by which these remarkable molecular machines are able to select overwhelmingly for K^+ ions over Na^+ ions, yet at the same time allow an incredibly rapid conductance of K^+ ions through the membrane. We will also see that the mechanisms of ion selectivity and conductance in this bacterial channel are virtually identical to those operating in the much larger mammalian channels. Evidently, the basic challenges in operating an ion channel were solved
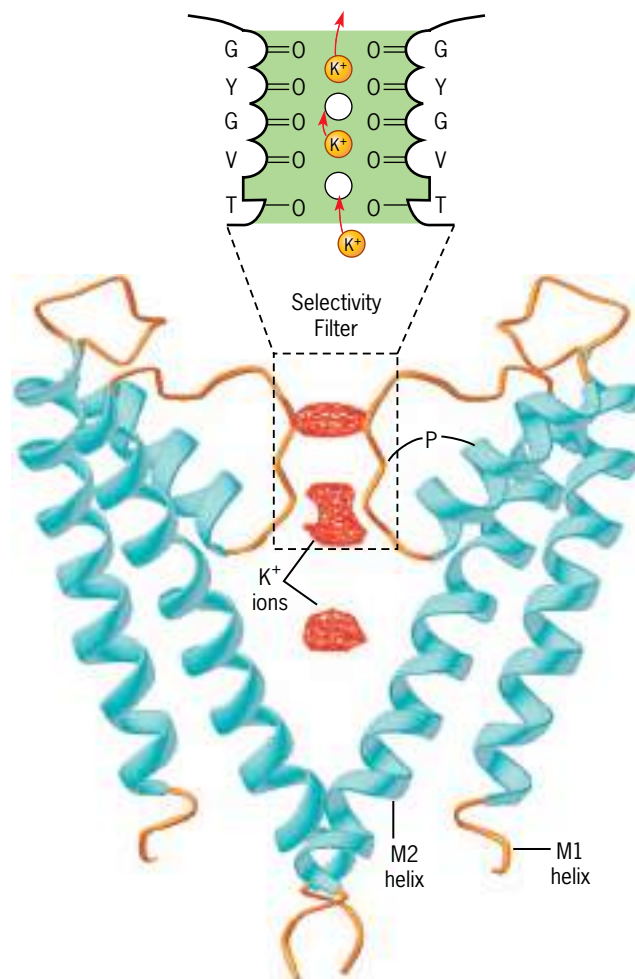
FIGURE 4.39 Three-dimensional structure of the bacterial KcsA channel and the selection of K^+ ions. This K^+ ion channel consists of four subunits, two of which are shown here. Each subunit is comprised of M1 and M2 helices joined by a P (pore) segment consisting of a short helix and a nonhelical portion that lines the channel through which the ions pass. A portion of each P segment contains a conserved pentapeptide (GYGVT) whose residues line the selectivity filter that screens for K^+ ions. The oxygen atoms of the carbonyl groups of these residues project into the channel where they can interact selectively with K^+ ions (indicated by the red mesh objects) within the filter. As indicated in the top inset, the selectivity filter contains four rings of carbonyl O atoms and one ring of threonyl O atoms; each of these five rings contains four O atoms, one donated by each subunit. The diameter of the rings is just large enough so that eight O atoms can coordinate a single K^+ ion, replacing its normal water of hydration. Although four K^+ binding sites are shown, only two are occupied at one time. (FROM RODERICK MACKINNON, REPRINTED WITH PERMISSION FROM NATURE MED. 5:1108, 1999; COPYRIGHT 1999, MACMILLAN MAGAZINES LIMITED.)

relatively early in evolution, although many refinements appeared over the following one or two billion years.

The KcsA channel consists of four subunits, two of which are shown in Figure 4.39. Each subunit of Figure 4.39 is seen to contain two membrane-spanning helices (M1 and M2) and a pore region (P) at the extracellular end of the channel. P

consists of a short pore helix that extends approximately one-third the width of the channel and a nonhelical loop (colored light brown in Figure 4.39) that forms the lining of a narrow *selectivity filter*, so named because of its role in allowing only the passage of K^+ ions.

The lining of the selectivity filter contains a highly conserved pentapeptide—Gly-Tyr-Gly-Val-Thr (or GYGVT in single-letter nomenclature). The X-ray crystal structure of the KcsA channel shows that the backbone carbonyl (C=O) groups from the conserved pentapeptide (see the backbone structure on page 50) create five successive rings of oxygen atoms (four rings are made up of carbonyl oxygens from the polypeptide backbone, and one ring consists of oxygen atoms from the threonine side chain). Each ring contains four oxygen atoms (one from each subunit) and has a diameter of approximately 3 Å, which is slightly larger than the 2.7 Å diameter of a K^+ ion that has lost its normal shell of hydration. Consequently, the electronegative O atoms that line the selectivity filter can substitute for the shell of water molecules that are displaced as each K^+ ion enters the pore. In this model, the selectivity filter contains four potential K^+ ion binding sites. As indicated in the top inset of Figure 4.39, a K^+ ion bound at any of these four sites would occupy the center of a "box" having four O atoms in a plane above the ion and four O atoms in a plane below the atom. As a result, each K^+ ion in one of these sites could coordinate with eight O atoms of the selectivity filter. Whereas the selectivity filter is a precise fit for a dehydrated K^+ ion, it is much larger than the diameter of a dehydrated Na^+ ion (1.9 Å). Consequently, a Na^+ ion cannot interact optimally with the eight oxygen atoms necessary to stabilize it in the pore. As a result, the smaller Na^+ ions cannot overcome the higher energy barrier required to penetrate the pore.

Although there are four potential K^+ ion binding sites, only two are occupied at any given time. Potassium ions are thought to move, two at a time—from sites 1 and 3 to sites 2 and 4—as indicated in the top inset of Figure 4.39. The entry of a third K^+ ion into the selectivity filter creates an electrostatic repulsion that ejects the ion bound at the opposite end of the line. Studies indicate that there is virtually no energy barrier for an ion to move from one binding site to the next, which accounts for the extremely rapid flow of ions across the membrane. Taken together, these conclusions concerning K^+ ion selectivity and conductance provide a superb example of how much can be learned about biological function through an understanding of molecular structure.

The KcsA channel depicted in Figure 4.39 has a gate, just like eukaryotic channels. The opening of the gate of the KcsA channel in response to very low pH was illustrated in Figure 4.22. The structure of KcsA shown in Figure 4.39 is actually the closed conformation of the protein (despite the fact that it contains ions in its channel). It has not been possible to crystallize the KcsA channel in its open conformation, but the structure of a homologous prokaryotic K^+ channel (called MthK) in the open conformation has been crystallized and its structure determined. Comparison of the open structure of MthK and the closed structure of the homologous protein KcsA strongly
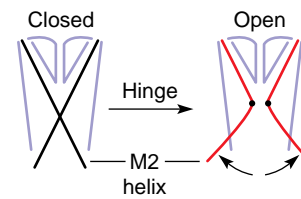
FIGURE 4.40 Schematic illustration of the hinge-bending model for the opening of the bacterial KcsA channel. The M2 helices from each subunit bend outward at a specific glycine residue, which opens the intracellular end of the channel to K^+ ions. (FROM B. L. KELLY AND A. GROSS, REPRINTED WITH PERMISSION FROM NATURE STRUCT. BIOL. 10:280, 2003; COPYRIGHT 2003, MACMILLAN MAGAZINES LIMITED.)

suggested that gating of these molecules is accomplished by conformational changes of the cytoplasmic ends of the inner (M2) helices. In the closed conformation, as seen in Figure 4.39 and Figure 4.40, left drawing, the M2 helices are straight and cross over one another to form a "helix bundle" that seals the cytoplasmic face of the pore. In the model shown in Figure 4.40, the channel opens when the M2 helices bend at a specific hinge point where a glycine residue is located.

Now that we have seen how these prokaryotic K^+ channels operate, we are in a better position to understand the structure and function of the more complex eukaryotic versions, which are thought to perform in a similar manner. Genes that encode a variety of distinct voltage-gated K^+ (or Kv) channels have been isolated and the molecular anatomy of their proteins scrutinized. The Kv channels of plants play an important role in salt and water balance and in regulation of cell volume. The Kv channels of animals are best known for their role in muscle and nerve function, which is explored at the end of the chapter. Eukaryotic Kv channel subunits contain six membrane-associated helices, named S1–S6, which are shown two-dimensionally in Figure 4.41. These six helices can be grouped into two functionally distinct domains:

1. a *pore domain*, which has the same basic architecture as that of the entire bacterial channel illustrated in Figure 4.39 and contains the selectivity filter that permits the selective passage of K^+ ions. Helices M1 and M2 and the P segment of the KcsA channel of Figure 4.39 are homologous to helices S5 and S6 and the P segment of the voltage-gated eukaryotic channel illustrated in Figure 4.41. Like the four M2 helices of KcsA, the four S6 helices line much of the pore, and their configuration determines whether the gate to the channel is open or closed.

2. a *voltage-sensing domain* consisting of helices S1–S4 that senses the voltage across the plasma membrane (as discussed below).
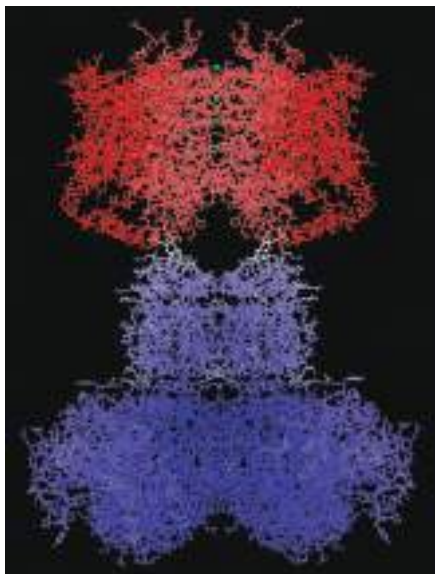
The three-dimensional crystal structure of a complete eukaryotic Kv channel purified from rat brain is shown in Figure 4.42. Determination of this structure was made possible by use of a mixture of detergent and lipid throughout the purification and crystallization process. The presence of negatively charged phospholipids is thought to be important in

FIGURE 4.41 **The structure of one subunit of a eukaryotic, voltage-gated K⁺ channel.** A two-dimensional portrait of a K⁺ channel subunit showing its six transmembrane helices and a portion of the polypeptide (called the pore helix or P) that dips into the protein to form part of the channel's wall. The inset shows the sequence of amino acids of the positively charged S4 helix of the *Drosophila* K⁺ *Shaker* ion channel, which serves as a voltage sensor. The positively charged side chains are situated at every third residue along the otherwise hydrophobic helix. This member of the Kv family is called a *Shaker* channel because flies with certain mutations in the protein shake vigorously when anesthetized with ether. The *Shaker* channel was the first K⁺ channel to be identified and cloned in 1987.
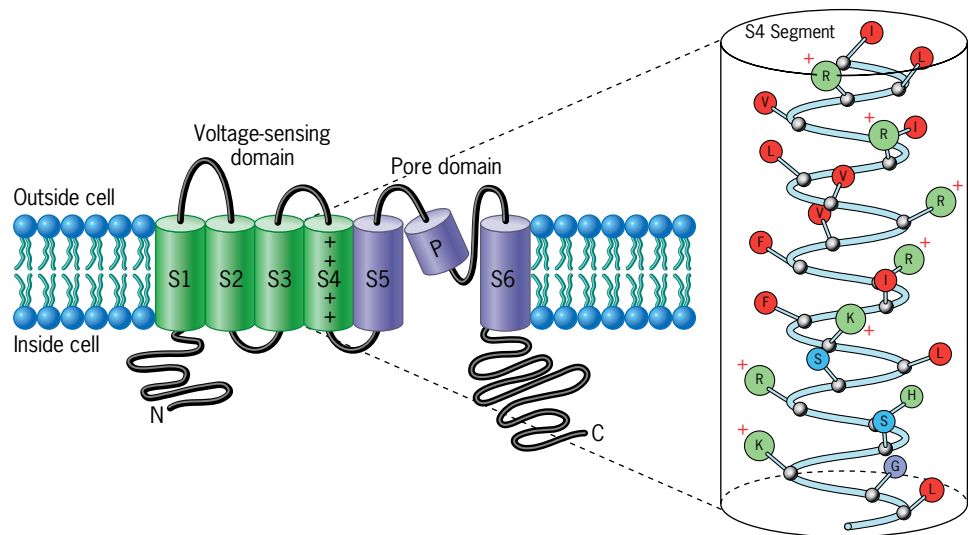

maintaining the native structure of the membrane protein and promoting its function as a voltage-gated channel. Like the KcsA channel, a single eukaryotic Kv channel consists of four homologous subunits arranged symmetrically around the central ion-cond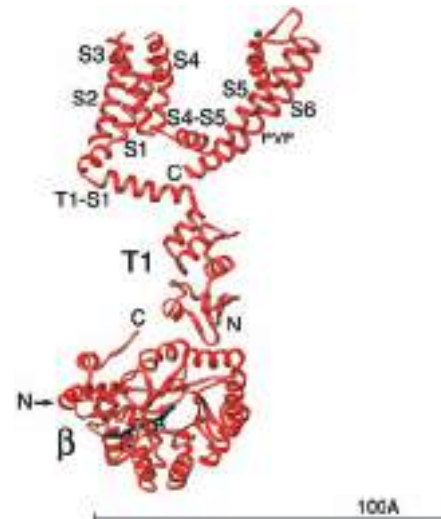ucting pore. The selectivity filter, and thus the presumed mechanism of K⁺ ion selection, is virtually identical in the prokaryotic KcsA and eukaryotic Kv proteins. The gate leading into a Kv channel is formed by the inner ends of the S6 helices and is thought to open and close in a manner roughly similar to that of the M2 helices of the bacterial
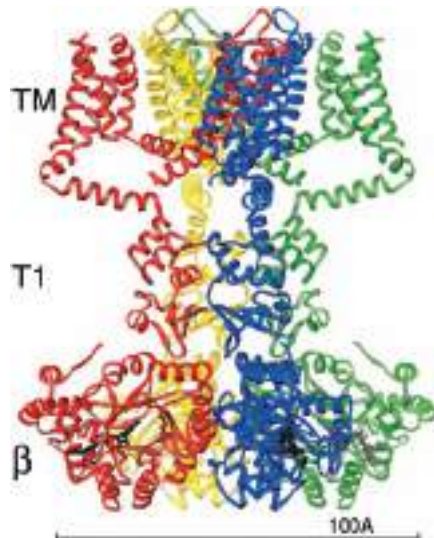
(a) (b) (c)

FIGURE 4.42 **Three-dimensional structure of a voltage-gated mammalian K⁺ channel.** (*a*) The crystal structure of the entire tetrameric Kv1.2 channel, a member of the *Shaker* family of K⁺ ion channels found in nerve cells of the brain. The transmembrane portion is shown in red, and the cytoplasmic portion in blue. The potassium ion binding sites are indicated in green. (*b*) Ribbon drawing of the same channel shown in *a*, with the four subunits that make up the channel shown in different colors. If you focus on the red subunit, you can see (1) the spatial separation between the voltage-sensing and pore domains of the subunit and (2) the manner in which the voltage-sensing domain from each subunit is present on the outer edge of the pore domain of a neighboring subunit. The cytoplasmic portion of this particular channel consists of a T1 domain, which is part of the channel polypeptide itself, and a separate β polypeptide. (*c*) Ribbon drawing of a single subunit showing the spatial orientation of the six membrane-spanning helices (S1–S6) and also the presence of the S4–S5 linker helix, which connects the voltage-sensing and pore domains. This linker transmits the signal from the S4 voltage sensor that opens the channel. The inner surface of the channel below the pore domain is lined by the S6 helix (roughly similar to the M2 helix of the bacterial channel shown in Figure 4.39). The channel shown here is present in the open configuration with the S6 helices curved outward (compare to Figure 4.40) at the site marked PVP (standing for Pro-Val-Pro, which is likely the amino acid sequence of the "hinge"). (REPRINTED WITH PERMISSION FROM STEPHEN B. LONG ET AL., SCIENCE 309:867, 899, 2005, COURTESY OF RODERICK MACKINNON; COPYRIGHT 2005, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

channel (shown in Figure 4.40). The protein depicted in Figure 4.42 represents the open state of the channel.

The S4 helix, which contains several positively charged amino acid residues spaced along the polypeptide chain (inset of Figure 4.41), acts as the key element of the voltage sensor. The voltage-sensing domain is seen to be connected to the pore domain by a short linker helix denoted as S4-S5 in the model in Figure 4.42. Under resting conditions, the negative potential across the membrane (page 160) keeps the gate closed. A change in the potential to a more positive value (a depolarization, page 161) exerts an electric force on the S4 helix. This force is thought to cause the transmembrane S4 helix to move in such a way that its positively charged residues shift from a position where they were exposed to the cytoplasm to a new position where they are exposed to the outside of the cell. Voltage sensing is a dynamic process whose mechanism cannot be resolved by a single static view of the protein such as that shown in Figure 4.42. In fact, several competing models describing the mechanism of action of the voltage sensor are currently debated. However it occurs, the movement of the S4 helix in response to membrane depolarization initiates a series of conformational changes within the protein that opens the gate at the cytoplasmic end of the channel.

Once opened, more than ten million potassium ions can pass through the channel per second, which is nearly the rate that would occur by free diffusion in solution. Because of the large ion flux, the opening of a relatively small number of K^+ channels has significant impact on the electrical properties of the membrane. After the channel is open for a few milliseconds, the movement of K^+ ions is "automatically" stopped by a process known as inactivation. To understand channel inactivation, we have to consider an additional portion of a Kv channel besides the two transmembrane domains discussed above.

Eukaryotic Kv channels typically contain a large cytoplasmic structure whose composition varies among different channels. As indicated in Figure 4.43a inactivation of the channel is accomplished by movement of a small inactivation peptide that dangles from the cytoplasmic portion of the protein. The inactivation peptide is thought to gain access to the cytoplasmic mouth of the pore by snaking its way through one of four "side windows" indicated in the figure. When one of these dangling peptides moves up into the mouth of the pore (Figure 4.43a), the passage of ions is blocked, and the channel is inactivated. At a subsequent stage of the cycle, the inactivation peptide is released, and the gate to the channel is closed. It follows from this discussion that the potassium channel can exist in three different states—open, inactivated, and closed—which are illustrated schematically in Figure 4.43b.

Potassium channels come in many different varieties. It is remarkable that *C. elegans*, a nematode worm whose body consists of only about 1000 cells, contains more than 90 different genes that encode K^+ channels. It is evident that a single cell—whether in a nematode, human, or plant—is likely to possess a variety of different K^+ channels that open and close in response to different voltages. In addition, the voltage required to open or close a particular K^+ channel can vary depending on whether or not the channel protein is phosphorylated, which in turn is regulated by hormones and other factors. It is appar-
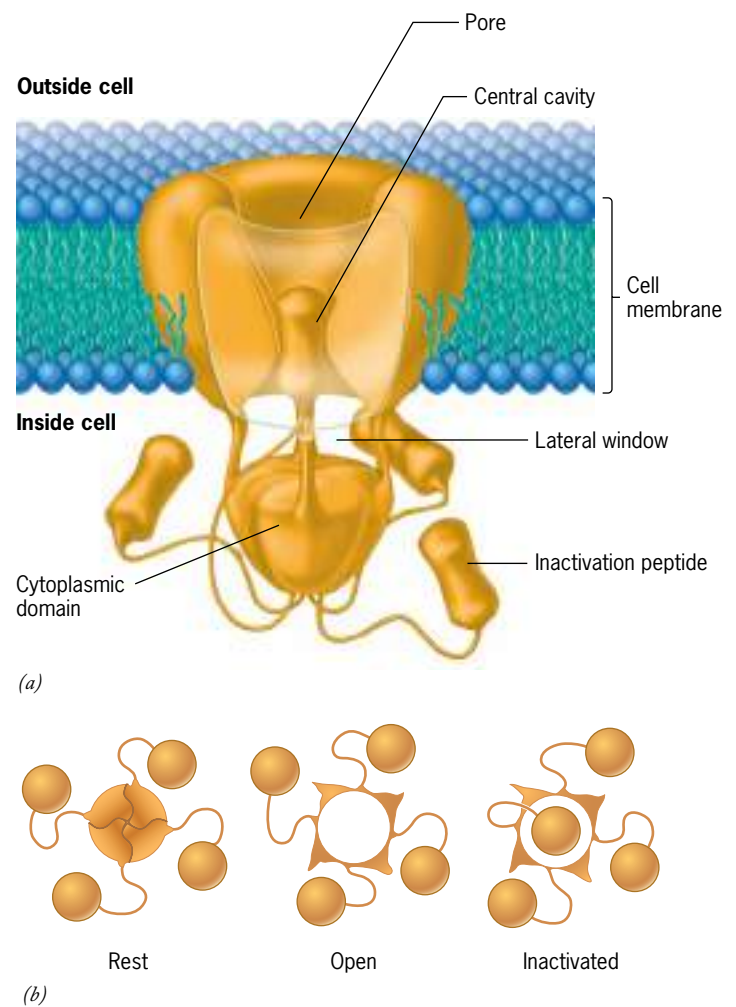
FIGURE 4.43 Conformational states of a voltage-gated K^+ ion channel. (*a*) Three-dimensional model of a eukaryotic K^+ ion channel. Inactivation of channel activity occurs as one of the inactivation peptides, which dangle from the cytoplasmic portion of the complex, fits into the cytoplasmic opening of the channel. (*b*) Schematic representation of a view into a K^+ ion channel, perpendicular to the membrane from the cytoplasmic side, showing the channel in the closed (resting), open, and inactivated state. (B: REPRINTED FROM NEURON, VOL. 20, C. M. ARMSTRONG AND B. HILLE, VOLTAGE-GATED ION CHANNELS AND ELECTRICAL EXCITABILITY, PAGE 377; COPYRIGHT 1998, WITH PERMISSION FROM ELSEVIER SCIENCE.)

ent that ion channel function is under the control of a diverse and complex set of regulatory agents. The structure and function of a very different type of ion channel, the ligand-gated nicotinic acetylcholine receptor, is the subject of the Experimental Pathways section at the end of this chapter.

Facilitated Diffusion

Substances always diffuse across a membrane from a region of higher concentration on one side to a region of lower concentration on the other side, but they do not always diffuse through the lipid bilayer or through a channel. In many cases, the diffusing substance first binds selectively to a membrane-spanning protein, called a **facilitative transporter,** that facilitates the

diffusion process. The binding of the solute to the facilitative transporter on one side of the membrane is thought to trigger a conformational change in the protein, exposing the solute to the other surface of the membrane, from where it can diffuse down its concentration gradient. This mechanism is illustrated in Figure 4.44. Because they operate passively, that is, without being coupled to an energy-releasing system, facilitated transporters can mediate the movement of solutes equally well in both directions. The direction of net flux depends on the relative concentration of the substance on the two sides of the membrane.

Facilitated diffusion, as this process is called, is similar in many ways to an enzyme-catalyzed reaction. Like enzymes, facilitative transporters are specific for the molecules they transport, discriminating, for example, between D and L stereoisomers (page 43). In addition, both enzymes and transporters exhibit saturation-type kinetics (Figure 4.45). Unlike ion channels, which can conduct millions of ions per second, most facilitative transporters can move only hundreds to thousands of solute molecules per second across the membrane. Another important feature of facilitative transporters is that, like enzymes and ion channels, their activity can be regulated. Facilitated diffusion is particularly important in mediating the entry and exit of polar solutes, such as sugars and amino acids, that do not penetrate the lipid bilayer. This is illustrated in the following section.

The Glucose Transporter: An Example of Facilitated Diffusion

Glucose is the body's primary source of direct energy, and most mammalian cells contain a membrane protein that facilitates the diffusion of glucose from the bloodstream into the cell (as depicted in Figures 4.44 and 4.49). A gradient favoring the continued diffusion of glucose into the cell is maintained by phosphorylating the sugar after it enters the cytoplasm, thus lowering the intracellular glucose concentration. Humans have at least five related proteins (isoforms) that act as facilitative glucose transporters. These isoforms, termed GLUT1 to GLUT5, are distinguished by the tissues in which they are located, as well as their kinetic and regulatory characteristics.

Insulin is a hormone produced by endocrine cells of the pancreas and plays a key role in maintaining proper blood sugar levels. An increase in blood glucose levels triggers the secretion of insulin, which stimulates the uptake of glucose into various target cells, most notably skeletal muscle and fat cells (adipocytes). Insulin-responsive cells share a common isoform of the facilitative glucose transporter, specifically GLUT4. When insulin levels are low, these cells contain relatively few glucose transporters on their plasma membrane. Instead, the transporters are present within the membranes of cytoplasmic vesicles. Rising insulin levels act on target cells to stimulate the fusion of the cytoplasmic vesicles to the plasma membrane, which moves transporters to the cell surface where they can bring glucose into the cell (see Figure 15.24).

Active Transport

Life cannot exist under equilibrium conditions (page 91). Nowhere is this more apparent than in the imbalance of ions across the plasma membrane. The differences in concentration of the major ions between the outside and inside of a typical
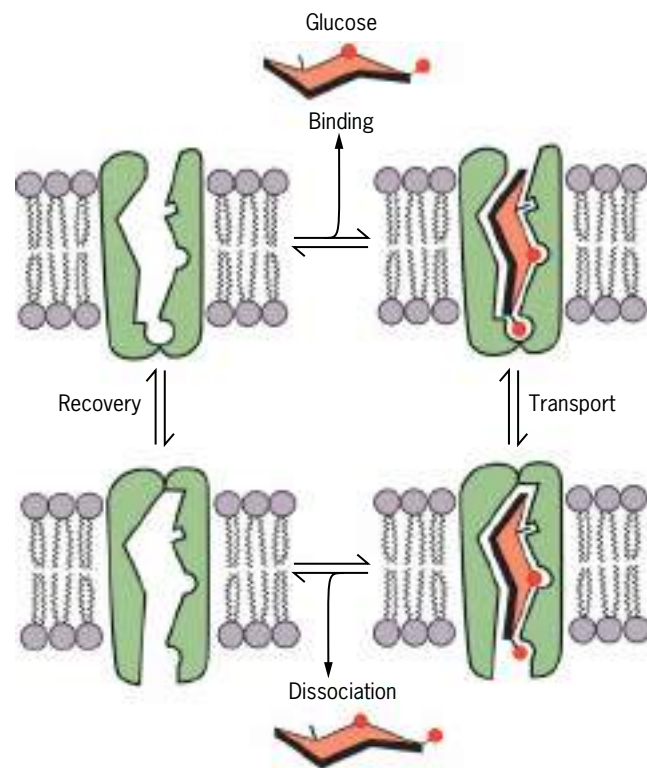
FIGURE 4.44 Facilitated diffusion. A schematic model for the facilitated diffusion of glucose depicts the alternating conformation of a carrier that exposes the glucose binding site to either the inside or outside of the membrane. (AFTER S. A. BALDWIN AND G. E. LIENHARD, TRENDS BIOCHEM. SCI. 6:210, 1981.)

mammalian cell are shown in Table 4.3. The ability of a cell to generate such steep concentration gradients across its plasma membrane cannot occur by either simple or facilitated diffusion. Rather, these gradients must be generated by **active transport**.

Like facilitated diffusion, active transport depends on integral membrane proteins that selectively bind a particular solute and move it across the membrane in a process driven by changes in the protein's conformation. Unlike facilitated diffusion, however, movement of a solute against a gradient
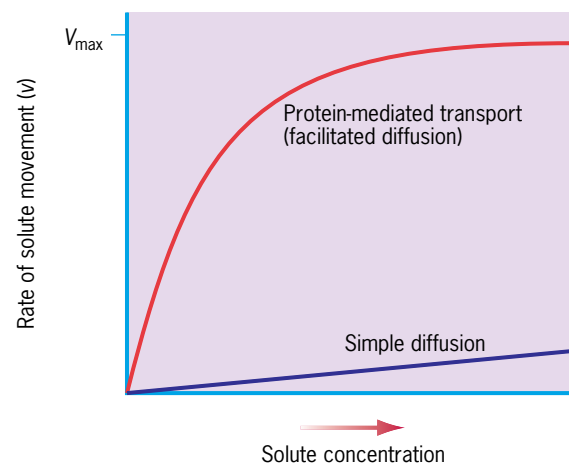
FIGURE 4.45 The kinetics of facilitated diffusion as compared to that of simple physical diffusion.

TABLE 4.3 Ion Concentrations Inside and Outside of a Typical Mammalian Cell

	Extracellular concentration	Intracellular concentration	Ionic gradient
Na^+	150 mM	10 mM	15×
K^+	5 mM	140 mM	28×
Cl^-	120 mM	10 mM	12×
Ca^{2+}	10^{-3} M	10^{-7} M	10,000×
H^+	$10^{-7.4}$ M (pH of 7.4)	$10^{-7.2}$ M (pH of 7.2)	Nearly 2×

The ion concentrations for the squid axon are given on page 172.

requires the coupled input of energy. Consequently, the endergonic movement of ions or other solutes across the membrane against a concentration gradient is coupled to an exergonic process, such as the hydrolysis of ATP, the absorbance of light, the transport of electrons, or the flow of other substances down their gradients. Proteins that carry out active transport are often referred to as "pumps."

Coupling Active Transport to ATP Hydrolysis   In 1957, Jens Skou, a Danish physiologist, discovered an ATP-hydrolyzing

enzyme in the nerve cells of a crab that was only active in the presence of both Na^+ and K^+ ions. Skou proposed, and correctly so, that this enzyme, which was responsible for ATP hydrolysis, was the same protein that was active in transporting the two ions; the enzyme was called the Na^+/K^+-ATPase, or the *sodium–potassium pump*.

Unlike the protein-mediated movement of a facilitated diffusion system, which will carry the substance equally well in either direction, active transport drives the movement of ions in only one direction. It is the Na^+/K^+-ATPase that is responsible for the large excess of Na^+ ions outside of the cell and the large excess of K^+ ions inside the cell. The positive charges carried by these two cations are balanced by negative charges carried by various anions so that the extracellular and intracellular compartments are, for the most part, electrically neutral. Cl^- ions are present at greater concentration outside of cells, where they balance the extracellular Na^+ ions. The abundance of intracellular K^+ ions is balanced primarily by excess negative charges carried by proteins and nucleic acids.

A large number of studies have indicated that the ratio of Na^+:K^+ pumped by the Na^+/K^+-ATPase is not 1:1, but 3:2 (see Figure 4.46). In other words, for each ATP hydrolyzed, three sodium ions are pumped out as two potassium ions are pumped in. Because of this pumping ratio, the

(a)   (b)
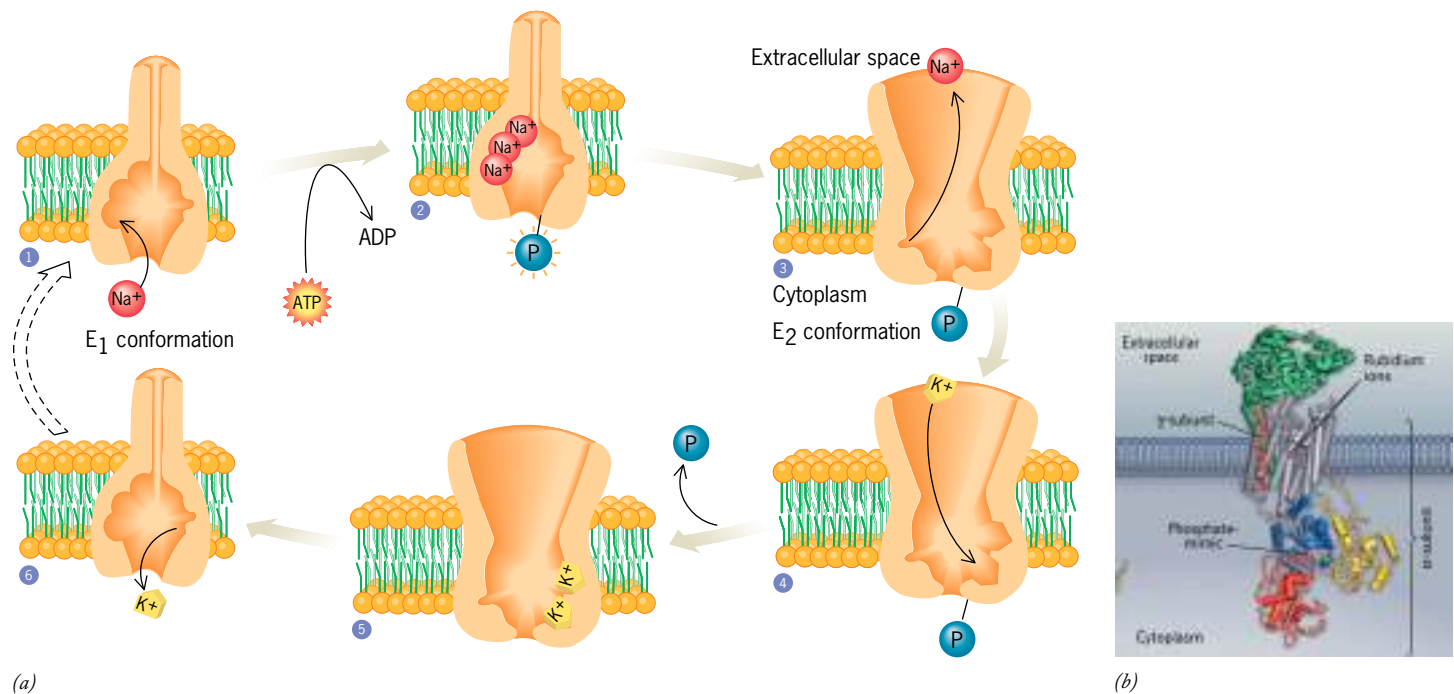
FIGURE 4.46 The Na^+/K^+-ATPase. (*a*) Simplified schematic model of the transport cycle. Sodium ions (1) bind to the protein on the inside of the membrane. ATP is hydrolyzed, and the phosphate is transferred to the protein (2), changing its conformation (3) and allowing sodium ions to be expelled to the external space. Potassium ions then bind to the protein (4), and the phosphate group is subsequently lost (5), which causes the protein to snap back to its original conformation, allowing the potassium ions to diffuse into the cell (6). The cation binding sites are located deep within the transmembrane domain, which consists of 10 membrane-spanning helices. Note that the actual

Na^+/K^+-ATPase is composed of at least two different membrane-spanning subunits: a larger α subunit, which carries out the transport activity, and a smaller β subunit, which functions primarily in the maturation and assembly of the pump within the membrane. A third (γ) subunit may also be present. (Steps where ATP binds to the protein prior to hydrolysis are not included.) (*b*) A model of the E_2 conformation of the protein based on a recent X-ray crystallographic study. The two rubidium ions are located where the potassium ions would normally be bound. (B: FROM AYAKO TAKEUCHI COURTESY OF DAVID C. GADSBY, NATURE 450:958, 2007, COPYRIGHT 2007, MACMILLAN MAGAZINES LIMITED.)

Na^+/K^+-ATPase is *electrogenic*, which means that it contributes directly to the separation of charge across the membrane. The Na^+/K^+-ATPase is an example of a P-type ion pump. The "P" stands for phosphorylation, indicating that, during the pumping cycle, the hydrolysis of ATP leads to the transfer of the released phosphate group to an aspartic acid residue of the transport protein, which in turn causes an essential conformational change within the protein. Conformational changes are necessary to change the affinity of the protein for the two cations that are transported. Consider the activity of the protein. It must pick up sodium or potassium ions from a region of low concentration, which means that the protein must have a relatively high affinity for the ions. Then the protein must release the ions on the other side of the membrane into a much greater concentration of each ion. To do this, the affinity of the protein for that ion must decrease. Thus, the affinity for each ion on the two sides of the membrane must be different. This is achieved by phosphorylation, which changes the shape of the protein molecule. The change in shape of the protein also serves to expose the ion binding sites to different sides of the membrane, as discussed in the following paragraph.

A general scheme for the pumping cycle of the Na^+/K^+-ATPase is shown in Figure 4.46*a*. When the protein binds three Na^+ ions on the inside of the cell (step 1) and becomes phosphorylated (step 2), it shifts from the E_1 conformation to the E_2 conformation (step 3). In doing so, the binding site becomes exposed to the extracellular compartment, and the protein loses its affinity for Na^+ ions, which are then released outside the cell. Once the three sodium ions have been released, the protein picks up two potassium ions (step 4), becomes dephosphorylated (step 5), and shifts back to the original E_1 conformation (step 6). In this state, the binding site opens to the internal surface of the membrane and loses its affinity for K^+ ions, leading to the release of these ions into the cell. The cycle is then repeated. A model of the Na^+/K^+-ATPase structure based on a recent X-ray crystallographic study is shown in Figure 4.46*b*.

The importance of the sodium–potassium pump becomes evident when one considers that it consumes approximately one-third of the energy produced by most animal cells and two-thirds of the energy produced by nerve cells. Digitalis, a steroid obtained from the foxglove plant that has been used for 200 years as a treatment for congestive heart disease, binds to the Na^+/K^+-ATPase. Digitalis strengthens the heart's contraction by inhibiting the Na^+/K^+ pump, which leads to a chain of events that increases Ca^{2+} availability inside the muscle cells of the heart.

Other Ion Transport Systems   The best studied P-type pump is the Ca^{2+}-ATPase whose three-dimensional structure has been determined at several stages of the pumping cycle. The calcium pump is present in the membranes of the endoplasmic reticulum, where it actively transports calcium ions out of the cytosol into the lumen of this organelle. The transport of calcium ions by the Ca^{2+}-ATPase is accompanied by large conformational changes, which couple the hy-

drolysis of ATP to changes in access and affinity of the ion binding sites.

The sodium–potassium pump is found only in animal cells. This protein is thought to have evolved in primitive animals as the primary means to maintain cell volume and as the mechanism to generate the steep Na^+ and K^+ gradients that play such a key role in the formation of impulses in nerve and muscle cells. Plant cells have a H^+-transporting, P-type, plasma membrane pump. In plants, this proton pump plays a key role in the secondary transport of solutes (discussed later), in the control of cytosolic pH, and possibly in control of cell growth by means of acidification of the plant cell wall.

The epithelial lining of the stomach also contains a P-type pump, the H^+/K^+-ATPase, which secretes a solution of concentrated acid (up to 0.16 N HCl) into the stomach chamber. In the resting state, these pump molecules are situated in cytoplasmic membranes of the parietal cells of the stomach lining and are nonfunctional (Figure 4.47). When food enters the stomach, a hormonal message is transmitted to the parietal cells that causes the pump-containing membranes to move to the apical cell surface, where they fuse with the plasma membrane and begin secreting acid (Figure 4.47). In addition to functioning in digestion, stomach acid can also lead to heartburn. Prilosec is a widely used drug that prevents heartburn by inhibiting the stomach's H^+/K^+-ATPase. Other acid-blocking heartburn medications (e.g., Zantac, Pepcid, and Tagamet) do not inhibit the H^+/K^+-ATPase directly, but block a receptor on the surface of the parietal cells, thereby stopping the cells from becoming activated by the hormone.

Unlike P-type pumps, V-type pumps utilize the energy of ATP without forming a phosphorylated protein intermediate. V-type pumps actively transport hydrogen ions across the walls of cytoplasmic organelles and vacuoles (hence the designation V-type). They occur in the membranes that line lysosomes, secretory granules, and plant cell vacuoles where they maintain the low pH of the contents. V-type pumps have also been found in the plasma membranes of a variety of cells. For example, a V-type pump in the plasma membranes of kidney tubules helps maintain the body's acid–base balance by secreting protons into the forming urine. V-type pumps are large multisubunit complexes similar to that of the ATP synthase shown in Figure 5.23.

Another diverse group of proteins that actively transport ions is the *ATP-binding cassette* (*ABC*) *transporters*, so called because all of the members of this superfamily share a homologous ATP-binding domain. The best studied ABC transporter is described in the accompanying Human Perspective.

Using Light Energy to Actively Transport Ions   *Halobacterium salinarium* (or *H. halobium*) is an archaebacterium that lives in extremely salty environments, such as that found in the Great Salt Lake. When grown under anaerobic conditions, the plasma membranes of these prokaryotes take on a purple color due to the presence of one particular protein, *bacteriorhodopsin*. As shown in Figure 4.48, bacteriorhodopsin contains retinal, the same prosthetic group present in rhodopsin, the light-absorbing protein of the rods of the ver-
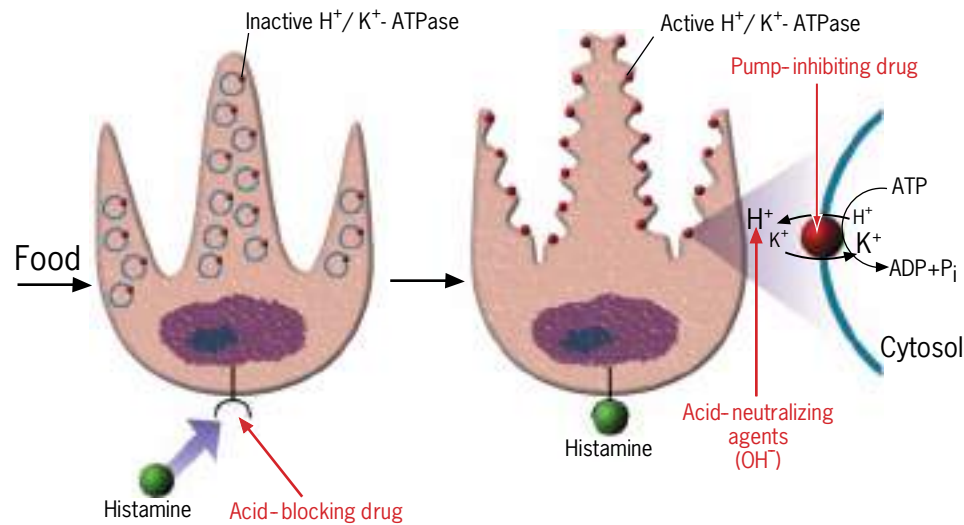
FIGURE 4.47 Control of acid secretion in the stomach. In the resting state, the H^+/K^+-ATPase molecules are present in the walls of cytoplasmic vesicles. Food entering the stomach triggers a cascade of hormone-stimulated reactions in the stomach wall leading to the release of histamine, which binds to a receptor on the surface of the acid-secreting parietal cells. Binding of histamine to its receptor stimulates a response that causes the H^+/K^+-ATPase-containing vesicles to fuse to the plasma membrane, forming deep folds, or canaliculi. Once at the surface, the transport protein is activated and pumps protons into the stomach cavity against a concentration gradient (indicated by the size of the letters). The heartburn drug Prilosec blocks acid secretion by directly inhibiting the H^+/K^+-ATPase, whereas several other acid-blocking medications interfere with activation of the parietal cells. Acid-neutralizing medications provide basic anions that combine with the secreted protons.

tebrate retina. The absorption of light energy by the retinal group induces a series of conformational changes in the protein that cause a proton to move from the retinal group, through a channel in the protein, to the cell exterior (Figure 4.48). The proton donated by the photo-excited retinal is replaced by another proton transferred to the protein from the cytoplasm. In effect, this process results in the translocation of protons from the cytoplasm to the external environment, thereby generating a steep H^+ gradient across the plasma membrane. This gradient is subsequently used by an ATP-synthesizing enzyme to phosphorylate ADP, as described in the next chapter.

FIGURE 4.48 Bacteriorhodopsin: a light-driven proton pump. The protein contains seven membrane-spanning helices and a centrally located retinal group (shown in purple), which serves as the light-absorbing element (chromophore). Absorption of a photon of light causes a change in the electronic structure of retinal, leading to the transfer of a proton from the —NH^+ group to a closely associated, negatively charged aspartic acid residue (#85) (step 1). The proton is then released to the extracellular side of the membrane (step 2) by a relay system consisting of several amino acid residues (Asp82, Glu204, and Glu194). The spaces between these residues are filled with hydrogen-bonded water molecules that help shuttle protons along the pathway. The deprotonated retinal is returned to its original state (step 3) when it accepts a proton from an undissociated aspartic acid residue (Asp96) located near the cytoplasmic side of the membrane. Asp96 is then reprotonated by a H^+ from the cytoplasm (step 4). Asp85 is deprotonated (step 5) prior to receiving a proton from retinal in the next pumping cycle. As a result of these events, protons move from the cytoplasm to the cell exterior through a central channel in the protein. (FROM HARTMUT LUECKE ET AL., COURTESY OF JANOS K. LANYI, SCIENCE 286:255, 1999; COPYRIGHT 1999, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)
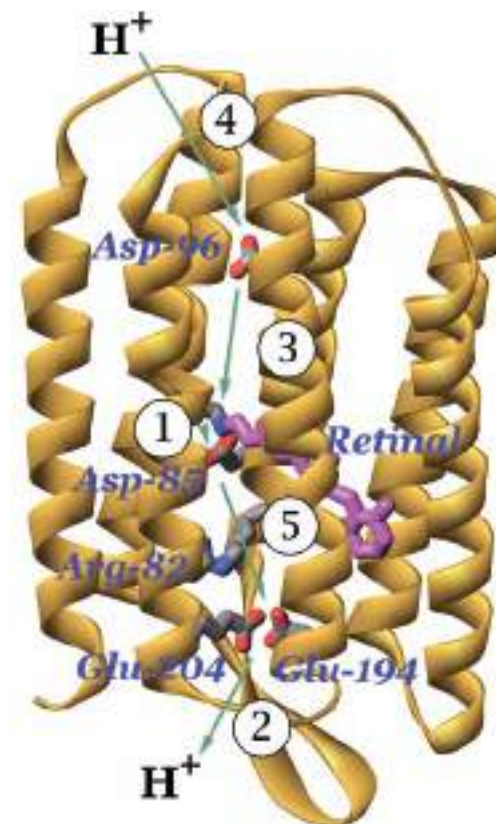
THE HUMAN PERSPECTIVE

Defects in Ion Channels and Transporters as a Cause of Inherited Disease

Several severe, inherited disorders have been traced to mutations in genes that encode ion channel proteins (Table 1). Most of the disorders listed in Table 1 affect the movement of ions across the plasma membranes of excitable cells (i.e., muscle, nerve, and sensory cells), reducing the ability of these cells to develop or transmit impulses (page 162). In contrast, cystic fibrosis, the best studied and most common inherited ion channel disorder, results from a defect in the ion channels of epithelial cells.

On average, 1 out of every 25 persons of Northern European descent carries one copy of the mutant gene that can cause cystic fibrosis. Because they show no symptoms of the mutant gene, most heterozygotes are unaware that they are carriers. Consequently, approximately 1 out of every 2500 infants in this Caucasian population $(1/25 \times 1/25 \times 1/4)$ is homozygous recessive at this locus and born with cystic fibrosis (CF). Although cystic fibrosis affects various organs, including the intestine, pancreas, sweat glands, and reproductive tract, the respiratory tract usually exhibits the most severe effects. Victims of CF produce a thickened, sticky mucus that is very hard to propel out of the airways. Afflicted individuals typically suffer from chronic lung infections and inflammation, which progressively destroy pulmonary function.

The gene responsible for cystic fibrosis was isolated in 1989. Once the sequence of the CF gene was determined and the amino acid sequence of the corresponding polypeptide was deduced, it was apparent that the polypeptide was a member of the ABC transporter superfamily. The protein was named *cystic fibrosis transmembrane conductance regulator* (*CFTR*), an ambiguous term that reflected the fact that researchers weren't sure of its precise function. The

question was thought to be answered after the protein was purified, incorporated into artificial lipid bilayers, and shown to act as a cyclic AMP-regulated chloride channel, not a transporter. But subsequent studies have added numerous complications to the story as it has been shown that, in addition to functioning as a chloride channel, CFTR also (1) conducts bicarbonate (HCO_3^-) ions, (2) suppresses the activity of an epithelial Na^+ ion channel (ENaC), and (3) stimulates the activity of a family of epithelial chloride/bicarbonate exchangers. As the role of CFTR has become more complex, it has become difficult to establish precisely how a defect in this protein leads to the development of chronic lung infections. While there is considerable debate, many researchers would agree with the following statements.

Because the movement of water out of epithelial cells by osmosis follows the movement of salts, abnormalities in the flux of Cl^-, HCO_3^-, and/or Na^+ caused by CFTR deficiency leads to a decrease in the fluid that bathes the epithelial cells of the airways (Figure 1). A reduction in volume of the surface liquid, and a resulting increase in viscosity of the secreted mucus, impair the function of the cilia that push mucus and bacteria out of the respiratory tract. At the present time, the most promising candidates for improving the quality of life of CF patients are drugs, such as Bronchitol, that increase the volume of the airway surface liquid and improve ciliary function. Bronchitol is a fine, dry powder consisting of the sugar mannitol, which is delivered by a simple inhaler. When inhaled into the lungs, the dissolved mannitol increases the osmolarity of the airway surface liquid, which causes water to move osmotically out of the epithelial cells and into the extracellular fluid. Clinical trials are also being car-

TABLE 1

Inherited disorder	Type of channel	Gene	Clinical consequences
Familial hemiplegic migraine (FHM)	Ca^{2+}	CACNL1A4	Migraine headaches
Episodic ataxia type-2 (EA-2)	Ca^{2+}	CACNL1A4	Ataxia (lack of balance and coordination)
Hypokalemic periodic paralysis	Ca^{2+}	CACNL1A3	Periodic myotonia (muscle stiffness) and paralysis
Episodic ataxia type-1	K^+	KCNA1	Ataxia
Benign familial neonatal convulsions	K^+	KCNQ2	Epileptic convulsions
Nonsyndromic dominant deafness	K^+	KCNQ4	Deafness
Long QT syndrome	K^+	HERG KCNQ1, or	Dizziness, sudden death from ventricular fibrillation
	Na^+	SCN5A	
Hyperkalemic periodic paralysis	Na^+	SCN4A	Periodic myotonia and paralysis
Liddle Syndrome	Na^+	B-ENaC	Hypertension (high blood pressure)
Myasthenia gravis	Na^+	nAChR	Muscle weakness
Dent's disease	Cl^-	CLCN5	Kidney stones
Myotonia congenita	Cl^-	CLC-1	Periodic myotonia
Bartter's syndrome type IV	Cl^-	CLC-Kb	Kidney dysfunction, deafness
Cystic fibrosis	Cl^-	CFTR	Lung congestion and infections
Cardiac arrhythmias	Na^+ K^+ Ca^{2+}	many different genes	Irregular or rapid heartbeat

See *Nature Cell Biol.* 6:1040, 2004, or *Nature* 440:444, 2006, for a more complete list.
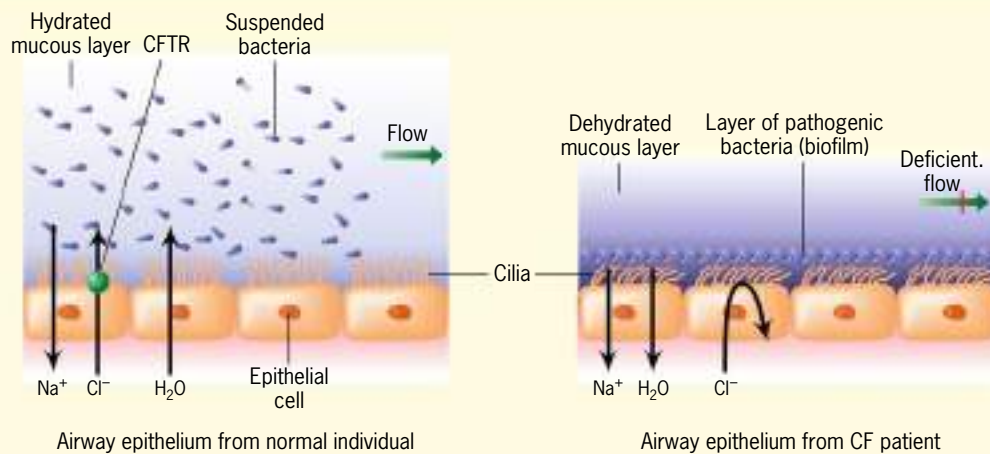
FIGURE 1 An explanation for the debilitating effects on lung function from the absence of the CFTR protein. In the airway epithelium of a normal individual, water flows out of the epithelial cells in response to the outward movement of ions, thus hydrating the surface mucous layer. The hydrated mucous layer, with its trapped bacteria, is readily moved out of the airways. In the airway epithelium of a person with cystic fibrosis, the abnormal movement of ions causes water to flow in the opposite direction, thus dehydrating the mucous layer. As a result, trapped bacteria cannot be moved out of the airways, which allows them to proliferate as a biofilm (page 12) and cause chronic infections.

ried out on compounds that have the potential to increase the volume of surface fluid by altering the movements of ions in and out of the epithelial cells (Figure 1). These compounds include Na^+-channel inhibitors (e.g., Denufosol) with the potential to reduce Na^+ ion absorption from the airway fluid into the epithelium and activators of Cl^- channels (other than CFTR) (e.g., Moli1901) with the potential to increase Cl^- ion conductance from the epithelium into the airway fluid.

In the past decade, researchers have identified more than 1000 different mutations that give rise to cystic fibrosis. However, approximately 70 percent of the alleles responsible for cystic fibrosis in the United States contain the same genetic alteration (designated ΔF508)—they are all missing three base pairs of DNA that encode a phenylalanine at position 508, within one of the cytoplasmic domains of the CFTR polypeptide. Subsequent research has revealed that CFTR polypeptides lacking this particular amino acid fail to be processed normally within the membranes of the endoplasmic reticulum and, in fact, never reach the surface of epithelial cells. As a result, CF patients who are homozygous for the ΔF508 allele completely lack the CFTR channel in their plasma membranes and have a severe form of the disease. When cells from these patients are grown in culture at lower temperature, the mutant protein is transported to the plasma membrane where it functions quite well. This finding has prompted a number of drug companies to screen for small molecules that can bind to these mutant CFTR molecules, thereby preventing their destruction in the cytoplasm and allowing them to reach the cell surface. Several promising candidates have been identified, but none has yet to be proven effective in clinical trials.
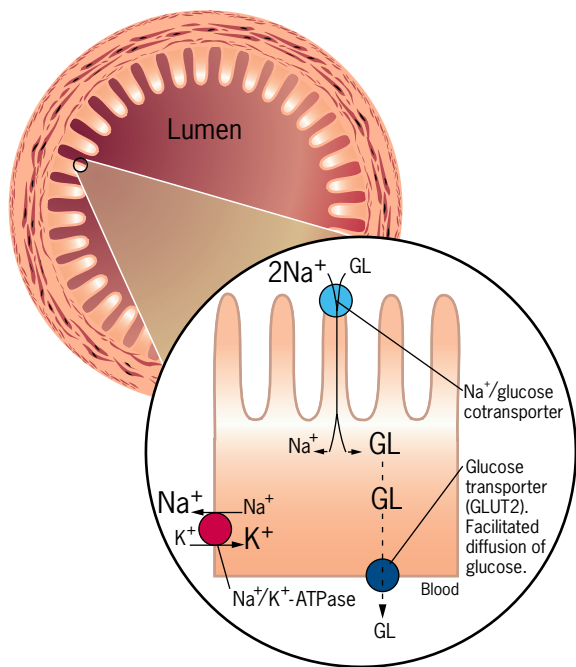
According to one estimate, the ΔF508 mutation had to have originated more than 50,000 years ago to have reached such a high frequency in the population. The fact that the CF gene has reached this frequency suggests that heterozygotes may receive some selective advantage over those lacking a copy of the defective gene. It has been proposed that CF heterozygotes may be protected from the effects of cholera, a disease that is characterized by excessive fluid secretion by the wall of the intestine. One difficulty with this proposal is that there is no record of cholera epidemics in Europe until the 1820s. An alternate proposal suggests that heterozygotes are protected from typhoid fever because the bacterium responsible for this disease adheres poorly to the wall of an intestine having a reduced number of CFTR molecules.

Ever since the isolation of the gene responsible for CF, the development of a cure by gene therapy—that is, by replacement of the defective gene with a normal version—has been a major goal of CF researchers. Cystic fibrosis is a good candidate for gene therapy because the worst symptoms of the disease result from the defective activities of epithelial cells that line the airways and, therefore, are accessible to agents that can be delivered by inhalation of an aerosol. Clinical trials have been conducted using several different types of delivery systems. In one group of trials, the normal *CFTR* gene was incorporated into the DNA of a defective adenovirus, a type of virus that normally causes upper respiratory tract infections. The recombinant virus particles were then allowed to infect the cells of the airway, delivering the normal gene to the genetically deficient cells. The primary disadvantage in using adenovirus is that the viral DNA (along with the normal *CFTR* gene) does not become integrated into the chromosomes of the infected host cell so that the virus must be readministered frequently. As a result, the procedure often induces an immune response within the patient that eliminates the virus and leads to lung inflammation. Researchers are hesitant to employ viruses that integrate their genomes for fear of initiating the formation of cancers. In other trials, the DNA encoding the normal *CFTR* gene has been linked to positively charged liposomes (page 124) that can fuse with the plasma membranes of the airway cells, delivering their DNA contents into the cytoplasm. Lipid-based delivery has an advantage over viruses in being less likely to stimulate a destructive immune response following repeated treatments, but has the disadvantage of being less effective in achieving genetic modification of target cells. To date, none of the clinical trials of gene therapy has resulted in significant improvement of either physiologic processes or disease symptoms. The development of more effective DNA delivery systems, which are capable of genetically altering a greater percentage of airway cells, will be required if a treatment for CF based on gene therapy is to be achieved.

Cotransport: Coupling Active Transport to Existing Ion Gradients

The establishment of concentration gradients, such as those of Na^+, K^+, and H^+, provides a means by which free energy can be stored in a cell. The potential energy stored in ionic gradients is utilized by a cell in various ways to perform work, including the transport of other solutes. Consider the physiologic activity of the intestine. Within its lumen, enzymes hydrolyze high-molecular-weight polysaccharides into simple sugars, which are absorbed by the epithelial cells that line the intestine. The movement of glucose across the apical plasma membrane of the epithelial cells, against a concentration gradient, occurs by **cotransport** with sodium ions, as illustrated in Figure 4.49. The Na^+ concentration is kept very low within the cells by the action of a *primary* active transport system (the Na^+/K^+-ATPase), located in the basal and lateral plasma membrane, which pumps sodium ions out of the cell against a concentration gradient. The tendency for sodium ions to diffuse back across the apical plasma membrane down their concentration gradient is "tapped" by the epithelial cells to drive the cotransport of glucose molecules into the cell *against* a concentration gradient. The glucose molecules are said to be driven by *secondary active transport*. In this case, the

transport protein, called a *Na$^+$/glucose cotransporter*, moves two sodium ions and one glucose molecule with each cycle. Once inside, the glucose molecules diffuse through the cell and are moved across the basal membrane by facilitated diffusion (page 152).

To appreciate the power of an ion gradient in accumulating other types of solutes in cells, we can briefly consider the energetics of the Na^+/glucose cotransporter. Recall from page 144 that the free-energy change for the movement of a mole of Na^+ ions into the cell is equal to -3.1 kcal/mol, and thus 6.2 kcal for two moles of Na^+ ions, which would be available to transport one mole of glucose uphill into the cell. Recall also from page 143 that the equation for the movement of a nonelectrolyte, such as glucose, across the membrane is

$$\Delta G = RT \ln \frac{[C_i]}{[C_o]}$$

$$\Delta G = 2.303 \, RT \log_{10} \frac{[C_i]}{[C_o]}$$

Using this equation, we can calculate how steep a concentration gradient of glucose (X) that this cotransporter can generate. At 25°C,

$$-6.2 \text{ kcal/mol} = 1.4 \text{ kcal/mol} \cdot \log_{10} X$$

$$\log_{10} X = -4.43$$

$$X = \frac{1}{23,000}$$

This calculation indicates that the Na^+/glucose cotransporter is capable of transporting glucose into a cell against a concentration gradient greater than 20,000-fold.

Plant cells rely on secondary active transport systems to take up a variety of nutrients, including sucrose, amino acids, and nitrate. In plants, uptake of these compounds is coupled to the downhill, inward movement of H^+ ions rather than Na^+ ions. The secondary active transport of glucose into the epithelial cells of the intestine and the transport of sucrose into a plant cell are examples of *symport*, in which the two transported species (Na^+ and glucose or H^+ and sucrose) move in the same direction. Numerous secondary transport proteins have been isolated that engage in *antiport*, in which the two transported species move in opposite directions. For example, cells often maintain a proper cytoplasmic pH by coupling the inward, downhill movement of Na^+ with the outward movement of H^+. Proteins that mediate antiport are usually called *exchangers*.

FIGURE 4.49 Secondary transport: the use of energy stored in an ionic gradient. The Na^+/K^+-ATPase residing in the plasma membrane of the lateral surface maintains a very low cytosolic concentration of Na^+. The Na^+ gradient across the plasma membrane represents a storage of energy that can be tapped to accomplish work, such as the transport of glucose by a Na^+/glucose cotransporter located in the apical plasma membrane. Once transported across the apical surface into the cell, the glucose molecules diffuse to the basal surface where they are carried by a glucose facilitative transporter out of the cell and into the bloodstream. The relative size of the letters indicates the directions of the respective concentration gradients. Two Na^+ ions are transported for each glucose molecule; the 2:1 Na^+/glucose provides a much greater driving force for moving glucose into the cell than a 1:1 ratio.
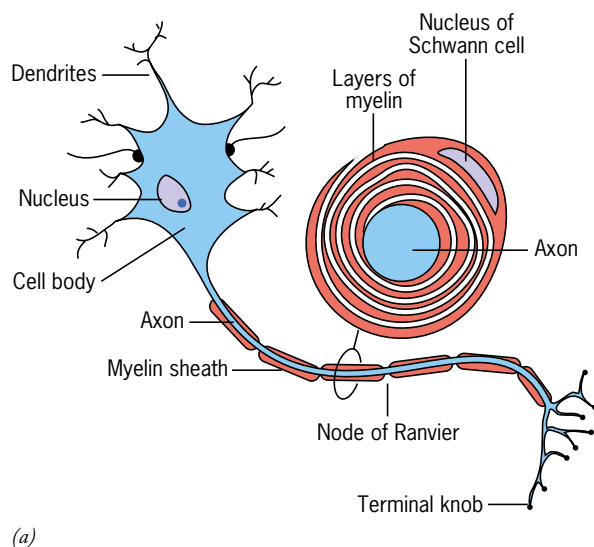
1. Compare and contrast the four basically different ways that a substance can move across the plasma membrane (as indicated in Figure 4.33).

2. Contrast the energetic difference between the diffusion of an electrolyte versus a nonelectrolyte across the membrane.

3. Describe the relationship between partition coefficient and molecular size with regard to membrane permeability.

4. Explain the effects of putting a cell into a hypotonic, hypertonic, or isotonic medium.

5. Describe two ways in which energy is utilized to move ions and solutes against a concentration gradient.

6. How does the Na^+/K^+-ATPase illustrate the asymmetry of the plasma membrane?

7. What is the role of phosphorylation in the mechanism of action of the Na^+/K^+-ATPase?

8. What is the structural relation between the parts of the prokaryotic KcsA K^+ channel and the eukaryotic voltage-regulated K^+ channel? Which part of the channel is involved in ion selectivity, which part in channel gating, and which part in channel inactivation? How does each of these processes (ion selectivity, gating, and inactivation) occur?

9. Because of its smaller size, one would expect Na^+ ions to be able to penetrate any pore large enough for a K^+ ion. How does the K^+ channel select for this specific ion?

4.8 MEMBRANE POTENTIALS AND NERVE IMPULSES

All organisms respond to external stimulation, a property referred to as *irritability*. Even a single-celled amoeba, if poked with a fine glass needle, responds by withdrawing its pseudopodia, rounding up, and moving off in another direction. Irritability in an amoeba depends on the same basic properties of membranes that lead to the formation and propagation of nerve impulses, which is the subject of the remainder of the chapter.
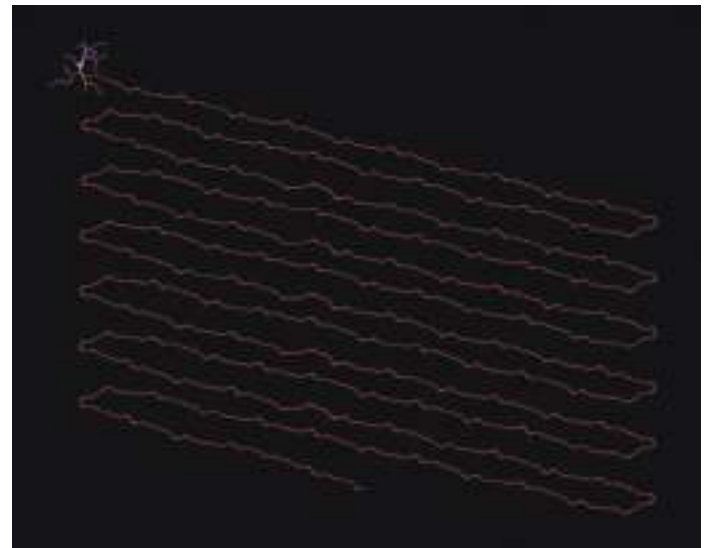
Nerve cells (or *neurons*) are specialized for the collection, conduction, and transmission of information, which is coded in the form of fast-moving electrical impulses. The basic parts of a typical neuron are illustrated in Figure 4.50). The nucleus of the neuron is located within an expanded region called the *cell body*, which is the metabolic center of the cell and the site where most of its material contents are manufactured. Extending from the cell bodies of most neurons are a number of fine extensions, called **dendrites**, which receive *incoming* information from external sources, typically other neurons. Also emerging from the cell body is a single, more prominent extension, the **axon**, which conducts *outgoing* impulses away from the cell body and toward the target cell(s). Although some axons may be only a few micrometers in length, others extend for many meters in the body of a large vertebrate, such as a giraffe or whale. Most axons split near their ends into smaller processes, each ending in a *terminal knob*—a specialized site where impulses are transmitted from neuron to target cell. Many neurons in the brain end in thousands of terminal knobs, allowing these brain cells to communicate with thousands of potential targets. As discussed on page 162, most neurons in the vertebrate body are wrapped in a lipid-rich **myelin sheath**, whose function is described below.

The Resting Potential

A voltage (or electric potential difference) between two points, such as the inside and outside of the plasma membrane, results when there is an excess of positive ions at one point and an excess of negative ions at the other point. Voltages across plasma membranes can be measured by inserting

(a)

(b)

FIGURE 4.50 The structure of a nerve cell. (*a*) Schematic drawing of a simple neuron with a myelinated axon. As the inset shows, the myelin sheath comprises individual Schwann cells that have wrapped themselves around the axon. The sites where the axon lacks myelin wrapping are called nodes of Ranvier. (*Note:* Myelin-forming cells within the central nervous system are called oligodendrocytes rather than Schwann cells.) (*b*) A composite micrograph of a single rat hippocampal neuron with cell body and dendrites (purple) and an axon 1 cm in length (red). Motor nerve cells in larger mammals can be 100 times this length. (B: FROM CARLOS F. IBÁÑEZ, TRENDS CELL BIOL. 17:520, 2007, COPYRIGHT 2007, WITH PERMISSION FROM ELSEVIER SCIENCE.)

one fine glass electrode (or *microelectrode*) into the cytoplasm of a cell, placing another electrode in the extracellular fluid outside the cell, and connecting the electrodes to a voltmeter, an instrument that measures a difference in charge between two points (Figure 4.51). When this experiment was first carried out on a giant axon of the squid, a potential difference of approximately 70 millivolts (mV) was recorded, the inside being negative with respect to the outside (indicated with a minus sign, -70 mV). The presence of a membrane potential is not unique to nerve cells; such potentials are present in all types of cells, the magnitude varying between about -15 and -100 mV. For nonexcitable cells, that is, those other than neurons and muscle cells, this voltage is simply termed the **membrane potential**. In a nerve or muscle cell, this same potential is referred to as the **resting potential** because it is subject to dramatic change, as discussed in the following section.

The magnitude and direction of the voltage across the plasma membrane are determined by the differences in concentrations of ions on either side of the membrane and their relative permeabilities. As described earlier in the chapter, the Na^+/K^+-ATPase pumps Na^+ out of the cell and K^+ into the cell, thereby establishing steep gradients of these two ions across the plasma membrane. Because of these gradients, you might expect that potassium ions would leak out of the cell and sodium ions would leak inward through their respective ion channels. However, the vast majority of the ion channels that are open in the plasma membrane of a *resting* nerve cell are selective for K^+; they are often referred to as K^+ *leak channels*. K^+ leak channels are thought to be members of a family of K^+ channels that lack the S4 voltage sensor (page 151) and fail to respond to changes in voltage.

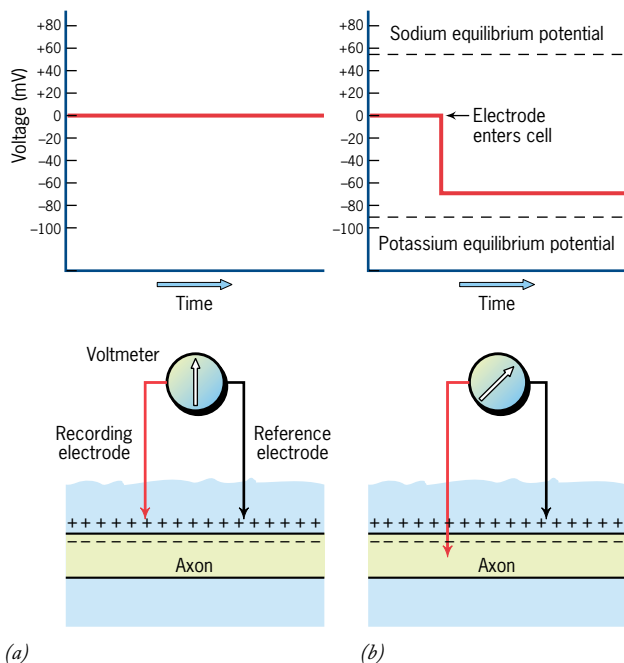Because K^+ ions are the only charged species with significant permeability in a resting nerve cell, their outflow through the membrane leaves an excess of negative charges on the cytoplasmic side of the membrane. Although the concentration gradient across the membrane favors continued efflux of K^+, the electrical gradient resulting from the excess negative charge on the inside of the membrane favors the retention of K^+ ions inside the cell. When these two opposing forces are balanced, the system is at equilibrium, and there is no further *net* movement of K^+ ions across the membrane. Using the following equation, which is called the Nernst equation, one can calculate the membrane potential (V_m) that would be measured at equilibrium if the plasma membrane of a nerve cell were permeable only to K^+ ions.[6] In this case, V_m would be equal to the potassium equilibrium potential (E_K):

$$E_K = 2.303 \frac{RT}{zF} \cdot \log_{10} \frac{[K_o^+]}{[K_i^+]}$$

For a squid giant axon, the internal $[K_i^+]$ is approximately 350 mM, while the external $[K_o^+]$ is approximately 10 mM; thus at 25°C (298 K) and $z = +1$ (for the univalent K^+ ion),

$$E_K = 59 \log_{10} 0.028 = -91 \text{ mV}$$

A similar calculation of the Na^+ equilibrium potential (E_{Na}) would produce a value of approximately $+55$ mV. Because measurements of the voltage across the resting nerve membrane are similar in sign and magnitude (-70 mV) to the potassium equilibrium potential just calculated, the movement of potassium ions across the membrane is considered the most important factor in determining the resting potential. The difference between the calculated K^+ equilibrium potential (-91 mV) and the measured resting potential (-70 mV, Figure 4.51) is due to a slight permeability of the membrane to Na^+ through a recently described Na^+ leak channel.

The Action Potential

Our present understanding of membrane potentials and nerve impulses rests on a body of research carried out on the giant axons of the squid in the late 1940s and early 1950s by a group of British physiologists, most notably Alan Hodgkin, Andrew Huxley, and Bernard Katz. These axons, which are approximately 1 mm in diameter, carry impulses at high speeds, enabling the squid to escape rapidly from predators. If the membrane of a resting squid axon is stimulated by poking it with a fine needle or jolting it with a very small electric current, some of its sodium channels open, allowing a limited number of sodium ions to diffuse into the cell. This opportunity for positively charged ions to move into the cell reduces the membrane

(a)   (b)

FIGURE 4.51   **Measuring a membrane's resting potential.** A potential is measured when a difference in charge is detected between the reference and recording electrodes. In (*a*), both electrodes are on the outside of the cell, and no potential difference (voltage) is measured. As one electrode penetrates the plasma membrane of the axon in (*b*), the potential immediately drops to -70 mV (inside negative), which approaches the equilibrium potential for potassium ions, that is, the potential that would result if the membrane were impermeable to all ions except potassium.

---

[6]The Nernst equation is derived from the equation provided on page 144, by setting ΔG at zero, which is the case when the movement of the ions is at equilibrium. Walther Nernst was a German physical chemist who won the 1920 Nobel Prize.

potential, making it less negative. Because the positive change in membrane voltage causes a *decrease* in the polarity between the two sides of the membrane, it is called a **depolarization**.

If the stimulus causes the membrane to depolarize by only a few millivolts, say from −70 to −60 mV, the membrane rapidly returns to its resting potential as soon as the stimulus has ceased (Figure 4.52a, left box). If, however, the stimulus depolarizes the membrane beyond a certain point, called the **threshold**, which occurs at about −50 mV, then a new series of events is launched. The change in voltage causes the voltage-gated sodium channels to open. As a result, sodium ions diffuse freely into the cell (Figure 4.52a, middle box) down both their concentration and electric gradients. The increased permeability of the membrane to Na^+ ions and the corresponding movement of positive charge into the cell causes the membrane to reverse potential briefly (Figure 4.52b), becoming positive at about +40 mV, which approaches the equilibrium potential for Na^+ (Figure 4.51).

After approximately 1 msec, the sodium channels spontaneously inactivate, blocking further influx of Na^+ ions. According to the prevailing view, inactivation results from the random diffusion of an inactivation peptide into the opening of the channel pore in a manner similar to that described for K^+ channels on page 151. Meanwhile, the change in membrane potential caused by Na^+ influx triggers the opening of the voltage-gated potassium channels (Figure 4.52a, right box) discussed on page 151. As a result, potassium ions diffuse freely out of the cell down their steep concentration gradient. The decreased permeability of the membrane to Na^+ and the increased permeability to K^+ cause the membrane potential to swing back to a negative value of about −80 mV, approaching that of the K^+ equilibrium potential (Figure 4.51). The large negative membrane potential causes the voltage-gated potassium channels to close (see Figure 4.43b), which returns the membrane to its resting state. Collectively, these changes in membrane potential are called an **action potential** (Figure 4.52b). The entire series of changes during an action potential takes only about 5 msec in the squid axon and less than

1 msec in a myelinated mammalian nerve cell. Following an action potential, the membrane enters a brief *refractory period* during which it cannot be restimulated. The refractory period occurs because the sodium channels that were inactivated during the initial stage of the action potential must close before they can be reopened in response to another stimulus. As depicted in Figure 4.43, the transformation of the ion channel from the inactivated to the closed conformation can only occur after the inactivating peptide has swung out of the opening of the pore.

Although the action potential changes membrane voltage dramatically, only a minute percentage of the ions on the two sides of the membrane are involved in any given action potential. The striking changes in membrane potential seen in Figure 4.52b are not caused by changes in Na^+ and K^+ ion *concentrations* on the two sides of the membrane (such changes are

(a)

(b)
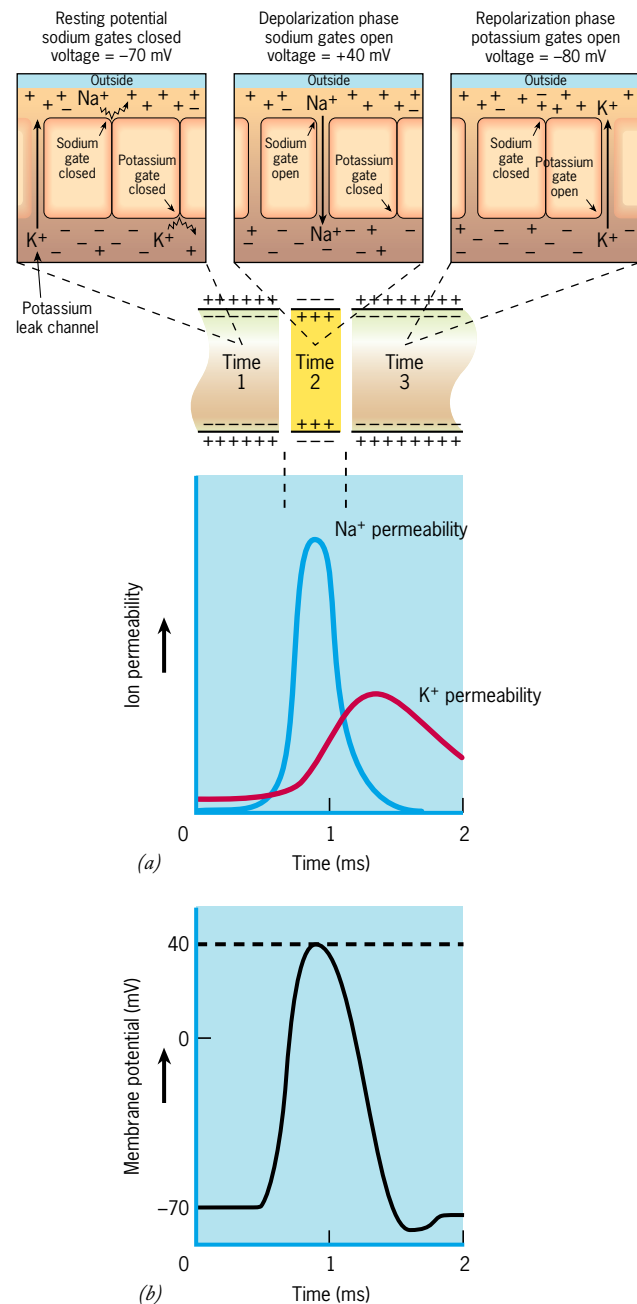
FIGURE 4.52 Formation of an action potential. (*a*) Time 1, upper left box: The membrane in this region of the nerve cell exhibits the resting potential, in which only the K^+ leak channels are open and the membrane voltage is approximately −70 mV. Time 2, upper middle box, shows the depolarization phase: The membrane has depolarized beyond the threshold value, opening the voltage-regulated sodium gates, leading to an influx of Na^+ ions (indicated in the permeability change in the lower graph). The increased Na^+ permeability causes the membrane voltage to temporarily reverse itself, reaching a value of approximately +40 mV in the squid giant axon (time 2). It is this reversal of membrane potential that constitutes the action potential. Time 3, upper right box, shows the repolarization phase: Within a tiny fraction of a second, the sodium gates are inactivated and the potassium gates open, allowing potassium ions to diffuse across the membrane (lower part of the drawing) and establish an even more negative potential at that location (−80 mV) than that of the resting potential. Almost as soon as they open, the potassium gates close, leaving the potassium leak channels as the primary path of ion movement across the membrane and reestablishing the resting potential. (*b*) A summary of the voltage changes that occur during an action potential, as described in part *a*.

insignificant). Rather, they are caused by the movements of charge in one direction or the other that result from the fleeting changes in permeability to these ions. Those Na^+ and K^+ ions that do change places across the membrane during an action potential are eventually pumped back by the Na^+/K^+-ATPase. Even if the Na^+/K^+-ATPase is inhibited, a neuron can often continue to fire thousands of impulses before the ionic gradients established by the pump's activity are dissipated.

Once the membrane of a neuron is depolarized to the threshold value, a full-blown action potential is triggered without further stimulation. This feature of nerve cell function is known as the *all-or-none law*. There is no in-between; subthreshold depolarization is incapable of triggering an action potential, whereas threshold depolarization automatically elicits a maximum response. It is also noteworthy that an action potential is not an energy-requiring process, but one that results from the flow of ions down their respective electrochemical gradients. Energy is required by the N^+/K^+-ATPase to generate the steep ionic gradients across the plasma membrane, but once that is accomplished, the various ions are poised to flow through the membrane as soon as their ion channels are opened.

The movements of ions across the plasma membrane of nerve cells form the basis for neural communication. Certain *local* anesthetics, such as procaine and novocaine, act by closing ion channels in the membranes of sensory cells and neurons. As long as these ion channels remain closed, the affected cells are unable to generate action potentials and thus unable to inform the brain of events occurring at the skin or teeth.

Propagation of Action Potentials as an Impulse

Up to this point, we have restricted the discussion to events occurring at a particular site on the nerve cell membrane where experimental depolarization has triggered an action potential. Once an action potential has been initiated, it does not remain localized at a particular site but is *propagated* as a **nerve impulse** down the length of the cell to the nerve terminals.

Nerve impulses are propagated along a membrane because an action potential at one site has an effect on the adjacent site. The large depolarization that accompanies an action potential creates a difference in charge along the inner and outer surfaces of the plasma membrane (Figure 4.53). As a result, positive ions move toward the site of depolarization on the outer surface of the membrane and away from that site on the inner surface (Figure 4.53). This local flow of current causes the membrane in the region just ahead of the action potential to become depolarized. Because the depolarization accompanying the action potential is very large, the membrane in the adjacent region is readily depolarized to a level greater than the threshold value, which opens the sodium channels in this adjacent region, generating another action potential. Thus, once triggered, a succession of action potentials passes down the entire length of the neuron without any loss of intensity, arriving at its target cell with the same strength it had at its point of origin.

Because all impulses traveling along a neuron exhibit the same strength, stronger stimuli cannot produce "bigger" impulses than weaker stimuli. Yet, we are clearly capable of detecting differences in the strength of a stimulus. The ability to
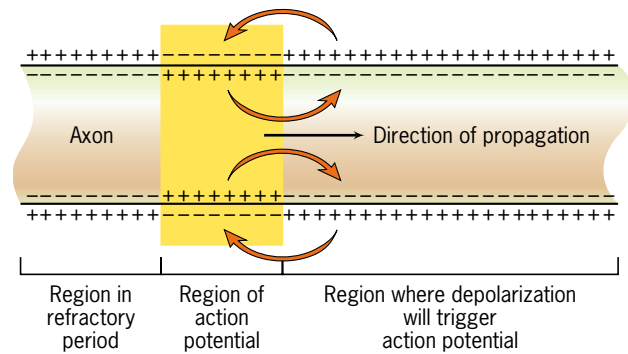

FIGURE 4.53 **Propagation of an impulse results from the local flow of ions.** An action potential at one site on the membrane depolarizes an adjacent region of the membrane, triggering an action potential at the second site. The action potential can only flow in the forward direction because the portion of the membrane that has just experienced an action potential remains in a refractory period.

make sensory discriminations depends on several factors. For example, a stronger stimulus, such as scalding water, activates more nerve cells than does a weaker stimulus, such as warm water. It also activates "high-threshold" neurons that would remain at rest if the stimulus were weaker. Stimulus strength is also encoded in the pattern and frequency by which action potentials are launched down a particular neuron. In most cases, the stronger the stimulus, the greater the number of impulses generated.

Speed Is of the Essence   The greater the diameter of an axon, the less the resistance to local current flow and the more rapidly an action potential at one site can activate adjacent regions of the membrane. Some invertebrates, such as squid and tube worms, have evolved giant axons that facilitate the animal's escape from danger. There is, however, a limit to this evolutionary approach. Because the speed of conduction increases with the square root of the increase in diameter, an axon that is 480 μm in diameter can conduct an action potential only four times faster than one that is 30 μm in diameter.

During the evolution of vertebrates, an increase in conduction velocity was achieved when the axon became wrapped in a myelin sheath (see Figures 4.5 and 4.50). Because it is composed of many layers of lipid-containing membranes, the myelin sheath is ideally suited to prevent the passage of ions across the plasma membrane. In addition, nearly all of the Na^+ ion channels of a myelinated neuron reside in the unwrapped gaps, or *nodes of Ranvier*, between adjacent Schwann cells or oligodendrocytes that make up the sheath (see Figure 4.50). Consequently, the nodes of Ranvier are the only sites where action potentials can be generated. An action potential at one node triggers an action potential at the next node (Figure 4.54), causing the impulse to jump from node to node without having to activate the intervening membrane. Propagation of an impulse by this mechanism is called **saltatory conduction**. Impulses are conducted along a myelinated axon at speeds up to 120 meters per second, which is more than 20 times faster than the speed that impulses travel in an unmyelinated neuron of the same diameter.
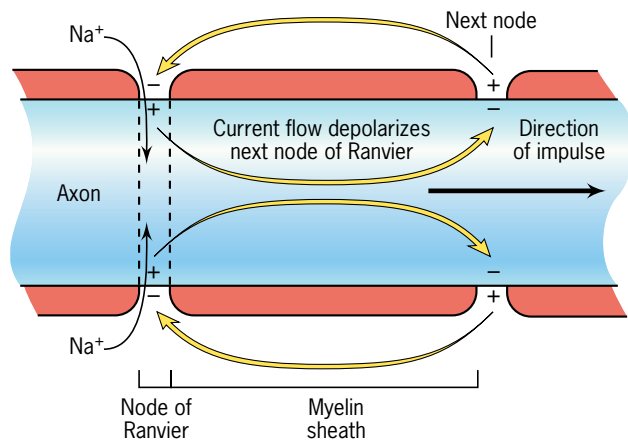
FIGURE 4.54 Saltatory conduction. During saltatory conduction, only the membrane in the nodal region of the axon becomes depolarized and capable of forming an action potential. This is accomplished as current flows directly from an activated node to the next resting node along the axon.

The importance of myelination is dramatically illustrated by multiple sclerosis (MS), a disease associated with deterioration of the myelin sheath that surrounds axons in various parts of the nervous system. Manifestations of the disease usually begin in young adulthood; patients experience weakness in their hands, difficulty in walking, and problems with their vision.

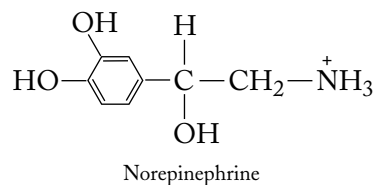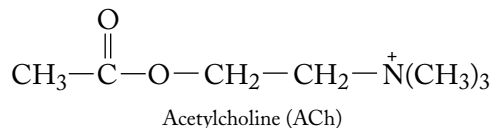## Neurotransmission: Jumping the Synaptic Cleft

Neurons are linked with their target cells at specialized junctions called **synapses**. Careful examination of a synapse reveals that the two cells do not make direct contact but are separated from each other by a narrow gap of about 20 to 50 nm. This gap is called the **synaptic cleft**. A **presynaptic cell** (a receptor cell or a neuron) conducts impulses toward a synapse, and a **postsynaptic cell** (a neuron, muscle, or gland cell) always lies on the receiving side of a synapse. Figure 4.55 shows a number of synapses between the terminal branches of an axon and a skeletal muscle cell; synapses of this type are called **neuromuscular junctions**.
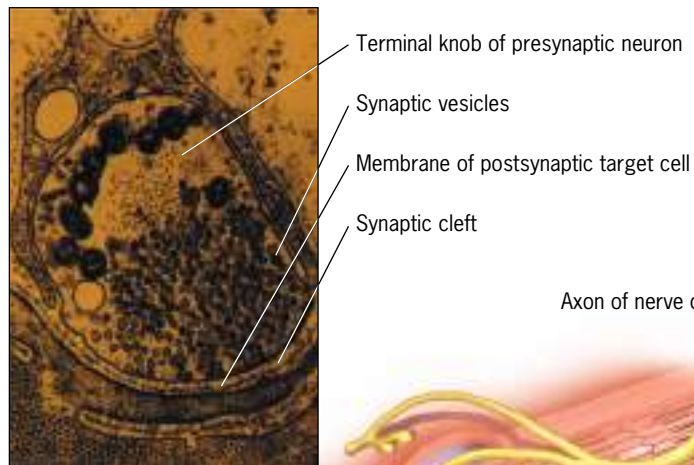
FIGURE 4.55 The neuromuscular junction is the site where branches from a motor axon form synapses with the muscle fibers of a skeletal muscle. The left inset shows the synaptic vesicles residing within the terminal knob of the axon and the narrow synaptic cleft between the terminal knob and the postsynaptic target cell. The right inset shows the terminal knob pressed closely against the muscle cell plasma membrane. Neurotransmitter molecules (red) released from synaptic vesicles of the presynaptic neuron are binding to receptors (orange) on the surface of the muscle cell (blue). (LEFT INSET FROM VU/T. REESE AND DON W. FAWCETT/VISUALS UNLIMITED.)
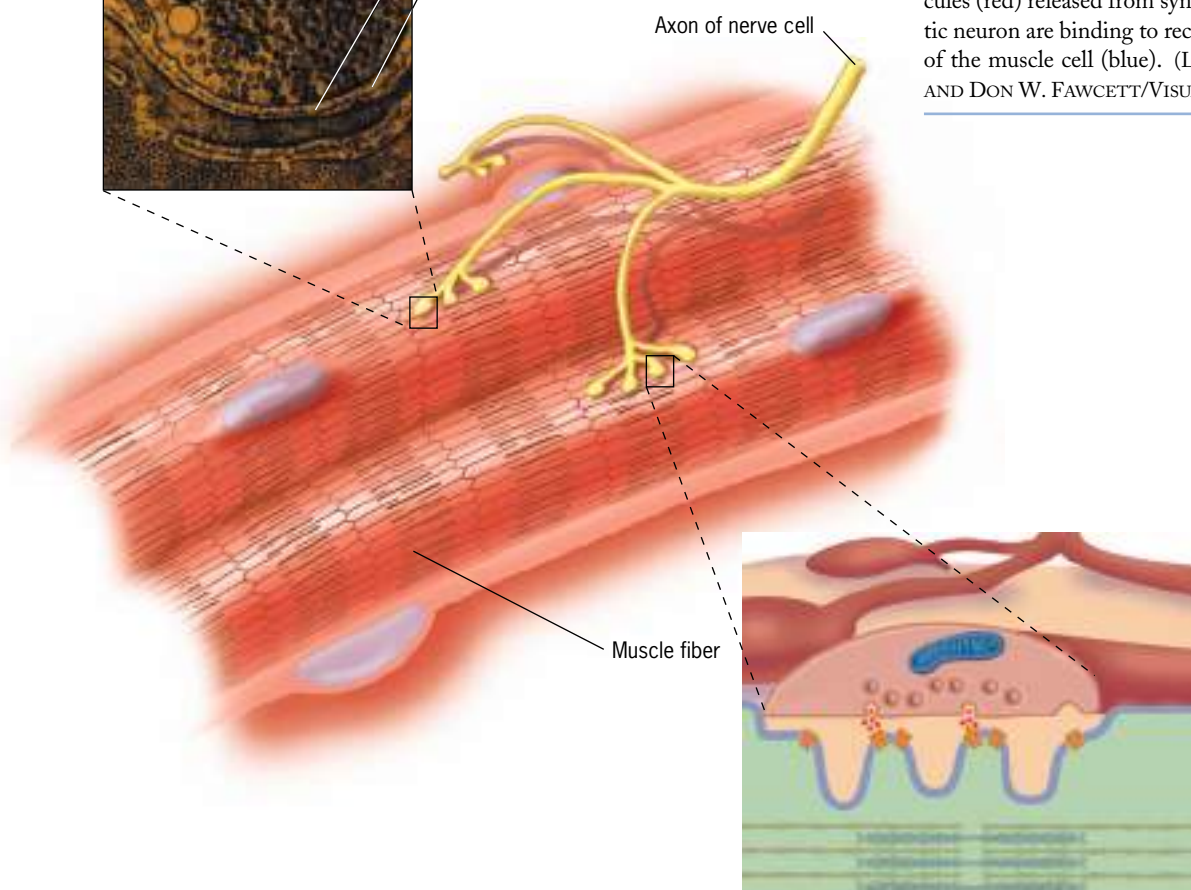
How does an impulse in a presynaptic neuron jump across the synaptic cleft and affect the postsynaptic cell? Studies carried out decades ago indicated that a chemical substance is involved in the transmission of an impulse from one cell to another (page 166). The very tips (terminal knobs) of the branches of an axon appear in the electron microscope to contain large numbers of **synaptic vesicles** (Figure 4.55, left inset) that serve as storage sites for the chemical transmitters that act on postsynaptic cells. Two of the best studied **neurotransmitters** are acetylcholine and norepinephrine,

$$CH_3-\overset{\overset{\displaystyle O}{\|}}{C}-O-CH_2-CH_2-\overset{+}{N}(CH_3)_3$$

Acetylcholine (ACh)

$$HO-\underset{}{\bigcirc}\overset{OH}{\underset{OH}{\overset{|}{\underset{|}{C}}}}-CH_2-\overset{+}{N}H_3$$

Norepinephrine

which transmit impulses to the body's skeletal and cardiac muscles.

The sequence of events during synaptic transmission can be summarized as follows (Figure 4.56). When an impulse reaches a terminal knob (step 1, Figure 4.56), the accompanying depolarization induces the opening of a number of voltage-gated Ca^{2+} channels in the plasma membrane of this part of the presynaptic nerve cell (step 2, Figure 4.56). Calcium ions are normally present at very low concentration within the neuron (about 100 nM), as in all cells. When the gates open, calcium ions diffuse from the extracellular fluid into the terminal knob of the neuron, causing the $[Ca^{2+}]$ to rise more than a thousandfold within localized microdomains near the channels. The elevated $[Ca^{2+}]$ triggers the rapid fusion of one or a few nearby synaptic vesicles with the plasma membrane, causing the release of neurotransmitter molecules into the synaptic cleft (step 3, Figure 4.56).

Once released from the synaptic vesicles, the neurotransmitter molecules diffuse across the narrow gap and bind selectively to receptor molecules that are concentrated directly across the cleft in the postsynaptic plasma membrane (step 4, Figure 4.56). A neurotransmitter molecule can have one of two opposite effects depending on the type of receptor on the target cell membrane to which it binds:[7]

1. The bound transmitter can trigger the opening of cation-selective channels in the membrane, leading primarily to an influx of sodium ions and a less negative (more positive) membrane potential. This depolarization of the

postsynaptic membrane *excites* the cell, making the cell more likely to respond to this or subsequent stimuli by generating an action potential of its own (Figure 4.56, steps 5a and 6).

2. The bound transmitter can trigger the opening of anion-selective channels, leading mainly to an influx of chloride ions, and a more negative (hyperpolarized) membrane potential. Hyperpolarization of the postsynaptic membrane makes it less likely the cell will generate an action potential because greater sodium influx is subsequently required to reach the membrane's threshold (Figure 4.56, step 5b).

Most nerve cells in the brain receive both excitatory and inhibitory signals from many different presynaptic neurons. It is the summation of these opposing influences that determine whether or not an impulse will be generated in the postsynaptic neuron.

All terminal knobs of a given neuron release the same neurotransmitter(s). However, a given neurotransmitter may have a stimulatory effect on one particular postsynaptic membrane and an inhibitory effect on another. Acetylcholine, for example, inhibits contractility of the heart but stimulates contractility of skeletal muscle. Within the brain, glutamate serves as the primary excitatory neurotransmitter and gamma-aminobutyric acid (GABA) as the primary inhibitory neurotransmitter. A number of general anesthetics, as well as Valium and its derivatives, act by binding to the GABA receptor and enhancing the activity of the brain's primary "off" switch.

Actions of Drugs on Synapses   It is important that a neurotransmitter has only a short half-life following its release from a presynaptic neuron; otherwise the effect of the neurotransmitter would be extended, and the postsynaptic neuron would not recover. A neurotransmitter is eliminated from the synapse in two ways: by enzymes that destroy neurotransmitter molecules in the synaptic cleft and by proteins that transport neurotransmitter molecules back to the presynaptic terminals—a process called *reuptake*. Because of the destruction or reuptake of neurotransmitter molecules, the effect of each impulse lasts no more than a few milliseconds.

Interfering with the destruction or reuptake of neurotransmitters can have dramatic physiologic and behavioral effects. Acetylcholinesterase is an enzyme localized within the synaptic cleft that hydrolyzes acetylcholine. If this enzyme is inhibited by exposure to the nerve gas DFP, for example, the skeletal muscles of the body undergo violent contraction due to the continued presence of high concentrations of acetylcholine.

Many drugs act by inhibiting the transporters that sweep neurotransmitters out of the synaptic cleft. A number of widely prescribed antidepressants, including Prozac, inhibit the reuptake of serotonin, a neurotransmitter implicated in mood disorders. Cocaine, on the other hand, interferes with the reuptake of the neurotransmitter dopamine that is released by certain nerve cells in a portion of the brain known as the limbic system. The limbic system contains the brain's

---

[7]It is important to note that this discussion ignores an important class of neurotransmitter receptors that are not ion channels and thus do not *directly* affect membrane voltage. This other group of receptors are members of a class of proteins called GPCRs, which are discussed at length in Section 15.3. When a neurotransmitter binds to one of these receptors, it can initiate a variety of responses, which often includes the opening of ion channels by an indirect mechanism.
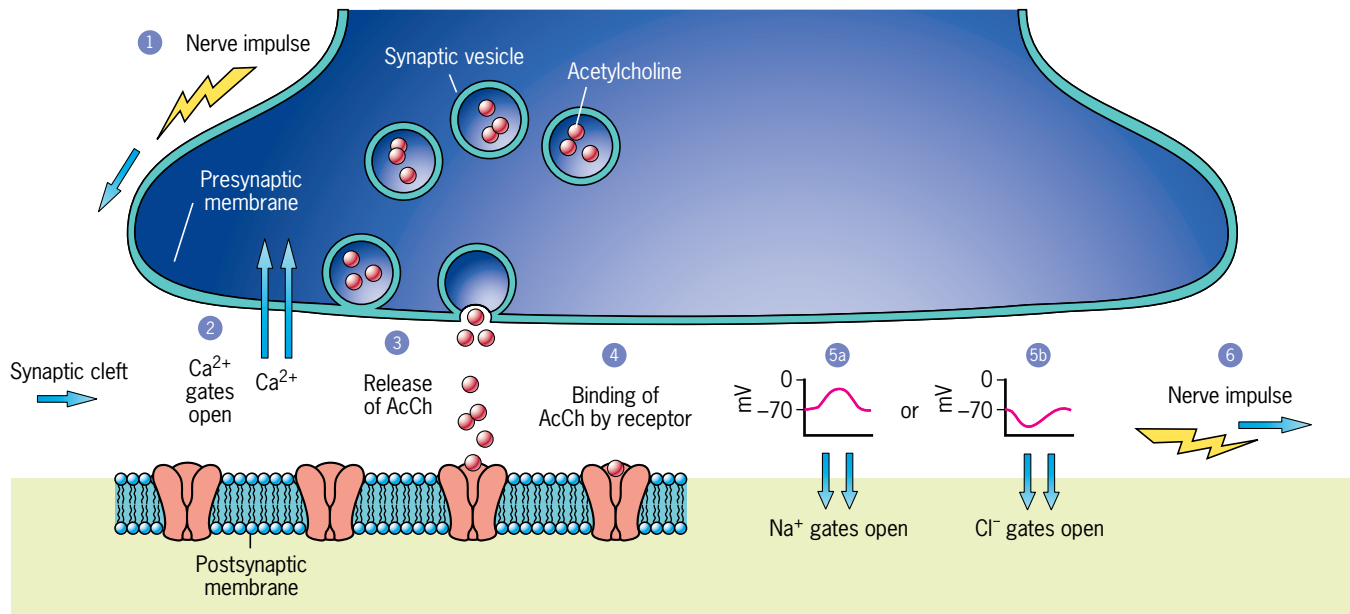
FIGURE 4.56 The sequence of events during synaptic transmission with acetylcholine as the neurotransmitter. During steps 1–4, a nerve impulse reaches the terminal knob of the axon, calcium gates open leading to an influx of Ca^{2+}, and acetylcholine is released from synaptic vesicles and binds to receptors on the postsynaptic membrane. If the binding of the neurotransmitter molecules causes a depolarization of the postsynaptic membrane (as in 5a), a nerve impulse may be generated there (6). If, however, the binding of neurotransmitter causes a hyperpolarization of the postsynaptic membrane (5b), the target cell is inhibited, making it more difficult for an impulse to be generated in the target cell by other excitatory stimulation. The breakdown of the neurotransmitter by acetylcholinesterase is not shown.

"pleasure" or "reward" centers. The sustained presence of dopamine in the synaptic clefts of the limbic system produces a short-lived feeling of euphoria, as well as a strong desire to repeat the activity. With repeated use, the pleasurable effects of the drug are increasingly reduced, but its addictive properties are enhanced. Amphetamines also act on dopamine-releasing neurons; they are thought to stimulate the excessive release of dopamine from presynaptic terminals and interfere with the reuptake of the neurotransmitter molecules from the synaptic cleft. Mice that have been genetically engineered to lack the dopamine transporter (DAT)—the protein responsible for dopamine reuptake—show behavior similar to that of normal mice that have been given cocaine or amphetamines. Administration of cocaine or amphetamines has no additional behavioral effects on animals lacking the *DAT* gene.

The active compound in marijuana (Δ^9-tetrahydrocannabinol) acts by a totally different mechanism. It binds to cannabinoid (CB1) receptors located on the *presynaptic* terminals of certain neurons of the brain, which reduces the likelihood that these neurons will release neurotransmitters. CB1 receptors normally interact with compounds produced in the body called *endocannabinoids*. Endocannabinoids are produced by postsynaptic neurons following depolarization. These substances diffuse "backwards" across the synaptic cleft to the presynaptic membrane, where they bind to CB1 receptors, suppressing synaptic transmission. CB1 receptors are located in many areas of the brain, including the hippocampus, cerebellum, and hypothalamus, which explains the effects of marijuana on memory, motor coordination, and appetite,

respectively. If marijuana increases appetite by binding to CB1 receptors, it follows that blocking these receptors might decrease appetite. This line of reasoning led to the development of a CB1-blocking weight-loss drug called Acomplia, which has been pulled from the market because of side effects.

Synaptic Plasticity Synapses are more than simply connecting sites between adjacent neurons; they are key determinants in the routing of impulses through the nervous system. The human brain is thought to contain at least one hundred trillion synapses. These synapses act like gates stationed along the various pathways, allowing some pieces of information to pass from one neuron to another, while holding back other pieces or rerouting them in another direction. While synapses are often perceived as fixed, unchanging structures, they can display a remarkable dynamic quality known as "synaptic plasticity." Synaptic plasticity is particularly important during infancy and childhood, when the neuronal circuitry of the brain achieves it mature configuration.

Synaptic plasticity is most readily observed in studies on neurons from the hippocampus, a portion of the brain that is vitally important in learning and short-term memory. When hippocampal neurons are repeatedly stimulated over a short period of time, the synapses that connect these neurons to their neighbors become "strengthened" by a process known as long-term potentiation (LTP), which may last for days, weeks, or even longer. Research into LTP has focused on the NMDA receptor, which is one of several receptor types that bind the excitatory neurotransmitter glutamate. When glutamate

binds to a postsynaptic NMDA receptor, it opens an internal cation channel within the receptor that allows the influx of Ca^{2+} ions into the postsynaptic neuron, triggering a cascade of biochemical changes that lead to synaptic strengthening. Synapses that have undergone LTP are able to transmit weaker stimuli and evoke stronger responses in postsynaptic cells. These changes are thought to play a major role as newly learned information or memories are encoded in the neural circuits of the brain. When laboratory animals are treated with drugs that inhibit LTP, such as those that interfere with the activity of the NMDA receptor, their ability to learn new information is greatly reduced.

There are numerous other reasons the study of synapses is so important. For example, a number of diseases of the nervous system, including myasthenia gravis, Parkinson's disease, schizophrenia, and even depression, are thought to have their roots in synaptic dysfunction.

REVIEW ?

1. What is a resting potential? How is it established as a result of ion flow?
2. What is an action potential? What are the steps that lead to its various phases?
3. How is an action potential propagated along an axon? What is saltatory conduction, and how does such a process occur?
4. What is the role of the myelin sheath in conduction of an impulse?
5. Describe the steps between the time an impulse reaches the terminal knob of a presynaptic neuron and an action potential is initiated in a postsynaptic cell.
6. Contrast the roles of ion pumps and channels in establishing and using ion gradients, particularly as it applies to nerve cells.

EXPERIMENTAL PATHWAYS

The Acetylcholine Receptor

In 1843, at the age of 30, Claude Bernard moved from a small French town, where he had been a pharmacist and an aspiring playwright, to Paris, where he planned to pursue his literary career. Instead, Bernard enrolled in medical school and went on to become the foremost physiologist of the nineteenth century. Among his many interests was the mechanism by which nerves stimulate the contraction of skeletal muscles. His studies included the use of curare, a highly toxic drug isolated from tropical plants and utilized for centuries by native South American hunters to make poisonous darts. Bernard found that curare paralyzed a skeletal muscle without interfering with either the ability of nerves to carry impulses to that muscle or the ability of the muscle to contract on direct stimulation. Bernard concluded that curare somehow acted on the region of contact between the nerve and muscle.

This conclusion was confirmed and extended by John Langley, a physiologist at Cambridge University. Langley was studying the ability of nicotine, another substance derived from plants, to stimulate the contraction of isolated frog skeletal muscles and the effect of curare in inhibiting nicotine action. In 1906, Langley concluded that "the nervous impulse should not pass from nerve to muscle by an electric discharge, but by the secretion of a special substance on the end of the nerve."[1] Langley proposed that this "chemical transmitter" was binding to a "receptive substance" on the surface of the muscle cells, the same site that bound nicotine and curare. These proved to be farsighted proposals.

Langley's suggestion that the stimulus from nerve to muscle was transmitted by a chemical substance was confirmed in 1921 in an ingenious experiment by the Austrian-born physiologist, Otto Loewi, the design of which came to Loewi during a dream. The heart rate of a vertebrate is regulated by input from two opposing (antagonistic) nerves. Loewi isolated a frog's heart with both nerves intact. When he stimulated the inhibitory (*vagus*) nerve, a chemical was released from the heart preparation into a salt solution, which

was allowed to drain into the medium bathing a second isolated heart. The rate of the second heart slowed dramatically, as though its own inhibitory nerve had been activated.[2] Loewi called the substance responsible for inhibiting the frog's heart "Vagusstoff." Within a few years, Loewi had shown that the chemical and physiologic properties of Vagusstoff were identical to acetylcholine, and he concluded that acetylcholine (ACh) was the substance released by the tips of the nerve cells that made up the vagus nerve.

In 1937, David Nachmansohn, a neurophysiologist at the Sorbonne, was visiting the World's Fair in Paris where he observed several living electric fish of the species *Torpedo marmorata* that were on display. These rays have electric organs that deliver strong shocks (40–60 volts) capable of killing potential prey. At the time, Nachmansohn was studying the enzyme acetylcholinesterase, which acts to destroy ACh after its release from the tips of motor nerves. Nachmansohn was aware that the electric organs of these fish were derived from modified skeletal muscle tissue (Figure 1), and he asked if he could have a couple of the fish for study once the fair had ended. The results of the first test showed the electric organ was an extraordinarily rich source of acetylcholinesterase.[3] It was also a very rich source of the nicotinic acetylcholine receptor (nAChR),* the receptor present on the postsynaptic membranes of skeletal muscle cells that binds ACh molecules released from the tips

---

*The receptor is described as nicotinic because it can be activated by nicotine as well as by acetylcholine. This contrasts with muscarinic acetylcholine receptors of the parasympathetic nerve synapses, which can be activated by muscarine, but not nicotine, and are inhibited by atropine, but not curare. Smokers' bodies become accustomed to high levels of nicotine, and they experience symptoms of withdrawal when they stop smoking because the postsynaptic neurons that possess nAChRs are no longer stimulated at their usual level. The drug Chantix, which is marketed as an aid to stop smoking, acts by binding to the most common version of brain nAChR (one with $\alpha 4 \beta 2$ subunits). Once bound, the Chantix molecule partially stimulates the receptor while preventing binding of nicotine.
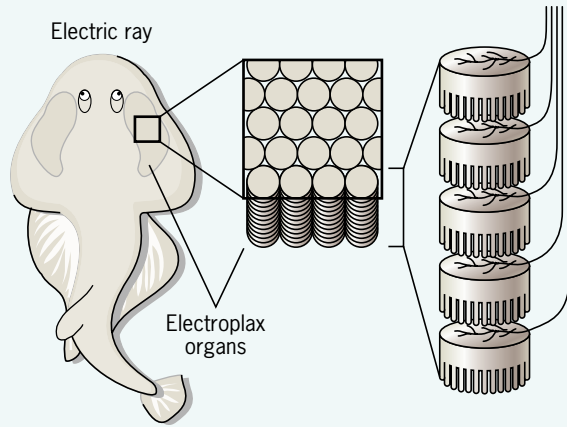
FIGURE 1 The electric organs of *Torpedo* consist of stacks of modified neuromuscular junctions located on each side of the body. (FROM Z. W. HALL, AN INTRODUCTION TO NEUROBIOLOGY, SINAUER ASSOCIATES, INC., SUNDERLAND, MA © 1992.)

of a motor nerve. Finding an ideal system can prove invaluable in the study of a particular aspect of cell structure or function. As will be evident from this discussion, the electric organs of fish have been virtually the sole source of material for the study of the nAChR.

The nAChR is an integral membrane protein, and it wasn't until the 1970s that techniques were developed for the isolation of such

proteins. As discussed in Chapter 18, the purification of a particular protein requires a suitable assay to determine the amount of that protein present in any particular fraction. The ideal assay for nAChR was a compound that bound selectively and tightly to this particular protein. Such a compound was discovered in 1963 by Chen-Yuan Lee and his colleagues of the National Taiwan University. The compound was α-bungarotoxin, a substance present in the venom of a Taiwanese snake. The α-bungarotoxin causes paralysis by binding tightly to the nAChRs on the postsynaptic membrane of skeletal muscle cells, blocking the response of the muscle to ACh.[4]

Equipped with labeled α-bungarotoxin to use in an assay, electric organs as a source, and a detergent capable of solubilizing membrane proteins, a number of investigators were able to isolate the acetylcholine receptors in the early 1970s. In one of these studies,[5] the membranes containing the nAChR were isolated by homogenizing the electric organs in a blender and centrifuging the suspension to pellet the membrane fragments. Membrane proteins were extracted from the membrane fragments using Triton X-100 (page 129), and the mixture was passed through a column containing tiny beads coated with a synthetic compound whose end bears a structural resemblance to ACh (Figure 2*a*). As the mixture of dissolved proteins passed through the column, two proteins that have binding sites for acetylcholine, nAChR and acetylcholinesterase (AChE), stuck to the beads. The remaining 90 percent of the protein in the extract failed to bind to the beads and simply passed through the column and was collected (Figure 2*b*). Once this group of proteins had passed through, a solution of 10^{-3} M flaxedil was passed through the column, which selectively removed the nAChR from
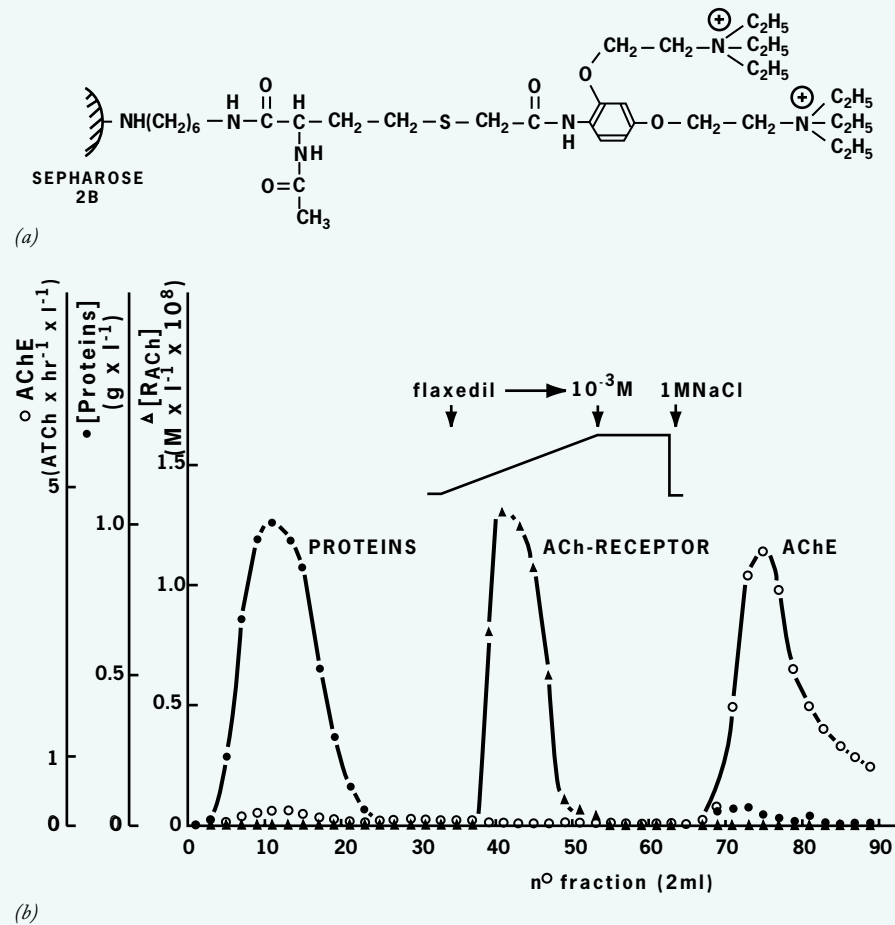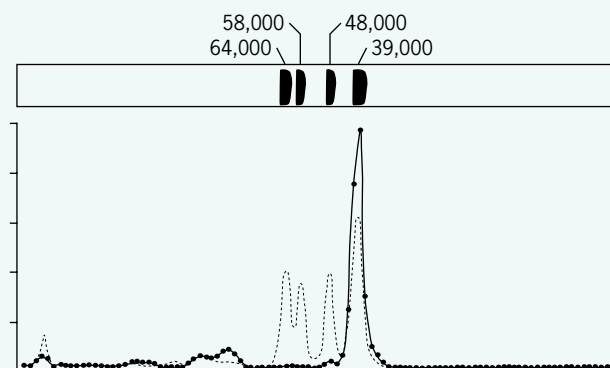

FIGURE 2 Steps used in the isolation of the nAChR. (*a*) Structure of a synthetic compound, CT5263, that was attached to sepharose beads to form an affinity column. The ends of the compound projecting from the beads resemble acetylcholine, causing both acetylcholinesterase (AChE) and the nicotinic acetylcholine receptor (nAChR) to bind to the beads. (*b*) When the Triton X-100 extract was passed through the column, both of the acetylcholine-binding proteins stuck to the beads, while the remaining dissolved protein (about 90 percent of the total protein in the extract) passed directly through the column. Subsequent passage of a solution of 10^{-3} M flaxedil through the column released the bound nAChR, without disturbing the bound AChE (which was subsequently eluted with 1 M NaCl). (FROM R. W. OLSEN, J.-C. MEUNIER, AND J.-P. CHANGEUX, FEBS LETT. 28:99, 1972.)

the beads, leaving the AChE behind. Using this procedure, the acetylcholine receptor as measured by bungarotoxin binding was purified by more than 150-fold in a single step. This type of procedure is known as *affinity chromatography*, and its general use is discussed in Section 18.7.

The next step was to determine something about the structure of the acetylcholine receptor. Studies in the laboratory of Arthur Karlin at Columbia University determined that the nAChR was a pentamer, a protein consisting of five subunits. Each receptor contained two copies of a subunit called α, and one copy each of three other subunits. The subunits could be distinguished by extracting membrane proteins in Triton X-100, purifying the nAChR by affinity chromatography, and then subjecting the purified protein to electrophoresis through a polyacrylamide gel (SDS–PAGE, as discussed in Section 18.7), which separates the individual polypeptides according to size (Figure 3).[6] The four different subunits proved to be homologous to one another, each subunit containing four homologous transmembrane helices (M1–M4).

Another important milestone in the study of the nAChR was the demonstration that the purified receptor acted as both a site for binding ACh and a channel for the passage of cations. It had been postulated years earlier by Jean-Pierre Changeux of the Pasteur Institute in Paris that the binding of ACh to the receptor caused a conformational change that opened an ion channel within the protein. The inward flux of Na^+ ions through the channel could then lead to a depolarization of the membrane and the activation of the muscle cell. During the last half of the 1970s, Changeux and his colleagues succeeded in incorpora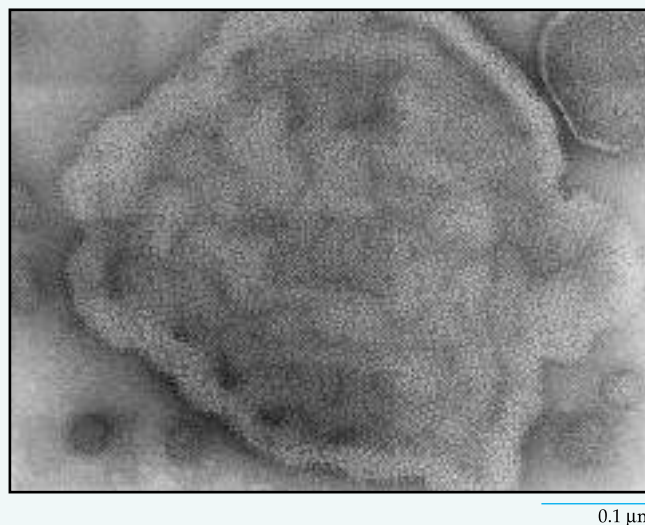ting purified nAChR molecules into artificial lipid vesicles.[7] Using vesicles containing various concentrations of labeled sodium and potassium ions, they demonstrated that binding of ACh to the receptors in the lipid bilayer initiated a flux of cations through the "membrane." It was evident that "the pure protein does indeed contain all the structural elements needed for the chemical transmission of an electrical signal—namely, an acetylcholine binding site, an ion channel and a mechanism for coupling their activity."
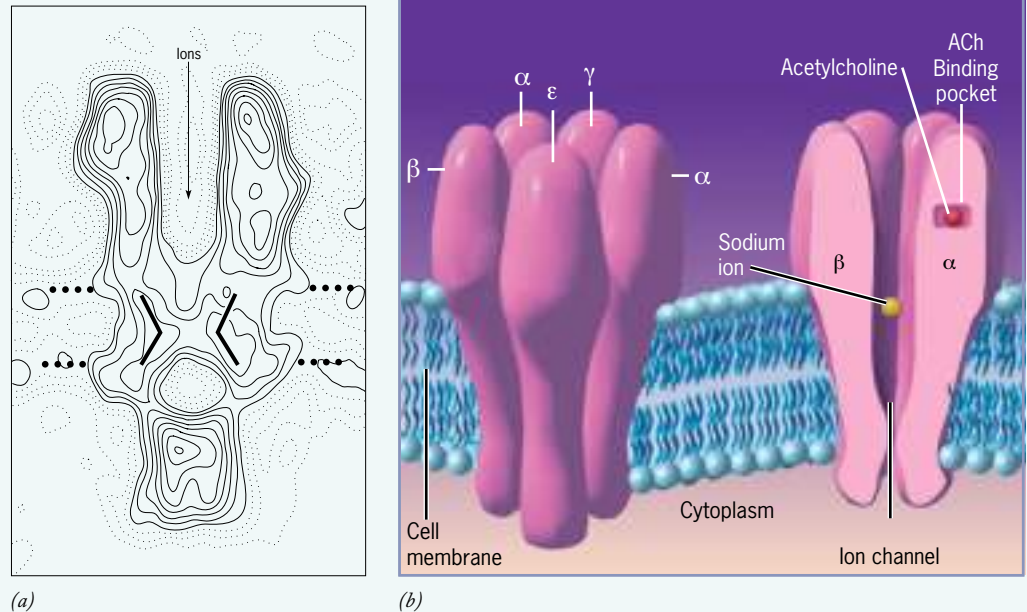
During the past two decades, researchers have focused on determining the structure of the nAChR and the mechanism by which binding of acetylcholine induces the opening of the ion gate. Analysis of the structure has taken several different paths. In one approach, scientists have used purified genes, determination of amino acid sequences, and site-directed mutagenesis to determine the specific parts of the polypeptides that span the membrane, or bind the neurotransmitter, or form the ion channel. These noncrystallographic studies on the molecular anatomy of a protein are similar in principle to those described in Section 4.4.

Another approach employed the electron microscope. The first glimpses of the nAChR were seen in electron micrographs of the membranes of the electric organs (Figure 4).[8] The receptors appeared to be ring shaped, with a diameter of 8 nm and a central channel of 2-nm diameter, and protruded out from the lipid bilayer into the external space. An increasingly detailed model of the nAChR has been developed by Nigel Unwin and his colleagues of the Medical Research Council of England.[9–13] Using the technique of electron crystallography (Section 18.8), which involves a mathematical analysis of electron micrographs of frozen membranes from the electric organs, Unwin was able to analyze the structure of the nAChR as it exists in its native lipid environment. He described the arrangement of the five subunits around a central channel (Figure 5). The ion channel consists of a narrow pore lined by a wall composed of five inner (M2) α helices, one from each of the surrounding subunits. The gate to the pore lies near the middle of the membrane, where each of the M2 α helices bends inward (at the apex of the V-

FIGURE 3 The top portion of the figure shows an SDS–polyacrylamide gel following electrophoresis of a preparation of the purified nAChR. The receptor consists of four different subunits whose molecular weights are indicated. Prior to electrophoresis, the purified receptor preparation was incubated with a radioactive compound (^{3}H-MBTA) that resembles acetylcholine and binds to the acetylcholine-binding site of the nAChR. Following electrophoresis, the gel was sliced into 1-mm sections and the radioactivity of each slice determined. All of the radioactivity was bound to the 39,000 dalton subunit, indicating this subunit contains the ACh-binding site. The dotted line indicates the light absorbance of each fraction, which provides a measure of the total amount of protein present in that fraction. The heights of the peaks provide a measure of the relative amounts of each of the subunits in the protein. All of the subunits are present in equal numbers except the smallest subunit (the α subunit, which contains the ACh-binding site), which is present in twice the number of copies. (FROM C. L. WEILL, M. G. MCNAMEE, AND A. KARLIN, BIOCHEM. BIOPHYS. RES. COMMUN. 61:1002, 1974.)

FIGURE 4 Electron micrograph of negatively stained, receptor-rich membranes from the electric organ of an electric fish showing the dense array of nAChR molecules. Each receptor molecule is seen as a small whitish circle with a tiny black dot in its center; the dot corresponds to the central channel, which has collected electron-dense stain. (FROM WERNER SCHIEBLER AND FERDINAND HUCHO, EUR. J. BIOCHEM. 85:58, 1978.)

FIGURE 5 (*a*) An electron density map of a slice through the nAChR obtained by analyzing electron micrographs of tubular crystals of *Torpedo* membranes embedded in ice. These analyses have allowed researchers to reconstruct the three-dimensional appearance of a single nAChR protein as it resides within the membrane. The continuous contours indicate lines of similar density greater than that of water. The two dark, bar-shaped lines represent the α helices that line the channel at its narrowest point. (*b*) Schematic diagram of the nAChR showing the arrangement of the subunits and a cross-sectional representation of the protein. Each of the five subunits contains four membrane-spanning helices (M1–M4). Of these, only the inner helix (M2) lines the pore, and is the subject of the remainder of the discussion. (A: FROM N. UNWIN, J. MOL. BIOL. 229:1118, 1993; COPYRIGHT 1993, BY PERMISSION OF THE PUBLISHER ACADEMIC PRESS, ELSEVIER SCIENCE.)

shaped bars in Figure 5*a*) to form a kink in the unactivated receptor. The side chain of a leucine residue projects inward from each kink. In this model, the leucine residues from the five inner helices form a tight hydrophobic ring that prevents the ions from crossing the membrane. The gate opens following the binding of two ACh molecules, one per α subunit. Each ACh binds to a site located within a pocket of an α subunit (Figure 5*b*).

To study the changes in the nAChR during channel opening, Unwin carried out the following experiment.[11] Preparations of nAChR-rich membranes were applied to a support grid that was allowed to fall into a bath of liquid nitrogen-cooled ethane, which freezes the membranes. Approximately 5 msec before they reached the surface of the freezing bath, the grids were sprayed with a solution of ACh, which bound to the receptors and triggered the conformational change required to open the channel. By comparing electron micrographs of nAChRs trapped in the open versus closed state, Unwin found that ACh binding triggers a small conformational change in the extracellular domains of the receptor subunits near the two ACh binding sites. According to the model depicted in Figure 6, this conformational change is propagated down the protein, causing a small (15°) rotation of the M2 helices that line the ion-conduction pore. The rotation of these inner helices breaks apart the hydrophobic gate, which allows Na$^+$ ions to enter the cell.[12,13] Other models for channel opening have been proposed by other laboratories, and a more definitive statement about the mechanism of gating will probably require a high-resolution X-ray crystal structure of the protein in both the open and closed states. Discussion of various models and the bases for them can be found in References 14–16.

FIGURE 6 **Ribbon drawings illustrating the proposed changes that occur within the nAChR upon binding of acetylcholine.** Only the two alpha subunits of the receptor are shown. In the closed state (left) the pore is blocked (pink patch) by the close apposition of a ring of hydrophobic residues (the valine and leucine side chains of these residues on the alpha subunits are indicated by the small ball-and-stick models at the site of pore constriction). The diameter of the pore at its narrowest point is about 6 Å, which is not sufficient for a hydrated Na$^+$ ion to pass. Although it is wide enough for passage of a dehydrated Na$^+$ ion, the wall of the channel lacks the polar groups that would be required to substitute for the displaced shell of water molecules (as occurs in the selectivity filter of the K$^+$ channel, Figure 4.39). A tryptophan side chain in the cytoplasmic domains of the subunits indicates the approximate site of acetylcholine binding. Following binding of ligand to each cytoplasmic domain (which is seen to consist largely of β sheets), a conformational change is proposed, which leads to a small rotation of the inner β sheets in the cytoplasmic domain (curved arrows on left figure). This, in turn, induces a rotational movement of the inner transmembrane helices of the subunits, expanding the diameter of the pore, which allows the flow of Na$^+$ ions through the open state of the channel (right). The relevant moving parts are shown in blue. (FROM N. UNWIN, FEBS LETT. 555:94, 2003, COPYRIGHT 2003, WITH PERMISSION FROM ELSEVIER SCIENCE.)
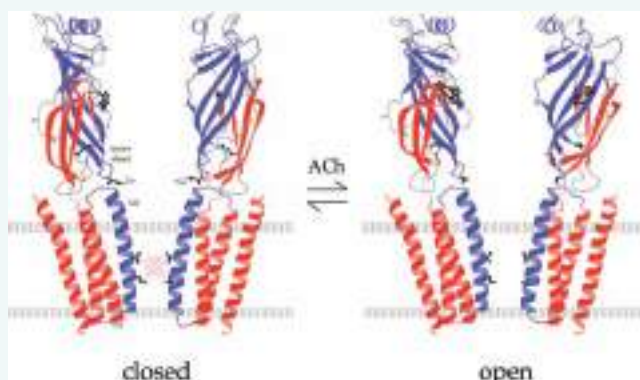
References

1. LANGLEY, J. N. 1906. On nerve endings and on special excitable substances in cells. *Proc. R. Soc. London B Biol. Sci.* 78:170–194.

2. LOEWI, O. 1921. Uber humorale ubertragbarkeit der herznervenwirkung. *Pfluger's Arch.* 214:239–242. (A review of Loewi's work written by him in English can be found in *Harvey Lect.* 28:218–233, 1933.)

3. MARNAY, A. 1937. Cholinesterase dans l'organe électrique de la torpille. *Compte Rend.* 126:573–574. (A review of Nachmansohn's work written in English can be found in his book, *Chemical and Molecular Basis of Nerve Action*, 2nd ed., Academic Press, 1975.)

4. CHANG, C. C. & LEE, C.-Y. 1963. Isolation of the neurotoxins from the venom of *Bungarus multicinctus* and their modes of neuromuscular blocking action. *Arch. Int. Pharmacodyn. Ther.* 144:241–257.

5. OLSEN, R. W., MEUNIER, J. C., & CHANGEUX, J. P. 1972. Progress in the purification of the cholinergic receptor protein from *Electrophorus electricus* by affinity chromatography. *FEBS Lett.* 28:96–100.

6. WEILL, C. L., McNAMEE, M. G., & KARLIN, A. 1974. Affinity-labeling of purified acetylcholine receptor from *Torpedo californica. Biochem. Biophys. Res. Commun.* 61:997–1003.

7. POPOT, J. L., CARTAUD, J., & CHANGEUX, J.-P. 1981. Reconstitution of a functional acetylcholine receptor. *Eur. J. Biochem.* 118:203–214.

8. SCHIEBLER, W. & HUCHO, F. 1978. Membranes rich in acetylcholine receptor: Characterization and reconstitution to excitable membranes from exogenous lipids. *Eur. J. Biochem.* 85:55–63.

9. BRISSON, A. & UNWIN, N. 1984. Tubular crystals of acetylcholine receptor. *J. Cell Biol.* 99:1202–1211.

10. UNWIN, N. 1993. Acetylcholine receptor at 9 Å resolution. *J. Mol. Biol.* 229:1101–1124.

11. UNWIN, N. 1995. Acetylcholine receptor channel imaged in the open state. *Nature* 373:37–43.

12. MIYAZAWA, A., FUJIYOSHI, Y. & UNWIN, N. 2003. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* 423:949–955.

13. UNWIN, N. 2005. Refined structure of the nicotinic acetylcholine receptor at 4.0 Å resolution. *J. Mol. Biol.* 346:967–989.

14. CORRY, B. 2006. An energy-efficient gating mechanism in the acetylcholine receptor channel suggested by molecular and brownian dynamics. *Biophys. J.* 90:799–810.

15. CYMES, G. D. & GROSMAN, C. 2008. Pore-opening mechanism of the nicotinic acetylcholine receptor evinced by proton transfer. *Nature Struct. Mol. Biol.* 15:389–396.

16. CHANGEUX, J.-P. & TALY, A. 2008. Nicotinic receptors, allosteric proteins and medicine. *Trends. Mol. Med.* 14:93–102.

SYNOPSIS

Plasma membranes are remarkably thin, delicate structures, yet they play a key role in many of a cell's most important functions. The plasma membrane separates the living cell from its environment; it provides a selectively permeable barrier that allows the exchange of certain substances, while preventing the passage of others; it contains the machinery that physically transports substances from one side of the membrane to another; it contains receptors that bind to specific ligands in the external space and relay the information to the cell's internal compartments; it mediates interactions with other cells; it provides a framework in which components can be organized; it is a site where energy is transduced from one type to another. (p. 117)

Membranes are lipid–protein assemblies in which the components are held together in a thin sheet by noncovalent bonds. The membrane is held together as a cohesive sheet by a lipid bilayer consisting of a bimolecular layer of amphipathic lipids, whose polar head groups face outward and hydrophobic fatty acyl tails face inward. Included among the lipids are phosphoglycerides, such as phosphatidylcholine; sphingosine-based lipids, such as the phospholipid sphingomyelin and the carbohydrate-containing cerebrosides and gangliosides (glycolipids); and cholesterol. The proteins of the membrane can be divided into three groups: integral proteins that penetrate into and through the lipid bilayer, with portions exposed on both the cytoplasmic and extracellular membrane surfaces; peripheral proteins that are present wholly outside the lipid bilayer, but are noncovalently associated with either the polar head groups of the lipid bilayer or with the surface of an integral protein; and lipid-anchored proteins that are outside the lipid bilayer but covalently linked to a lipid that is part of the bilayer. The transmembrane segments of integral proteins occur typically as an α helix, composed predominantly of hydrophobic residues. (p. 122)

Membranes are highly asymmetric structures whose two leaflets have very different properties. As examples, all of the membrane's carbohydrate chains face away from the cytosol; many of the integral proteins bear sites on their external surface that interact with extracellular ligands and sites on their internal surface that interact with peripheral proteins that form part of an inner membrane skeleton; and the phospholipid content of the two halves of the bilayer is highly asymmetric. The organization of the proteins within the membrane is best revealed in freeze-fracture replicas in which the cells are frozen, their membranes are split through the center of the bilayer by a fracture plane, and the exposed internal faces are visualized by formation of a metal replica. (p. 125)

The physical state of the lipid bilayer has important consequences for the lateral mobility of both phospholipids and integral proteins. The viscosity of the bilayer and the temperature at which it undergoes phase transition depend on the degree of unsaturation and the length of the fatty acyl chains of the phospholipids. Maintaining a fluid membrane is important for many cellular activities, including signal transduction, cell division, and formation of specialized membrane regions. The lateral diffusion of proteins within the membrane was originally demonstrated by cell fusion, and it can be quantitated by techniques that follow the movements of proteins tagged with fluorescent compounds or electron-dense markers. Measurement of the diffusion coefficients of integral proteins suggests that most are subject to restraining influences that inhibit their mobility. Proteins may be restrained by association with other integral proteins or with peripheral proteins located on the membrane surface. Because of these various types of restraint, membranes achieve a considerable measure of organizational stability in which particular membrane regions are differentiated from one another. (p. 133)

The erythrocyte plasma membrane contains two major integral proteins, band 3 and glycophorin A, and a well-defined inner skeleton composed of peripheral proteins. Each band 3 subunit spans the membrane at least a dozen times and contains an internal channel through which bicarbonate and chloride anions are exchanged. Glycophorin A is a heavily glycosylated protein of uncer-

tain function containing a single transmembrane domain consisting of a hydrophobic α helix. The major component of the membrane skeleton is the fibrous protein spectrin, which interacts with other peripheral proteins to provide support for the membrane and restrain the diffusion of its integral proteins. (p. 140)

The plasma membrane is a selectively permeable barrier that allows solute passage by several mechanisms, including simple diffusion through either the lipid bilayer or membrane channels, facilitated diffusion, and active transport. Diffusion is an energy-independent process in which a solute moves down an electrochemical gradient, dissipating the free energy stored in the gradient. Small inorganic solutes, such as O_2, CO_2, and H_2O, penetrate the lipid bilayer readily, as do solutes with large partition coefficients (high lipid solubility). Ions and polar organic solutes, such as sugars and amino acids, require special transporters to enter or leave the cell. (p. 143)

Water moves by osmosis through the semipermeable plasma membrane from a region of lower solute concentration (the hypotonic compartment) to a region of higher solute concentration (the hypertonic compartment). Osmosis plays a key role in a multitude of physiologic activities. In plants, for example, the influx of water generates turgor pressure against the cell wall that helps support nonwoody tissues. (p. 144)

Ions diffuse through a plasma membrane by means of special protein-lined channels that are often specific for particular ions. Ion channels are usually gated and are controlled by either voltage or chemical ligands, such as neurotransmitters. Analysis of a bacterial ion channel (KcsA) has revealed how oxygen atoms from the backbone of the polypeptide are capable of replacing the water molecules normally assocated with K^+ ions, allowing the protein to selectively conduct K^+ ions through its central channel. Voltage-gated K^+ channels contain a charged helical segment that moves in response to the membrane's voltage leading to the opening or closing of the gate. (p. 146)

Facilitated diffusion and active transport involve integral membrane proteins that combine specifically with the solute to be transported. Facilitative transporters act without the input of energy and move solutes down a concentration gradient in either direction across the membrane. They are thought to act by changing conformation, which alternately exposes the solute binding site to the two sides of the membrane. The glucose transporter is a facilitative transporter whose presence in the plasma membrane is stimulated by increasing levels of insulin. Active transporters require the input of energy and move ions and solutes against a concentration gradient. P-type active transporters, such as the Na^+/K^+-ATPase, are driven by the transfer of a phosphate group from ATP to the transporter, changing its affinity toward the transported ion. Secondary active-transport systems tap the energy stored in an ionic gradient to transport a second solute against a gradient. For example, the active transport of glucose across the apical surface of an intestinal epithelial cell is driven by the cotransport of Na^+ down its electrochemical gradient. (p. 152)

The resting potential across the plasma membrane is due largely to the limited permeability of the membrane to K^+ and is subject to dramatic change. The resting potential of a typical nerve or muscle cell is about -70 mV (inside negative). When the membrane of an excitable cell is depolarized past a threshold value, events are initiated that lead to the opening of the gated Na^+ channels and the influx of Na^+, which is measured as a reversal in the voltage across the membrane. Within milliseconds after they have opened, the Na^+ gates close and gated potassium channels open, which leads to an efflux of K^+ and a restoration of the resting potential. The series of dramatic changes in membrane potential following depolarization constitutes an action potential. (p. 159)

Once an action potential has been initiated, it becomes self-propagating. Action potentials propagate because the depolarization that accompanies an action potential at one site on the membrane is sufficient to depolarize the adjacent membrane, which initiates an action potential at that site. In a myelinated axon, an action potential at one node in the sheath is able to depolarize the membrane at the next node, allowing the action potential to hop rapidly from node to node. When the action potential reaches the terminal knobs of an axon, the calcium gates in the plasma membrane open, allowing an influx of Ca^{2+}, which triggers the fusion of the membranes of neurotransmitter-containing secretory vesicles with the overlying plasma membrane. The neurotransmitter diffuses across the synaptic cleft and binds to receptors on the postsynaptic membrane, inducing either the depolarization or hyperpolarization of the target cell. (p. 162)

ANALYTIC QUESTIONS

1. What types of integral proteins would you expect to reside in the plasma membrane of an epithelial cell that might be absent from that of an erythrocyte? How do such differences relate to the activities of these cells?

2. Many different types of cells possess receptors that bind steroid hormones, which are lipid-soluble molecules. Where in the cell do you think such receptors might reside? Where in the cell would you expect the insulin receptor to reside? Why?

3. When the trilaminar appearance of the plasma membrane was first reported, the images (e.g., Figure 4.1*a*) were taken as evidence to support the Davson-Danielli model of plasma membrane structure. Why do you think these micrographs might have been interpreted in this way?

4. Suppose you were planning to use liposomes in an attempt to deliver drugs to a particular type of cell in the body, for example, a fat or muscle cell. Is there any way you might be able to construct the liposome to increase its target specificity?

5. How is it that, unlike polysaccharides such as starch and glycogen, the oligosaccharides on the surface of the plasma membrane can be involved in specific interactions? How is this feature illustrated by determining a person's blood type prior to receiving a transfusion?

6. Trypsin is an enzyme that can digest the hydrophilic portions of membrane proteins, but it is unable to penetrate the lipid bilayer and enter a cell. Because of these properties, trypsin has been used in conjunction with SDS–PAGE to determine which proteins have an extracellular domain. Describe an experiment using trypsin to determine the sidedness of proteins of the erythrocyte membrane.

7. Look at the scanning electron micrograph of erythrocytes in Figure 4.32*a*. These cells, which are flattened and have circular depressions on each side, are said to have a biconcave shape. What is the physiologic advantage of a biconcave erythocyte over a spherical cell with respect to O_2 uptake?

8. Suppose you were culturing a population of bacteria at 15°C and then raised the temperature of the culture to 37°C. What effect do you think this might have on the fatty acid composition of the membrane? on the transition temperature of the lipid bilayer? on the activity of membrane desaturases?

9. Looking at Figure 4.6, which lipids would you expect to have the greatest rate of flip-flop? which the least? Why? If you determined experimentally that phosphatidylcholine actually exhibited the greatest rate of flip-flop, how might you explain this finding? How would you expect the rate of phospholipid flip-flop to compare with that of an integral protein? Why?

10. What is the difference between a two-dimensional and a three-dimensional representation of a membrane protein? How are the different types of profiles obtained, and which is more useful? Why do you think there are so many more proteins whose two-dimensional structure is known?

11. If you were to inject a squid giant axon with a tiny volume of solution containing 0.1 M NaCl and 0.1 M KCl in which both the Na^+ and K^+ ions were radioactively labeled, which of the radioactively labeled ions would you expect to appear most rapidly within the seawater medium while the neuron remained at rest? after the neuron had been stimulated to conduct a number of action potentials?

12. It has been difficult to isolate proteins containing water channels (i.e., aquaporins) due to the high rate of diffusion of water through the lipid bilayer. Why would this make aquaporin isolation difficult? Is there any way you might be able to distinguish diffusion of water through the lipid bilayer versus that through aquaporins? The best approach to studying aquaporin behavior has been to express the aquaporin genes in frog oocytes. Is there any reason why the oocytes of a pond-dwelling amphibian might be particularly well suited for such studies?

13. How is it that diffusion coefficients measured for lipids within membranes tend to be closer to that expected for free diffusion than those measured for integral proteins in the same membranes?

14. Assume that the plasma membrane of a cell was suddenly permeable to the same degree to both Na^+ and K^+ and that both ions were present at a concentration gradient of the same magnitude. Would you expect these two ions to move across the membrane at the same rates? Why or why not?

15. Most marine invertebrates show no loss or gain of water by osmosis, whereas most marine vertebrates experience continual water loss in their high-salt environment. Speculate on the basis for this difference and how it might reflect different pathways of evolution of the two groups.

16. How would you expect the concentrations of solute inside a plant cell to compare to that of its extracellular fluids? Would you expect the same to be true of the cells of an animal?

17. What would be the consequence for impulse conduction if the Na^+ channels were able to reopen immediately after they had closed during an action potential?

18. What would be the value of the potassium equilibrium potential if the external K^+ concentration was 200 mM and the internal concentration was 10 mM at 25°C? at 37°C?

19. As discussed on page 158, the Na^+/glucose cotransporter transports two Na^+ ions for each glucose molecule. What if this ratio were 1:1 rather than 2:1; how would this affect the glucose concentration that the transporter could work against?

20. A transmembrane protein usually has the following features: (1) the portion that transits the membrane bilayer is at least 20 amino acids in length, which are largely or exclusively nonpolar residues; (2) the portion that anchors the protein on the external face has two or more consecutive acidic residues; and (3) the portion that anchors the protein on the cytoplasmic face has two or more consecutive basic residues. Consider the transmembrane protein with the following sequence:

NH_2-MLSTGVKRKGAVLLILLFPWMVAGGPLFWLA ADESTYKGS-COOH

Draw this protein as it would reside in the plasma membrane. Make sure you label the N- and C-termini and the external and cytoplasmic faces of the membrane. (Single-letter code for amino acids is given in Figure 2.26.)

21. Many marine invertebrates, such as squid, have extracellular fluids that resemble seawater and therefore have much higher intracellular ion concentrations than mammals. For a squid neuron, the ionic concentrations are roughly

ion	intracellular concentration	extracellular concentration
K^+	350 mM	10 mM
Na^+	40 mM	440 mM
Cl^-	100 mM	560 mM
Ca^{2+}	2×10^{-4} mM	10 mM
pH	7.6	8.0

If the resting potential of the plasma membrane, V_m, is -70 mV, are any of the ions at equilibrium? How far out of equilibrium, in mV, is each ion? What is the direction of net movement of each ion through an open channel permeable to that ion?

22. The membrane potential of a cell is determined by the relative permeability of the membrane to various ions. When acetylcholine binds to its receptors on the postsynaptic muscle membrane, it causes a massive opening of channels that are equally permeable to sodium and potassium ions. Under these conditions,

$$V_m = (V_{K^+} + V_{Na^+})/2$$

If $[K^+_{in}] = 140$ mM and $[Na^+_{in}] = 10$ mM for the muscle cell, and $[Na^+_{out}] = 150$ mM and $[K^+_{out}] = 5$ mM, what is the membrane potential of the neuromuscular junction of an acetylcholine-stimulated muscle?

23. Transmembrane domains consist of individual α helices or a β sheet formed into a barrel. Looking at Figures 2.30 and 2.31, why is a single α helix more suited to spanning the bilayer than a single β strand?

24. Knowing how the K^+ channel selects for K^+ ions, suggest a mechanism by which the Na^+ channel is able to select for its ion.

25. How would you compare the rate of movement of ions passing through a channel versus those transported actively by a P-type pump? Why?

5
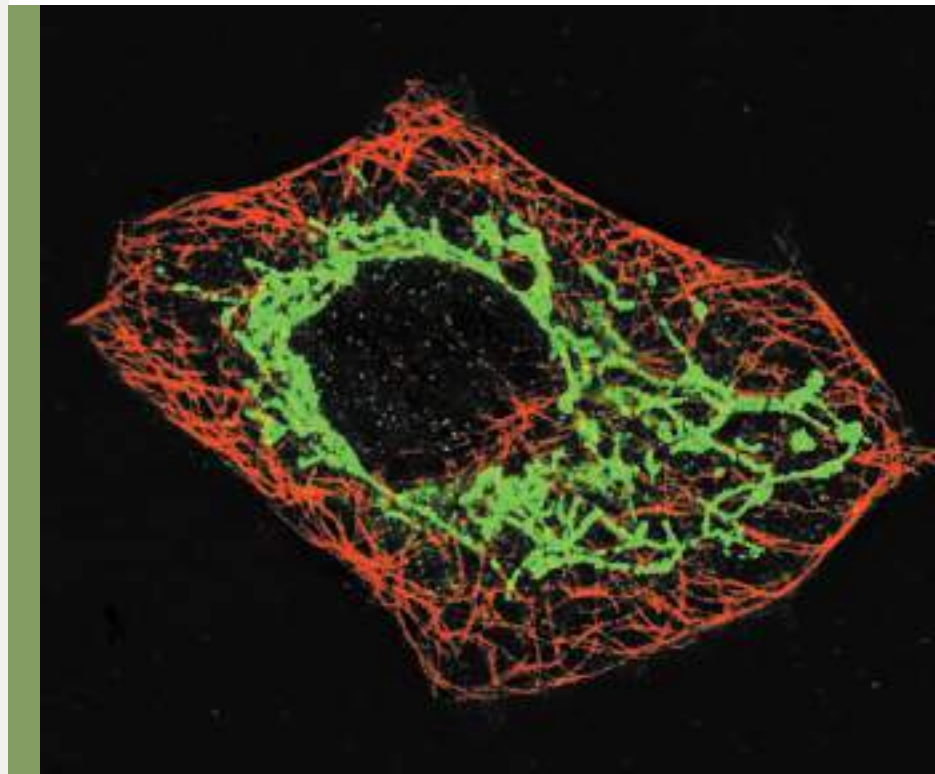
Aerobic Respiration and the Mitochondrion

D uring the first two billion years or so that life existed on Earth, the atmosphere consisted largely of reduced molecules, such as molecular hydrogen (H_2), ammonia (NH_3), and H_2O. The Earth of this period was populated by **anaerobes**—organisms that captured and utilized energy by means of oxygen-independent (anaerobic) metabolism, such as glycolysis and fermentation (Figures 3.24 and 3.29). Then, between 2.4 and 2.7 billion years ago, the cyanobacteria appeared—a new kind of organism that carried out a new type of photosynthetic process, one in which water molecules were split apart and molecular oxygen (O_2) was released. At some point following the appearance of the cyanobacteria, the Earth's atmosphere began to accumulate significant levels of oxygen, which set the stage for a dramatic change in the types of organisms that would come to inhabit the planet.

As discussed on page 34, molecular oxygen can be a very toxic substance, taking on extra electrons and reacting with a variety of biological molecules. The presence of oxygen in the atmosphere must have been a powerful agent for natural selection. Over time, species evolved that were not only protected from the damaging effects of molecular oxygen, but possessed metabolic pathways that utilized the molecule to great advantage. Without the ability to use oxygen, organisms could

Micrograph of a mammalian fibroblast that has been fixed and stained with fluorescent antibodies that reveal the distribution of the mitochondria (green) and the microtubules of the cytoskeleton (red). The mitochondria are seen as an extensive network or reticulum that extends through much of the cell. (From Michael P. Yaffe, University of California, San Diego, Reprinted with permission from Science 283:1493, 1999; Copyright 1999, American Association for the Advancement of Science.)

only extract a limited amount of energy from their foodstuffs, excreting energy-rich products such as lactic acid and ethanol, which they were unable to metabolize further. In contrast, organisms that incorporated O_2 into their metabolism could completely oxidize such compounds to CO_2 and H_2O and, in the process, extract a much larger percentage of their energy content. These organisms that became dependent on oxygen were Earth's first **aerobes**, and they eventually gave rise to all of the oxygen-dependent prokaryotes and eukaryotes living today. In eukaryotes, the utilization of oxygen as a means of energy extraction takes place in a specialized organelle, the **mitochondrion**. As discussed in Chapter 1, a massive body of data indicates that mitochondria have evolved from an ancient aerobic bacterium that took up residence inside the cytoplasm of an anaerobic host cell. ■

5.1 MITOCHONDRIAL STRUCTURE AND FUNCTION

Mitochondria are large enough to be seen in the light microscope (Figure 5.1*a*), and their presence within cells has been known for over a hundred years. Depending on the cell type, mitochondria can have a very different overall structure. At one end of the spectrum, mitochondria can appear as individual, bean-shaped organelles (Figure 5.1*b*), ranging from 1 to 4 μm in length. At the other end of the spectrum, mitochondria can appear as a highly branched, interconnected tubular network. This type of mitochondrial structure can be seen in the chapter-opening micrograph.

Observations of fluorescently labeled mitochondria within living cells have shown them to be dynamic organelles capable of dramatic changes in shape. Most importantly, mitochondria can fuse with one another, or split in two (Figure 5.2). Our understanding of mitochondrial fission and fusion has improved in recent years with the development of in vitro assays for their study and the identification of proteins required for both events. The balance between fusion and fission is likely a major determinant of mitochondrial number, length, and degree of interconnection. When fusion becomes more frequent than fission, the mitochondria tend to become more elongated and interconnected, whereas a predominance of fission leads to the formation of more numerous and distinct mitochondria. A number of inherited neurologic diseases are caused by mutations in genes that encode components of the mitochondrial fusion machinery.

Mitochondria occupy 15 to 20 percent of the volume of an average mammalian liver cell and contain more than a thousand different proteins. These organelles are best known for their role in generating the ATP that is used to run most of the cell's energy-requiring activities. To accomplish this function, mitochondria are often associated with fatty acid-containing oil droplets from which they derive raw materials to be oxidized. A particularly striking arrangement of mitochondria occurs in sperm cells, where they are often located in the midpiece, just behind the nucleus (Figure 5.1*c*). The movements of a sperm are powered by ATP produced in these mitochondria. Mitochondria are also prominent in many plant cells where they are the primary suppliers of ATP in nonphotosynthetic tissues, as well as being a source of ATP in photosynthetic leaf cells during periods of dark.

(a)   (b)   (c)
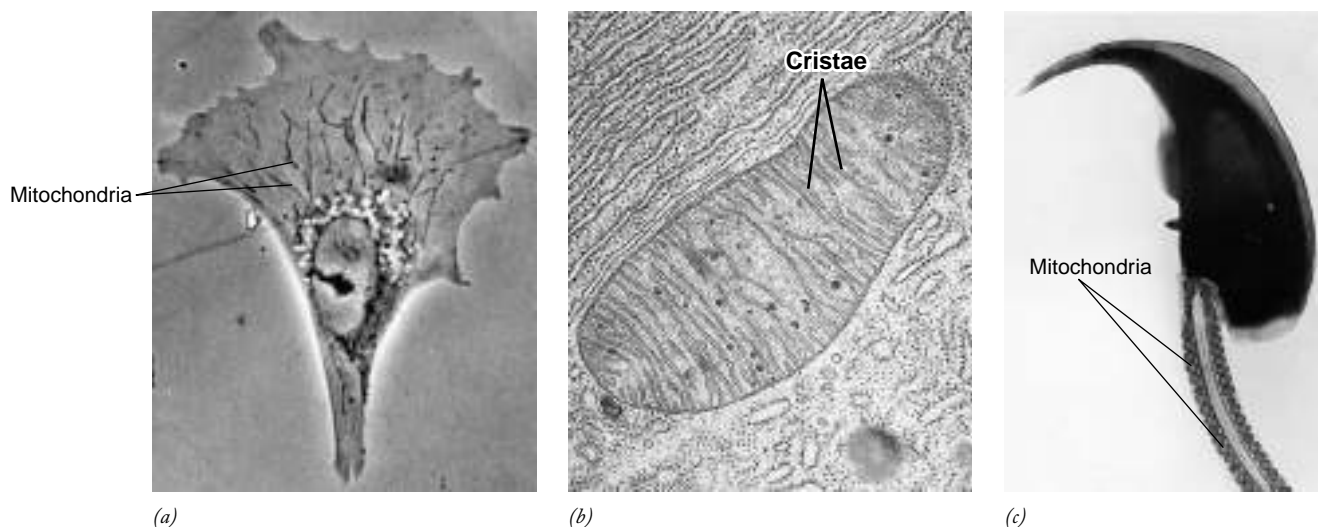
FIGURE 5.1 Mitochondria. (*a*) A living fibroblast viewed with a phase-contrast microscope. Mitochondria are seen as elongated, dark bodies. (*b*) Transmission electron micrograph of a thin section through a mitochondrion revealing the internal structure of the organelle, particularly the membranous cristae of the inner membrane. (*c*) Localization of mitochondria in the sperm midpiece surrounding the proximal portion of the flagellum. (A: COURTESY OF NORMAN K. WESSELS. B: COURTESY OF K. R. PORTER/PHOTO RESEARCHERS; C: DON W. FAWCETT/VISUALS UNLIMITED.)
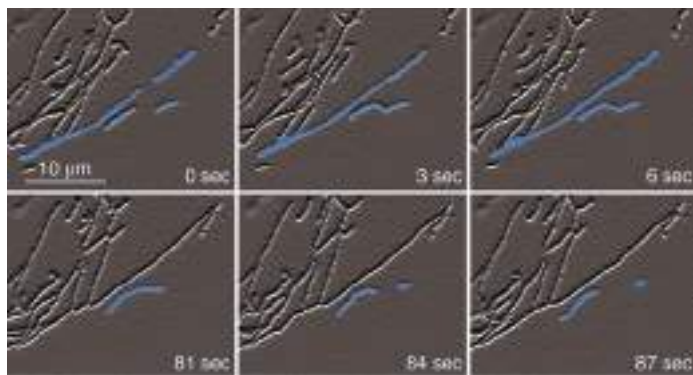
FIGURE 5.2 Mitochondrial fusion and fission. The dynamic nature of these organelles is captured in the frames of this movie, which shows a portion of a mouse fibroblast whose mitochondria have been labeled with a fluorescent protein. In the first three frames, two pairs of mitochondria (which have been artificially colored) contact end-to-end and immediately fuse. In the last three frames, the lower fusion product undergoes fission, and the daughter mitochondria move apart. (FROM DAVID C. CHAN, CELL 125:1242, 2006; BY PERMISSION OF CELL PRESS.)

While energy metabolism has been the focus of interest in the study of mitochondria, these organelles are also involved in other, unrelated activities. For example, mitochondria are the sites of synthesis of numerous substances, including certain amino acids and the heme groups shown in Figures 2.34 and 5.12. Mitochondria also play a vital role in the uptake and release of calcium ions. Calcium ions are essential triggers for cellular activities (Section 15.5), and mitochondria (along with the endoplasmic reticulum) play an important role in regulating the Ca^{2+} concentration of the cytosol. The process of cell death, which plays an enormous role in the life of all multicellular animals, is also regulated to a large extent by events that occur within mitochondria (Section 15.8).

Mitochondrial Membranes

If you take a close look at the electron micrograph of Figure 5.1*b*, it can be seen that the outer boundary of a mitochondrion contains two membranes: the *outer mitochondrial membrane* and the *inner mitochondrial membrane*. The outer mitochondrial membrane completely encloses the mitochondrion, serving as its outer boundary. The inner mitochondrial membrane is subdivided into two interconnected domains. One of these domains, called the *inner boundary membrane*, lies just inside the outer mitochondrial membrane, forming a double-membrane outer envelope. The inner boundary membrane is particularly rich in the proteins responsible for the import of mitochodrial proteins (see Figure 8.47). The other domain of the inner mitochondrial membrane is present within the interior of the organelle as a series of invaginated membranous sheets, called **cristae**. The role of mitochondria as energy transducers is intimately tied to the membranes of the cristae that are so prominent in electron micrographs of these organelles. The cristae contain a large amount of membrane surface, which houses the machinery needed for aerobic respiration and ATP formation (see Figure 5.21). The organization of the cristae are shown in a clearer profile in the scanning electron micrograph of Figure 5.3*a* and the three-dimensional reconstruction of Figure 5.3*b*. The inner boundary membrane and internal cristal membranes are joined to one another by narrow connections, or *cristae junctions*, as shown in the schematic illustrations of Figure 5.3*c*.

The membranes of the mitochondrion divide the organelle into two aqueous compartments, one within the interior of the mitochondrion, called the **matrix**, and a second between the outer and inner membrane, called the **intermembrane space**. The matrix has a gel-like consistency owing to the presence of a high concentration (up to 500 mg/ml) of water-soluble proteins. The proteins of the intermembrane space are best known for their role in initiating cell suicide, a subject discussed in Section 15.8.

The outer and inner membranes have very different properties. The outer membrane is composed of approximately 50 percent lipid by weight and contains a curious mixture of enzymes involved in such diverse activities as the oxidation of epinephrine, the degradation of tryptophan, and the elongation of fatty acids. In contrast, the inner membrane contains more than 100 different polypeptides and has a very high protein/lipid ratio (more than 3 : 1 by weight, which corresponds to about one protein molecule for every 15 phospholipids). The inner membrane is virtually devoid of cholesterol and rich in an unusual phospholipid, cardiolipin (diphosphatidylglycerol; see Figure 4.6 for the structure), which are characteristics of bacterial plasma membranes, from which the inner mitochondrial membrane has presumably evolved. Cardiolipin plays an important role in facilitating the activity of the proteins involved in ATP synthesis. The outer mitochondrial membrane is thought to be homologous to an outer membrane present as part of the cell wall of certain bacterial cells. The outer mitochondrial membrane and outer bacterial membrane both contain *porins*, integral proteins that have a relatively large internal channel (e.g., 2–3 nm) surrounded by a barrel of β strands (Figure 5.4). The porins of the outer mitochondrial membrane are not static structures as was once thought but can undergo reversible closure in response to conditions within the cell. When the porin channels are wide open, the outer membrane is freely permeable to molecules such as ATP, NAD, and coenzyme A, which have key roles to play in energy metabolism within the mitochondrion. In contrast, the inner mitochondrial membrane is highly impermeable; virtually all molecules and ions require special membrane transporters to gain entrance to the matrix.

As will be discussed in following sections, the composition and organization of the inner mitochondrial membrane are the keys to the bioenergetic activities of the organelle. The architecture of the inner membrane and the apparent fluidity of its bilayer facilitate the interactions of components that are required for ATP formation.

(a)

0.2 µm

(b)
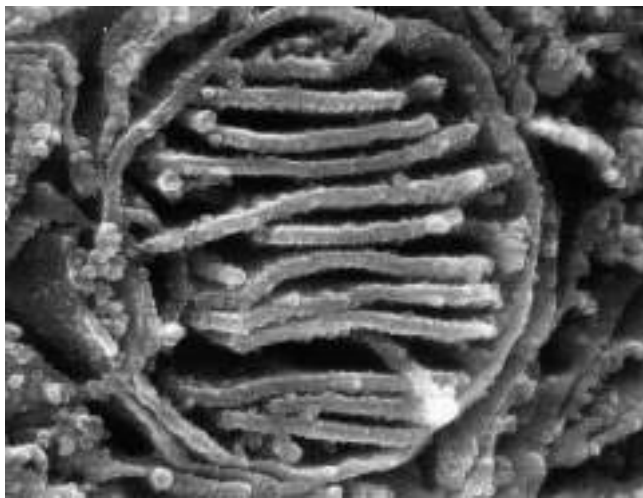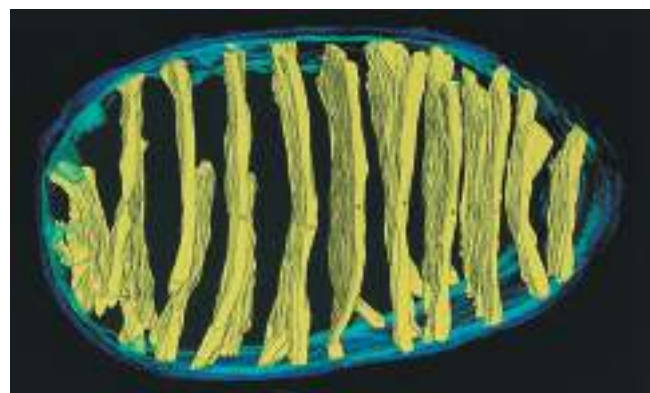
1 µm

FIGURE 5.3 **The structure of a mitochondrion.** (*a*) Scanning electron micrograph of a macerated mitochondrion, showing the internal matrix enclosed by folds of the inner membrane. (*b*) Three-dimensional reconstruction of a mitochondrion based on a series of micrographs taken with a high-voltage electron microscope of a single thick section of brown fat tissue that had been tilted at various angles. High-voltage instruments accelerate electrons at velocities that allow them to penetrate thicker tissue sections (up to 1.5 µm). This technique suggests that the cristae are present as flattened sheets (lamellae) that communicate with the intermembrane space by way of narrow tubular openings, rather than "wide open" channels as is typically depicted. In this reconstruction, the inner mitochondrial membrane is shown in blue at the peripheral regions and in yellow when it penetrates into the matrix to form the cristae. (*c*) Schematic diagrams showing the three-dimensional internal structure (top) and a thin section (bottom) of a mitochondrion from bovine heart tissue. (A: FROM K. TANAKA AND T. NAGURO, INT. REV. CYTOL. 68:111, 1980. B: FROM G. A. PERKINS ET AL., J. BIOEN. BIOMEMB. 30:436, 1998.)

(c)
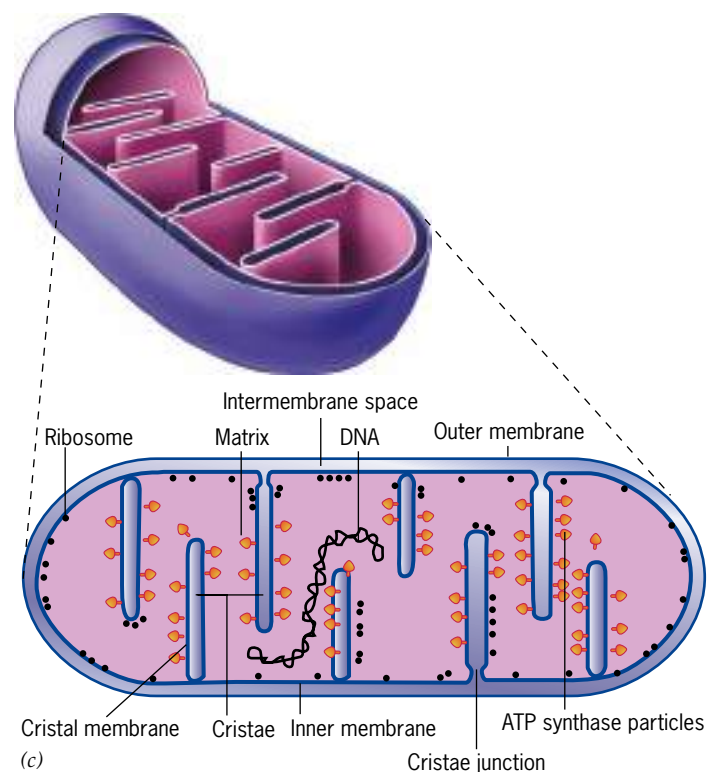
The Mitochondrial Matrix

In addition to an array of enzymes, the mitochondrial matrix also contains ribosomes (of considerably smaller size than those found in the cytosol) and several molecules of DNA, which is circular in higher plants and animals (Figure 5.3*c*). Thus, mitochondria possess their own genetic material and the machinery to manufacture their own RNAs and proteins. This nonchromosomal DNA is important because it encodes a small number of mitochondrial polypeptides (13 in humans) that are tightly integrated into the inner mitochondrial membrane along with polypeptides encoded by genes residing

FIGURE 5.4 **Porins.** Gram-negative bacteria have a lipid-containing outer membrane outside of their plasma membrane as part of their cell wall. This outer membrane contains proteins, called porins, that consist of a barrel of β sheet and form an opening through which moderate-sized molecules can penetrate. This image shows the protein OmpW embedded in the outer membrane of *E. coli*. The porin contains a small hydrophobic compound within its central channel. A variety of porins having different sized channels and selectivities are also found in the outer mitochondrial membrane in eukaryotic cells. (FROM HEEDEOK HONG ET AL., J. BIOL. CHEM. 281, COVER OF #11, 2006; COURTESY OF BERT VAN DEN BERG.)
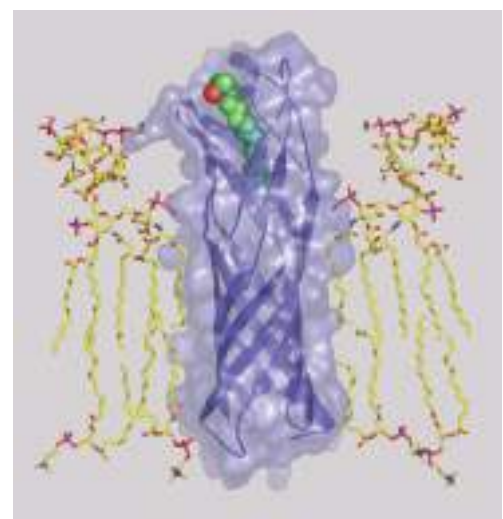
within the nucleus. Human mitochondrial DNA also encodes 2 ribosomal RNAs and 22 tRNAs that are used in protein synthesis within the organelle. Mitochondrial DNA (mtDNA) is a relic of ancient history. It is thought to be the legacy from a single aerobic bacterium that took up residence in the cytoplasm of a primitive cell that ultimately became an ancestor of all eukaryotic cells (page 25). Most of the genes of this ancient symbiont were either lost or transferred over the course of evolution to the nucleus of the host cell, leaving only a handful of genes to encode some of the most hydrophobic proteins of the inner mitochondrial membrane. It is interesting to note that the RNA polymerase that synthesizes the various mitochondrial RNAs is not related to the multisubunit enzyme found in prokaryotic and eukaryotic cells. Instead, the mitochondrial RNA polymerase is a single subunit enzyme similar in many respects to certain bacterial viruses (bacteriophage) from which it appears to have evolved. For a number of reasons, mtDNA is well suited for use in the study of human migration and evolution. A number of companies, for example, employ mtDNA sequences to trace the ancestry of clients who are searching for their ethnic or geographic roots.

We will return later in the chapter to discuss the molecular architecture of mitochondrial membranes, but first let us consider the role of these organelles in the basic oxidative pathways of eukaryotic cells, which is summarized in Figure 5.5. It might help to examine this overview figure and read the accompanying legend before moving on to the following detailed descriptions of these pathways.

REVIEW

1. Describe the metabolic changes in oxidative metabolism that must have accompanied the evolution and success of the cyanobacteria.

2. Compare the properties and evolutionary history of the inner and outer mitochondrial membranes; the intermembrane space and the matrix.

5.2 OXIDATIVE METABOLISM IN THE MITOCHONDRION

In Chapter 3 we described the initial stages in the oxidation of carbohydrates. Beginning with glucose, the first steps of the oxidation process are carried out by the enzymes of glycolysis, which are located in the cytosol (Figure 5.5). The 10 reactions that constitute the glycolytic pathway were illustrated in Figure
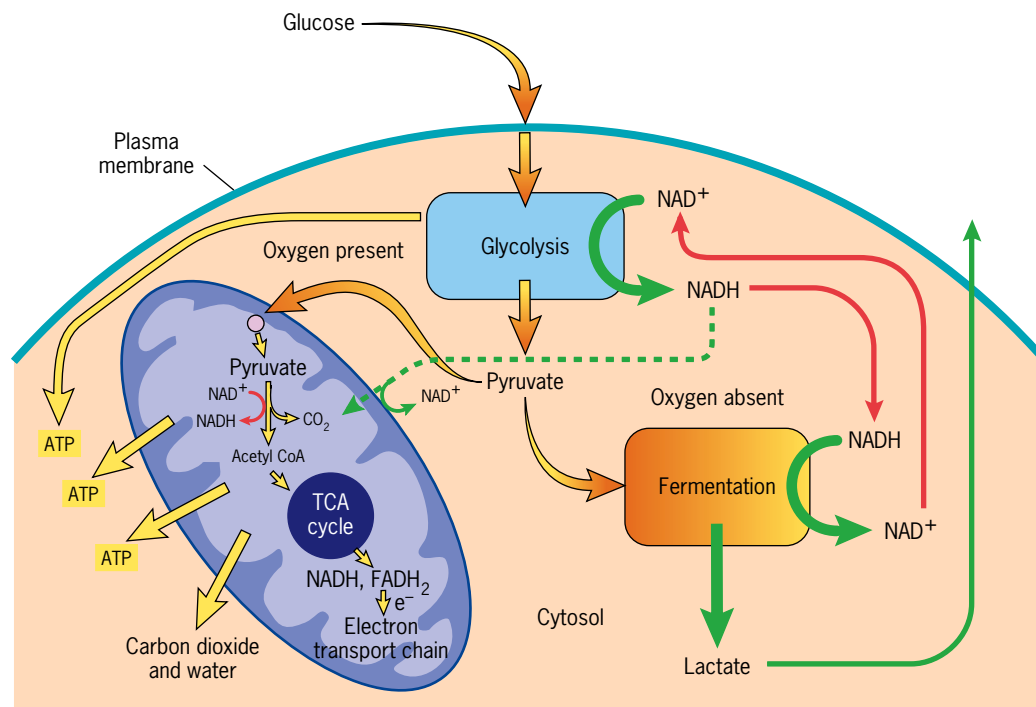
FIGURE 5.5 **An overview of carbohydrate metabolism in eukaryotic cells.** The reactions of glycolysis generate pyruvate and NADH in the cytosol. In the absence of O_2, the pyruvate is reduced by NADH to lactate (or another product of fermentation, such as ethanol in yeast; see Figure 3.29 for details). The NAD^+ formed in the reaction is reutilized in the continuation of glycolysis. In the presence of O_2, the pyruvate moves into the matrix (facilitated by a membrane transporter), where it is decarboxylated and linked to coenzyme A (CoA), a reaction that generates NADH. The NADH produced during glycolysis donates its high-energy electrons to a compound that crosses the inner mitochondrial membrane (as shown in Figure 5.9). The acetyl CoA passes through the TCA cycle (as shown in Figure 5.7), which generates NADH and $FADH_2$. The electrons in these various NADH and $FADH_2$ molecules are passed along the electron-transport chain, which is made up of carriers that are embedded in the inner mitochondrial membrane, to molecular oxygen (O_2). The energy released during electron transport is used in the formation of ATP by a process discussed at length later in the chapter. If all of the energy from electron transport were to be utilized in ATP formation, approximately 36 ATPs could be generated from a single molecule of glucose.

FIGURE 5.6 An overview of glycolysis showing some of the key steps. These include the two reactions in which phosphate groups are transferred from ATP to the six-carbon sugar to produce fructose 1,6-bisphosphate (steps 1, 3); the oxidation and phosphorylation of glyceraldehyde 3-phosphate to produce 1,3-bisphosphoglycerate and NADH (step 6); and the transfer of phosphate groups from three-carbon phosphorylated substrates to ADP to produce ATP by substrate phosphorylation (steps 7 and 10). Keep in mind that two molecules of glyceraldehyde 3-phosphate are formed from each molecule of glucose, so reactions 6 through 10 shown here occur twice per glucose molecule oxidized.
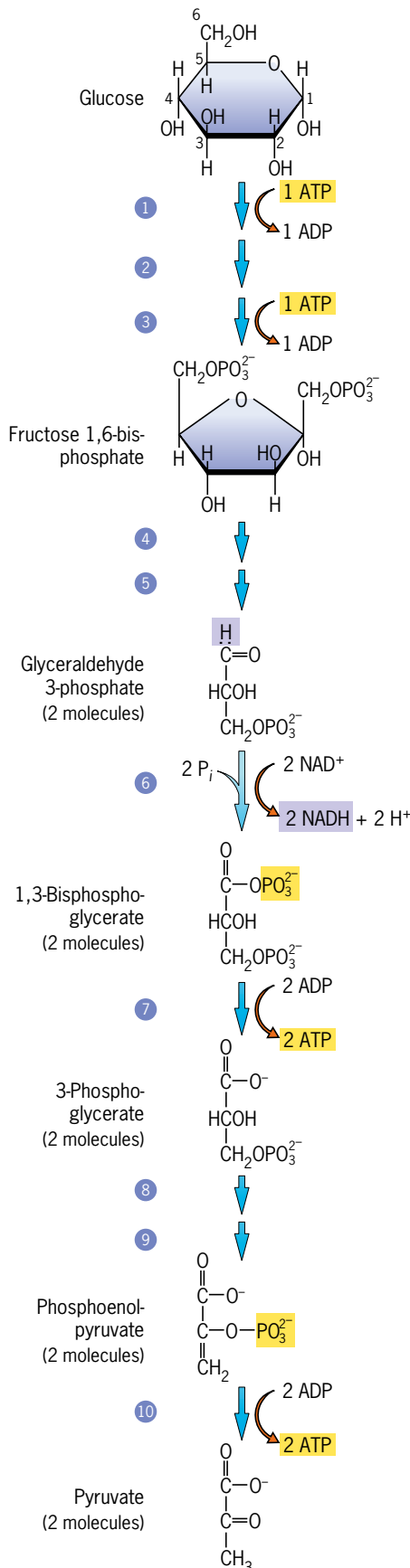
Glucose is phosphorylated at the expense of one ATP, rearranged structurally to form fructose phosphate, and then phosphorylated again at the expense of a second ATP. The two phosphate groups are situated at the two ends (C1, C6) of the fructose chain.

The six-carbon bisphosphate is split into two three-carbon monophosphates.

The three-carbon aldehyde is oxidized to an acid as the electrons removed from the substrate are used to reduce the coenzyme NAD^+ to NADH. In addition, the C1 acid is phosphorylated to form an acyl phosphate, which has a high phosphate group-transfer potential (denoted by the yellow shading).

The phosphate group from C1 is transferred to ADP forming ATP by substrate-level phosphorylation. Two ATPs are formed per glucose oxidized.

These reactions result in the rearrangement and dehydration of the substrate to form an enol phosphate at the C2 position that has a high phosphate group-transfer potential.

The phosphate group is transferred to ADP forming ATP by substrate-level phosphorylation, generating a ketone at the C2 position. Two ATPs are formed per glucose oxidized.

NET REACTION:

Glucose + 2 NAD^+ + 2 ADP + 2 P_i $\longrightarrow$ 2 Pyruvate + 2 ATP + 2 NADH + 2 H^+ + 2 H_2O
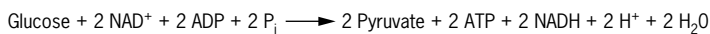
3.23; the major steps of the pathway are summarized in Figure 5.6. Only a small fraction of the free energy available in glucose is made available to the cell during glycolysis—enough for the net synthesis of only two molecules of ATP per molecule of glucose oxidized (Figure 5.6). Most of the energy remains stored in pyruvate. Each molecule of NADH produced during the oxidation of glyceraldehyde 3-phosphate (reaction 6, Figure 5.6) also carries a pair of high-energy electrons.[1] The two products of glycolysis—pyruvate and NADH—can be metabolized in two very different ways, depending on the type of cell in which they are formed and the presence or absence of oxygen.

In the presence of O_2, aerobic organisms are able to extract large amounts of additional energy from the pyruvate and NADH produced during glycolysis—enough to synthesize more than 30 additional ATP molecules. This energy is extracted in mitochondria (Figure 5.5). We will begin with pyruvate and return to consider the fate of the NADH later in the discussion. Each pyruvate molecule produced by glycolysis is transported across the inner mitochondrial membrane and into the matrix, where it is decarboxylated to form a two-carbon acetyl group ($—CH_3COO^-$). The acetyl group is trans-

---

[1]The terms *high-energy electrons* and *low-energy electrons* are not always well received by biochemists, but they convey a useful image. As discussed on page 182, high-energy electrons are ones that are held with lower affinity and are more readily transferred from a donor to an acceptor than are low-energy electrons.

ferred to *coenzyme A* (a complex organic compound derived from the vitamin pantothenic acid) to produce acetyl CoA.

$$Pyruvate + HS—CoA + NAD^+ \rightarrow$$
$$acetyl\ CoA + CO_2 + NADH + H^+$$

The decarboxylation of pyruvate and transfer of the acetyl group to CoA (Figures 5.5 and 5.7) is catalyzed by the giant multienzyme complex pyruvate dehydrogenase, whose structure was shown in Figure 2.39.

FIGURE 5.7 The tricarboxylic acid (TCA) cycle, also called the Krebs cycle after the person who formulated it, or the citric acid cycle after the first compound formed in it. The cycle begins with the condensation of oxaloacetate (OAA) and acetyl CoA (reaction 12). The carbons of these two compounds are marked with numbers or letters. The two carbons lost during passage through the cycle are derived from oxaloacetate. The standard free energies (in kcal/mol) and names of enzymes are also provided. Five pairs of electrons are removed from substrate molecules by pyruvate dehydrogenase and the enzymes of the TCA cycle. These high-energy electrons are transferred to NAD^+ or FAD and then passed down the electron-transport chain for use in ATP production. The reactions shown here begin with number 11 because the pathway continues where the last reaction of glycolysis (number 10 of Figure 5.6) leaves off.
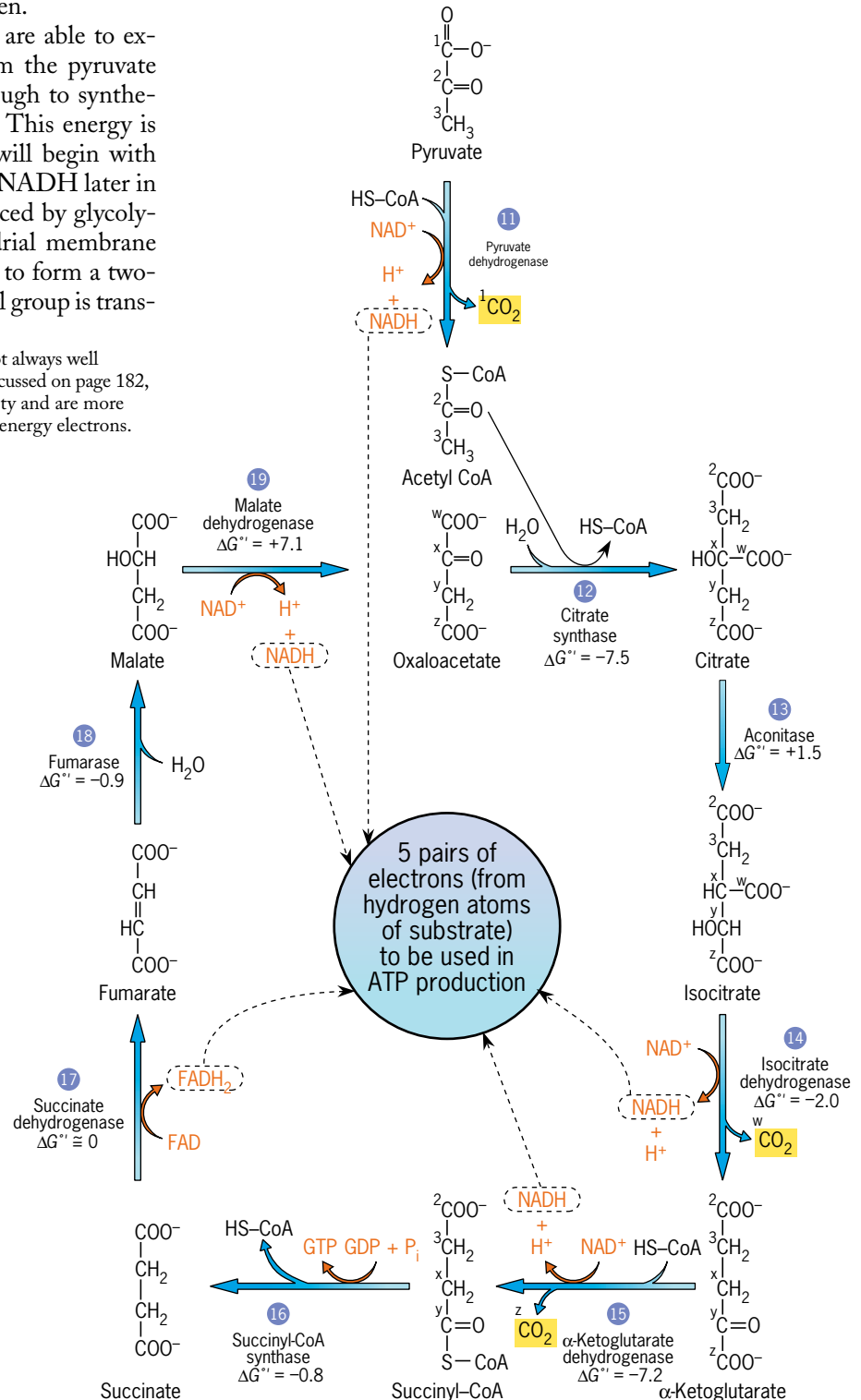
## The Tricarboxylic Acid (TCA) Cycle

Once formed, acetyl CoA is fed into a cyclic pathway, called the **tricarboxylic acid (TCA) cycle**, where the substrate is oxidized and its energy conserved. Other than succinate dehydrogenase, which is bound to the inner membrane, all the enzymes of the TCA cycle reside in the soluble phase of the matrix (Figure 5.5). The TCA cycle is also referred to as the Krebs cycle after the British biochemist Hans Krebs, who worked out the pathway in the 1930s. When Krebs first obtained sufficient evidence to support the idea of a metabolic cycle, he submitted the results to the British journal *Nature*. The paper was returned several days later accompanied by a letter of rejection. Krebs quickly published the paper in another journal.

The first step in the TCA cycle is the condensation of the two-carbon acetyl group with a four-carbon oxaloacetate to form a six-carbon citrate molecule (Figure 5.7, step 12). During the cycle, the citrate molecule is decreased in chain length, one carbon at a time, regenerating the four-carbon oxaloacetate molecule, which can condense with another acetyl CoA. It is the two carbons that are removed during the TCA cycle (which are not the same ones that were brought in with the acetyl group) that are completely oxidized to carbon dioxide.

During the TCA cycle, four reactions occur in which a pair of electrons are transferred from a substrate to an electron-accepting coenzyme. Three of the reactions reduce NAD^+ to NADH, and one reaction reduces FAD to $FADH_2$ (Figure 5.7). The net equation for the reactions of the TCA cycle can be written as

$$\text{Acetyl CoA} + 2\,H_2O + FAD + 3\,NAD^+ + GDP + P_i$$
$$\rightarrow 2\,CO_2 + FADH_2 + 3\,NADH + 3\,H^+ +$$
$$GTP + HS\text{---}CoA$$
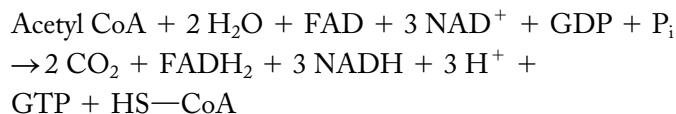
The TCA cycle is a critically important metabolic pathway. If the position of the TCA cycle in the overall metabolism of the cell is considered (Figure 5.8; see also Figure 3.22), the metabolites of this cycle are found to be the same compounds generated by most of the cell's catabolic pathways. Acetyl CoA, for example, is an important end product of a number of catabolic paths, including the disassembly of fatty acids, which are degraded within the matrix of the mitochondrion into two-carbon units (Figure 5.8a). These two-carbon compounds enter the TCA cycle as acetyl CoA. The catabolism of amino acids also generates metabolites of the TCA cycle (Figure 5.8b), which enter the matrix by means of special transport systems in the inner mitochondrial membrane. It
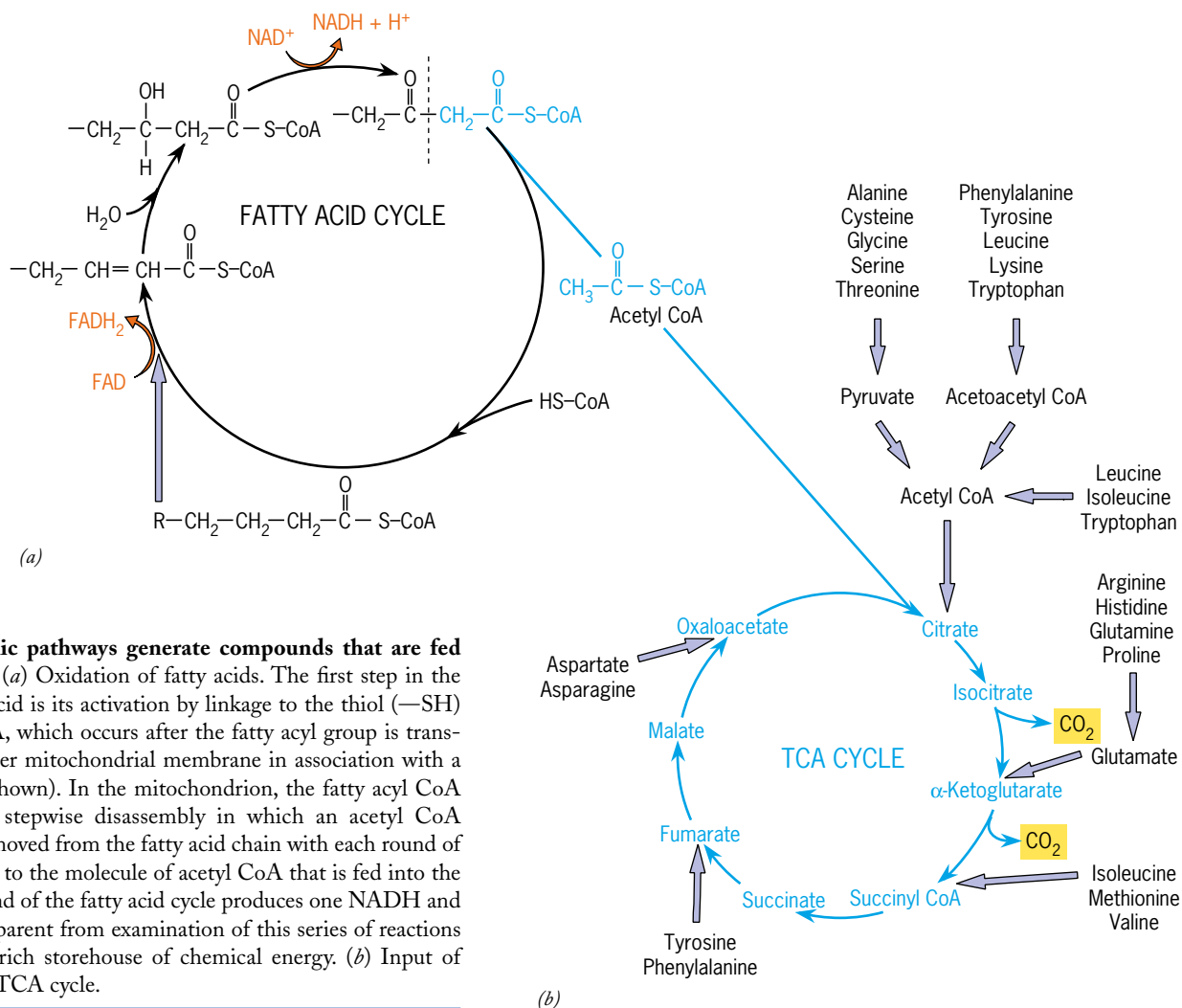
FIGURE 5.8 Catabolic pathways generate compounds that are fed into the TCA cycle. (a) Oxidation of fatty acids. The first step in the oxidation of a fatty acid is its activation by linkage to the thiol (—SH) group of coenzyme A, which occurs after the fatty acyl group is transported across the inner mitochondrial membrane in association with a carrier protein (not shown). In the mitochondrion, the fatty acyl CoA molecule undergoes stepwise disassembly in which an acetyl CoA (shown in blue) is removed from the fatty acid chain with each round of the cycle. In addition to the molecule of acetyl CoA that is fed into the TCA cycle, each round of the fatty acid cycle produces one NADH and one $FADH_2$. It is apparent from examination of this series of reactions why fats are such a rich storehouse of chemical energy. (b) Input of amino acids into the TCA cycle.

is apparent that all of a cell's energy-providing macromolecules (polysaccharides, fats, and proteins) are broken down into metabolites of the TCA cycle. Thus, the mitochondrion becomes the focus for the final energy-conserving steps in metabolism regardless of the nature of the starting material.

The Importance of Reduced Coenzymes in the Formation of ATP

It is evident from the net equation of the TCA cycle that the primary products of the pathway are the reduced coenzymes $FADH_2$ and NADH, which contain the high-energy electrons removed from various substrates as they are oxidized. NADH was also one of the products of glycolysis (along with pyruvate). Mitochondria are not able to import the NADH formed in the cytosol during glycolysis. Instead, the electrons of NADH are used to reduce a low-molecular-weight metabolite that can either (1) enter the mitochondrion (by a pathway called the malate–aspartate shuttle) and reduce NAD^+ to NADH, or (2) transfer its electrons to FAD (by a pathway called the glycerol phosphate shuttle, which is shown in Figure 5.9) to produce $FADH_2$. Both mechanisms allow the electrons from cytosolic NADH to be fed into the mitochondrial electron-transport chain and used for ATP formation.

Now that we have accounted for the formation of NADH and $FADH_2$ by both glycolysis and the TCA cycle, we can turn to the steps that utilize these reduced coenzymes to produce ATP. The overall process can be divided into two steps, which can be summarized as follows:
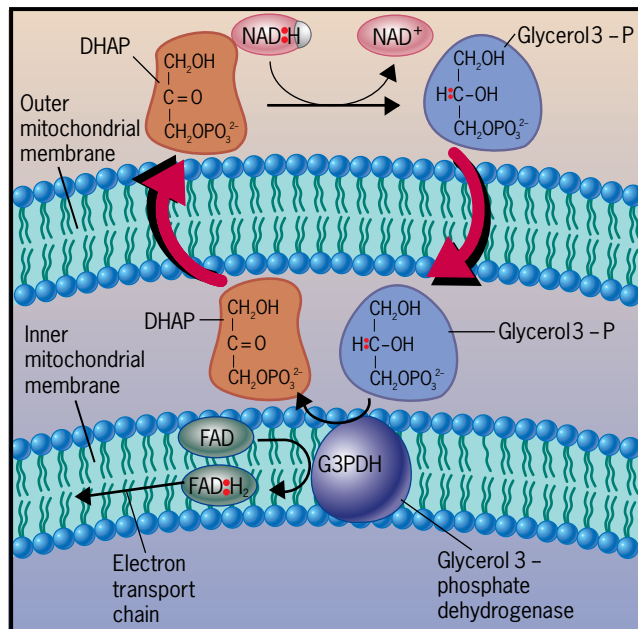
FIGURE 5.9 The glycerol phosphate shuttle. In the glycerol phosphate shuttle, electrons are transferred from NADH to dihydroxyacetone phosphate (DHAP) to form glycerol 3-phosphate, which shuttles them into the mitochondrion. These electrons then reduce FAD at the inner mitochondrial membrane, forming $FADH_2$, which can transfer the electrons to a carrier of the electron-transport chain.

Step 1 (Figure 5.10). *High-energy electrons are passed from $FADH_2$ or NADH* to the first of a series of electron carriers that make up the electron-transport chain, which is located in the inner mitochondrial membrane. Electrons pass along the electron-transport chain in energy-releasing reactions. These reactions are coupled to energy-requiring conformational changes in electron carriers that move protons outward across the inner mitochondrial membrane. As a result, the energy released during electron transport is stored in the form of an electrochemical gradient of protons across the membrane. Eventually, the low-energy electrons are transferred to the terminal electron acceptor, namely, molecular oxygen (O_2), which becomes reduced to water.

Step 2 (Figure 5.10). *The controlled movement of protons back across the membrane* through an ATP-synthesizing enzyme provides the energy required to phosphorylate ADP to ATP. The importance of proton movements in the formation of ATP was first proposed in 1961 by Peter Mitchell of the University of Edinburgh. The experiments that led to the acceptance of the **chemiosmotic mechanism**, as Mitchell called it, are discussed in the Experimental Pathways, which can be found at www.wiley.com/college/karp on the Web. Further analysis of the two steps summarized above will occupy us for much of the remainder of the chapter.
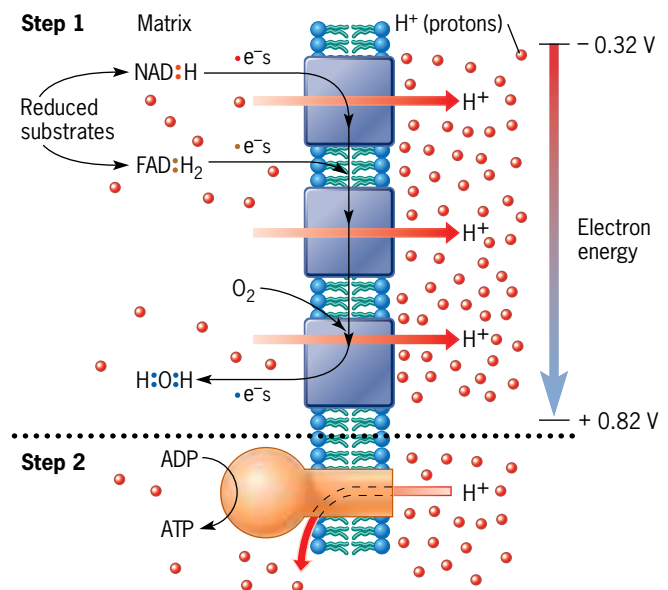
FIGURE 5.10 A summary of the process of oxidative phosphorylation. In the first step of the process, substrates such as isocitrate and succinate are oxidized (Figure 5.7) and the electrons are transferred to the coenzymes NAD^+ or FAD to form NADH or $FADH_2$. These high-energy electrons are then transferred through a series of electron carriers of the electron transport chain. The energy released is used to translocate protons from the matrix to the intermembrane space, establishing a proton electrochemical gradient across the inner mitochondrial membrane. In step 2, the protons move down the electrochemical gradient, through an ATP-synthesizing complex. The energy stored in the gradient is used to synthesize ATP. These two essential steps of oxidative phosphorylation form the basis of the chemiosmotic mechanism proposed by Peter Mitchell in 1961.

Each pair of electrons transferred from NADH to oxygen by means of the electron-transport chain releases sufficient energy to drive the formation of approximately three molecules of ATP. Each pair donated by $FADH_2$ releases sufficient energy for the formation of approximately two molecules of ATP. If one adds up all of the ATPs formed from one molecule of glucose completely catabolized by means of glycolysis and the TCA cycle, the net gain is about 36 ATPs (which includes the GTP formed by each round of the TCA cycle, step 16, Figure 5.7). The actual number of ATPs formed per molecule of glucose oxidized depends on the particular activities in which the cell is engaged. The relative importance of glycolysis versus the TCA cycle, that is, of anaerobic versus aerobic oxidative metabolism, in human skeletal muscle function is discussed in the accompanying Human Perspective.

REVIEW ?

1. How are the two products of glycolysis connected to the reactions of the TCA cycle?
2. Why is the TCA cycle considered to be the central pathway in the energy metabolism of a cell?
3. Describe the mechanism by which NADH produced in the cytosol by glycolysis is able to feed electrons into the TCA cycle.

5.3 THE ROLE OF MITOCHONDRIA IN THE FORMATION OF ATP

Mitochondria are often described as miniature power plants. Like power plants, mitochondria extract energy from organic materials and store it, temporarily, in the form of electrical energy. More specifically, the energy extracted from substrates is utilized to generate an ionic gradient across the inner mitochondrial membrane. An ionic gradient across a membrane represents a form of energy that can be tapped to perform work. We saw in Chapter 4 how intestinal cells utilize an ionic gradient across their plasma membrane to transport sugars and amino acids out of the intestinal lumen, just as nerve cells use a similar gradient to conduct neural impulses. The use of ionic gradients as a form of energy currency requires several components, including a system to generate the gradient, a membrane capable of maintaining the gradient, and the machinery to utilize the gradient to perform work.

Mitochondria utilize an ionic gradient across their inner membrane to drive numerous energy-requiring activities, most notably the synthesis of ATP. When ATP formation is driven by energy that is released from electrons removed during substrate oxidation, the process is called **oxidative phosphorylation** and is summarized in Figure 5.10. Oxidative phosphorylation can be contrasted with substrate-level phosphorylation, as discussed on page 110, in which ATP is formed directly by transfer of a phosphate group from a substrate molecule to ADP. According to one estimate, oxidative phosphorylation accounts for the production of more than 2

$\times 10^{26}$ molecules (>160 kg) of ATP in our bodies per day. Unraveling the basic mechanism of oxidative phosphorylation has been one of the crowning achievements in the field of cell and molecular biology; filling in the remaining gaps continues to be an active area of research. To understand the mechanism of oxidative phosphorylation, it is first necessary to consider how substrate oxidation is able to release free energy.

Oxidation–Reduction Potentials

If one compares a variety of oxidizing agents, they can be ranked in a series according to their affinity for electrons: the greater the affinity, the stronger the oxidizing agent. Reducing agents can also be ranked according to their affinity for electrons: the lower the affinity (the more easily electrons are released), the stronger the reducing agent. To put this into quantifiable terms, reducing agents are ranked according to **electron-transfer potential**; those substances having a high electron-transfer potential, such as NADH, are strong reducing agents, whereas those with a low electron-transfer potential, such as H_2O, are weak reducing agents. Oxidizing and reducing agents occur as *couples*, such as NAD^+ and NADH, which differ in their number of electrons. Strong reducing agents are coupled to weak oxidizing agents, and vice versa. For example, NAD^+ (of the NAD^+-NADH couple) is a weak oxidizing agent, whereas O_2 (of the O_2-H_2O couple) is a strong oxidizing agent.

Because the movement of electrons generates a separation of charge, the affinity of substances for electrons can be measured by instruments that detect voltage (Figure 5.11). What is measured for a given couple is an **oxidation–reduction potential** (or **redox potential**) relative to the potential for some standard couple. The standard couple has arbitrarily been
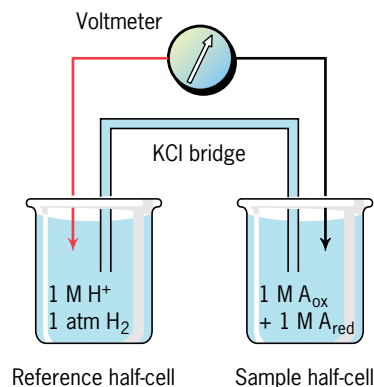
FIGURE 5.11 Measuring the standard oxidation–reduction (redox) potential. The sample half-cell contains the oxidized and reduced members of the couple, both present at 1 M. The reference half-cell contains a 1 M solution of H^+ that is in equilibrium with hydrogen gas at 1 atm pressure. An electric circuit is formed by connecting the half-cells through a voltmeter and a salt bridge. If electrons flow preferentially from the sample half-cell to the reference half-cell, then the standard redox potential (E_0) of the sample couple is negative; if the electron flow is in the opposite direction, the standard redox potential of the sample couple is positive. The salt bridge, which consists of a saturated KCl solution, provides a path for counter-ions to move between the half-cells and maintain electrical neutrality in the two compartments.

THE HUMAN PERSPECTIVE

The Role of Anaerobic and Aerobic Metabolism in Exercise

Muscle contraction requires large amounts of energy. Most of the energy is utilized in causing actin- and myosin-containing filaments to slide over one another, as discussed in Chapter 9. The energy that drives muscle contraction is derived from ATP. The rate of ATP hydrolysis increases more than 100-fold in a skeletal muscle undergoing maximal contraction compared to the same muscle at rest. It is estimated that the average human skeletal muscle has enough ATP available to fuel a 2- to 5-second burst of vigorous contraction. Even as ATP is hydrolyzed, it is important that additional ATP be produced; otherwise the ATP/ADP ratio would fall and with it the free energy available to fuel contraction. Muscle cells contain a store of creatine phosphate (CrP), one of the compounds with a phosphate transfer potential greater than that of ATP (see Figure 3.28) and thus can be utilized to generate ATP in the following reaction:

$$CrP + ADP \rightarrow Cr + ATP$$
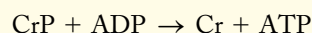
Skeletal muscles typically have enough stored creatine phosphate to maintain elevated ATP levels for approximately 15 seconds. Because muscle cells have a limited supply of both ATP and creatine phosphate, it follows that intense or sustained muscular activity requires the formation of additional quantities of ATP, which must be obtained by oxidative metabolism.

Human skeletal muscles consist of two general types of fibers (Figure 1): fast-twitch fibers, which can contract very rapidly (e.g., 15–40 msec), and slow-twitch fibers, which contract more slowly (e.g., 40–100 msec). Fast-twitch fibers are seen under the electron microscope to be nearly devoid of mitochondria, which indicates that these cells are unable to produce much ATP by aerobic respiration. Slow-twitch fibers, on the other hand, contain large numbers of mitochondria. These two types of skeletal muscle fibers are suited for different types of activities. For example, lifting weights or running sprints depends primarily on fast-twitch fibers, which are able to generate more force than their slow-twitch counterparts. Fast-twitch fibers produce nearly all of their ATP anaerobically as a result of glycolysis. Even though glycolysis only produces about 5 percent as much ATP per molecule of glucose oxidized compared with aerobic respiration, the reactions of glycolysis occur much more rapidly than those of the TCA cycle and electron transport;
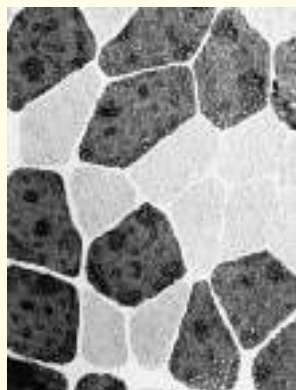
thus, the rate of anaerobic ATP production is actually higher than that achieved by aerobic respiration. The problems with producing ATP by glycolysis are the rapid use of the fiber's available glucose (stored as glycogen) and the production of an undesirable end product, lactic acid. Let's consider this latter aspect further.

Recall that the continuation of glycolysis requires the regeneration of NAD^+, which occurs by fermentation (page 111). Muscle cells regenerate NAD^+ by reducing pyruvate—the end product of glycolysis—to lactic acid. Most of the lactic acid diffuses out of the active muscle cells into the blood, where it is carried to the liver and converted back to glucose. Glucose produced in the liver is released into the blood, where it can be returned to the active muscles to continue to fuel the high levels of glycolysis. However, the formation of lactic acid is associated with a drop in pH within the muscle tissue (from about pH 7.00 to 6.35), which may produce the pain and cramps that accompany vigorous exercise. Increased acidity, together with the depletion of glycogen stores, probably accounts for the sensation of muscle fatigue that accompanies anaerobic exercises.[1]

If, instead of trying to use your muscles to lift weights or run sprints, you were to engage in an aerobic exercise, such as bicycling or fast walking, you would be able to continue to perform the activity for much longer periods of time without feeling muscle pain or fatigue. Aerobic exercises, as their name implies, are designed to allow your muscles to continue to perform aerobically, that is, to continue to produce the necessary ATP by electron transfer and oxidative phosphorylation. Aerobic exercises depend largely on the contraction of the slow-twitch fibers of the skeletal muscles. Although these fibers generate less force, they can continue to function for long periods of time due to the continuing aerobic production of ATP without production of lactic acid.

Aerobic exercise is initially fueled by the glucose molecules stored as glycogen in the muscles themselves, but after a few minutes, the muscles depend more and more on free fatty acids released into the blood from adipose (fat) tissue. The longer the period of exercise, the greater the dependency on fatty acids. After 20 minutes of vigorous aerobic exercise, it is estimated that about 50 percent of the calories being consumed by the muscles are derived from fat. Aerobic exercise, such as jogging, fast walking, swimming, or bicycling, is one of the best ways of reducing the body's fat content.

The ratio of fast-twitch to slow-twitch fibers varies from one particular muscle to another. For example, postural muscles in the back that are needed to allow a person to stand consist of a higher proportion of slow-twitch fibers than arm muscles used to throw or lift an object. The precise ratio of fast-twitch to slow-twitch fibers in a particular muscle is genetically determined and quite variable from person to person, allowing a particular individual to excel in certain types of physical activities. For example, world-class sprinters and weight lifters tend to have a higher proportion of fast-twitch fibers in their muscles than long-distance runners. In addition, training for sports such as weight lifting leads to a disproportionate enlargement of the fast-twitch fibers.

The muscle tissue of the heart must also increase its level of activity during vigorous exercise, but unlike skeletal muscle tissue, the heart can only produce ATP by aerobic metabolism. In fact, approximately 40 percent of the cytoplasmic space of a human heart muscle cell is occupied by mitochondria.

FIGURE 1 **Skeletal muscles** contain a mix of fast-twitch (or type II) fibers (darkly stained) and slow-twitch (or type I) fibers (lightly stained). (COURTESY OF DUNCAN MACDOUGALL.)

---

[1]An alternate view can be found in *Science* 305:1112, 2004.

chosen as hydrogen (H^+-H_2). As with free-energy changes, where the standard free-energy change, ΔG°, was used, a similar assignment is employed for redox couples. The standard redox potential, E_0, for a given couple is designated as the voltage produced by a half-cell (with only members of one couple present) in which each member of the couple is present under standard conditions (as in Figure 5.11). Standard conditions are 1.0 M for solutes and ions and 1 atm pressure for gases (e.g., H_2) at 25°C. The standard redox potential for the oxidation–reduction reaction involving hydrogen (2 H^+ + 2 electrons → H_2) is 0.00 V. Table 5.1 gives the redox potentials of some biologically important couples. The value for the hydrogen couple in the table is not 0.00, but -0.42 V. This number represents the value when the concentration of H^+ is 10^{-7} M (pH of 7.0) rather than 1.0 M (pH of 0.0), which would be of little physiologic use. When calculated at pH 7, the standard redox potential is indicated by the symbol E_0' rather than E_0. The assignment of sign (positive or negative) to couples is arbitrary and varies among different disciplines. We will consider the assignment in the following way. Those couples whose reducing agents are better donors of electrons are assigned more negative redox potentials. For example, the standard redox potential for the NAD^+-NADH couple is -0.32 V (Table 5.1). Acetaldehyde is a stronger reducing agent than NADH, and the acetate-acetaldehyde couple has a standard redox potential of -0.58 V. Those couples whose oxidizing agents are better electron acceptors than NAD^+, that is, have greater affinity for electrons than NAD^+, have more positive redox potentials.

Just as any other spontaneous reaction is accompanied by a loss in free energy, so too are oxidation–reduction reactions. The standard free-energy change during a reaction of the type

$$A_{(ox)} + B_{(red)} \rightleftharpoons A_{(red)} + B_{(ox)}$$
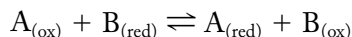
can be calculated from the standard redox potentials of the two couples involved in the reaction according to the equation

$$\Delta G^{\circ'} = -nF \Delta E_0'$$

where n is the number of electrons transferred in the reaction, F is the Faraday constant (23.063 kcal/V · mol), and $\Delta E_0'$ is the difference in volts between the standard redox potentials of the two couples. The greater the difference in standard redox potential between two couples, the farther the reaction proceeds under standard conditions to the formation of products before an equilibrium state is reached. Let's consider the reaction in which NADH, a strong reducing agent, is oxidized by molecular oxygen, a strong oxidizing agent.

$$NADH + \tfrac{1}{2}O_2 + H^+ \rightarrow H_2O + NAD^+$$
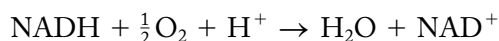
The standard redox potentials of the two couples can be written as

$$\tfrac{1}{2}O_2 + 2 H^+ + 2 e^- \rightarrow H_2O \qquad E_0' = +0.82 \text{ V}$$
$$NAD^+ + 2 H^+ + 2 e^- \rightarrow NADH + H^+ \quad E_0' = -0.32 \text{ V}$$
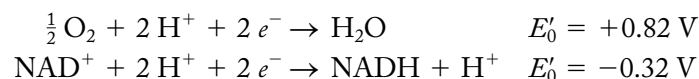
The voltage change for the total reaction is equal to the difference between the two E_0' values ($\Delta E_0'$):

$$\Delta E_0' = +0.82 \text{ V} - (-0.32 \text{ V}) = 1.14 \text{ V}$$

TABLE 5.1 Standard Redox Potentials of Selected Half-Reactions

Electrode equation	E_0'(V)
Succinate + CO_2 + 2 H^+ + 2 $e^- \rightleftharpoons$ α-ketoglutarate + H_2O	-0.670
Acetate + 2 H^+ + 2 $e^- \rightleftharpoons$ acetaldehyde	-0.580
2 H^+ + 2 $e^- \rightleftharpoons H_2$	-0.421
α-Ketoglutarate + CO_2 + 2 H^+ + 2 $e^- \rightleftharpoons$ isocitrate	-0.380
Cystine + 2 H^+ + 2 $e^- \rightleftharpoons$ 2 cysteine	-0.340
NAD^+ + 2 H^+ + 2 $e^- \rightleftharpoons$ NADH + H^+	-0.320
$NADP^+$ + 2 H^+ + 2 $e^- \rightleftharpoons$ NADPH + H^+	-0.324
Acetaldehyde + 2 H^+ + 2 $e^- \rightleftharpoons$ ethanol	-0.197
Pyruvate + 2 H^+ + 2 $e^- \rightleftharpoons$ lactate	-0.185
Oxaloacetate + 2 H^+ + 2 $e^- \rightleftharpoons$ malate	-0.166
FAD + 2 H^+ + 2 $e^- \rightleftharpoons$ $FADH_2$ (in flavoproteins)	$+0.031$
Fumarate + 2 H^+ + 2 $e^- \rightleftharpoons$ succinate	$+0.031$
Ubiquinone + 2 H^+ + 2 $e^- \rightleftharpoons$ ubiquinol	$+0.045$
2 cytochrome $b_{(ox)}$ + 2 $e^- \rightleftharpoons$ 2 cytochrome $b_{(red)}$	$+0.070$
2 cytochrome $c_{(ox)}$ + 2 $e^- \rightleftharpoons$ 2 cytochrome $c_{(red)}$	$+0.254$
2 cytochrome $a_{3(ox)}$ + 2 $e^- \rightleftharpoons$ 2 cytochrome $a_{3(red)}$	$+0.385$
$\tfrac{1}{2}O_2$ + 2 H^+ + 2 $e^- \rightleftharpoons H_2O$	$+0.816$

which is a measure of the free energy released when NADH is oxidized by molecular oxygen under standard conditions. Substituting this value into the above equation,

$$\Delta G^{\circ'} = (-2)(23.063 \text{ kcal/V} \cdot \text{mol})(1.14 \text{ V})$$
$$= -52.6 \text{ kcal/mol of NADH oxidized}$$

the standard free-energy difference ($\Delta G^{\circ'}$) is -52.6 kcal/mol. As with other reactions, the actual ΔG values depend on the relative concentrations of reactants and products (oxidized and reduced versions of the compounds) present in the cell at a given instant. Regardless, it would appear that the drop in free energy of a pair of electrons as they pass from NADH to molecular oxygen ($\Delta G^{\circ'} = -52.6$ kcal/mol) should be sufficient to drive the formation of several molecules of ATP ($\Delta G^{\circ'} = +7.3$ kcal/mol) even under conditions in the cell, where ATP/ADP ratios are much higher than those at standard conditions. The transfer of this energy from NADH to ATP within the mitochondrion occurs in a series of small, energy-releasing steps, which will be the primary topic of discussion in the remainder of the chapter.

Electrons are transferred to NAD^+ (or FAD) within the mitochondrion from several substrates of the TCA cycle, namely, isocitrate, α-ketoglutarate, malate, and succinate (reactions 14, 15–16, 19, and 17 of Figure 5.7, respectively). The first three of these intermediates have redox potentials of relatively high negative values (Table 5.1)—sufficiently high to transfer electrons to NAD^+ *under conditions that prevail in the cell.*[2] In contrast, the oxidation of succinate to fumarate, which

[2]As indicated in Table 5.1, the standard redox potential (E_0') of the oxaloacetate-malate couple is more positive than that of the NAD^+-NADH couple. Thus, the oxidation of malate to oxaloacetate has a positive $\Delta G^{\circ'}$ (reaction 19, Figure 5.7) and can proceed to the formation of oxaloacetate only when the ratio of products to reactants is kept below that of standard conditions. The ΔG of this reaction is kept negative by maintenance of very low oxaloacetate concentrations, which is possible because the next reaction in the cycle (reaction 12, Figure 5.7) is highly exergonic and is catalyzed by one of the major rate-controlling enzymes of the TCA cycle.

has a more positive redox potential, proceeds by the reduction of FAD, a coenzyme of greater electron affinity than NAD^+.

Electron Transport

Five of the nine reactions illustrated in Figure 5.7 are catalyzed by dehydrogenases, enzymes that transfer pairs of electrons from substrates to coenzymes. Four of these reactions generate NADH and one produces $FADH_2$. The NADH molecules, which are formed in the mitochondrial matrix, dissociate from their respective dehydrogenases and bind to NADH dehydrogenase, an integral protein of the inner mitochondrial membrane (see Figure 5.17). Unlike the other enzymes of the TCA cycle, succinate dehydrogenase, the enzyme that catalyzes the formation of $FADH_2$ (Figure 5.7, reaction 17), is a component of the inner mitochondrial membrane. In either case, the high-energy electrons associated with NADH or $FADH_2$ are transferred through a series of specific electron carriers that constitute the **electron-transport chain** (or **respiratory chain**) of the inner mitochondrial membrane.

Types of Electron Carriers

The electron-transport chain contains five types of membrane-bound electron carriers: flavoproteins, cytochromes, copper atoms, ubiquinone, and iron-sulfur proteins. With the exception of ubiquinone, all of the redox centers within the respiratory chain that accept and donate electrons are *prosthetic groups*, that is, non-amino acid components that are tightly associated with proteins.

- **Flavoproteins** consist of a polypeptide bound tightly to one of two related prosthetic groups, either flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN) (Figure 5.12a). The prosthetic groups of flavoproteins are derived from riboflavin (vitamin B_2), and each is capable
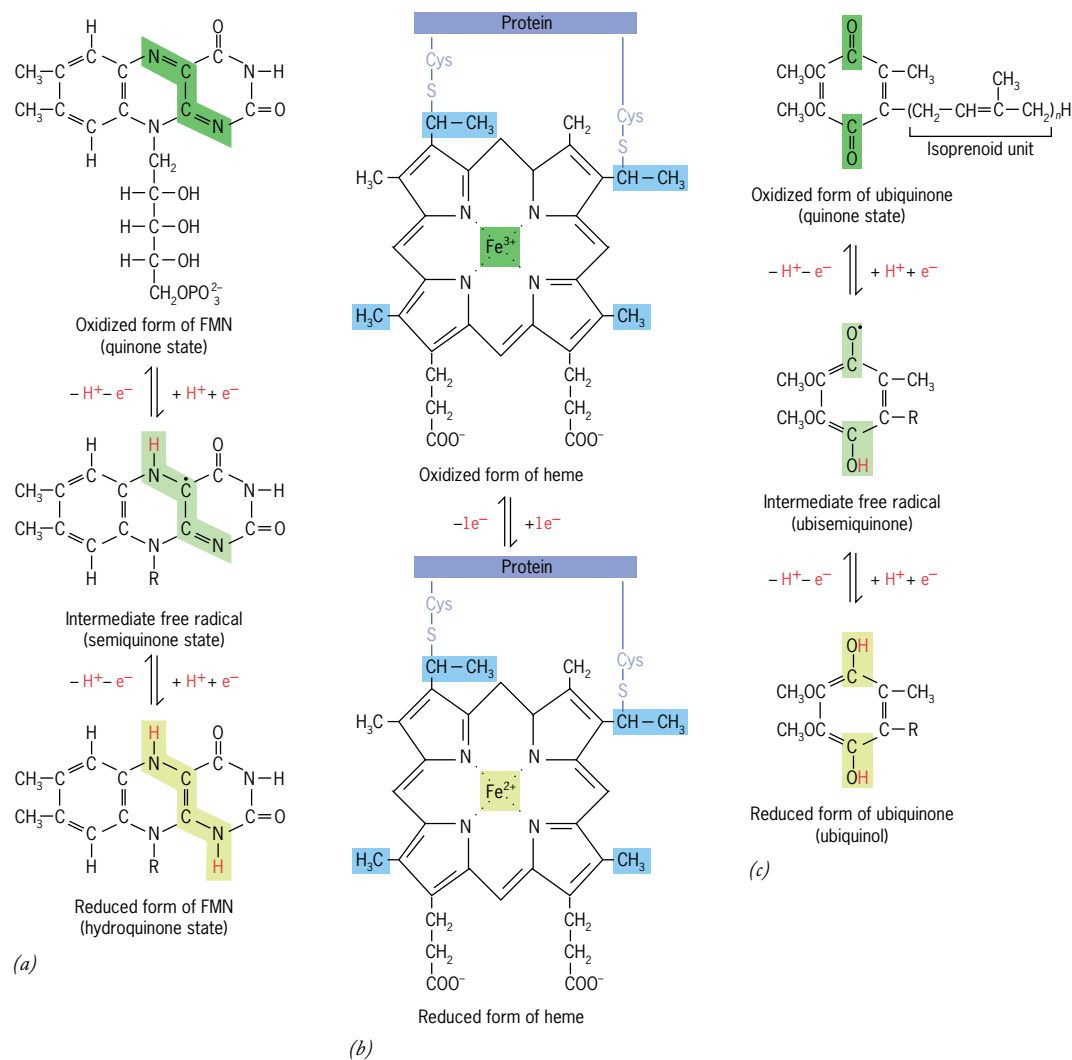
FIGURE 5.12 Structures of the oxidized and reduced forms of three types of electron carriers. (*a*) FMN of NADH dehydrogenase, (*b*) the heme group of cytochrome *c*, and (*c*) ubiquinone (coenzyme Q). The heme groups of the various cytochromes of the electron-transport chain differ in the substitutions on the porphyrin rings (indicated by the blue shading) and type of linkage to the protein. Cytochromes can accept only one electron, whereas FMN and quinones can accept two electrons and two protons in successive reactions, as shown. FAD differs from FMN in having an adenosine group bonded to the phosphate.

of accepting and donating two protons and two electrons. The major flavoproteins of the mitochondria are NADH dehydrogenase of the electron-transport chain and succinate dehydrogenase of the TCA cycle.

■ **Cytochromes** are proteins that contain *heme* prosthetic groups (such as that described for myoglobin on page 57). The iron atom of a heme undergoes reversible transition between the Fe^{3+} and Fe^{2+} oxidation states as a result of the acceptance and loss of a single electron (Figure 5.12*b*). There are three distinct cytochrome types—*a*, *b*, and *c*—present in the electron-transport chain, which differ from one another by substitutions within the heme group (indicated by the blue shaded portions of Figure 5.12*b*).

■ **Three copper atoms**, all located within a single protein complex of the inner mitochondrial membrane (see Figure 5.19), accept and donate a single electron as they alternate between the Cu^{2+} and Cu^{1+} states.

■ **Ubiquinone** (UQ, or coenzyme Q) is a lipid-soluble molecule containing a long hydrophobic chain composed of five-carbon isoprenoid units (Figure 5.12*c*). Like the flavoproteins, each ubiquinone is able to accept and donate two electrons and two protons. The partially reduced molecule is the free radical ubisemiquinone, and the fully reduced molecule is ubiquinol (UQH_2). UQ/UQH_2 remains within the lipid bilayer of the membrane, where it is capable of rapid lateral diffusion.

■ **Iron-sulfur proteins** are iron-containing proteins in which the iron atoms are not located within a heme group but instead are linked to inorganic sulfide ions as part of an *iron-sulfur center*. The most common centers contain either two or four atoms of iron and sulfur—designated [2Fe-2S] and [4Fe-4S]—linked to the protein at cysteine residues (Figure 5.13). Even though a single center may have several iron atoms, the entire complex is capable of accepting and donating only a single electron. The redox potential of an iron-sulfur center depends on the hydrophobicity and charge of the amino acid residues that make up its local environment. As a group, iron-sulfur proteins have potentials ranging from about −700 mV to about +300 mV, corresponding to a major portion of the span over which electron transport occurs. More than a dozen distinct iron-sulfur centers have been identified within mitochondria.

The carriers of the electron-transport chain are arranged in order of increasingly positive redox potential (Figure 5.14). Each carrier is reduced by the gain of electrons from the preceding carrier in the chain and is subsequently oxidized by the loss of electrons to the carrier following it. Thus, electrons are passed from one carrier to the next, losing energy as they move "downhill" along the chain. The final acceptor of this electron "bucket brigade" is O_2, which accepts the energy-depleted electrons and is reduced to water. The specific sequence of carriers that constitute the electron-transport chain was worked out by Britton Chance and co-workers at the University of Pennsylvania using a variety of inhibitors that blocked electron transport at specific sites along the route. The concept of
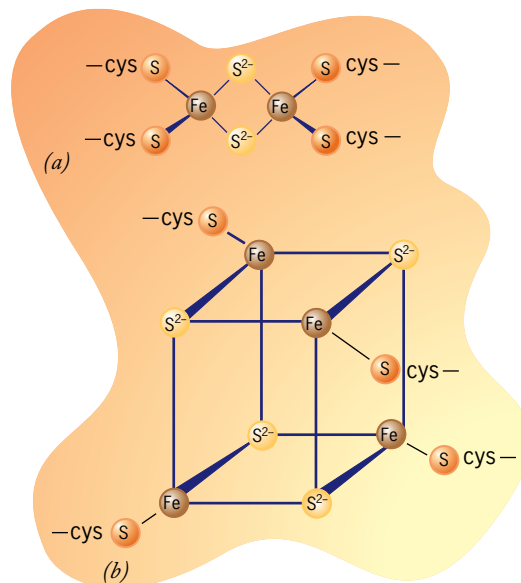
FIGURE 5.13 Iron-sulfur centers. Structure of a [2Fe-2S] (*a*) and a [4Fe-4S] (*b*) iron-sulfur center. Both types of iron-sulfur centers are joined to the protein by linkage to a sulfur atom (shown in orange) of a cysteine residue. Inorganic sulfide ions (S^{2-}) are shown in yellow. Both types of iron-sulfur centers accept only a single electron, whose charge is distributed among the various iron atoms.

these experiments is depicted by analogy in Figure 5.15. In each case, an inhibitor was added to cells, and the oxidation state of the various electron carriers in the inhibited cells was determined. This determination can be made using a spec-
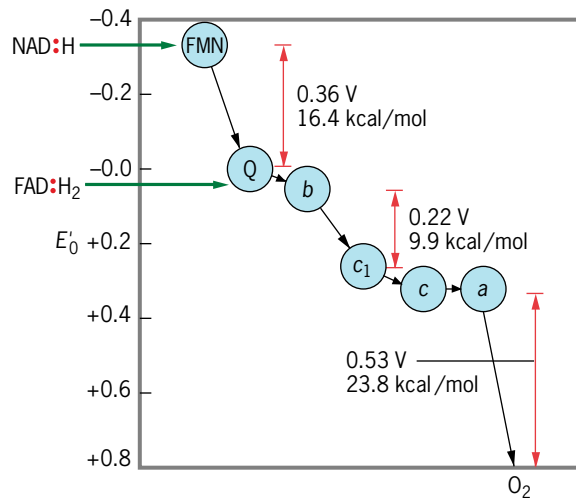
FIGURE 5.14 The arrangement of several carriers in the electron-transport chain. The diagram illustrates the approximate redox potential of the carriers and the decline in free energy as electron pairs move along the respiratory chain to molecular oxygen. The numerous iron-sulfur centers are not indicated in this figure for the sake of simplicity. As discussed in the following section, each of the three electron transfers that are denoted by red arrows yields sufficient energy to move protons across the inner mitochondrial membrane, which in turn provides the energy required to generate ATP from ADP. (AFTER A. L. LEHNINGER, BIOCHEMISTRY, 2D ED., 1975; WORTH PUBLISHERS, NEW YORK.)
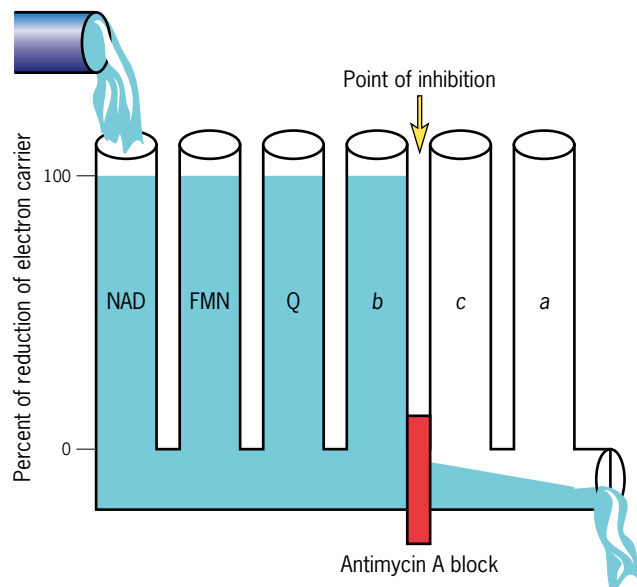
FIGURE 5.15 **Experimental use of inhibitors to determine the sequence of carriers in the electron-transport chain.** In this hydraulic analogy, treatment of mitochondria with the inhibitor antimycin A leaves those carriers on the upstream (NADH) side of the point of inhibition in the fully reduced state and those carriers on the downstream (O_2) side of inhibition in the fully oxidized state. Comparison of the effects of several inhibitors revealed the order of the carriers within the chain. (AFTER A. L. LEHNINGER, BIOCHEMISTRY, 2D ED., 1975; WORTH PUBLISHERS, NEW YORK.)

FIGURE 5.16 **Electron-tunneling pathways for the yeast cytochrome c-cytochrome c peroxidase complex.** The heme group of cytochrome c is blue, and that of cytochrome c peroxidase (which is not a carrier of the mitochondrial electron-transport chain but provides an analogous electron acceptor for which high-resolution crystal structures are available) is red. Several defined pathways (yellow) exist for the movement of electrons from one heme to the other. Each of the pathways carries the electrons through several amino acid residues situated between the heme groups. (It can be noted that other mechanisms of electron transfer have been proposed.) (BY JEFFREY J. REGAN AND J. N. ONUCHIC, FROM DAVID N. BERATAN ET AL., REPRINTED WITH PERMISSION FROM SCIENCE 258:1741, 1992. COPYRIGHT 1992; AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

trophotometer that measures the absorption of light of a sample at various wavelengths. This measurement reveals whether a particular carrier is in the oxidized or reduced state. In the case depicted in Figure 5.15, addition of antimycin A blocked electron transport at a site that left NADH, $FMNH_2$, QH_2, and cyt b in the reduced state, and left cytochromes c and a in the oxidized state. This result indicates that NAD, FMN, Q, and cyt b are situated "upstream" of the block. In contrast, an inhibitor (e.g., rotenone) that acts between FMN and Q would leave only NADH and $FMNH_2$ in the reduced state. Thus, by identifying reduced and oxidized components in the presence of different inhibitors, the sequences of the carriers could be determined.

The tendency for electrons to be transferred from one carrier to the next depends on the potential difference between the two redox centers, but the rate of transfer depends on the catalytic activities of the proteins involved. This distinction between thermodynamics and kinetics is similar to that discussed on page 93 regarding the activity of enzymes. Studies indicate that electrons may travel considerable distances (10–20 Å) between adjacent redox centers and that electrons probably flow through special "tunneling pathways" consisting of a series of covalent and hydrogen bonds that stretch across parts of several amino acid residues. An example of such a pathway involving cytochrome c is shown in Figure 5.16.

Electron-Transport Complexes   When the inner mitochondrial membrane is disrupted by detergent, the various electron carriers can be isolated as part of four distinct, asymmetric, membrane-spanning complexes, identified as complexes I, II, III, and IV (Figure 5.17). Each of these four complexes can be assigned a distinct function in the overall oxidation pathway. Two components of the electron-transport chain, cytochrome c and ubiquinone, are not part of any of the four complexes. Ubiquinone occurs as a pool of molecules dissolved in the lipid bilayer, and cytochrome c is a soluble protein in the intermembrane space. Ubiquinone and cytochrome c are thought to move within or along the membrane, shuttling electrons between the large, relatively immobile protein complexes. Once within one of the large, multiprotein complexes, electrons travel along defined pathways (of the type illustrated in Figure 5.16) between adjacent redox centers whose relative positions are fixed in place.
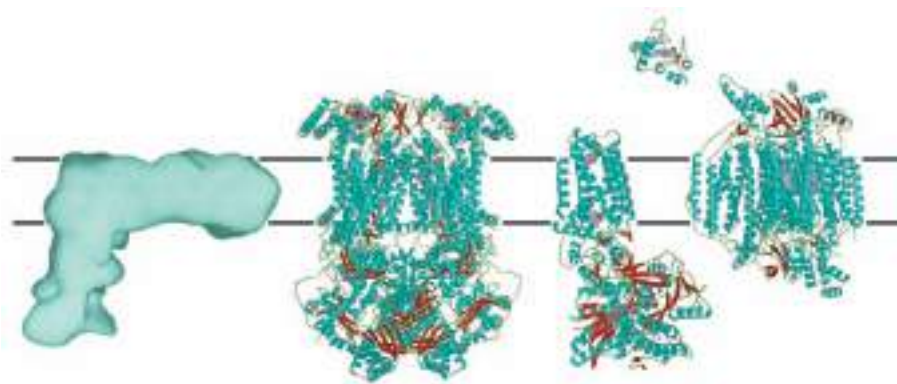
When NADH is the electron donor, electrons enter the respiratory chain by means of complex I, which transfers electrons to ubiquinone, generating ubiquinol (Figures 5.12 and 5.17). Unlike NADH, which can diffuse away from its soluble dehydrogenases, $FADH_2$ remains covalently bound to succinate dehydrogenase, a component of complex II. When $FADH_2$ is the donor, electrons are passed directly to ubiquinone, bypassing the upstream end of the chain, which has too negative a redox potential to accept the less energetic electrons of the flavin nucleotide (Figure 5.14).

If one examines the redox potentials of successive carriers in Figure 5.14, it is apparent that there are three places in which the transfer of electrons is accompanied by a major release of free energy. Each of these *coupling sites*, as they are called, occurs between carriers that are part of one of the three
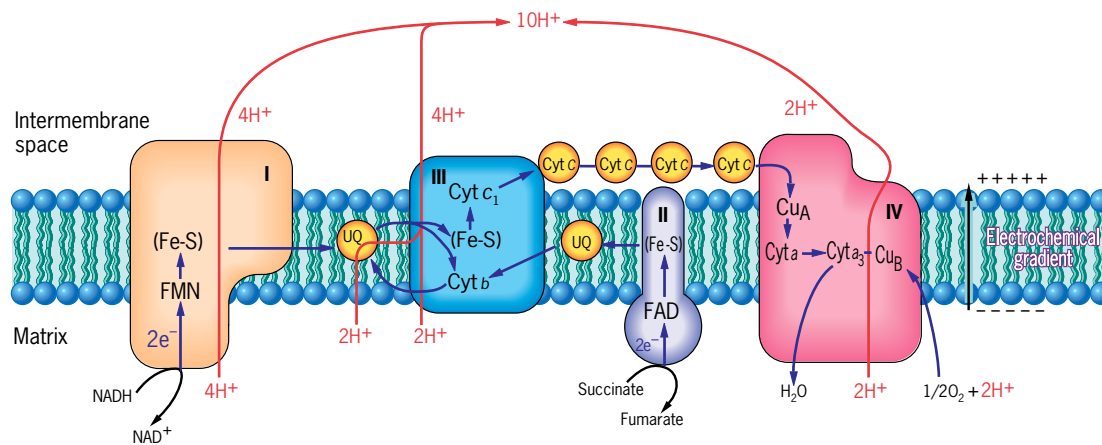
|	Complex I	Complex III	Complex II	Complex IV
	NADH dehydrogenase (Mammalian)	Cytochrome bc_1	Succinate dehydrogenase	Cytochrome c oxidase
SUBUNITS				
mtDNA	7	1	0	3
nDNA	39	10	4	10
TOTAL	46	11	4	13
Molecular mass (Da)	>900,000	~240,000	~125,000	~200,000

(a)

(b)

FIGURE 5.17 The electron-transport chain of the inner mitochondrial membrane. (*a*) The respiratory chain consists of four complexes of electron carriers and two other carriers (ubiquinone and cytochrome *c*) that are independently disposed. Electrons enter the chain from either NADH (via complex I) or $FADH_2$ (a part of complex II). Electrons are passed from either complex I or II to ubiquinone (UQ), which exists as a pool within the lipid bilayer. Electrons are subsequently passed from reduced ubiquinone (ubiquinol) to complex III and then to the peripheral protein cytochrome *c*, which is thought to be mobile. Electrons are transferred from cytochrome *c* to complex IV (cytochrome oxidase) and then to O_2 to form H_2O. The sites of proton translocation from the matrix to the cytosol are indicated. The precise number of protons translocated at each site remains controversial; the number indicated is a general consensus. Keep in mind that the proton numbers shown are those generated by each pair of electrons transported, which is sufficient to reduce only one-half of an O_2 molecule. (The translocation of protons by complex III occurs by way of a Q cycle, whose details are shown in Figure 4 of the Experimental Pathway on the Web. The Q cycle can be divided into two steps, each of which leads to the release of two protons into the cytosol.) (*b*) Structures of the protein components of the electron-transport chain (either mitochondrial or bacterial versions). The tertiary structure of Complex I has yet to be determined, but its overall shape is indicated. (B: FROM BRIAN E. SCHULTZ AND SUNNEY I. CHAN, REPRINTED WITH PERMISSION FROM THE ANNUAL REVIEW OF BIOPHYSICS BIOMOLECULAR STRUCTURE, VOLUME 30, © 2001, BY ANNUAL REVIEWS, INC.)

complexes, I, III, and IV. The free energy, which is released as electrons are passed across these three sites, is conserved by translocation of protons from the matrix across the inner membrane into the intermembrane space. These three protein complexes are often described as *proton pumps*. The translocation of protons by these electron-transporting complexes establishes the proton gradient that drives ATP synthesis. The ability of these remarkable molecular machines to act as independent proton-translocating units can be demonstrated by purifying each of them and incorporating them individually into artificial lipid vesicles. When given an appropriate electron donor, these protein-containing vesicles are capable of
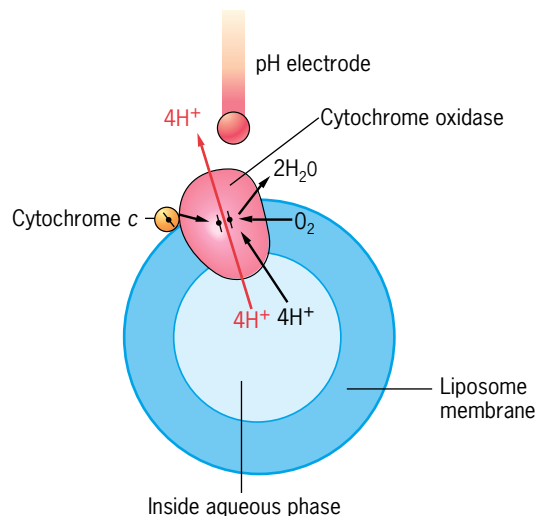
FIGURE 5.18 Experimental demonstration that cytochrome oxidase is a proton pump. When purified cytochrome oxidase is incorporated into the artificial bilayer of a liposome, the medium becomes acidified following the addition of reduced cytochrome c. This indicates that as electrons are transferred from cytochrome c to cytochrome oxidase, and O_2 is being reduced to water, protons are being translocated from the compartment within the vesicle to the external medium. This experiment was originally carried out in the late 1960s by Mårten Wikström and colleagues. (REPRINTED WITH PERMISSION FROM M. I. VERKHOVSKY ET AL. NATURE 400:481, 1999; COPYRIGHT 1999, MACMILLAN MAGAZINES LIMITED.)

accepting electrons and using the released energy to pump protons across the vesicle membrane (see Figure 5.18).

Great strides have been made in the past few years in elucidating the molecular architecture of all the protein complexes of the inner membrane (Figure 5.17b). Researchers are no longer asking what these proteins look like but are using the wealth of new structural information to understand how they work. We will briefly examine the mammalian versions of each of the four electron-transport complexes, which together contain about 70 different polypeptides. The bacterial versions are considerably simpler than their mammalian counterparts and contain far fewer subunits. The additional subunits of the mammalian complexes do not contain redox centers and are thought to function in either the regulation or assembly of the complex rather than electron transport. In other words, the basic process of electron transport during respiration has remained virtually unchanged since its evolution billions of years ago in our prokaryotic ancestors.

Complex I (or NADH dehydrogenase) Complex I is the gateway to the electron-transport chain, catalyzing the transfer of a pair of electrons from NADH to ubiquinone (UQ) to form ubiquinol (UQH$_2$). The mammalian version of complex I is a huge L-shaped conglomerate, containing at least 45 different subunits and representing a molecular mass of nearly one million daltons. Seven of the subunits—all hydrophobic, membrane-spanning polypeptides—are encoded by mitochondrial genes and are homologues of bacterial polypeptides.

As indicated in Figure 5.17b, complex I is the only member of the electron-transport chain whose three-dimensional structure has not been solved by X-ray crystallography at the time of this writing. Complex I includes an FMN-containing flavoprotein that oxidizes NADH, at least eight distinct iron-sulfur centers, and two bound molecules of ubiquinone. The passage of a pair of electrons through complex I is thought to be accompanied by the movement of four protons from the matrix into the intermembrane space. The importance of complex I dysfunction as a cause of neurodegeneration is discussed on page 202.

Complex II (or succinate dehydrogenase) Complex II consists of four polypeptides: two hydrophobic subunits that anchor the protein in the membrane and two hydrophilic subunits that comprise the TCA cycle enzyme, succinate dehydrogenase. Complex II provides a pathway for feeding lower-energy electrons (ones close to 0 mV) from succinate to FAD to ubiquinone (Figures 5.14 and 5.17). The path from FADH$_2$ at the enzyme catalytic site to ubiquinone takes the electrons over a distance of 40 Å through three iron-sulfur clusters. Complex II also contains a heme group, which is thought to attract escaped electrons, thereby preventing the formation of destructive superoxide radicals (page 34). Electron transfer through complex II is not accompanied by proton translocation.

Complex III (or cytochrome bc$_1$) Complex III catalyzes the transfer of electrons from ubiquinol to cytochrome c. Experimental determinations suggest that four protons are translocated across the membrane for every pair of electrons transferred through complex III. The protons are released into the intermembrane space in two separate steps powered by the energy released as a pair of electrons are separated from one another and passed along different pathways through the complex. Two protons are derived from the molecule of ubiquinol that entered the complex. Two additional protons are removed from the matrix and translocated across the membrane as part of a second molecule of ubiquinol. These steps are described in further detail in Figure 4 of the Experimental Pathways, which can be seen on the Web at www.wiley.com/college/karp. Three of the subunits of complex III contain redox groups: cytochrome b contains two heme b molecules with different redox potentials, cytochrome c_1, and an iron-sulfur protein. Cytochrome b is the only polypeptide of the complex encoded by a mitochondrial gene.

Complex IV (or cytochrome c oxidase) The final step of electron transport in a mitochondrion is the successive transfer of electrons from reduced cytochrome c to oxygen according to the reaction
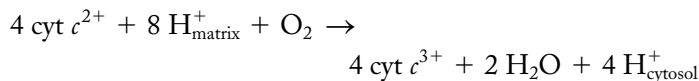
$$2 \text{ cyt } c^{2+} + 2 \text{ H}^+ + \tfrac{1}{2} O_2 \rightarrow 2 \text{ cyt } c^{3+} + H_2O$$
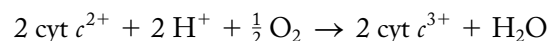
To reduce a whole molecule of O_2,

$$4 \text{ cyt } c^{2+} + 4 \text{ H}^+ + O_2 \rightarrow 4 \text{ cyt } c^{3+} + 2 H_2O$$
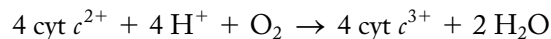
The reduction of O_2 is catalyzed by complex IV, a huge assembly of polypeptides referred to as **cytochrome oxidase**.

Cytochrome oxidase was the first component of the electron-transport chain shown to act as a redox-driven proton pump. This was demonstrated in the experiment depicted in Figure 5.18 in which the purified enzyme was incorporated into vesicles containing an artificial lipid bilayer (liposomes). Addition of reduced cytochrome c to the medium was accompanied by the expulsion of H^+ ions from the vesicles, which was measured as a drop in the surrounding pH. Whether inside a liposome or the inner mitochondrial membrane, translocation of protons is coupled to conformational changes generated by the release of energy that accompanies the transfer of electrons. For every molecule of O_2 reduced by cytochrome oxidase, eight protons are thought to be taken up from the matrix. Four of these protons are consumed in the formation of two molecules of water as indicated above, and the other four protons are translocated across the membrane and released into the intermembrane space (Figure 5.17). Consequently, the overall reaction can be written:

$$4 \text{ cyt } c^{2+} + 8 \text{ H}^+_{\text{matrix}} + O_2 \rightarrow$$
$$4 \text{ cyt } c^{3+} + 2 \text{ H}_2O + 4 \text{ H}^+_{\text{cytosol}}$$

Humans produce about 300 ml of "metabolic water" from this reaction per day. It can be noted that a number of potent respiratory poisons, including carbon monoxide (CO), azide (N_3^-), and cyanide (CN^-), have their toxic effect by binding to the heme a_3 site of cytochrome oxidase. (Carbon monoxide also binds to the heme group of hemoglobin.)

A Closer Look at Cytochrome Oxidase   Cytochrome oxidase consists of 13 subunits, three of which are encoded by the mitochondrial genome and contain all four of the protein's redox centers. The mechanism of electron transfer through complex IV has been intensively studied, most notably in the laboratories of Mårten Wikström at the University of Helsinki in Finland and Gerald Babcock at Michigan State University. A major challenge for investigators is to explain how carriers that are only able to transfer single electrons can reduce a molecule of O_2 to two molecules of H_2O, a process that requires four electrons (together with four protons). Most importantly, the process must occur very efficiently because the cell is dealing with very dangerous substances; the "accidental" release of partially reduced oxygen species has the potential to damage virtually every macromolecule in the cell (page 34).

The movement of electrons between redox centers of cytochrome oxidase is shown in Figure 5.19. Electrons are transferred one at a time from cytochrome c through a bimetallic copper center (Cu_A) of subunit II to a heme (heme a) of subunit I. From there, electrons are passed to a redox center located in subunit I that contains a second heme (heme a_3) and another copper atom (Cu_B) situated less than 5 Å apart. Acceptance of the first two electrons reduces the a_3-Cu_B binuclear center according to the reaction

$$Fe^{3+}_{a_3} + Cu^{2+}_B + 2 e^- \rightarrow Fe^{2+}_{a_3} + Cu^{1+}_B$$
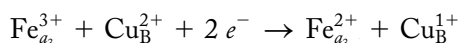
Once the binuclear center accepts its second electron, an O_2 molecule binds to the center. The O=O double bond is then
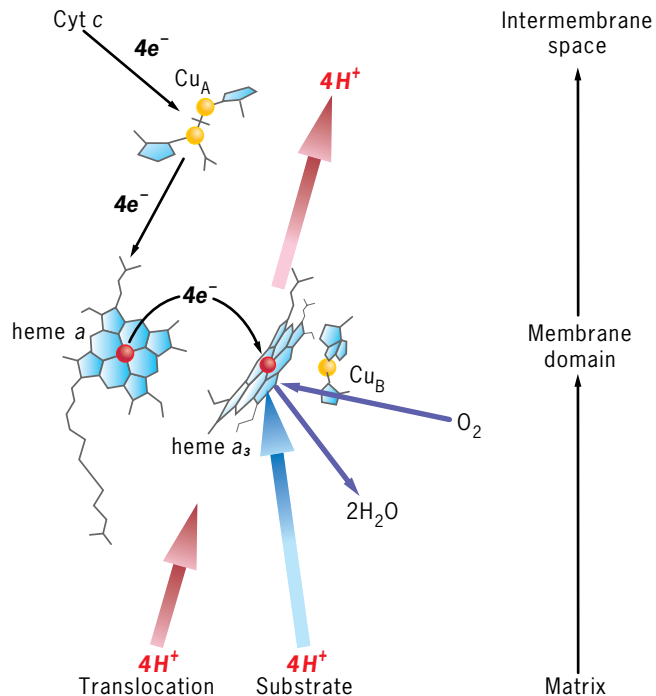
FIGURE 5.19   A summary of cytochrome oxidase activity. A model showing the flow of electrons through the four redox centers of cytochrome oxidase. Iron atoms are shown as red spheres, copper atoms as yellow spheres. Electrons are thought to be passed one at a time from cytochrome c to the dimeric copper center (Cu_A), then on to the iron of the heme group of cytochrome a, then on to the binuclear redox center consisting of a second iron (of the heme group of cytochrome a_3) and copper ion (Cu_B). The structures and suggested orientations of the redox centers are indicated. (FROM M. WIKSTRÖM ET AL., BIOCHIM. BIOPHYS. ACTA 1459:515, 2000.)

broken as the O atoms accept a pair of electrons from the reduced a_3-Cu_B binuclear center, likely forming a reactive peroxy anion O_2^{2-}.

$$Fe^{2+}\!-\!O \!\!\diagdown\!\!_{O \, Cu^{1+}} \rightarrow Fe^{3+}\!-\!O^- \!\!\diagdown\!\!_{O^- Cu^{2+}}$$
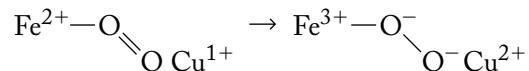
The peroxy ion is the highest energy component in the reaction sequence and reacts quickly to extract a third electron, either from the heme itself or from a nearby amino acid residue. At the same time, the binuclear center accepts two protons from the matrix, cleaving the covalent O—O bond and reducing one of the O atoms.

$$Fe^{4+}\!=\!O^{2-} \qquad Cu^{2+}\!-\!OH_2$$

The passage of a fourth electron and the entry of two additional protons from the matrix lead to the formation of two molecules of water:

$$Fe^{3+} + Cu^{2+} + 2 \text{ H}_2O$$
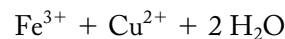
For each proton removed from the matrix, an excess negative charge (in the form of an OH^-) is left behind, thus contribut-

ing directly to the electrochemical gradient across the inner mitochondrial membrane.

As noted above, protons taken up from the matrix are utilized in two very different ways. For every molecule of O_2 reduced to 2 H_2O by cytochrome oxidase: (1) four H^+ ions are consumed in this chemical reaction and (2) four additional H^+ ions are translocated across the inner mitochondrial membrane. The first four protons can be referred to as "substrate" protons and the latter four as "pumped" protons. Movement of both groups of protons contributes to the electrochemical gradient across the membrane.

With the publication of the three-dimensional crystal structure of cytochrome oxidase, it became possible to search for pathways through which substrate and pumped protons might move through the huge protein. Unlike other ions (e.g., Na^+ or Cl^-) that have to diffuse by themselves across the entire distance being traversed, H^+ ions can "hop" through a channel by exchanging themselves with other protons present along the pathway. Such proton-conduction pathways (or "proton wires") can be identified because they consist of strings of acidic residues, hydrogen-bonded residues, and trapped water molecules. Researchers have identified potential proton conduits within the molecule, but their role has not been experimentally established. Unfortunately, static structural models, by themselves, cannot reveal the dynamic movements that take place within the protein as it functions. The energy released by O_2 reduction is presumably used to drive conformational changes that alter the ionization states and precise locations of amino acid side chains within these channels. These changes would, in turn, promote the movement of H^+ ions through the protein.

REVIEW ?

1. Describe the steps by which the transport of electrons down the respiratory chain leads to the formation of a proton gradient.

2. Of the five types of electron carriers, which has the smallest molecular mass? Which has the greatest ratio of iron atoms to electrons carried? Which has a component that is located outside the lipid bilayer? Which are capable of accepting protons and electrons and which only electrons?

3. Describe the relationship between a compound's affinity for electrons and its ability to act as a reducing agent. What is the relationship between the electron-transfer potential of a reducing agent and the ability of the other member of its couple to act as an oxidizing agent?

4. Look at Figure 5.12 and describe how the semiquinone states of ubiquinone and FMN are similar.

5. What is meant by the binuclear center of cytochrome oxidase? How does it function in the reduction of O_2?

6. What are two different ways that cytochrome oxidase contributes to the proton gradient?

7. Why do some electron transfers result in a greater release of energy than other transfers?

5.4 TRANSLOCATION OF PROTONS AND THE ESTABLISHMENT OF A PROTON-MOTIVE FORCE

We have seen that the free energy released during electron transport is utilized to move protons from the matrix to the intermembrane space and cytosol. Translocation of protons across the inner membrane is electrogenic (i.e., voltage producing) because it results in a greater number of positive charges in the intermembrane space and cytosol and a greater number of negative charges within the matrix. Thus, there are two components of the proton gradient to consider. One of the components is the concentration difference between hydrogen ions on one side of the membrane versus the other side; this is a pH gradient (**ΔpH**). The other component is the voltage (ψ) that results from the separation of charge across the membrane. A gradient that has both a concentration (chemical) and an electrical (voltage) component is an *electrochemical gradient* (page 144). The energy present in both components of the proton electrochemical gradient can be combined and expressed as the **proton-motive force (Δp)**, which is measured in millivolts. Thus,

$$\Delta p = \psi - 2.3\ (RT/F)\ \Delta pH$$

Because 2.3 RT/F is equal to 59 mV at 25°C, the equation[3] can be rewritten as

$$\Delta p = \psi - 59\ \Delta pH$$

The contribution to the proton-motive force made by the electric potential versus the pH gradient depends on the permeability properties of the inner membrane. For example, if the outward movement of protons during electron transport is accompanied by negatively charged chloride ions, then the electric potential (ψ) is reduced without affecting the proton gradient (ΔpH). Measurements made in various laboratories suggest that actively respiring mitochondria generate a proton-motive force of approximately 220 mV across their inner membrane. In mammalian mitochondria, approximately 80 percent of the free energy of Δp is represented by the voltage component and the other 20 percent by the proton concentration difference (approximately 0.5 to 1 pH unit difference). If the proton concentration difference were much greater than this it would likely affect the activity of cytoplasmic enzymes. The transmembrane voltage across the inner mitochondrial membrane can be visualized by the use of positively charged, lipid-soluble dyes that distribute themselves across membranes in proportion to the electric potential (Figure 5.20).

When cells are treated with certain lipid-soluble agents, most notably 2,4-dinitrophenol (DNP), they continue to oxidize substrates without being able to generate ATP. In other words, DNP *uncouples* glucose oxidation and ADP phospho-

---

[3]In other words, a difference of one pH unit, which represents a tenfold difference in H^+ concentration across the membrane, is equivalent to a potential difference of 59 millivolts, which is equivalent to a free-energy difference of 1.37 kcal/mol.
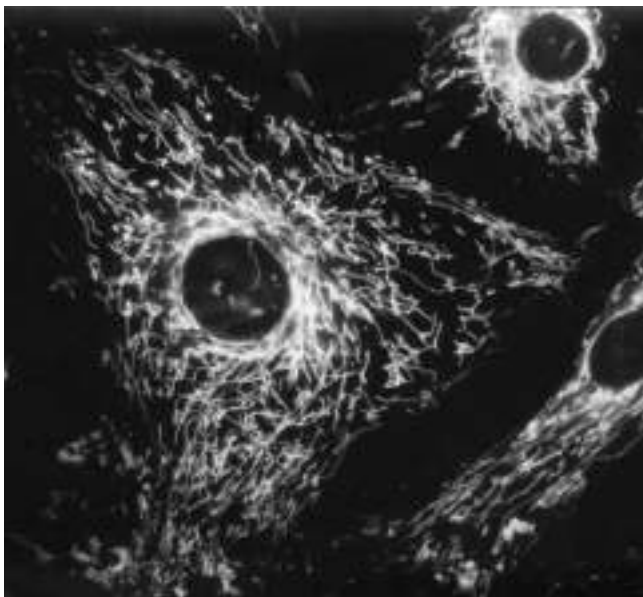
FIGURE 5.20 Visualizing the proton-motive force. Fluorescence micrograph of a cultured cell stained with the fluorescent, cationic dye rhodamine. When the cell is active, the voltage generated across the inner mitochondrial membrane (inside negative) leads to the accumulation of the cationic dye within the mitochondria, causing the organelles to fluoresce. (FROM L. V. JOHNSON ET AL., *J. CELL BIOL.* 88:528, 1981, COURTESY OF LAN BO CHEN; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

rylation. Because of this property, DNP has become widely used in the laboratory to inhibit ATP formation. During the 1920s, a few physicians even prescribed DNP as a diet pill. When exposed to the drug, the cells of obese patients continued to oxidize their stores of fat in a vain attempt to maintain normal ATP levels. The practice ceased with the deaths of a number of patients taking the drug. With the formulation of the chemiosmotic theory and the determination that mitochondria generate a proton gradient, the mechanism of DNP action became understood. This drug can uncouple oxidation and phosphorylation because it combines with protons and, because of its lipid solubility, carries the protons across the inner mitochondrial membrane down their electrochemical gradient.

The maintenance of a proton-motive force requires that the inner mitochondrial membrane remain highly impermeable to protons. If not, the gradient established by electron transport is quickly dissipated by the leakage of protons back into the matrix, leading to the release of energy as heat. It came as a surprise to discover that the inner mitochondrial membrane of certain cells contains proteins that act as natural (endogenous) uncouplers. These proteins, called *uncoupling proteins* (or *UCPs*), are especially abundant in brown adipose tissue of mammals, which functions as a source of heat production during exposure to cold temperatures. Human infants also rely on brown fat deposits to maintain body temperature. These brown fat cells are largely lost as we grow up, which

makes us dependent on muscle contraction (shivering) to generate body heat. UCP1 is the focus of a number of drug companies that are hoping to develop products that stimulate this protein as part of a weight-loss regimen. Other UCP isoforms (UCP2-UCP5) are present in a variety of tissues, especially in cells of the nervous system, but their roles are debated. According to one hypothesis, UCPs in the inner mitochondrial membrane prevent the build-up of an excessively large proton-motive force. If such a high-energy state were to form, it could block the passage of electrons through the respiratory complexes, leading to the leakage of electrons and the production of reactive oxygen radicals.

The experiments that led to the acceptance of the proton-motive force as an intermediate in the formation of ATP are discussed in the Experimental Pathways, which can be found at www.wiley.com/college/karp on the Web.

REVIEW ?

1. What are the two components of the proton-motive force, and how can their relative contribution vary from one cell to another?

2. What is the effect of dinitrophenol on ATP formation by mitochondria? Why is this the case?

5.5 THE MACHINERY FOR ATP FORMATION

Now that we have seen how the transport of electrons generates a proton electrochemical gradient across the inner mitochondrial membrane, we can turn to the molecular machinery that utilizes the energy stored in this gradient to drive the phosphorylation of ADP.

During the early 1960s, Humberto Fernandez-Moran of Massachussetts General Hospital was examining isolated mitochondria by the recently developed technique of negative staining. Fernandez-Moran discovered a layer of spheres attached to the inner (matrix) side of the inner membrane, projecting from the membrane and attached to it by stalks (Figure 5.21). A few years later, Efraim Racker of Cornell University isolated the inner membrane spheres, which he called *coupling factor 1*, or simply F_1. Racker discovered that the F_1 spheres behaved like an enzyme that hydrolyzed ATP, that is, an ATPase. At first glance this seems to be a peculiar finding. Why should mitochondria possess a predominant enzyme that hydrolyzes the substance it is supposed to produce?

If one considers that the hydrolysis of ATP is the reverse reaction from its formation, the function of the F_1 sphere becomes more evident; it contains the catalytic site at which ATP formation normally occurs. Recall that

1. Enzymes do not affect the equilibrium constant of the reaction they catalyze.

2. Enzymes are capable of catalyzing both the forward and reverse reactions.

10 nm
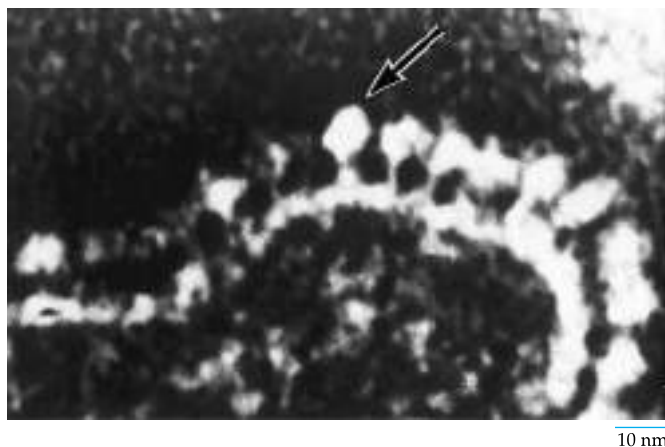
FIGURE 5.21 The machinery for ATP synthesis. Electron micrograph of a small portion of a beef heart mitochondrion that has been air-dried and negatively stained. At magnifications of approximately one-half million, spherical particles are seen (arrow) attached by a thin stalk to the inner surface of the cristae membranes. (FROM HUMBERTO FERNANDEZ-MORAN ET AL., J. CELL BIOL. 22:73, 1964; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

Consequently, the direction of an enzyme-catalyzed reaction at any given time depends on the prevailing conditions. This is nicely shown in experiments with other ATPases, such as the Na^+/K^+-ATPase of the plasma membrane (page 153). When this enzyme was discussed in Chapter 4, it was described as an enzyme that used the energy obtained from ATP hydrolysis to export Na^+ and import K^+ against their respective gradients. In the cell this is the only function of the enzyme. However, under experimental conditions, this enzyme can catalyze the formation of ATP rather than its hydrolysis
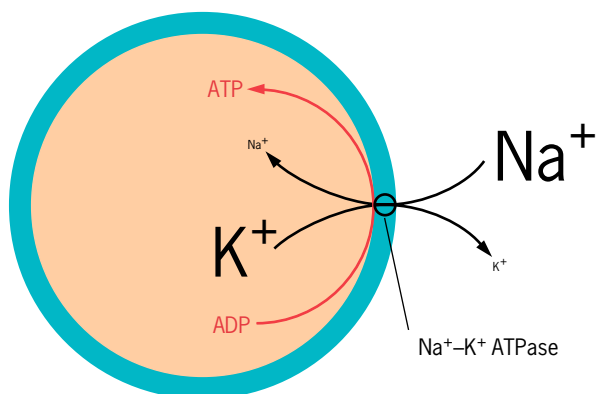
FIGURE 5.22 An experiment to drive ATP formation in membrane vesicles reconstituted with the Na^+/K^+-ATPase. By causing these vesicles to have very high internal K^+ concentration and very high external Na^+ concentration, the reaction is made to run in the opposite direction from that of its normal course in the plasma membrane. In the process, ATP is formed from ADP and P_i. The size of the letters indicates the direction of the concentration gradients.

(Figure 5.22). To obtain such conditions, red blood cell ghosts (page 141) are prepared with a *very* high internal K^+ concentration and a *very* high external Na^+ concentration, higher than normally exists in the body. Under these conditions, K^+ moves out of the "cell" and Na^+ moves into the "cell." Both ions are moving down their respective gradients, rather than against them as would normally occur in a living cell. If ADP and P_i are present within the ghost, the movement of the ions causes ATP to be synthesized rather than hydrolyzed. Experiments such as this illustrate the reality of what would be expected on the basis of the theoretical reversibility of enzyme-catalyzed reactions. They also illustrate how an ionic gradient can be used to drive a reaction in which ADP is phosphorylated to ATP, which is precisely what occurs in the mitochondrion. The driving force is the proton-motive force established by electron transport.

The Structure of ATP Synthase

Although the F_1 sphere is the catalytic portion of the enzyme that manufactures ATP in the mitochondrion, it is not the whole story. The ATP-synthesizing enzyme, which is called **ATP synthase**, is a mushroom-shaped protein complex (Figure 5.23*a*) composed of two principal components: a spherical F_1 head (about 90 Å diameter) and a basal section, called F_0, embedded in the inner membrane. High-resolution electron micrographs indicate that the two portions are connected by both a central and a peripheral stalk as shown in Figure 5.23*b*. A typical mammalian liver mitochondrion has roughly 15,000 copies of the ATP synthase. Homologous versions of the ATP synthase are found in the plasma membrane of bacteria, the thylakoid membrane of plant chloroplasts, and the inner membrane of mitochondria.

The F_1 portions of the bacterial and mitochondrial ATP synthases are highly conserved; both contain five different polypeptides with a composition of $\alpha_3\beta_3\delta\gamma\epsilon$. The α and β subunits are arranged alternately within the F_1 head in a manner resembling the segments of an orange (Figure 5.23*b*; see also 5.26*b*). Two aspects can be noted for later discussion: (1) each F_1 contains three catalytic sites for ATP synthesis, one on each β subunit, and (2) the γ subunit runs from the outer tip of the F_1 head through the central stalk and makes contact with the F_0 basepiece. In the mitochondrial enzyme, all five polypeptides of F_1 are encoded by the nuclear DNA, synthesized in the cytosol, and then imported (see Figure 8.47).

The F_0 portion of the ATP synthase resides within the membrane and consists of three different polypeptides with a stoichiometry of ab_2c_{10-14} (Figure 5.23*b*). The number of subunits in the *c* ring is written as 10–14 because structural studies have revealed that this number can vary depending on the source of the enzyme. Both the yeast mitochondrial and *E. coli* ATP synthase, for example, have 10 *c* subunits, whereas a chloroplast enzyme has 14 *c* subunits (Figure 5.24). The F_0 base contains a channel through which protons are conducted from the intermembrane space to the matrix. The presence of a transmembrane channel was first demonstrated by experi-
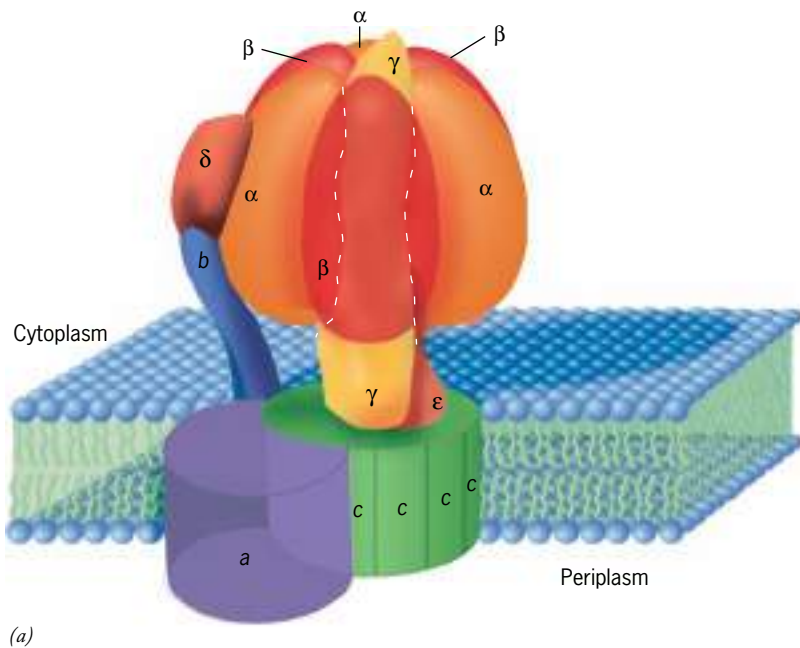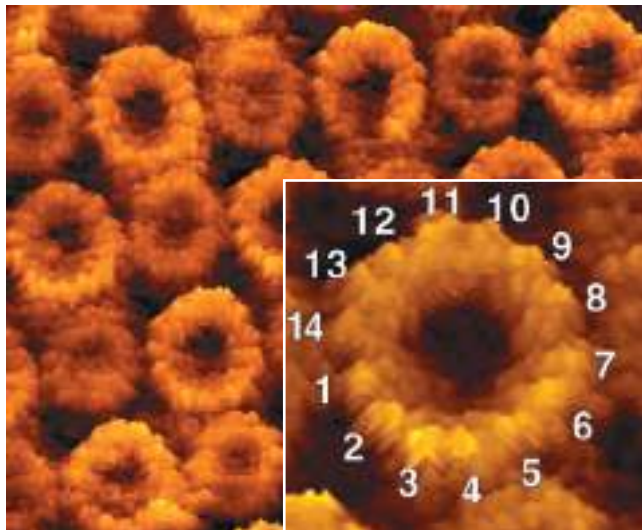
(a)
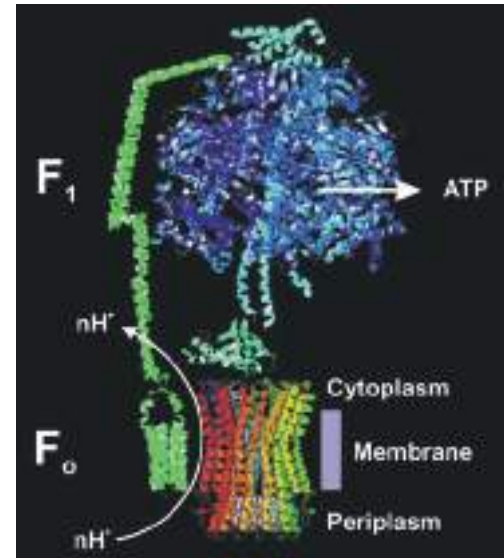
(b)

FIGURE 5.23 **The structure of the ATP synthase.** (*a*) Schematic diagram of the bacterial ATP synthase. The enzyme consists of two major portions, called F_1 and F_0. The F_1 head consists of five different subunits in the ratio $3\alpha : 3\beta : 1\delta : 1\gamma : 1\epsilon$. The α and β subunits are organized in a circular array to form the spherical head of the particle; the γ subunit runs through the core of the ATP synthase from the tip of F_1 down to F_0 to form a central stalk; the ϵ subunit helps attach the γ subunit to the F_0 base. The F_0 base, which is embedded in the bacterial plasma membrane, consists of three different subunits in the apparent ratio $1a : 2b : 10{-}14c$. As discussed later, the *c* subunits form a rotating ring within the membrane; the paired *b* subunits of the F_0 base and the

δ subunit of the F_1 head form a peripheral stalk that holds the α/β subunits in a fixed position; and the *a* subunit contains the proton channel that allows protons to traverse the membrane. The mammalian enzyme contains seven to nine small additional subunits whose functions are not well established. (*b*) Three-dimensional structure of the bacterial ATP synthase. This image is composed of several partial structures of the enzyme from various organisms. An animation of the enzyme can be found at www.biologie.uni-osnabrueck.de/biophysik/junge (B: FROM WOLFGANG JUNGE AND NATHAN NELSON, REPRINTED WITH PERMISSION FROM SCIENCE 308:643, 2005. COPYRIGHT 2005, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

ments in which the inner mitochondrial membrane was broken into fragments that form membrane vesicles, called *submitochondrial particles* (Figure 5.25*a*). Intact submitochondrial particles, which contain the ATP synthase embedded in the vesicle membrane, are capable of oxidizing substrates, generating a proton gradient, and synthesizing ATP (Figure 5.25*b*). If, however, the F_1 spheres are removed from the particles, the vesicle membrane can no longer maintain a proton gradient despite continuing substrate oxidation and electron transport. Protons that are translocated across the membrane during electron transport simply cross back through the "beheaded" ATP synthase, and the energy is dissipated.

5 nm

FIGURE 5.24 **Visualizing the oligomeric *c* ring of a chloroplast ATP synthase.** Atomic force microscopy of a "field" of *c* rings that have been isolated from chloroplast ATP synthases and reconstituted as a two-dimensional array within an artificial lipid layer. Rings of two different diameters are present in the field because the oligomers are thought to lie in both possible orientations within the "artificial membrane." (*Inset*) Higher-resolution view of one of the rings, showing that it is composed of 14 subunits. (REPRINTED WITH PERMISSION FROM HOLGER SEELERT ET AL., COURTESY OF DANIEL J. MÜLLER AND ANDREAS ENGEL, NATURE 405:419, 2000; COPYRIGHT 2000, MACMILLAN MAGAZINES LIMITED.)
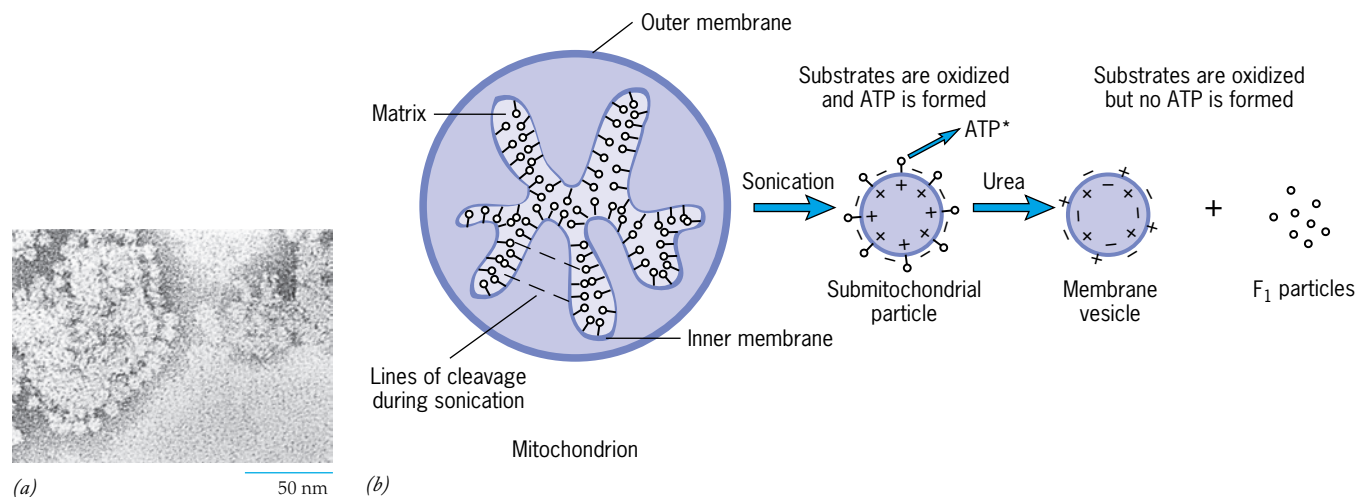
FIGURE 5.25 ATP formation in experiments with submitochondrial particles. (*a*) Electron micrograph of submitochondrial particles, which are fragments of the inner mitochondrial membrane that have become closed vesicles with the F_1 spheres projecting outward into the medium. (*b*) Outline of an experiment showing that intact submitochondrial particles are capable of substrate oxidation, proton gradient formation (indicated by the separation of charge across the membrane), and ATP formation. In contrast, submitochondrial particles lacking the F_1 heads, which are removed by treatment with urea, are capable of substrate oxidation but are not able to maintain a proton gradient (indicated by the lack of charge separation), and thus they are unable to form ATP. (A: COURTESY OF EFRAIM RACKER.)

The Basis of ATP Formation According to the Binding Change Mechanism

How does a proton electrochemical gradient provide the energy required to drive the synthesis of ATP? To answer this question, Paul Boyer of UCLA published an innovative hypothesis in 1979, called the *binding change mechanism*, which has since gained wide acceptance. Overall, the binding change hypothesis has several distinct components, which we will discuss in sequence.

1. **The energy released by the movement of protons is not used to drive ADP phosphorylation directly but principally to change the binding affinity of the active site for the ATP product.** We are used to thinking of cellular reactions occurring in an aqueous environment in which the concentration of water is 55 M and the reactants and products are simply dissolved in the medium. Under these conditions, energy is required to drive the formation of the covalent bond linking ADP with inorganic phosphate to form ATP. It has been demonstrated, however, that once ADP and P_i are bound within the catalytic site of the ATP synthase, the two reactants readily condense to form a tightly bound ATP molecule without the input of additional energy. In other words, even though the reaction

soluble ADP + soluble P_i → soluble ATP + H_2O
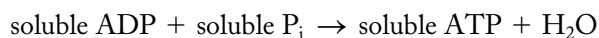
may require the input of considerable energy (7.3 kcal/mol under standard conditions), the reaction

enzyme-bound ADP + enzyme-bound P_i →
enzyme-bound ATP + H_2O
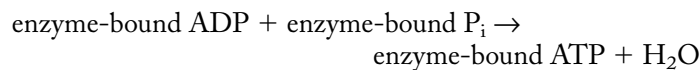
has an equilibrium constant close to 1 ($\Delta G^\circ = 0$) and thus

can occur spontaneously without the input of energy (page 88). This does not mean that ATP can be formed from ADP without energy expenditure. Instead, the energy is required for the release of the tightly bound product from the catalytic site, rather than the phosphorylation event itself.

2. **Each active site progresses successively through three distinct conformations that have different affinities for substrates and product.** Recall that the F_1 complex has three catalytic sites, one on each of the three β subunits. When investigators examined the properties of the three catalytic sites in a *static* enzyme (one that was not engaged in enzymatic turnover), the different sites exhibited different chemical properties. Boyer proposed that, at any one time, the three catalytic sites are present in different conformations, which causes them to have different affinity for nucleotides. At any given instant, one site is in the "loose" or L conformation in which ADP and P_i are loosely bound; a second site is in the "tight" or T conformation in which nucleotides (ADP + P_i substrates, or ATP product) are tightly bound; and the third site is in the "open" or O conformation, which, because it has a very low affinity for nucleotides, allows release of ATP.

Although there were differences among the three catalytic sites in a static enzyme, Boyer obtained evidence that all of the ATP molecules produced by an active enzyme were synthesized by the same catalytic mechanism. It seemed, in other words, that all three of the catalytic sites of the enzyme were operating identically. To explain the apparent contradiction between the asymmetry of the enzyme's structure and the uniformity of its catalytic mechanism, Boyer proposed that each of the three catalytic sites passed sequentially through the same L, T, and O conformations (see Figure 5.27).

3. ATP is synthesized by rotational catalysis in which one part of the ATP synthase rotates relative to another part. To explain the sequential changes in the conformation of each of the catalytic sites, Boyer proposed that the α and β subunits, which form a hexagonal ring of subunits within the F_1 head (Figure 5.23), rotate *relative* to the central stalk. In this model, which is referred to as *rotational catalysis*, rotation is driven by the movement of protons through the membrane via the channel in the F_0 base. Thus, according to this model, electrical energy stored in the proton gradient is transduced into mechanical energy of a rotating stalk, which is transduced into chemical energy stored in ATP.

Evidence to Support the Binding Change Mechanism and Rotary Catalysis The publication in 1994 of a detailed atomic model of the F_1 head by John Walker and his colleagues of the Medical Research Council in England provided a remarkable body of structural evidence to support Boyer's binding change mechanism. First, it revealed the structure of each of the catalytic sites in a static enzyme, confirming that they differ in their conformation and their affinity for nucleotides. Structures corresponding to the L, T, and O conformations were identified in the catalytic sites of the three β subunits. Second, it revealed that the γ subunit of the enzyme is perfectly positioned within the ATP synthase to transmit conformational changes from the F_0 membrane sector to the

F_1 catalytic sites. The γ subunit could be seen to extend as a shaft from the F_0 sector through the stalk and into a central cavity within the F_1 sphere (Figure 5.26*a*), where it contacts each of the three β subunits differently (Figure 5.26*b*).

The apical end of the γ subunit is highly asymmetric, and at any given instant, different faces of the γ subunit interact with the different β subunits, causing them to adopt different (L, T, and O) conformations. As it rotates, each binding site on the γ subunit interacts successively with the three β subunits of F_1. During a single catalytic cycle, the γ subunit rotates a full 360°, causing each catalytic site to pass sequentially through the L, T, and O conformations. This mechanism is illustrated schematically in Figure 5.27 and discussed in detail in the accompanying legend. As indicated in Figure 5.27*a*, condensation of ADP and P_i to form ATP occurs while each subunit is in the T conformation. As indicated in Figure 5.23*b*, the ε subunit is associated with the "lower" portion of the γ subunit, and the two subunits rotate together as a unit.

Direct evidence for the rotation of the γ subunit relative to the αβ subunits has been obtained in a variety of experiments. If "seeing is believing," then the most convincing demonstration came in 1997 from the work of Masasuke Yoshida and colleagues at the Tokyo Institute of Technology and Keio University in Japan. These researchers devised an ingenious system to watch individual enzyme molecules catalyze the reverse reaction from that normally operating in the cell. First, they prepared a genetically engineered version of the F_1

(a)

(b)
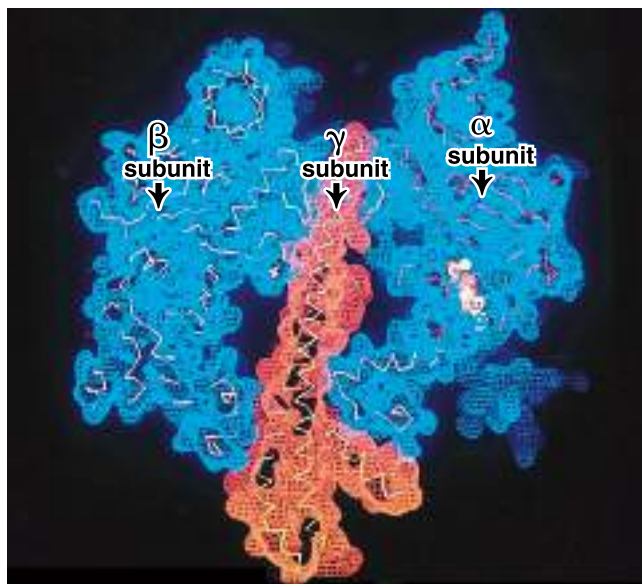
FIGURE 5.26   The structural basis of catalytic site conformation. (*a*) A section through the F_1 head shows the spatial organization of three of its subunits. The γ subunit is constructed of two extended and intertwined α helices. This helical stalk projects into the central cavity of the F_1, and between the α and β subunits on each side. The conformation of the catalytic site of the β subunit (shown on the left) is determined by its contact with the γ subunit. (*b*) A top view of the F_1 head showing the arrangement of the six α and β subunits (shown in red and yellow)
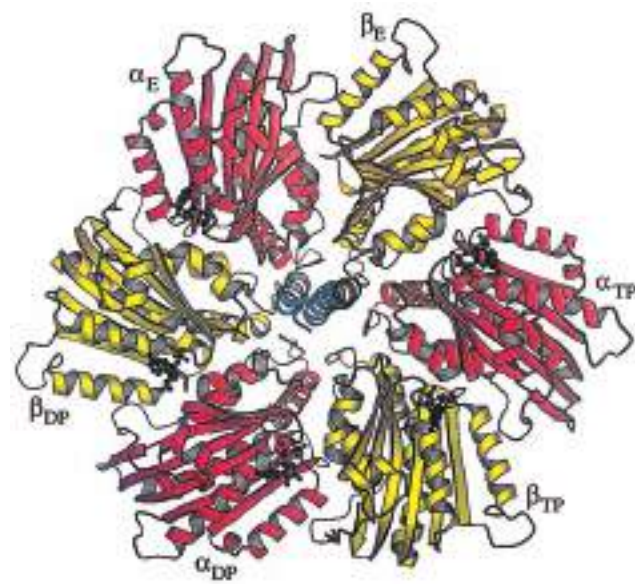
around the asymmetric γ subunit (shown in blue). The γ subunit is in position to rotate relative to the surrounding subunits. It is also evident that the γ subunit makes contact in a different way with each of the three β subunits, inducing each of them to adopt a different conformation. $β_E$ corresponds to the O conformation, $β_{TP}$ to the L conformation, and $β_{DP}$ to the T conformation. (REPRINTED WITH PERMISSION FROM J. P. ABRAHAMS ET AL., COURTESY OF JOHN E. WALKER, NATURE 370:624, 627, 1994. COPYRIGHT 1994, MACMILLAN MAGAZINES LIMITED.)

(a)

(b)
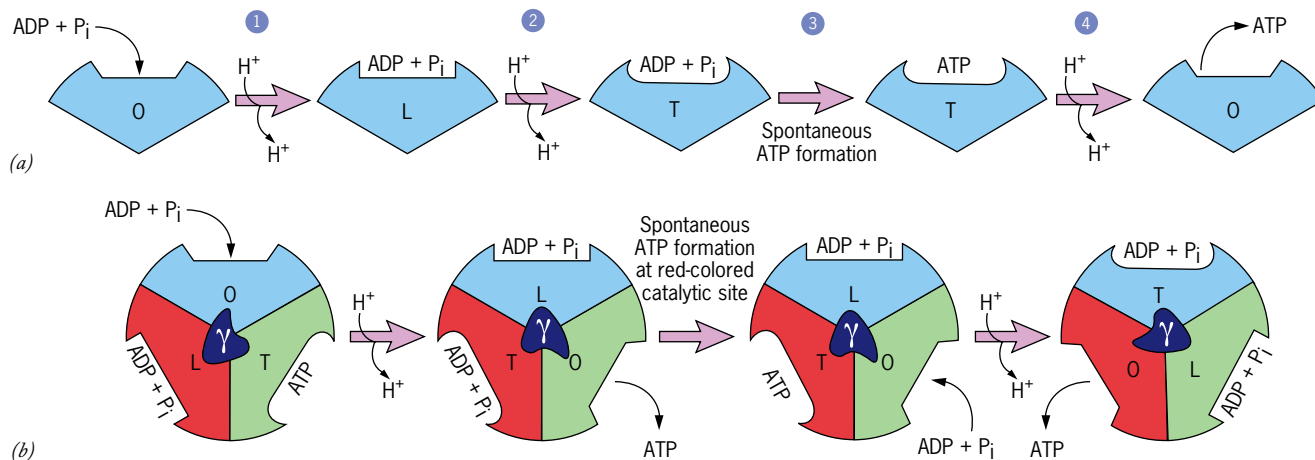
FIGURE 5.27 **The binding change mechanism for ATP synthesis.** (a) Schematic drawing showing changes in a single catalytic site during a cycle of catalysis. At the beginning of the cycle, the site is in the open (O) conformation, and substrates ADP and P_i are entering the site. In step 1, the movement of protons through the membrane induces a shift to the loose (L) conformation in which the substrates are loosely bound. In step 2, the movement of additional protons induces a shift to the tight (T) conformation, in which the affinity for substrates increases, causing them to be tightly bound to the catalytic site. In step 3, the tightly bound ADP and P_i spontaneously condense to form a tightly bound ATP; no change in conformation is required for this step. In step 4, the movement of additional protons induces a shift to the open (O) conformation, in which the affinity for ATP is greatly decreased, allowing the product to be released from the site. Once the ATP has dissociated, the catalytic site is available for substrate binding, and the cycle is repeated. (b) Schematic drawing showing changes at all three catalytic sites of the enzyme simultaneously. The movement of protons through the F_0 portion of the enzyme causes the rotation of the asymmetric γ subunit, which displays three different faces to the catalytic subunits. As the γ subunit rotates, it induces changes in the conformation of the catalytic site of the β subunits, causing each catalytic site to pass successively through the T, O, and L conformations.

portion of the ATP synthase, consisting of three α subunits, three β subunits, and a γ subunit ($\alpha_3\beta_3\gamma$) (Figure 5.28a). This polypeptide complex was then fixed to a glass coverslip by its head, and a short, fluorescently labeled actin filament was attached to the end of the γ subunit that projected into the medium (Figure 5.28a). When the preparation was incubated with ATP and observed under the microscope, the fluorescently labeled actin filament was seen to rotate like a microscopic propellor (Figure 5.28b), powered by the energy released as ATP molecules were bound and hydrolyzed by the catalytic sites of the β subunits.

More recently, a similar type of F_1 complex was "forced" to perform the more challenging activity that it normally carries out within the cell—the synthesis of ATP. In this study, a microscope magnetic bead was attached to the γ subunit of a single F_1 molecule and then caused to rotate in a clockwise direction by subjecting the preparation to a revolving magnetic field. When one of these F_1 molecules was placed within a tiny transparent microchamber in the presence of ADP and P_i, the forced rotation of the γ subunit led to the synthesis of ATP, which built up to μM concentrations. As would be expected from the model, three molecules of ATP were synthesized with each 360° turn. When the magnetic field was switched off, the γ subunit rotated spontaneously in the reverse direction, driven by the hydrolysis of the recently synthesized ATP. Taken together, these innovative bioengineering experiments demonstrate unequivocally that ATP synthase operates as a rotary motor.

Rotary machines are common in our industrialized society; we use rotary turbines, rotary drills, rotary wheels and

(a)

(b)
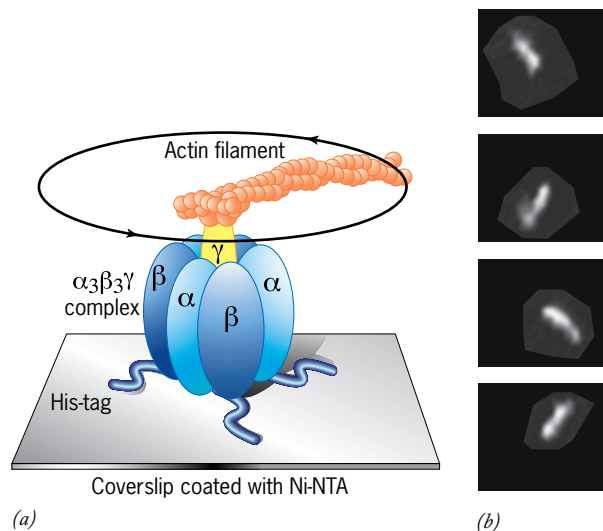
FIGURE 5.28 **Direct observation of rotational catalysis.** (a) To carry out the experiment, a modified version of a portion of the ATP synthase consisting of $\alpha_3\beta_3\gamma$ was prepared. Each β subunit was modified to contain 10 histidine (His) residues at its N-terminus, a site located on the outer (matrix) face of the F_1 head. The side chains of histidine have a high affinity for a substance (Ni-NTA), which was used to coat the coverslip. The γ subunit was modified by replacing one of the serine residues near the end of the stalk with a cysteine residue, which provided a means to attach the fluorescently labeled actin filament. In the presence of ATP, the actin filament was observed to rotate counterclockwise (when viewed from the membrane side). At low ATP concentrations, the actin filaments could be seen to rotate in steps of 120°. (b) A sequence of four frames from a video of a rotating actin filament. (REPRINTED WITH PERMISSION FROM H. NOJI ET AL., COURTESY OF MASASUKE YOSHIDA, NATURE 386:300, 1997; COPYRIGHT 1997, MACMILLAN MAGAZINES LIMITED.)

propellors, to name just a few. But rotary devices are extremely rare in living organisms. There are, for example, no known rotary organelles in eukaryotic cells, rotary joints in animals, or rotary feeding structures in the biological world. In fact, only two types of biological structures are known that contain rotating parts: ATP synthases (and related proteins that act as ionic pumps) and bacterial flagella (which are pictured in Figure 1.14*a*, inset), both of which can be described as rotary "nanomachines" because their size is measured in nanometers. Engineers have begun to invent nanoscale devices made of inorganic materials that may one day carry out various types of submicroscopic mechanical activities. The construction of nanosized motors poses a particular challenge, and attempts have already been made to use the ATP synthase to power simple, inorganic devices. Someday humans may be using ATP instead of electricity to power some of their most delicate instruments.

Using the Proton Gradient to Drive the Catalytic Machinery: The Role of the F_0 Portion of ATP Synthase By 1997, a detailed understanding of the working machinery of the F_1 complex was in hand, but major questions about the structure and function of the membrane-bound F_0 portion of the enzyme remained to be answered. Most important among them were: What is the path taken by protons as they move through the F_0 complex, and how does this movement lead to the synthesis of ATP? It had been postulated that:

1. The *c* subunits of the F_0 base were assembled into a ring that resides within the lipid bilayer (as in Figure 5.23*b*).

2. The *c* ring is physically bound to the γ subunit of the stalk.

3. The "downhill" movement of protons through the membrane drives the rotation of the ring of *c* subunits.

4. The rotation of the *c* ring of F_0 provides the twisting force (*torque*) that drives the rotation of the attached γ subunit, leading to the synthesis and release of ATP.

All of these presumptions have been confirmed. We will look at each of these aspects in more detail.

A body of evidence, including X-ray crystallography and atomic force microscopy, has shown that the *c* subunits are indeed organized into a circle to form a ring-shaped complex (Figure 5.24). High-resolution electron micrographs indicate that the two *b* subunits and the single *a* subunit of the F_0 complex reside outside the ring of *c* subunits, as shown in Figure 5.23. The *b* subunits are thought to be primarily structural components of the ATP synthase. The two elongated *b* subunits form a *peripheral stalk* that connects the F_0 and F_1 portions of the enzyme (Figure 5.23*b*) and, along with the δ subunit of F_1, are thought to hold the $\alpha_3\beta_3$ subunits in a fixed position while the γ subunit rotates within the center of the complex.

The rotation of the *c* ring of F_0 during ATP synthesis has been demonstrated in numerous experiments, confirming that both the *c* ring and γ subunit act as rotors during enzyme activity. How are these two "moving parts" connected? Each *c* subunit is built like a hairpin: it contains two transmembrane helices connected by a hydrophilic loop that projects toward

the F_1 head. The hydrophilic loops situated on the tops of the *c* subunits are thought to form a binding site for the bases of the γ and ε subunits, which together act like a "foot" that is firmly attached to the *c* ring (Figures 5.23*b* and 5.29). As a result of this attachment, rotation of the *c* ring drives rotation of the attached γ subunit.

The mechanism by which H^+ movements drive the rotation of the *c* ring is more complex and less well understood. Figure 5.29 provides a model as to how H^+ ions might flow through the F_0 complex. Keep in mind in the following discussion of this model (1) that the subunits of the *c* ring move successively past a *stationary a* subunit and (2) that protons are picked up from the intermembrane space one at a time by each *c* subunit and carried *completely around a circle* before they are released into the matrix. In this model, each *a* subunit has two half-channels that are physically separated (offset) from one another. One half-channel leads from the intermembrane (cytosolic) space into the middle of the *a* subunit and the other leads from the middle of the *a* subunit into the matrix. It is proposed that each proton moves from the in-
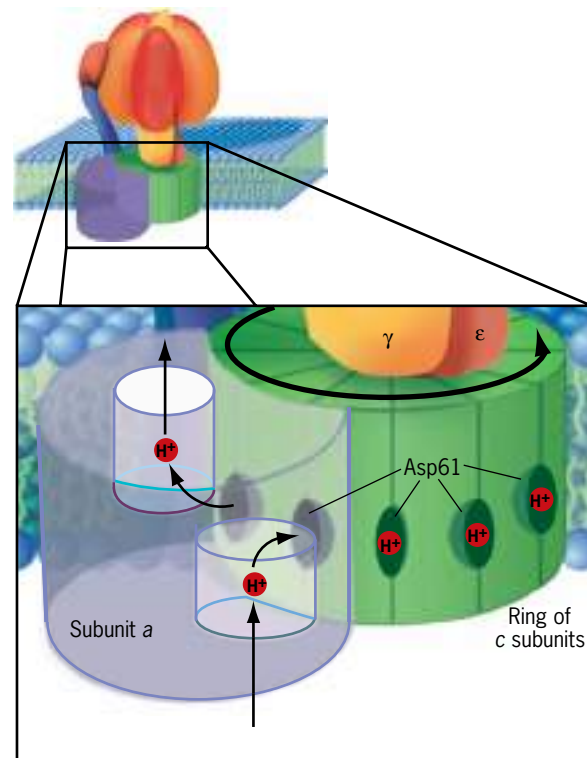
FIGURE 5.29 A model in which proton diffusion is coupled to the rotation of the *c* ring of the F_0 complex. As discussed in the text, it is proposed in this model that each proton from the intermembrane space enters a half-channel within the *a* subunit and then binds to an aspartic acid residue (Asp 61 in *E. coli*) that is accessible on one of the *c* subunits. Proton binding induces a conformational change that causes the ring to move by approximately 30°. The bound proton is carried in a full circle by the rotation of the *c* ring and is then released into a second half-channel that opens into the matrix. A succession of protons that engage in this activity causes the *c* ring to rotate in a counterclockwise direction as shown.

termembrane space through the half-channel and binds to a negatively charged aspartic acid residue situated at the surface of the adjoining *c* subunit (Figure 5.29). Binding of the proton to the carboxyl group generates a major conformational change in the *c* subunit that causes that subunit to rotate approximately 30° in a counterclockwise direction. The movement of the recently protonated *c* subunit brings the adjoining subunit in the ring (which was protonated at an earlier step) into alignment with the second half-channel of the *a* subunit. Once there, the aspartic acid releases its associated proton, which diffuses into the matrix. Following dissociation of the proton, the *c* subunit returns to its original conformation and is ready to accept another proton from the intermembrane space and repeat the cycle.

According to this model, the aspartic acid of each *c* subunit acts like a revolving proton carrier. A proton hops onto the carrier at a selected pick-up site and is then carried around a circle before being released at a selected drop-off site. Movement of the ring is driven by the conformational changes associated with the sequential protonation and deprotonation of the aspartic acid residue of each *c* subunit. It was noted on page 193 that the *c* ring has been shown to contain between 10 and 14 subunits, depending on the source. For the sake of simplicity, we will describe a *c* ring composed of 12 subunits, as illustrated in Figure 5.29. In this case, the association/dissociation of four protons in the manner described would move the ring 120°. Movement of the *c* ring 120°

would drive a corresponding rotation of the attached γ subunit 120°, which would lead to the release of one newly synthesized molecule of ATP by the F_1 complex. According to this stoichiometry, the translocation of 12 protons would lead to the full 360° rotation of the *c* ring and γ subunit, and the synthesis and release of three molecules of ATP. If the *c* ring were to contain greater or fewer than 12 subunits, the H^+/ATP ratio would change, but this can easily be accommodated within the basic model of proton-driven rotation shown in Figure 5.29.

Other Roles for the Proton-Motive Force in Addition to ATP Synthesis

Although the production of ATP may be the most important activity of mitochondria, these organelles are engaged in numerous other processes that require the input of energy. Unlike most organelles that rely primarily on ATP hydrolysis to power their activities, mitochondria rely on an alternate source of energy—the proton-motive force. For example, the proton-motive force drives the uptake of ADP and P_i into the mitochondria in exchange for ATP and H^+, respectively. These and other activities that take place during aerobic respiration are summarized in Figure 5.30. In other examples, the proton-motive force may be used as the source of energy to "pull" calcium ions into the mitochondrion; to drive the events of mitochondrial fusion; and to cause specifically targeted polypeptides to enter the mitochondrion from the matrix (Figure 8.47).
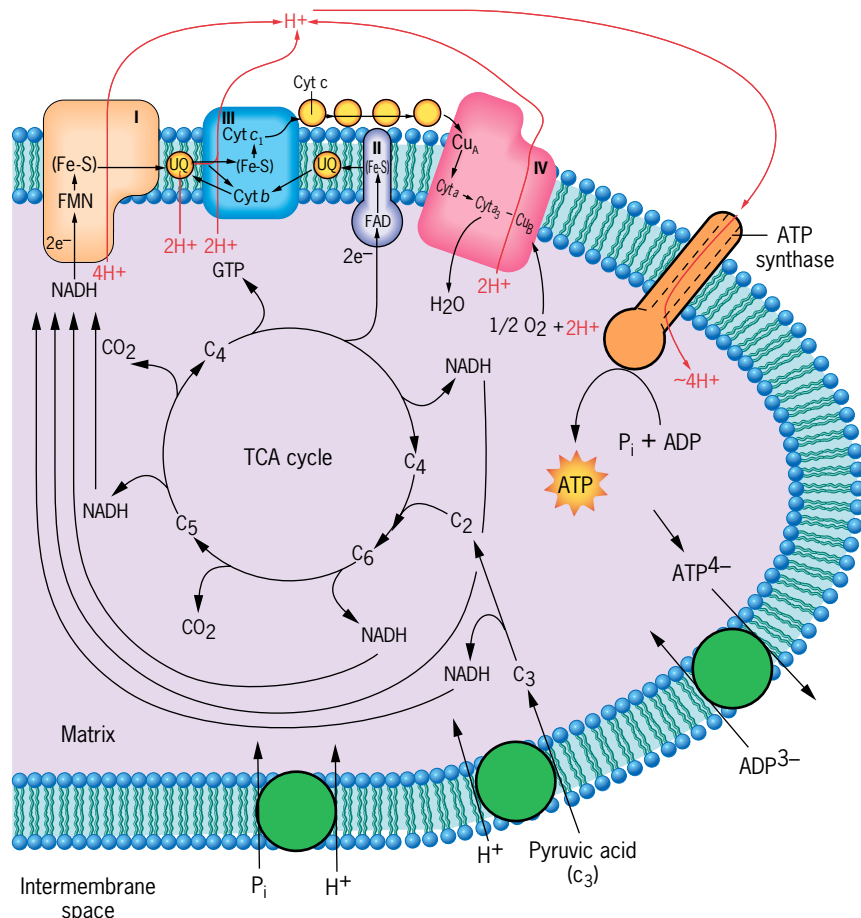
FIGURE 5.30 Summary of the major activities during aerobic respiration in a mitochondrion.

We saw in Chapter 3 how ATP levels play an important role in controlling the rate of glycolysis and the TCA cycle by regulating the activity of key enzymes. Within the mitochondrion, the ADP level is a key determinant of respiratory rate. When ADP levels are low, ATP levels are typically high, and there is no need for additional substrate to be oxidized to provide electrons for the respiratory chain. Under these conditions, ATP synthesis is low, so that protons are unable to reenter the mitochondrial matrix through the ATP synthase. This leads to a sustained build-up of the proton-motive force, which in turn inhibits the proton pumping reactions of the electron-transport chain and the consumption of oxygen by cytochrome oxidase. When the ATP/ADP ratio decreases, oxygen consumption abruptly increases. Many factors have been shown to influence the rate of mitochondrial respiration, but the pathways that govern respiratory control are poorly understood.

REVIEW ?

1. Describe the basic structure of the ATP synthase.
2. Describe the steps in ATP synthesis according to the binding change mechanism.
3. Describe some of the evidence that supports the binding change mechanism.
4. Describe a proposed mechanism whereby proton diffusion from the intermembrane space into the matrix drives the phosphorylation of ADP.

(a)

5.6 PEROXISOMES

Peroxisomes are simple, membrane-bound vesicles (Figure 5.31*a*) with a diameter of 0.1 to 1.0 μm that may contain a dense, crystalline core of oxidative enzymes. Peroxisomes (or *microbodies*, as they are also called) are multifunctional organelles, containing more than 50 enzymes involved in such diverse activities as the oxidation of very-long-chain fatty acids (VLCFAs, ones whose chain typically contains 24 to 26 carbons) and the synthesis of plasmalogens, which are an unusual class of phospholipids in which one of the fatty acids is linked to glycerol by an ether linkage rather than an ester linkage. Plasmalogens are very abundant in the myelin sheaths that insulate axons in the brain (Figure 4.5). Abnormalities in the synthesis of plasmalogens can lead to severe neurological dysfunction. The enzyme luciferase, which generates the light emitted by fireflies, is also a peroxisomal enzyme. Peroxisomes are considered in the present chapter because they share several properties with mitochondria: both mitochondria and peroxisomes form by splitting from preexisting organelles, using some of the same proteins to accomplish the feat; both types of organelles import preformed proteins from the cytosol (Section 8.9); and both engage in similar types of oxidative metabolism. In fact, at least one enzyme, alanine/glyoxylate aminotransferase, is found in the mitochondria of some mammals (e.g., cats and dogs) and in the peroxisomes of other mammals (e.g., rabbits and humans).

These organelles were named "peroxisomes" because they are the site of synthesis and degradation of hydrogen peroxide (H_2O_2), a highly reactive and toxic oxidizing agent. Hydrogen peroxide is produced by a number of peroxisomal enzymes, including urate oxidase, glycolate oxidase, and amino acid oxidases, that utilize molecular oxygen to oxidize their respective substrates (Figure 5.31*b*). The H_2O_2 generated in these reactions is rapidly broken down by the enzyme catalase, which is present in high concentrations in these organelles. The importance of peroxisomes in human metabolism is evident from the discussion in the accompanying Human Perspective.

(b)

FIGURE 5.31 The structure and function of peroxisomes. (*a*) Electron micrograph of a section of a rat liver cell, which had been stained for catalase, an enzyme localized in peroxisomes. Note the close association between the darkly stained peroxisomes and mitochondria. An electron micrograph of a peroxisome within a plant cell is shown in Figure 6.23. Bar equals 250 nm. (*b*) Peroxisomes contain enzymes that carry out the two-step reduction of molecular oxygen to water. In the first step, an oxidase removes electrons from a variety of substrates (RH_2), such as uric acid or amino acids. In the second step, the enzyme catalase converts the hydrogen peroxide formed in the first step to water. (A: FROM MICHAEL SCHRADER AND YISANG YOON, BIOESS. 29:1106, 2007, COPYRIGHT © 2007, JOHN WILEY & SONS, INC. REPRINTED BY PERMISSION OF JOHN WILEY & SONS.)

10 μm
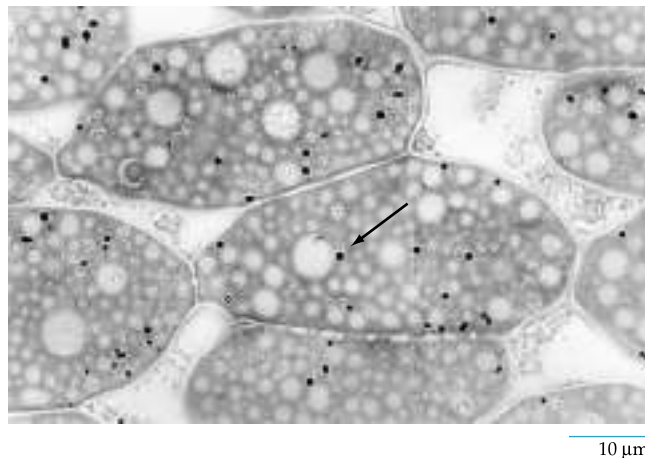
FIGURE 5.32 **Glyoxysome localization within plant seedlings.** Light micrograph of a section through cotyledons from imbibed cotton seeds. The glyoxysomes, which are seen as small dark structures (arrow), have been made visible by a cytochemical stain for the enzyme catalase. (FROM KENT D. CHAPMAN AND RICHARD N. TRELEASE, J. CELL BIOL. 115:998, 1991; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

Peroxisomes are also present in plants. Plant seedlings contain a specialized type of peroxisome, called a **glyoxysome** (Figure 5.32). Plant seedlings rely on stored fatty acids to provide the energy and material to form a new plant. One of the primary metabolic activities in these germinating seedlings is the conversion of stored fatty acids to carbohydrate. Disassembly of stored fatty acids generates acetyl CoA, which condenses with oxaloacetate (OAA) to form citrate, which is then converted into glucose by a series of enzymes of the glyoxylate cycle localized in the glyoxysome. The role of peroxisomes in the metabolism of leaf cells is discussed in Section 6.6.

REVIEW ?

1. What are some of the major activities of peroxisomes? What is the role of catalase in these activities?

2. In what ways are peroxisomes similar to mitochondria? In what ways are they unique?

THE HUMAN PERSPECTIVE

Diseases that Result from Abnormal Mitochondrial or Peroxisomal Function

Mitochondria

In 1962, Rolf Luft of the University of Stockholm in Sweden reported on a study of mitochondria isolated from a woman who had been suffering from long-standing fatigue and muscle weakness and who exhibited an elevated metabolic rate and body temperature. The mitochondria from this patient were somehow released from their normal respiratory control. Normally, when ADP levels are low, isolated mitochondria stop oxidizing substrate. In contrast, the mitochondria from this patient continued to oxidize substrate at a high rate, even in the absence of ADP to phosphorylate, producing heat rather than mechanical work. Since that initial report, a variety of disorders have been traced to mitochondrial dysfunction. Most of these disorders are characterized by degeneration of muscle or brain tissue, both of which utilize exceptionally large amounts of ATP. These conditions range in severity from diseases that lead to death during infancy; to disorders that produce seizures, blindness, deafness, and/or stroke-like episodes; to mild conditions characterized by intolerance to exercise or nonmotile sperm. Patients with serious conditions generally have abnormal skeletal muscle fibers, which contain large peripheral aggregates of mitochondria (Figure 1*a*). Closer examination of the mitochondria reveals large numbers of abnormal inclusions (Figure 1*b*).

More than 95 percent of the polypeptides of the respiratory chain are encoded by genes that reside in the nucleus. Yet the majority of mutations linked to mitochondrial-based diseases have been traced to mutations in mitochondrial DNA, or mtDNA. The most

serious disorders usually stem from mutations (or deletions) that affect genes encoding mitochondrial transfer RNAs, which are required for the synthesis of all 13 polypeptides produced in human mitochondria. Because all of the genes carried in mtDNA encode proteins required for oxidative phosphorylation, one would expect all types of mtDNA mutations to produce a similar clinical phenotype. Clearly, this is not the case.

The inheritance of mitochondrial disorders contrasts in several ways with the Mendelian inheritance of nuclear genes. The mitochondria present in the cells of a human embryo are derived exclusively from mitochondria that were present in the egg at the time of conception without any contribution from the fertilizing sperm. Consequently, mitochondrial disorders are inherited maternally. In addition, the mitochondria of a cell can contain a mixture of normal (i.e., wild-type) and mutant mtDNA, a condition known as *heteroplasmy*. The level of mutant mtDNA can vary from one organ to another within a single individual, and it is only when a preponderance of the mitochondria of a particular tissue contains defective genetic information that clinical symptoms usually appear. Because of this variability, members of a family carrying the same mtDNA mutation can exhibit strikingly different symptoms.

Because of its proximity to electron-transport processes, which lead to the release of mutagenic oxygen radicals (page 34), mtDNA is thought to come under a much greater level of attack than nuclear DNA. In addition, nuclear DNA is better protected from persistent damage by the presence of a greater variety of DNA repair systems

(a)

(b)
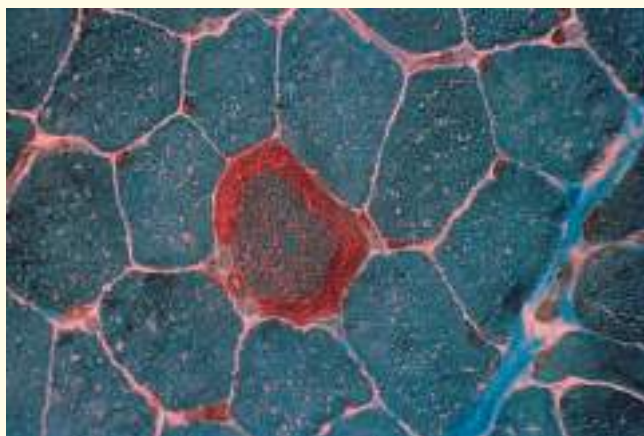
FIGURE 1 Mitochondrial abnormalities in skeletal muscle. (*a*) Ragged-red fibers. These degenerating muscle fibers are from a biopsy of a patient and show accumulations of red-staining "blotches" just beneath the cells' plasma membrane, which are due to abnormal proliferation of mitochondria. (*b*) Electron micrograph showing crystalline structures within the mitochondrial matrix from cells of a patient with abnormal mitochondria. (A: COURTESY OF DONALD R. JOHNS; B: FROM JOHN A. MORGAN-HUGHES AND D. N. LANDON, IN MYOLOGY, 2D ED., A. G. ENGEL AND C. FRANZINI-ARMSTRONG, EDS., MCGRAW-HILL, 1994.)
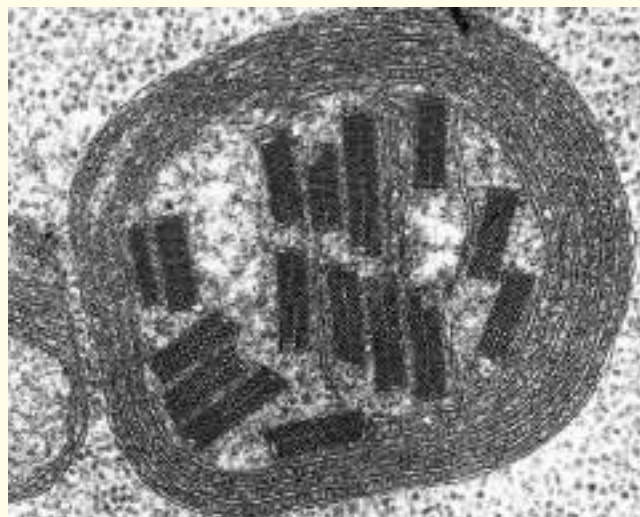
(Section 13.2). For these reasons, mtDNA experiences more than 10 times the mutation rate of nuclear DNA. Mitochondrial mutations are particularly likely to accumulate in cells that remain in the body for long periods of time, such as those of nerve and muscle tissue. The degenerative changes in mitochondrial function are thought to contribute to a number of common neurologic diseases with adult onset, most notably Parkinson's disease (PD). This possibility originally came to light in the early 1980s when a number of young drug addicts reported to hospitals with the sudden onset of severe muscular tremors that are characteristic of advanced Parkinson's disease in elderly adults. It was discovered that these individuals had intravenously injected themselves with a synthetic heroin that was contaminated with a compound called MPTP. Subsequent studies revealed that MPTP caused damage to complex I of the mitochondrial respiratory chain, leading to the death of nerve cells in the same region of the brain (the substantia nigra) that is affected in patients with PD. When cells from the substantia nigra of PD patients were then studied, they too were found to have a marked and selective decrease in complex I activity. This complex I deficiency is probably due to a documented increase in the mutation rate in the DNA of these cells compared to other cells. Studies have also implicated exposure to certain pesticides, particularly rotenone, a known complex I inhibitor, as an environmental risk factor for development of PD. Administration of rotenone to rats causes the destruction of dopamine-producing neurons, which is characteristic of the human disease. The possible link between rotenone and Parkinson's disease is being investigated.

It has long been speculated that the gradual accumulation of mutations in mtDNA is a major causal factor in human aging. This hypothesis is fueled by the fact that cells taken from older persons have an increased number of mtDNA mutations compared to the same cells from younger individuals. But are these mutations a cause of aging or simply a consequence of a more basic underlying aging process that affects many physiological properties? The experiments that may be the most relevant to this question have been performed on mice that are homozygous for a mutant gene (called *Polg*) that encodes the enzyme that replicates mtDNA. The mtDNA of these mice accumulates a much higher level of mutations than that of their normal littermates. These "mutator" mice appear normal for the first 6 to 9 months of age, but then rapidly develop signs of premature aging, such as hearing loss, graying hair, and osteoporosis (Figure 2). Whereas the control mice have a life span of more than two years, members of the experimental group live, on average, only about one year. Interestingly, the genetically modified animals show no evidence of increased oxidative (free radical) damage, so it is not clear

FIGURE 2 A premature-aging phenotype caused by increased mutations in mtDNA. The photograph shows a 13-month-old normal mouse and its "aged" littermate whose mitochondrial DNA harbors an abnormally high number of mutations. This premature-aging phenotype was caused by a mutation in the nuclear gene that encodes the DNA polymerase responsible for replication of mtDNA. (COURTESY OF JEFF MILLER, UNIVERSITY OF WISCONSIN-MADISON, PROVIDED BY G. C. KUJOTH.)

how the increased mtDNA mutation rate causes this phenotype. It is also noteworthy that mice that are heterozygous for the *Polg* mutation have an elevated mtDNA mutation rate (although far less than the homozygotes), yet they do not exhibit a shortened life span. If mtDNA mutations were a major contributor to the normal aging process, these heterozygotes should have a significantly shorter life. Taken together, these results suggest that mutations in mtDNA may be sufficient to cause an animal to age prematurely, but they are not necessarily required as part of the *normal* aging process. Whether these studies on mice bear directly on human aging remains an open question.

Peroxisomes

Zellweger syndrome (ZS) is a rare inherited disease characterized by a variety of neurologic, visual, and liver abnormalities leading to death during early infancy. In 1973, Sidney Goldfischer and his colleagues at Albert Einstein College of Medicine reported that liver and renal cells from these patients were lacking peroxisomes. Subsequent investigations revealed that peroxisomes were not entirely absent from the cells of these individuals, but, rather, they were present as empty membranous "ghosts"—that is, organelles lacking the enzymes that were normally found in peroxisomes. It is not that these individuals are unable to synthesize peroxisomal enzymes, but rather that the enzymes fail to be imported into the peroxisomes and remain largely in the cytosol where they are unable to carry out their normal functions. Genetic studies of cells from ZS patients showed that the disorder can arise from mutations in at least 12 different genes, all encoding proteins involved in the uptake of peroxisomal enzymes from the cytosol.

Unlike Zellweger syndrome, which affects a wide variety of peroxisomal functions, a number of inherited diseases are characterized by the absence of a single peroxisomal enzyme. One of these single-enzyme deficiency diseases is adrenoleukodystrophy (ALD), which was the subject of the 1993 movie *Lorenzo's Oil*. Boys with the disease are typically unaffected until mid-childhood, when symptoms of adrenal insufficiency and neurological dysfunction begin. The disease results from a defect in a membrane protein that transports very-long-chain fatty acids (VLCFAs) into the peroxisomes, where they are normally metabolized. In the absence of this protein, VLCFAs accumulate in the brain and destroy the myelin sheaths that insulate nerve cells. In *Lorenzo's Oil*, the parents of a boy stricken with ALD discover that a diet rich in certain fatty acids is able to retard the progress of the disease. Subsequent studies have generated contradictory results regarding the value of this diet. A number of ALD patients have been successfully treated by bone marrow transplantation, which provides normal cells capable of metabolizing VLCFAs, and by administration of drugs (e.g., lovastatin) that may lower VLCFA levels. Clinical studies employing gene therapy are also planned.

SYNOPSIS

Mitochondria are large organelles comprised of an outer porous membrane and a highly impermeable inner membrane that consists primarily of folds (cristae) that contain much of the machinery required for aerobic respiration. The porosity of the outer membrane results from integral proteins called porins. The architecture of the inner membrane and the apparent fluidity of its bilayer facilitate the interactions of components that are required during electron transport and ATP formation. The inner membrane surrounds a gel-like matrix that contains, in addition to proteins, a genetic system that includes DNA, RNA, ribosomes, and all the machinery necessary to transcribe and translate genetic information. Many of the properties of mitochondria can be accounted for by their presumed evolution from ancient symbiotic bacteria. (p. 174)

The mitochondrion is the center of oxidative metabolism in the cell, converting the products from carbohydrate, fat, and protein catabolism into chemical energy stored in ATP. Pyruvate and NADH are the two products of glycolysis. Pyruvate is transported across the inner mitochondrial membrane where it is decarboxylated and combined with coenzyme A to form acetyl CoA, which is condensed with oxaloacetate to form citrate, which is fed into the TCA cycle. As it traverses the reactions of the TCA cycle, two of the carbons from the citrate are removed and released as CO_2, which represents the most highly oxidized state of the carbon atom. Electrons removed from substrates are transferred to FAD and NAD^+ to form $FADH_2$ and NADH. Fatty acids are broken down into acetyl CoA, which is fed into the TCA cycle, and all 20 amino acids are broken down into either pyruvate, acetyl CoA, or TCA cycle intermediates. Thus, the TCA cycle is the pathway on which the cell's major catabolic pathways converge. (p. 177)

Electrons transferred from substrates to $FADH_2$ and NADH are passed along a chain of electron carriers to O_2, releasing energy that is used to generate an electrochemical gradient across the inner mitochondrial membrane. The controlled movement of protons back across the membrane through an ATP-synthesizing enzyme is used to drive the formation of ATP at the enzyme's catalytic site. Each pair of electrons from NADH releases sufficient energy to drive the formation of approximately three ATPs, whereas the energy released from a pair of electrons from $FADH_2$ accounts for the formation of approximately two ATPs. (p. 181)

The amount of energy released as an electron is transferred from a donor (reducing agent) to an acceptor (oxidizing agent) can be calculated from the difference in redox potential between the two couples. The standard redox potential of a couple is measured under standard conditions and compared to the H_2-H^+ couple. The standard redox potential of the NADH-NAD^+ couple is -0.32 V, reflecting the fact that NADH is a strong reducing agent, that is, one that readily transfers its electrons. The standard redox potential of the H_2O-O_2 couple is $+0.82$ V, reflecting the fact that O_2 is a strong oxidizing agent, that is, one that has a strong affinity for electrons. The difference between these two couples, which is equivalent to 1.14 V, provides a measure of the free energy released (52.6 kcal/mol) when a pair of electrons is passed from NADH along the entire electron-transport chain to O_2. (p. 182)

The electron-transport chain contains five different types of carriers: heme-containing cytochromes, flavin nucleotide-containing flavoproteins, iron-sulfur proteins, copper atoms, and quinones. Flavoproteins and quinones are capable of accepting and donating hydrogen atoms, while cytochromes, copper atoms, and iron-sulfur proteins are capable of accepting and donating only electrons. The carriers of the electron-transport chain are arranged in order of increasingly positive redox potential. The various carriers are organized into four large, multiprotein complexes. Cytochrome *c* and ubiquinone

are mobile carriers, shuttling electrons between the larger complexes. As pairs of electrons are passed through complexes I, III, and IV, protons are translocated from the matrix across the membrane into the intermembrane space. The translocation of protons by these electron-transporting complexes establishes the proton gradient in which energy is stored. The last of the complexes is cytochrome oxidase, which transfers electrons from cytochrome c to O_2, reducing it to form water, a step that also removes protons from the matrix, contributing to the proton gradient. (p. 185)

The translocation of protons creates a separation of charge across the membrane, in addition to a difference in proton concentration. Consequently, the proton gradient has two components—a voltage and a pH gradient—the magnitudes of which depend on the movement of other ions across the membrane. Together the two components constitute a proton-motive force (Δp). In mammalian mitochondria, approximately 80 percent of the free energy of Δp is represented by the voltage and 20 percent by the pH gradient. (p. 191)

The enzyme that catalyzes the formation of ATP is a large multiprotein complex called ATP synthase. ATP synthase contains two distinct parts: an F_1 head that projects into the matrix and includes the catalytic sites, and an F_0 base that is embedded in the lipid bilayer and forms a channel through which protons are conducted from the intermembrane space into the matrix. In the binding change hypothesis of ATP formation, which is now generally accepted, the controlled movement of protons through the F_0 portion of the enzyme induces the rotation of the γ subunit of the enzyme, which forms the stalk connecting the F_0 and F_1 portions of the enzyme. Rotation of the γ subunit is brought about by rotation of the c ring of the F_0 base, which is induced by the movement of protons through half-channels in the a subunit. The rotation of the γ subunit induces changes in the conformation of the F_1 catalytic sites, driving the formation of ATP. Evidence indicates that the energy-requiring step is not the actual phosphorylation of ADP, but the release of the ATP product from the active site, which occurs in response to the induced changes in conformation. In addition to the formation of ATP, the proton-motive force also provides the energy required for a number of transport activities, including the uptake of ADP into the mitochondrion in exchange for the release of ATP to the cytosol, the uptake of phosphate and calcium ions, and the import of mitochondrial proteins. (p. 192)

Peroxisomes are membrane-bound cytoplasmic vesicles that carry out a number of diverse metabolic reactions, including the oxidation of urate, glycolate, and amino acids, generating H_2O_2. (p. 200)

ANALYTIC QUESTIONS

1. Consider the reaction $A:H + B \rightleftharpoons B:H + A$. If, at equilibrium, the ratio of $[B:H]/[B]$ is equal to 2.0, one can conclude that (1) $B:H$ is the strongest reducing agent of the four compounds; (2) the $(A:H-A)$ couple has a more negative redox potential than the $(B:H-B)$ couple; (3) none of the four compounds are cytochromes; (4) the electrons associated with B are higher energy than those associated with A. Which of the previous statements are true? Draw one of the half-reactions for this redox reaction.

2. The membranous vesicle of a submitochondrial particle after removal of the F_1 spheres would be able to (1) oxidize NADH; (2) produce H_2O from O_2; (3) generate a proton gradient; (4) phosphorylate ADP. Which of the previous statements are true? How would this answer be different, if at all, if the objects being studied were intact submitochondrial particles that had been treated with dinitrophenol?

3. Protein A is a flavoprotein with a redox potential of -0.2 V. Protein B is a cytochrome with a redox potential of $+0.1$ V.

   a. Draw the half-reactions for each of these two electron carriers.

   b. Give the reaction that would occur if one were to mix reduced A molecules and oxidized B molecules.

   c. Which two compounds in the reaction of part b would be present at the highest concentration when the reaction reached equilibrium?

4. Of the following substances—ubiquinone, cytochrome c, NAD^+, NADH, O_2, H_2O—which is the strongest reducing agent? Which is the strongest oxidizing agent? Which has the greatest affinity for electrons?

5. If it were determined that the inner mitochondrial membrane was freely permeable to chloride ions, what effect might this have on the proton-motive force across the inner mitochondrial membrane?

6. Look at the drop in energy during electron transport shown in Figure 5.14. How would this profile be different if the original electron donor was $FADH_2$ rather than NADH?

7. Would you expect the import of P_i to lead to a decrease in the proton-motive force? Why?

8. In the chart of standard redox potentials in Table 5.1, the oxaloacetate-malate couple is less negative than the NAD^+-NADH couple. How are these values compatible with the transfer of electrons from malate to NAD^+ in the TCA cycle?

9. How many high-energy triphosphates are formed by substrate-level phosphorylation during each turn of the TCA cycle (consider the TCA cycle reactions only)? How many are formed as the result of oxidative phosphorylation? How many molecules of CO_2 are released? How many molecules of FAD are reduced? How many pairs of electrons are stripped from substrate?

10. The assembly of Fe-S clusters, which takes place in the mitochondrial matrix, is highly vulnerable to the presence of O_2. Given the importance of mitochondria in aerobic metabolism, does this seem to be an unlikely place for this process to occur?

11. The drop in $\Delta G^{\circ\prime}$ of a pair of electrons as they pass from NADH to O_2 is -52.6 kcal/mol, and the $\Delta G^{\circ\prime}$ of ATP formation is $+7.3$ kcal/mol. If three ATPs are formed for each pair of electrons removed from substrate, can you conclude that oxidative phosphorylation is only 21.9/52.6 or 42 percent efficient? Why or why not?

12. Would you expect metabolically active, isolated mitochondria to acidify or alkalize the medium in which they were suspended? Would your answer be different if you were working with submitochondrial particles rather than mitochondria? Why?

13. Suppose the movement of three protons were required for the synthesis of one ATP molecule. Calculate the energy released by the passage of three moles of protons into the matrix when the proton-motive force is measured at 220 mV (see p. 191 for information).

14. Would you expect isolated mitochondria to be able to carry out the oxidation of glucose with the accompanying production of ATP? Why or why not? What might be a good compound to add to a preparation of isolated mitochondria to achieve ATP synthesis?

15. Calculate the free energy released when $FADH_2$ is oxidized by molecular O_2 under standard conditions.

16. Using your understanding of the organization of the mitochondrion and how it generates ATP, answer the following questions. Give a brief explanation of each answer.

    a. What would be the effect on ATP production of adding a substance (protonophore) that makes the mitochondrial membrane leaky to protons?

    b. What would be the effect on the pH of the mitochondrial matrix of reducing the oxygen supply?

    c. Suppose that the glucose supply to the cell is greatly reduced. What would be the effect on CO_2 production by the mitochondria?

17. Suppose that you are able to manipulate the potential of the inner membrane of an isolated mitochondrion. You measure the pH of the mitochondrial matrix and find it to be 8.0. You measure the bathing solution and find its pH to be 7.0. You clamp the inner membrane potential at +59 mV—that is, you force the matrix to be 59 mV positive with respect to the bathing medium. Under these circumstances, can the mitochondrion use the proton gradient to drive the synthesis of ATP? Explain your answer.

18. Suppose you were able to synthesize an ATP synthase that was devoid of the γ subunit. How would the catalytic sites of the β subunits of such an enzyme compare to one another. Why?

19. In functional terms, the ATP synthase can be divided into two parts: a "stator" made up of subunits that do not move during catalysis, and a "rotor" made up of parts that do move. Which subunits of the enzyme make up each of these two parts. How are the immovable parts of the stator in F_0 related structurally to the immovable parts of the stator in F_1?

20. Cells contain a protein called IF that binds tightly to the catalytic site of the ATP synthase, inhibiting its activity under certain circumstances. Can you think of any circumstances where such an inhibitor might be useful within a cell?

21. Mitochondrial DNA is replicated by the enzyme DNA polymerase γ. The "mutator" mice discussed on page 202 had this phenotype because they possessed a polymerase γ that tended to make mistakes as it copied its mtDNA template. A couple of limitations in interpreting these studies in terms of the normal aging process were pointed out. What do you think could be learned by generating mice that had a super-accurate polymerase γ, that is, one that made fewer mistakes than the normal (wild-type) version?

6
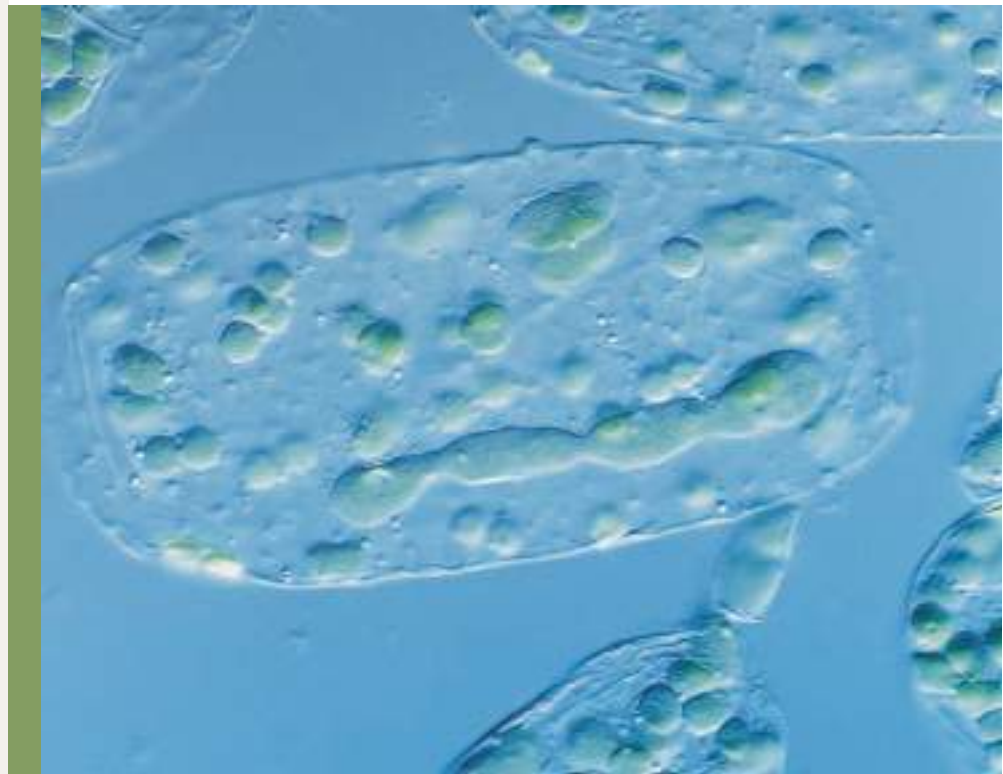
Photosynthesis and the Chloroplast

T he Earth's earliest life forms must have obtained their raw materials and energy from simple organic molecules dissolved in their aqueous environment. These organic molecules had to have formed *abiotically*, that is, as a result of nonbiological chemical reactions occurring in the primeval seas. Thus, just as we survive on nutrients taken in from our environment, so too must have the original life forms. Organisms that depend on an external source of organic compounds are called **heterotrophs**.

The number of heterotrophic organisms living on the primitive Earth must have been severely restricted because the spontaneous production of organic molecules occurs very slowly. The evolution of life on Earth received a tremendous boost with the appearance of organisms that employed a new metabolic strategy. Unlike their predecessors, these organisms could manufacture their own organic nutrients from the simplest types of inorganic molecules, such as carbon dioxide (CO_2) and hydrogen sulfide (H_2S). Organisms capable of surviving on CO_2 as their principal carbon source are called **autotrophs**.

Light micrograph of a living leaf epidermal cell from Arabidopsis thaliana. *The cell contains a number of chloroplasts—the organelles that house the plant's photosynthetic machinery. Chloroplasts normally divide by binary fission in which a single constriction splits the organelle into two equal daughters. This cell is from a mutant plant that is characterized by asymmetric fission. The highly elongated chloroplast has initiated fission asymmetrically at several sites, as indicated by the multiple constrictions. Mutants have proven invaluable in the identification of genes involved in all types of cellular processes. It is only when a gene malfunctions that its effects usually become visible, providing researchers with an idea of the gene's normal function—in this case a role in chloroplast fission.* (COURTESY OF KEVIN D. STOKES AND STANISLAV VITHA, MICHIGAN STATE UNIVERSITY.)

The manufacture of complex organic molecules from CO_2 requires the input of large amounts of energy. Over the course of evolution, two main types of autotrophs have evolved that can be distinguished by their source of energy. **Chemoautotrophs** utilize the chemical energy stored in inorganic molecules (such as ammonia, hydrogen sulfide, or nitrites) to convert CO_2 into organic compounds, while **photoautotrophs** utilize the radiant energy of the sun to achieve this result. Because all chemoautotrophs are prokaryotes, and their relative contribution to the formation of biomass on Earth is small, we will not consider their metabolic activities further. Photoautotrophs, on the other hand, are responsible for capturing the energy that fuels the activities of most organisms on Earth. Photoautotrophs include plants and eukaryotic algae, various flagellated protists, and members of several groups of prokaryotes. All of these organisms carry out **photosynthesis**, a process in which energy from sunlight is transformed into chemical energy that is stored in carbohydrates and other organic molecules.

During photosynthesis, relatively low-energy electrons are removed from a donor compound and converted into high-energy electrons using the energy absorbed from light.[1] These high-energy electrons are then employed in the synthesis of reduced biological molecules, such as starch and oils. It is likely that the first groups of photoautotrophs, which may have dominated Earth for two billion years, utilized hydrogen sulfide as their source of electrons for photosynthesis, carrying out the overall reaction

$$CO_2 + 2\,H_2S \xrightarrow{\text{light}} (CH_2O) + H_2O + 2\,S$$
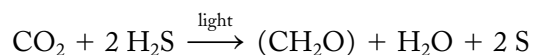
where (CH_2O) represents a unit of carbohydrate. There are numerous bacteria living today that carry out this type of photosynthesis; an example is shown in Figure 6.1. But, today, hydrogen sulfide is neither abundant nor widespread; consequently, organisms that depend on this compound as an electron source are restricted to habitats such as sulfur springs and deep-sea vents.

Approximately 2.7 to 2.4 billion years ago, a new type of photosynthetic prokaryote appeared on Earth that was able to utilize a much more abundant source of electrons, namely, water. Not only did the use of water allow these organisms—the cyanobacteria—to exploit a much more diverse array of habitats on Earth (see Figure 1.15), it produced a waste product of enormous consequence for all forms of life. The waste product was molecular oxygen (O_2), which is formed in the overall reaction

$$CO_2 + H_2O \xrightarrow{\text{light}} (CH_2O) + O_2$$
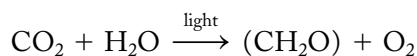
The switch from H_2S to H_2O as the substrate for photosynthesis is more difficult than switching one letter in the alphabet for another. The redox potential of the S-H_2S couple is -0.25 V as compared to $+0.816$ V for the O_2-H_2O couple (page 184). In other words, the sulfur atom in a molecule of H_2S has much less affinity for its electrons (and so is easier to oxidize) than does the oxygen atom in a molecule of H_2O.

---

[1]See the footnote on page 179 regarding use of the term *high-energy electrons*.
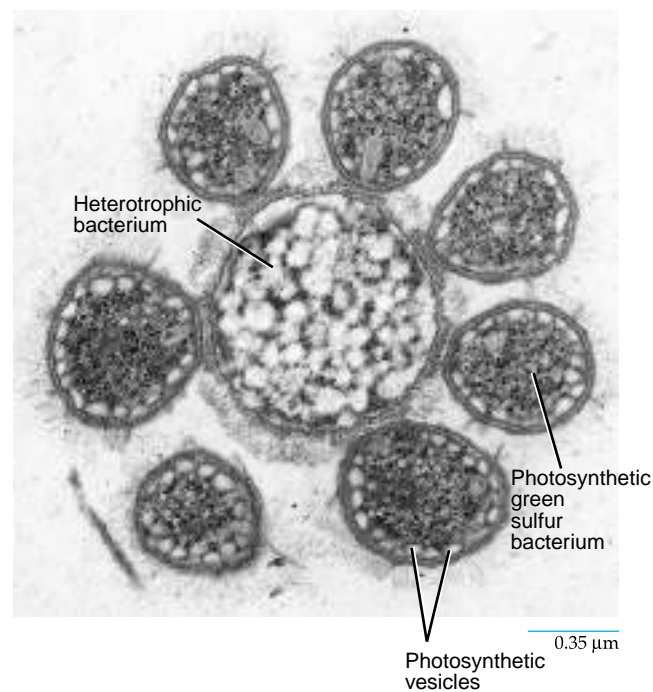
FIGURE 6.1 **Photosynthetic green sulfur bacteria** are present as a ring of peripheral cells that are living in a symbiotic relationship with a single anaerobic, heterotrophic bacterium in the center of the "colony." The heterotrophic bacterium receives organic matter produced by the photosynthetic symbionts. Photosynthetic vesicles containing the light-capturing machinery are visible in the green sulfur bacteria. (REPRINTED WITH PERMISSION FROM TOM FENCHEL, SCIENCE 296:1070, 2002, COURTESY OF KENNETH J. CLARKE; COPYRIGHT 2002, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

Thus, if an organism is going to carry out *oxygenic* (oxygen-releasing) photosynthesis, it has to generate a very strong oxidizing agent as part of its photosynthetic metabolism in order to pull tightly held electrons from water. The switch from H_2S (or other reduced substrates) to H_2O as the electron source for photosynthesis required an overhaul of the photosynthetic machinery.

At some point in time, one of these ancient, O_2-producing cyanobacteria took up residence inside a mitochondria-containing, nonphotosynthetic proeukaryotic cell. Over a long period of evolution, the symbiotic cyanobacterium was transformed from a separate organism living within a host cell into a cytoplasmic organelle, the **chloroplast**. As the chloroplast evolved, most of the genes that were originally present in the symbiotic cyanobacterium were either lost or transferred to the plant cell nucleus. As a result, the polypeptides found within modern-day chloroplasts are encoded by both the nuclear and chloroplast genomes. Extensive genetic analyses of chloroplast genomes suggest that all modern chloroplasts have arisen from a single, ancient symbiotic relationship. As a result of their common ancestry, chloroplasts and cyanobacteria share many basic characteristics, including a similar photosynthetic machinery, which will be discussed in detail in the following pages. ■

6.1 CHLOROPLAST STRUCTURE AND FUNCTION

Chloroplasts are located predominantly in the mesophyll cells of leaves. The structure of a leaf and the arrangement of chloroplasts around the central vacuole of a mesophyll cell are shown in Figure 6.2. Chloroplasts of higher plants are generally lens-shaped (Figure 6.3), approximately 2 to 4 μm wide and 5 to 10 μm long, and typically numbering 20 to 40 per cell. Their dimensions make chloroplasts giants among organelles—as large as an entire mammalian red blood cell. As illustrated in the chapter-opening micrograph, chloroplasts arise by fission from preexisting chloroplasts (or their nonpigmented precursors, which are called **proplastids**).

Chloroplasts were identified as the site of photosynthesis in 1881 in an ingenious experiment by the German biologist T. Engelmann. Engelmann illuminated cells of the green alga *Spirogyra* and found that actively moving bacteria would collect outside the cell near the site of the large ribbon-like chloroplast (see Figure 1.5). The bacteria were utilizing the minute quantities of oxygen released during photosynthesis in the chloroplast for their aerobic respiration.

The outer covering of a chloroplast consists of an envelope composed of two membranes separated by a narrow space (Figure 6.3). Like the outer membrane of a mitochondrion, the outer membrane of a chloroplast envelope contains several different porins (page 175). Although these proteins have relatively large channels (in the range of 1 nm), they exhibit some selectivity toward various solutes and thus may not be as freely permeable to key metabolites as is often described. The inner membrane of the envelope is highly impermeable; substances moving through this membrane do so only with the aid of a variety of transporters.

Much of the photosynthetic machinery of the chloroplast—including light-absorbing pigments, a complex chain of electron carriers, and an ATP-synthesizing apparatus—is part of an internal membrane system that is physically separate from the double-layered envelope. The internal membrane of the chloroplast, which contains the energy-transducing machinery, is organized into flattened membranous sacs, called **thylakoids**.

FIGURE 6.2 **The functional organization of a leaf.** The section of the leaf shows several layers of cells that contain chloroplasts distributed within their cytoplasm. These chloroplasts carry out photosynthesis, providing raw materials and chemical energy for the entire plant.
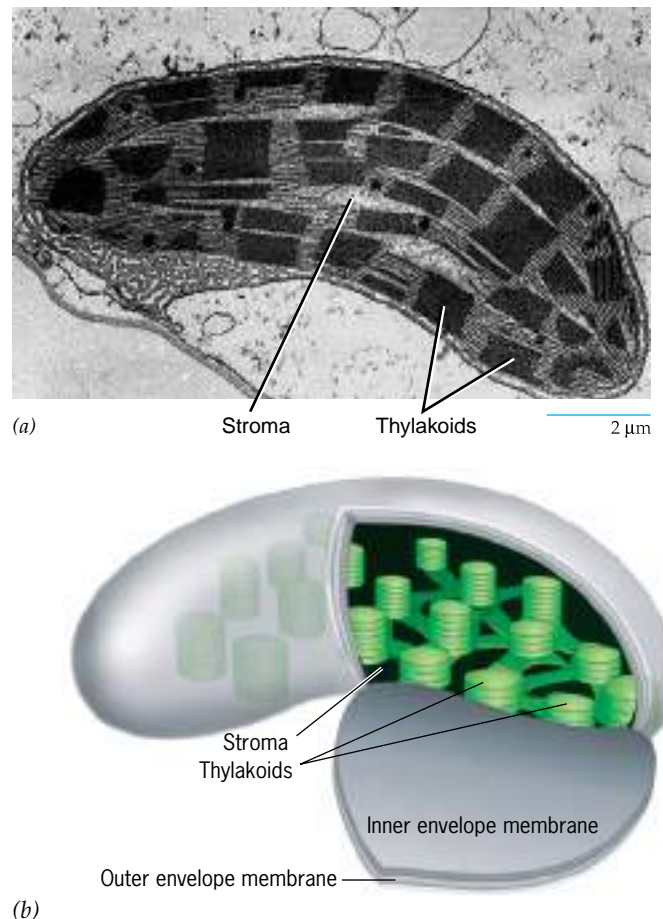
FIGURE 6.3 **The internal structure of a chloroplast.** (*a*) Transmission electron micrograph through a single chloroplast. The internal membrane is arranged in stacks (grana) of disk-like thylakoids that are physically separate from the outer double membrane that forms the envelope. (*b*) Schematic diagram of a chloroplast showing the outer double membrane and the thylakoid membranes. (A: COURTESY OF LESTER K. SHUMWAY.)

Grana thylakoid

Stroma thylakoid

0.3 µm
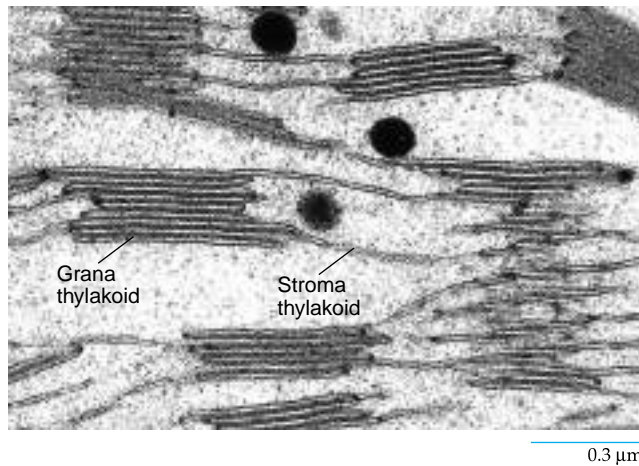
FIGURE 6.4 Thylakoid membranes. Electron micrograph of a section through a portion of a spinach chloroplast showing the stacked grana thylakoids, which are connected to one another by unstacked stroma thylakoids (or stroma lamellae). The dark spheres are osmium-stained lipid granules. (From L. A. Staehelin, in L. A. Staehelin, ed., Encyclopedia of Plant Physiology, vol. 19, p. 23, Springer-Verlag, 1986.)

Both fatty acids of these lipids contain several double bonds, which makes the lipid bilayer of thylakoid membranes highly fluid. The fluidity of the lipid bilayer facilitates lateral diffusion of protein complexes through the membrane during photosynthesis.

Thylakoids are arranged in orderly stacks called **grana** (Figures 6.3 and 6.4). The space inside a thylakoid sac is the *lumen*, and the space outside the thylakoid and within the chloroplast envelope is the **stroma**, which contains the enzymes responsible for carbohydrate synthesis.

Like the matrix of a mitochondrion, the stroma of a chloroplast contains small, double-stranded, circular DNA molecules and prokaryotic-like ribosomes. As discussed above, the DNA of the chloroplast is a relic of the genome of an ancient bacterial endosymbiont. Depending on the organism, chloroplast DNA contains between about 60 and 200 genes involved in either gene expression (e.g., tRNAs, rRNAs, ribosomal proteins) or photosynthesis. The vast majority of the estimated 2000–3500 polypeptides of a plant chloroplast are encoded by the DNA of the nucleus and synthesized in the cytosol. These proteins must be imported into the chloroplast by a specialized transport machinery (Section 8.9).

Thylakoid membranes have a high protein content and are unusual in having relatively little phospholipid. Instead, these membranes have a high percentage of galactose-containing glycolipids such as the following:

$CH_3-CH_2-CH=CH-CH_2-CH=CH-CH_2-CH=CH-(CH_2)_7-\overset{\overset{O}{\|}}{C}-OCH_2$

$CH_3-CH_2-CH=CH-CH_2-CH=CH-CH_2-CH=CH-(CH_2)_7-\overset{\overset{O}{\|}}{C}-OCH$
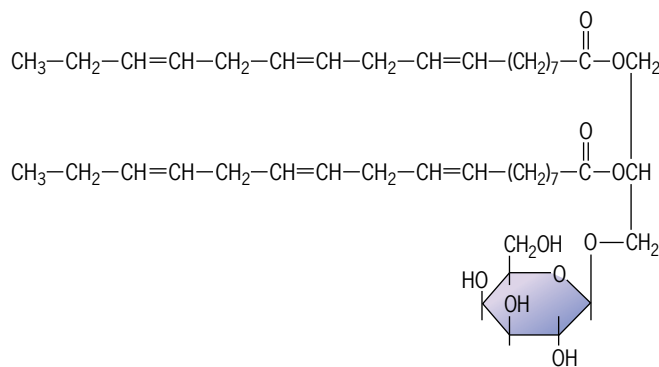
Monogalactosyl diacylglycerol

6.2 AN OVERVIEW OF PHOTOSYNTHETIC METABOLISM

A major advance in our understanding of the chemical reactions of photosynthesis came in the early 1930s with a proposal by C. B. van Niel, who was then a graduate student at Stanford University. Consider the overall equation of photosynthesis as given earlier:

$$CO_2 + H_2O \xrightarrow{light} (CH_2O) + O_2$$
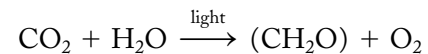
The prevailing belief in 1930 was that the energy of light was used to split CO_2, releasing molecular oxygen (O_2) and transferring the carbon atom to a molecule of water to form a unit of carbohydrate (CH_2O). In 1931, van Niel proposed an alternate mechanism based on his work with sulfur bacteria. It had been demonstrated that these organisms were able to reduce CO_2 to carbohydrate using the energy of light without the simultaneous production of molecular O_2. The proposed reaction for sulfur bacteria was

$$CO_2 + 2 H_2S \xrightarrow{light} (CH_2O) + H_2O + 2 S$$
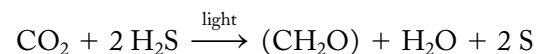
Postulating a basic similarity in the photosynthetic processes of all organisms, van Niel proposed a general reaction to include all these activities:

$$CO_2 + 2 H_2A \xrightarrow{light} (CH_2O) + H_2O + 2 A$$
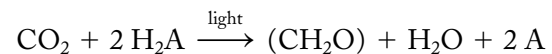
For the production of a hexose, such as glucose, the reaction would be

$$6 CO_2 + 12 H_2A \xrightarrow{light} C_6H_{12}O_6 + 6 H_2O + 12 A$$
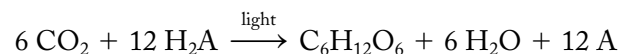
Van Niel recognized that photosynthesis was essentially a process of oxidation–reduction. In the preceding reaction, H_2A is an electron donor (reducing agent) and can be represented by H_2O, H_2S, or other reduced substrates utilized by various types of bacteria. Carbon dioxide, however, is an oxi-

dizing agent, which, in a plant cell, is reduced to form hexose in the following reaction:

$$6\ CO_2 + 12\ H_2O \xrightarrow{\text{light}} C_6H_{12}O_6 + 6\ H_2O + 6\ O_2$$
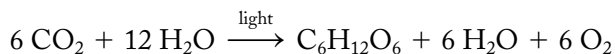
In this scheme, each molecule of oxygen is derived not from CO_2 but from the breakdown of two molecules of H_2O, a process driven by the absorption of light. The role of water in the formation of molecular oxygen was clearly established in 1941 by Samuel Ruben and Martin Kamen of the University of California, Berkeley. These investigators carried out experiments on suspensions of green algae using a specially labeled isotope of oxygen, ^{18}O, as a replacement for the common isotope, ^{16}O. One sample of algae was exposed to labeled $C[^{18}O_2]$ and unlabeled water, while another sample was exposed to unlabeled carbon dioxide and labeled $H_2[^{18}O]$. The investigators asked a simple question: Which of these two samples of photosynthetic organisms released labeled $^{18}O_2$? The algae given labeled water produced labeled oxygen, demonstrating that O_2 produced during photosynthesis derived from H_2O. The algae given labeled carbon dioxide produced unlabeled oxygen, confirming that O_2 was not being produced by a chemical splitting of CO_2. Contrary to popular belief, it wasn't carbon dioxide that was being split into its two atomic components, but water. Van Niel's hypothesis had been confirmed.

The van Niel proposal placed photosynthesis in a different perspective; it became, in essence, the reverse of mitochondrial respiration. Whereas respiration in mitochondria reduces oxygen to water, photosynthesis in chloroplasts oxidizes water to oxygen. The former process releases energy, so the latter process must require energy. An overview of the thermodynamics of photosynthesis and aerobic respiration is indicated in Figure 6.5. Many similarities between these two metabolic activities will be apparent in the following pages.

The events of photosynthesis can be divided into two series of reactions. During the first stage, the **light-dependent reactions**, energy from sunlight is absorbed and stored as chemical energy in two key biological molecules: ATP and NADPH. As discussed in Chapter 3, ATP is the cell's primary source of chemical energy, and NADPH is its primary source of reducing power. During the second stage, the **light-independent reactions** (or "dark reactions" as they are often called), carbohydrates are synthesized from carbon dioxide using the energy stored in the ATP and NADPH molecules produced in the light-dependent reactions. It is estimated that plant life on Earth converts approximately 500 trillion kg of CO_2 to carbohydrate each year, an amount roughly 10,000 times the world's yearly beef production.

We will begin with the light-dependent reactions, which are complex and remain incompletely understood.

REVIEW ?

1. In what ways is photosynthesis the reverse of respiration?

2. In what ways is nonoxygenic photosynthesis that uses H_2S as an electron source similar to oxygenic photosynthesis that uses H_2O as an electron source? In what ways are they different?

3. In general terms, how do the light-independent reactions differ from the light-dependent reactions? What are the primary products of the two types of reactions?
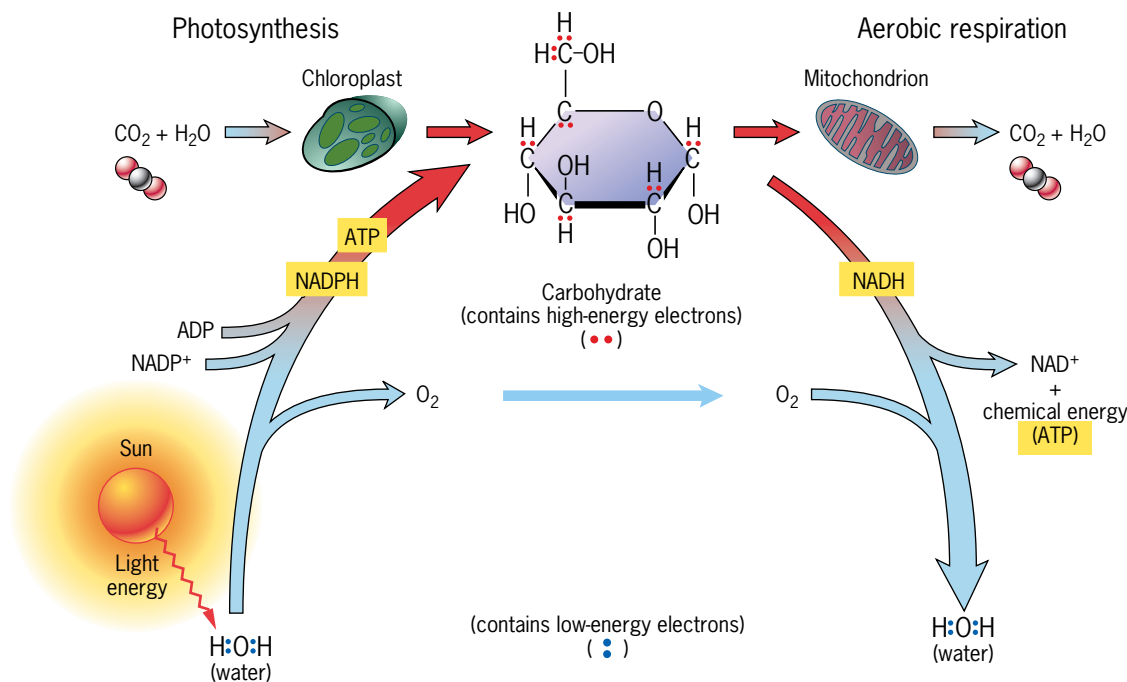
FIGURE 6.5 **An overview of the energetics of photosynthesis and aerobic respiration.**

6.3 THE ABSORPTION OF LIGHT

Light travels in packets (or quanta) of energy called **photons**, which can be thought of as "particles" of light.[2] The energy content of a photon depends on the wavelength of the light according to the equation

$$E = hc/\lambda$$

where h is Planck's constant (1.58×10^{-34} cal · sec), c is the speed of light in a vacuum, and λ is the wavelength of light. The shorter the wavelength, the higher the energy content. One mole (6.02×10^{23}) of photons of 680 nm wavelength, which is an important wavelength in photosynthesis, contains approximately 42 kcal of energy, which is equivalent to a change in redox potential of approximately 1.8 V (calculated by dividing 42 kcal by the Faraday constant of 23.06 kcal/V).

The absorption of light is the first step in any photochemical process. When a photon is absorbed by a molecule, an electron becomes sufficiently energetic to be pushed from an inner to an outer orbital. The molecule is said to have shifted from the **ground state** to an **excited state**. Because the number of orbitals in which an electron can exist is limited and each orbital has a specific energy level, it follows that any given atom or molecule can absorb only light of certain specific wavelengths.

The excited state of a molecule is unstable and can be expected to last only about 10^{-9} second. Several consequences can befall an excited electron, depending on the circumstances. Consider a molecule of **chlorophyll**, which is the most important light-absorbing photosynthetic pigment. If the electron of an excited chlorophyll molecule drops back to a lower orbital, the energy it had absorbed must be released. If the energy is released in the form of light (fluorescence or phosphorescence) or heat, the chlorophyll has returned to the original ground state and the energy of the absorbed photon has not been utilized. This is precisely what is observed when a preparation of *isolated chlorophyll* in solution is illuminated: the solution becomes strongly fluorescent because the absorbed energy is reemitted at a longer (i.e., lower energy) wavelength. However, if the same experiment is performed on a preparation of *isolated chloroplasts*, only a faint fluorescence is observed, indicating that very little of the absorbed energy is dissipated. Instead, the excited electrons of chlorophyll molecules are transferred to electron acceptors within the chloroplast membranes before they have a chance to drop back to lower energy orbitals. Thus, chloroplasts are able to harness the absorbed energy before it dissipates.

Photosynthetic Pigments

Pigments are compounds that appear colored because they only absorb light of particular wavelength(s) within the visible spectrum. Leaves are green because their chloroplasts contain large quantities of the pigment chlorophyll, which absorbs most strongly in the blue and red, leaving the intermediate green wavelengths to be reflected to our eyes. The structure of

chlorophyll is shown in Figure 6.6. Each molecule consists of two parts: (1) a porphyrin ring that functions in light absorption and (2) a hydrophobic phytol chain that keeps the chloro-
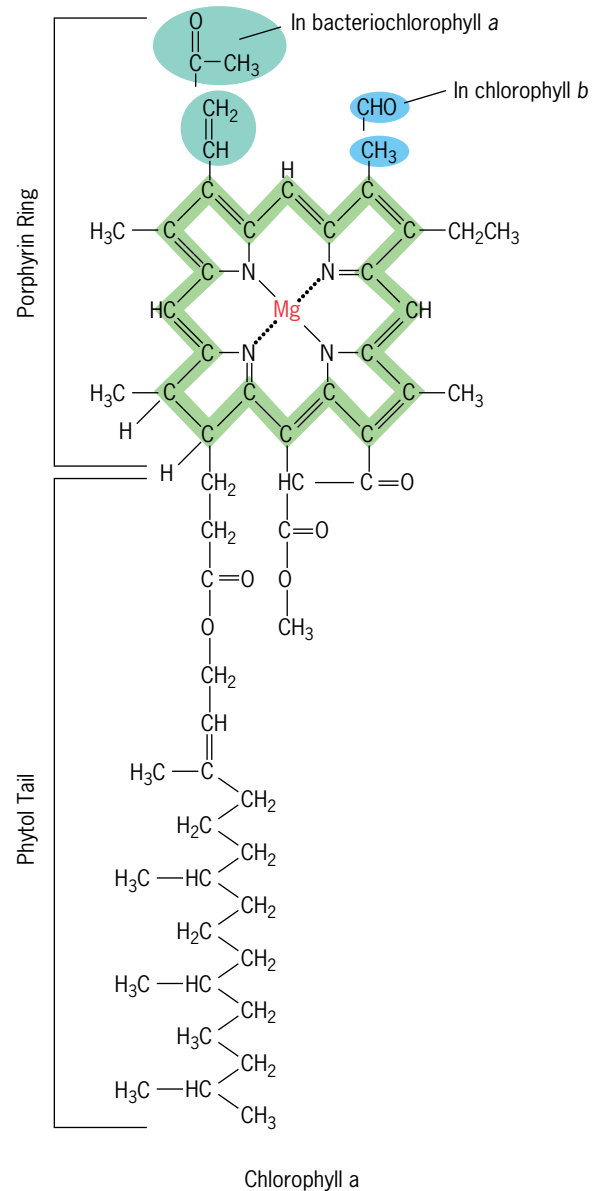
FIGURE 6.6 The structure of chlorophyll *a*. The molecule consists of a porphyrin ring (which in turn is constructed of four smaller pyrrole rings) with a magnesium ion at its center and a long hydrocarbon tail. The green shading around the edge of the porphyrin indicates the delocalization of electrons that form a cloud. The structure of the magnesium-containing porphyrin of chlorophyll can be compared to the iron-containing porphyrin of a heme shown in Figure 5.12. Chlorophyll *b* and bacteriochlorophyll *a* contain specific substitutions as indicated. For example, the —CH_3 group on ring II is replaced by a —CHO group in chlorophyll *b*. Chlorophyll *a* is present in all oxygen-producing photosynthetic organisms, but it is absent in the various sulfur bacteria. In addition to chlorophyll *a*, chlorophyll *b* is present in all higher plants and green algae. Others not shown are chlorophyll *c*, present in brown algae, diatoms, and certain protozoa, and chlorophyll *d*, found in red algae. Bacteriochlorophyll is found only in green and purple bacteria, organisms that do not produce O_2 during photosynthesis.

---

[2]The idea that electromagnetic radiation (e.g., light) has both wave-like and particle-like properties emerged at the start of the twentieth century from the work of Max Planck and Albert Einstein and marked the beginning of the study of quantum mechanics.

phyll embedded in the photosynthetic membrane. Unlike the red, iron-containing porphyrins (heme groups) of hemoglobin and myoglobin, the porphyrin in a chlorophyll molecule contains a magnesium atom. The alternating single and double bonds along the edge of the porphyrin ring delocalize electrons, which form a cloud around the ring (Figure 6.6). Systems of this type are said to be *conjugated* and are strong absorbers of visible light. The absorbed energy causes a redistribution of the electron density of the molecule, which in turn favors loss of an electron to a suitable acceptor. The conjugated bond system also broadens the absorption peaks, enabling individual molecules to absorb energy of a range of wavelengths. These features are evident in an absorption spectrum of purified chlorophyll molecules (Figure 6.7). An **absorption spectrum** is a plot of the intensity of light absorbed relative to its wavelength. The range of wavelengths absorbed by photosynthetic pigments situated within the thylakoids is further increased because the pigments are noncovalently associated with a variety of different polypeptides.

Several classes of chlorophyll, differing from one another in the side groups attached to the porphyrin ring, occur among photosynthetic organisms. The structures of these pigments are indicated in Figure 6.6. Chlorophylls are the primary light-absorbing photosynthetic pigments, but terrestrial plants also contain orange and red accessory pigments called *carotenoids*, including β-carotene, which contain a linear system of conjugated double bonds:
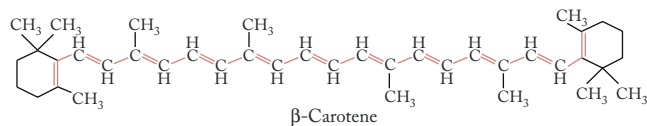
β-Carotene

Carotenoids absorb light primarily in the blue and green region of the spectrum (Figure 6.7), while reflecting those of the yellow, orange, and red regions. Carotenoids produce the

characteristic colors of carrots and oranges, and the leaves of some plants during the fall. Carotenoids have multiple functions: they act as secondary light collectors during photosynthesis, and they draw excess energy away from excited chlorophyll molecules and dissipate it as heat. If this excess energy were not absorbed by carotenoids, it could be transferred to oxygen, producing an ultrareactive form of the molecule called singlet oxygen ($^1O^*$) that can destroy biological molecules and cause cell death.

Because the light falling on a leaf is composed of a variety of wavelengths, the presence of pigments with varying absorption properties ensures that a greater percentage of incoming photons will stimulate photosynthesis. This can be seen by examining an **action spectrum** (Figure 6.8), which is a plot of the relative rate (or efficiency) of photosynthesis produced by light of various wavelengths. Unlike an absorption spectrum, which simply measures the wavelengths of light that are absorbed by particular pigments, an action spectrum identifies the wavelengths that are effective in bringing about a given physiologic response. The action spectrum for photosynthesis follows the absorption spectrum of chlorophylls and carotenoids fairly closely, reflecting the participation of these pigments in the photosynthetic process.

REVIEW ?

1. What is the difference between an absorption spectrum and an action spectrum?
2. Compare the structure, absorption, and function of chlorophylls and carotenoids.

FIGURE 6.7 Absorption spectrum for several photosynthetic pigments of higher plants. The background shows the colors that we perceive for the wavelengths of the visible spectrum. Chlorophylls absorb most strongly in the violet, blue, and red regions of the spectrum, while carotenoids (e.g., β-carotene) also absorb into the green region. Red algae and cyanobacteria contain additional pigments (phycobilins) that absorb in the middle bands of the spectrum.

FIGURE 6.8 Action spectrum for photosynthesis. The action spectrum (represented by the red-colored line) indicates the relative efficiency with which light of various wavelengths is able to promote photosynthesis in the leaves of a plant. An action spectrum can be generated by measuring the O_2 produced by the leaves following exposure to various wavelengths. The black lines indicate the absorption spectra of each of the major photosynthetic pigments. The green line shows the combined absorption spectrum of all pigments.

6.4 PHOTOSYNTHETIC UNITS AND REACTION CENTERS

In 1932, Robert Emerson and William Arnold of the California Institute of Technology carried out an experiment suggesting that not all of the chlorophyll molecules in a chloroplast were directly involved in conversion of light energy into chemical energy. Using suspensions of the green alga *Chlorella* and flashing lights of extremely short duration at saturating intensity, they determined the minimum amount of light needed to produce maximal O_2 production during photosynthesis. Based on the number of chlorophyll molecules present in the preparation, they calculated that one molecule of oxygen was being released during a brief flash of light for every 2500 molecules of chlorophyll present. Later, Emerson showed that a minimum of eight photons must be absorbed to produce one molecule of O_2, which meant that chloroplasts contain approximately 300 times as many chlorophyll molecules as would appear to be necessary to oxidize water and generate O_2.

One possible interpretation of this finding is that only a very small percentage of chlorophyll molecules is involved in photosynthesis. This is not the case, however. Rather, several hundred chlorophyll molecules act together as one **photosynthetic unit** in which only one member of the group—**the reaction-center chlorophyll**—actually transfers electrons to an electron acceptor. Even though the bulk of the pigment molecules do not participate *directly* in the conversion of light energy into chemical energy, they are responsible for light ab-

sorption. These pigment molecules form a light-harvesting **antenna** that absorbs photons of varying wavelength and transfers that energy (called *excitation energy*) very rapidly to the pigment molecule at the reaction center.

The transfer of excitation energy from one pigment molecule to another is very sensitive to the distance between the molecules. The chlorophyll molecules of an antenna are kept in close proximity (less than 1.5 nm separation) and proper orientation by noncovalent linkage to integral membrane polypeptides. A "rule" that operates among antenna pigments is that energy can only be transferred to an equal or less energy-requiring molecule. In other words, energy can only be passed to a pigment molecule that absorbs light of equal or longer wavelength (lower energy) than that absorbed by the donor molecule. As the energy "wanders" through a photosynthetic unit (Figure 6.9), it is repeatedly transferred to a pigment molecule that absorbs at a longer wavelength. The energy is ultimately transferred to a chlorophyll of the reaction center, which absorbs light of longer wavelength than any of its neighbors. Once the energy is received by the reaction center, the electron excited by light absorption can be transferred to its acceptor.

Oxygen Formation: Coordinating the Action of Two Different Photosynthetic Systems

The evolution of organisms capable of using H_2O as an electron source was accompanied by major changes in the photosynthetic machinery. The reason for these changes is revealed by considering the energetics of oxygenic (O_2-releasing) photosynthesis. The O_2-H_2O couple has a standard redox potential of $+0.82$ V, whereas that of the $NADP^+$-$NADPH$ couple is -0.32 V (Table 5.1, page 184). The difference between the redox potentials of these two couples (1.14 V) provides a measure of the minimal energy that must be absorbed by the system to remove an electron from H_2O and pass it to $NADP^+$ *under standard conditions*. However, cells don't operate under standard conditions, and the transfer of electrons from H_2O to $NADP^+$ requires more than the minimal input of energy. It is estimated that, during actual operations in the chloroplast, more than 2 V of energy are utilized to carry out this oxidation–reduction reaction. (This value of 2 V is estimated from the left-hand scale of Figure 6.10, which extends from below $+1$ V to above -1 V.) It was noted on page 211 that one mole of photons of 680 nm wavelength (red light) is equivalent to a change in redox potential of 1.8 V. Thus, while it is theoretically possible for one photon of red light to boost an electron to an energy level required to reduce $NADP^+$ under standard conditions (i.e., 1.14 V), the process is accomplished in the cell through the combined action of two different light-absorbing reactions.

The light-absorbing reactions of photosynthesis occur in large pigment–protein complexes called **photosystems**. Two types of photosystems are required to catalyze the two light-absorbing reactions utilized in oxygenic photosynthesis. One photosystem, **photosystem II (PSII)**, boosts electrons from an energy level below that of water to a midway point (Fig-

Antenna pigment molecules

Photon
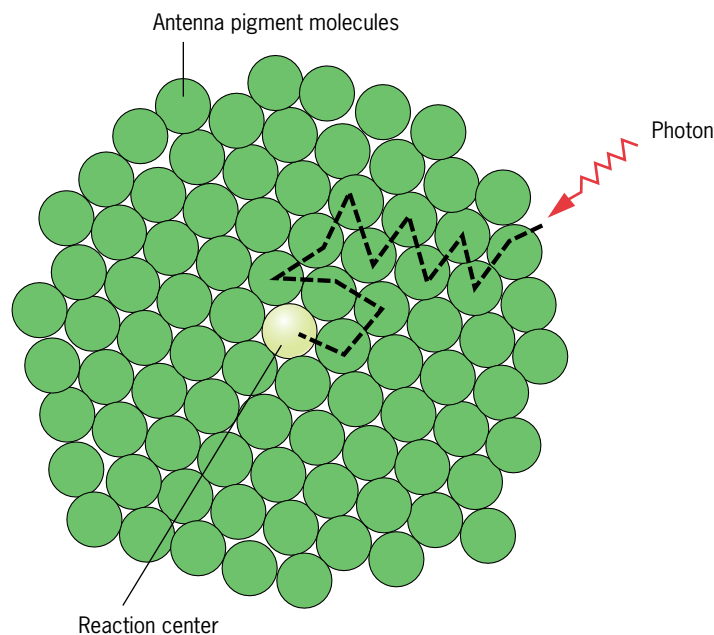
Reaction center

FIGURE 6.9 The transfer of excitation energy. Energy is transferred randomly through a network of pigment molecules that absorb light of increasingly longer wavelength until the energy reaches a reaction-center chlorophyll, which transfers an excited electron to a primary acceptor, as described later in the chapter.
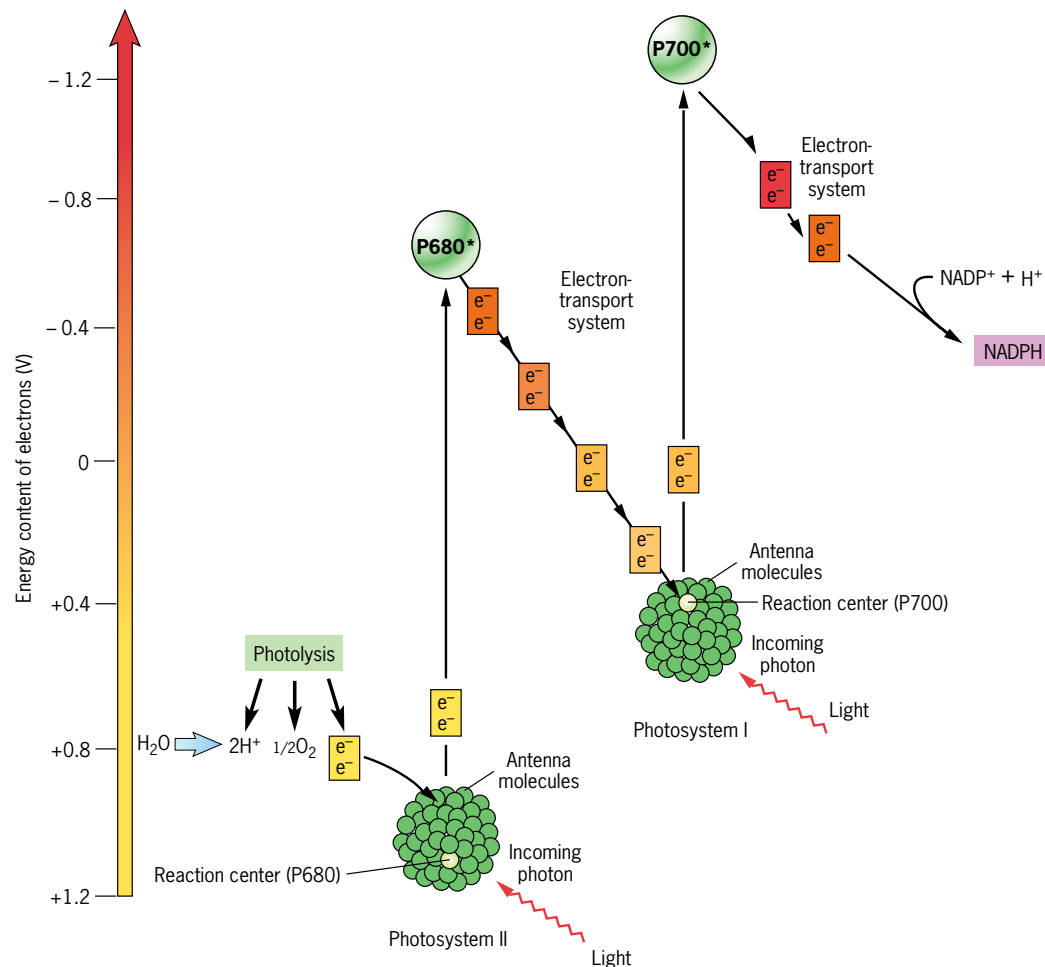
FIGURE 6.10 An overview of the flow of electrons during the light-dependent reactions of photosynthesis. The events depicted in this schematic drawing are described in detail in the following pages. The energy content of electrons is given in volts. To convert these values to calories, multiply by the Faraday constant, 23.06 kcal/V. For example, a difference of 2.0 V corresponds to an energy difference of 46 kcal/mol of electrons. This can be compared to the energy of red light (680 nm), which contains about 42 kcal/mol of photons.

ure 6.10). The other photosystem, **photosystem I (PSI)**, raises electrons from a midway point to an energy level well above that of $NADP^+$. The two photosystems act in series, that is, one after the other. Even though they mediate distinctly different photochemical reactions, the two types of photosystems in plants, as well as those of photosynthetic bacterial cells, exhibit marked similarities in protein composition and overall architecture. These shared properties suggest that all photosynthetic reaction centers have evolved from a common ancestral structure that has been conserved for more than three billion years.

The reaction center of photosystem II is a chlorophyll dimer referred to as **P680**, "P" standing for "pigment" and "680" standing for the wavelength of light that this particular pair of chlorophylls absorbs most strongly. The reaction center of photosystem I is also a chlorophyll dimer and is referred to as **P700** for comparable reasons. When sunlight strikes a thylakoid membrane, energy is absorbed by antenna pigments of both PSII and PSI and passed to the reaction centers of both photosystems. Electrons of both reaction-center pigments are boosted to an outer orbital, and each electron is transferred to a **primary electron acceptor**. After losing their electrons, the reaction-center chlorophylls of

PSII and PSI become positively charged pigments referred to as $P680^+$ and $P700^+$, respectively. The electron acceptors, in turn, become negatively charged. In essence, this separation of charge within the photosystems *is* the light reaction—the conversion of light energy into chemical energy. Positively charged reaction centers act as attractants for electrons, and negatively charged acceptors act as electron donors. Consequently, the separation of charge within each photosystem sets the stage for the flow of electrons along a chain of specific carriers.

In oxygenic photosynthesis, where the two photosystems act in series, electron flow occurs along three legs—from water to PSII, from PSII to PSI, and from PSI to $NADP^+$—an arrangement described as the *Z scheme*, which was first proposed by Robert Hill and Fay Bendall of the University of Cambridge. The broad outline of the Z scheme is illustrated in Figure 6.10; we will fill in the names of some of the specific components as we examine each of the major parts of the pathway. Like the members of the respiratory chain of mitochondria (Chapter 5), most of the electron carriers of the Z scheme are found as part of large membrane protein complexes (see Figure 6.16). Over the years, researchers have generated increas-

ingly detailed portraits of the structures of these complexes. These efforts have culminated in the past few years with the publication of X-ray crystallographic structures of both PSI and PSII by several laboratories at increasingly higher resolution.

As in mitochondria, electron transfer releases energy, which is used to establish a proton gradient, which in turn drives the synthesis of ATP. As discussed on page 223, ATP produced in the chloroplast is used primarily within the organelle in the synthesis of carbohydrates; ATP utilized outside the chloroplast is derived largely from that produced in plant cell mitochondria.

PSII Operations: Obtaining Electrons by Splitting Water
Photosystem II uses absorbed light energy for two interrelated activities: removing electrons from water and generating a proton gradient. The PSII of plant cells is a complex of more than 20 different polypeptides, most of which are embedded in the thylakoid membrane. Two of these proteins, designated D1 and D2, are particularly important because together they bind the P680 reaction-center chlorophyll dimer and all of the cofactors involved in electron transport through the photosystem (Figure 6.11).

The first step in PSII activation is the absorption of light by an antenna pigment. Most of the antenna pigments that collect solar energy for PSII reside within a separate pigment–protein complex, called **light-harvesting complex II** or, simply, **LHCII**. LHCII proteins bind both chlorophylls and carotenoids and are situated outside the core of the photosystem (Figure 6.11*a*). As discussed in the Experimental Pathways, which can be accessed on the Web at www.wiley.com/college/karp, LHCII is not always associated with PSII but, under appropriate conditions, can migrate along the thylakoid membrane and become associated with PSI, serving as a light-harvesting complex for the PSI reaction center as well.

The Flow of Electrons from PSII to Plastoquinone Excitation energy is passed from the outer-antenna pigments of LHCII

(a)   (b)
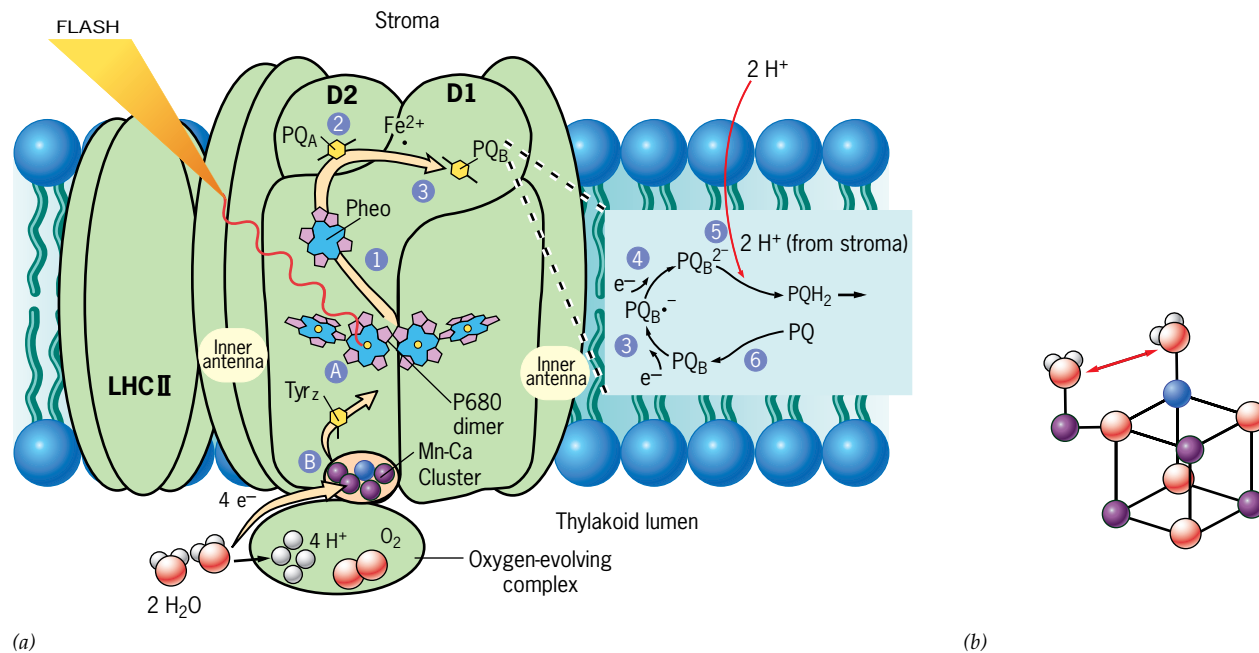
FIGURE 6.11 The functional organization of photosystem II. (*a*) A simplified schematic model of the huge protein–pigment complex, which catalyzes the light-driven oxidation of water and reduction of plastoquinone. The path taken by electrons is indicated by the yellow arrows. Events begin with the absorption of light by an antenna pigment in the outer light-harvesting complex (LHCII). Energy is transferred from LHCII through an inner antenna pigment–protein complex to a P680 reaction-center chlorophyll *a*, which is one of four closely spaced chlorophyll *a* molecules (the P680 dimer and two accessory chlorophyll *a* molecules). Absorption of this energy by P680 excites an electron, which is transferred to pheophytin (Pheo) (step 1), the primary electron acceptor of PSII. (Pheophytin is a chlorophyll molecule that lacks the Mg^{2+} ion.) The electron subsequently passes to a plastoquinone PQ_A (step 2) and then through a nonheme Fe^{2+} to PQ_B (step 3) to form a negatively charged free radical $PQ_B \cdot^-$. Absorption of a second photon sends a second electron along the same pathway, converting the acceptor to PQ_B^{2-} (step 4). Two protons then enter from the stroma (step 5) generating PQH_2, which is released into

the lipid bilayer and replaced by a new oxidized PQ_B molecule (step 6). As the above events are occurring, electrons are moving from H_2O by way of Tyr_z to the positively charged reaction-center pigment (steps B and A). Thus, overall, PSII catalyzes the transfer of electrons from water to plastoquinone. The oxidation of two molecules of H_2O to release one molecule of O_2 generates two molecules of PQH_2. Because the oxidation of water releases protons into the thylakoid lumen and the reduction of PQ_B^{2-} removes protons from the stroma, the operation of PSII makes a major contribution to formation of a H^+ gradient. The figure shows one monomer of a dimeric PSII complex. (*b*) A proposed organization of metal ions at the water-oxidation site based on spectroscopic and X-ray crystallographic data. In this model, three Mn ions and a Ca ion (and four oxygen atoms) are arranged as part of a cube-like cluster with a bridge to a fourth Mn ion at a nearby site. One of the substrate water molecules is bound to the fourth Mn ion, and the other substrate water molecule is bound to the Ca ion. Reaction between the oxygen atoms of the two water molecules (red arrow) leads to formation of an O=O bond. Mn, brown; Ca, blue; O, red.

to a small number of inner-antenna chlorophyll molecules situated within the core of PSII. From there, the energy is ultimately passed to the PSII reaction center. The excited reaction-center pigment (P680*) responds by transferring a single photoexcited electron to a closely associated chlorophyll-like pheophytin molecule (step 1, Figure 6.11a), which is the primary electron acceptor. This electron transfer generates a separation of charge in PSII between a positively charged donor (P680$^+$) and a negatively charged acceptor (Pheo$^-$). The importance of the formation of two oppositely charged species, P680$^+$ and Pheo$^-$, becomes more apparent when we consider the oxidation–reduction capabilities of these two species. P680$^+$ is electron-deficient and can accept electrons, making it an oxidizing agent. In contrast, Pheo$^-$ has an extra electron that it will readily lose, making it a reducing agent. This event—the light-driven formation of an oxidizing agent and a reducing agent—takes less than one billionth of a second and is the essential first step in photosynthesis.

Because they are oppositely charged, P680$^+$ and Pheo$^-$ exhibit an obvious reactivity with one another. Interaction between the oppositely charged species is prevented by moving the separated charges farther apart, ultimately to opposite sides of the membrane, by passage through several different sites. Pheo$^-$ first transfers its electron (step 2, Figure 6.11a) to a molecule of plastoquinone (designated PQ$_A$ in Figure 6.11a) bound near the outer (stromal) side of the membrane. Plastoquinone (PQ) is a lipid-soluble molecule (Figure 6.12) similar in structure to ubiquinone (see Figure 5.12c). The electron from PQ$_A$ is transferred (step 3, Figure 6.11a) to a second plastoquinone (designated PQ$_B$ in Figure 6.11) to produce a semireduced form of the molecule (PQ$_B$•$^-$) that remains firmly bound to the D1 protein of the reaction center. With each of these transfers, the electron moves closer to the stromal side of the membrane.

The positively charged pigment (P680$^+$) is reduced back to P680 (as described below), which primes the reaction center for the absorption of another photon. The absorption of a second photon sends a second energized electron along the path from P680 to pheophytin to PQ$_A$ to (PQ$_B$•$^-$), forming PQ$_B^{2-}$ (step 4, Figure 6.11), which combines with two protons to form plastoquinol, PQH$_2$ (step 5, Figure 6.11a; Figure 6.12). The protons utilized in the formation of PQH$_2$ are derived from the stroma, causing a decrease in H$^+$ concentration of the stroma, which contributes to formation of the proton gradient. The reduced PQH$_2$ molecule dissociates from the D1 protein and diffuses into the lipid bilayer. The displaced PQH$_2$ is replaced by a fully oxidized PQ molecule derived from a small "pool" of plastoquinone molecules in the bilayer (step 6, Figure 6.11a). We can summarize this portion of the PSII reaction with the equation

$$2e^- + PQ + 2\,H^+_{stroma} \xrightarrow{\text{2 photons}} PQH_2$$
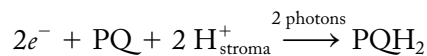
As we will see shortly, the oxidation of a molecule of water by PSII requires four photons, so we can better describe this portion of the PSII reaction with the equation

$$4\,e^- + 2\,PQ + 4\,H^+_{stroma} \xrightarrow{\text{4 photons}} 2\,PQH_2$$
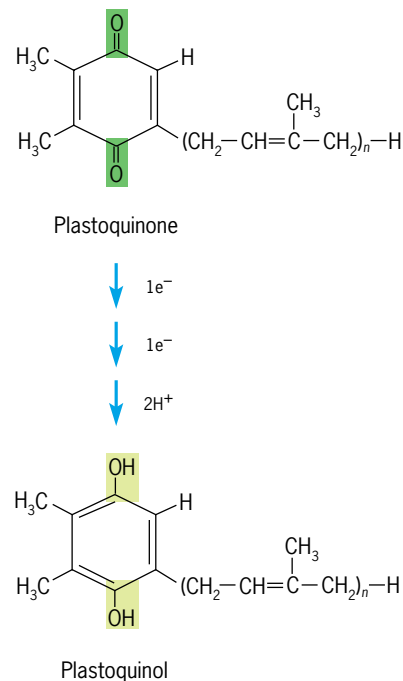
FIGURE 6.12 **Plastoquinone.** The acceptance of two electrons and two protons reduces PQ (plastoquinone) to PQH$_2$ (plastoquinol). The intermediates are similar to those shown in Figure 5.12c for ubiquinone of the mitochondrion.

We will follow the fate of the electrons (and protons) carried by PQH$_2$ in a following section.

The Flow of Electrons from Water to PSII   The least understood part of the flow of electrons from water to NADP$^+$ is the first step between H$_2$O and PSII. Water is a very stable molecule made up of tightly held hydrogen and oxygen atoms. In fact, the splitting of water is the most thermodynamically challenging (endergonic) reaction known to occur in living organisms. Splitting water in a laboratory requires the use of a strong electric current or temperatures approaching 2000°C. Yet a plant cell can accomplish this feat on a snowy mountainside using only the energy of visible light.

We saw in the previous section how the absorption of light by PSII leads to the formation of two charged molecules, P680$^+$ and Pheo$^-$. We have followed the route taken by the excited electron associated with Pheo$^-$; now we turn to the other species, P680$^+$, which is the most powerful oxidizing agent yet to be discovered in a biological system. The redox potential of the oxidized form of P680 is sufficiently strong to pull tightly held (low-energy) electrons from water (redox potential of +0.82 V), thus splitting the molecule. The splitting of water during photosynthesis is called **photolysis**. The formation of one molecule of oxygen during photolysis is thought to require the *simultaneous* loss of four electrons from two molecules of water according to the reaction

$$2\,H_2O \xrightarrow{\text{4 photons}} 4\,H^+_{lumen} + O_2 + 4\,e^-$$
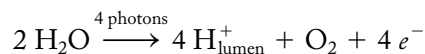
Yet one PSII reaction center can only generate one positive

charge ($P680^+$), or oxidizing equivalent, at a time. A solution to this problem was proposed around 1970 by Pierre Joliot and Bessel Kok as the S-state hypothesis, which allows the photosystem to accumulate the four oxidizing equivalents needed to oxidize water. Closely associated with the D1 protein of PSII at its luminal surface is a cluster of five metal ions—four manganese (Mn) ions and one calcium (Ca) ion—that is stabilized and protected by a number of peripheral proteins that form the *oxygen-evolving complex* (Figure 6.11*a*). A proposed organization of the Mn-Ca cluster is shown in Figure 6.11*b* and discussed in the accompanying legend. The Mn-Ca cluster accumulates four oxidizing equivalents by transferring four electrons, one at a time, to the nearby $P680^+$. The transfer of each electron from the Mn-Ca cluster to $P680^+$ (steps B and A of Figure 6.11) is accomplished by passage through an intermediate electron carrier, a tyrosine residue on the D1 protein, termed Tyr_z. After each electron is transferred to $P680^+$, regenerating P680, the pigment becomes reoxidized (back to $P680^+$) following the absorption of another photon by the photosystem. Thus, the stepwise accumulation of four oxidizing equivalents by the Mn-Ca cluster is driven by the successive absorption of four photons of light by the PSII photosystem. Once this has occurred, the system can now catalyze the removal of $4\ e^-$ from two closely bound H_2O molecules (Figure 6.11*b*), as indicated in the following manner:

$$4\,H^+ + O_2\ 2\,H_2O$$
$$\overset{\displaystyle \longleftarrow 4\,e^-\ \longleftarrow}{}$$
$$S_0 \overset{h\nu}{\longrightarrow} S_1 \overset{h\nu}{\longrightarrow} S_2 \overset{h\nu}{\longrightarrow} S_3 \overset{h\nu}{\longrightarrow} S_4$$
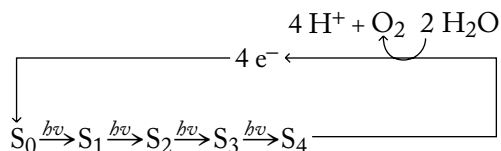
where the subscript on the S indicates the number of oxidizing equivalents stored by the Mn-Ca cluster. Evidence for the accumulation of successive oxidizing equivalents was first obtained by exposure of algal cells to very brief (1 μsec) flashes of light (Figure 6.13). It can be seen in this plot that O_2 production peaks after every fourth flash of light, indicating that the effect of four individual photoreactions must accumulate before O_2 can be released.

The protons produced in the photolysis reaction are retained in the thylakoid lumen (Figure 6.11), where they contribute to the proton gradient. The four electrons produced in the photolysis reaction serve to regenerate the fully reduced Mn-Ca cluster (S_0 state), while the O_2 is released as a waste product into the environment.

Before leaving the subject of PSII, it can be noted that the activity and integrity of this photosystem can be negatively affected by high light intensity. This phenomenon is termed *photoinhibition*. The formation of a very strong oxidizing agent, and the ever-present danger of forming highly toxic oxygen species, gives PSII the potential for its own self-destruction as the result of overexcitation of the system. Most of the damage appears to be targeted to the polypeptide (D1) that binds the active redox centers and Mn-Ca cluster of the photosystem. Chloroplasts contain a mechanism for the selective proteolytic degradation of D1 and its replacement by a newly synthesized polypeptide molecule.
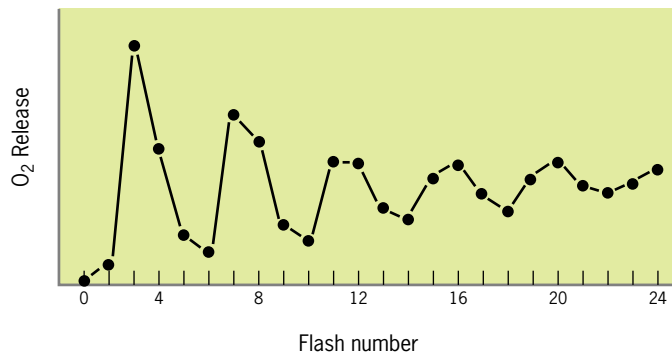
FIGURE 6.13 **Measuring the kinetics of O_2 release.** The plot shows the response by isolated chloroplasts that have been kept in the dark to a succession of light flashes of very short duration. The amount of oxygen released peaks with every fourth flash of light. The first peak occurs after three flashes (rather than four) because most of the manganese cluster is present in the S_1 state (one oxidizing equivalent) when kept in the dark. The oscillations become damped as the number of flashes increases.

From PSII to PSI It was described earlier how the successive absorption of two photons by the reaction center of PSII leads to the formation of a fully reduced PQH_2 molecule. Consequently, the production of a single O_2 molecule, which requires the absorption of four photons by PSII, leads to the formation of two molecules of PQH_2. PQH_2 is a mobile electron carrier that diffuses through the lipid bilayer of the thylakoid membrane and binds to a large multiprotein complex called *cytochrome $b_6 f$* (Figure 6.14). Each PQH_2 molecule donates its two electrons to cytochrome $b_6 f$, while its two protons are released into the lumen. Cytochrome $b_6 f$ is related in structure and function to cytochrome bc_1 of the electron transport chain of mitochondria (page 189). Both complexes have quinols as substrates and share similar redox groups, both can be poisoned by some of the same inhibitors, and both engage in a Q cycle that translocates $4\ H^+$ for each pair of electrons.
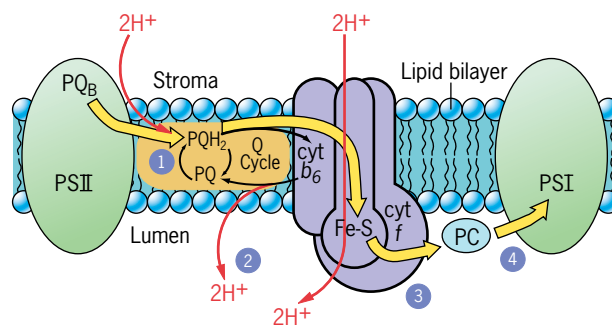
FIGURE 6.14 **Electron transport between PSII and PSI.** The flow of a pair of electrons is indicated by the yellow arrow. Cytochrome $b_6 f$ operates in a manner very similar to that of cytochrome bc_1 in the mitochondria and engages in a Q cycle (not discussed in the text) that translocates four protons for every pair of electrons that moves through the complex. PQH_2 and PC (plastocyanin) are mobile carriers that can transport electrons between distantly separated photosystems.

In this case, 2 H^+ are donated by PQH_2, and 2 additional H^+ are translocated through the complex from the stroma. Because all of these protons had originally been derived from the stroma, their release into the lumen constitutes a movement of protons across the thylakoid membrane (see Figure 6.16b). The electrons from cytochrome $b_6 f$ are passed to another mobile electron carrier, a water-soluble, copper-containing, peripheral membrane protein called *plastocyanin*, situated on the luminal side of the thylakoid membrane (Figure 6.14). Plastocyanin carries electrons to the luminal side of PSI, where they are transferred to $P700^+$, the positively charged reaction-center pigment of PSI.

Keep in mind that all of the electron transfers described in this discussion are exergonic and occur as electrons are passed to carriers with increasing affinity for electrons (more positive redox potentials, page 184). The need for mobile electron carriers, such as PQH_2 and plastocyanin, became evident when it was discovered that the two types of photosystems (PSII and PSI) are not situated in close proximity to one another in the membrane but rather are separated by distances on the order of 0.1 μm. The experiments that led to this discovery are described in detail in the Chapter 6 Experimental Pathways, which can be found on the Web at www.wiley.com/college/karp.

PSI Operations: The Production of NADPH

PSI of higher plants consists of a reaction-center core made up of 12–14 polypeptide subunits and a peripheral complex of protein-bound pigments called LHCI. Light energy is absorbed by the antenna pigments of LHCI and passed to the PSI reaction-center pigment P700, which is a chlorophyll *a* dimer (Figure 6.15). Following energy absorption, an excited reaction-center pigment (P700*) transfers an electron to a separate monomeric chlorophyll *a* molecule (designated A_0), which acts as the primary electron acceptor (step 1, Figure 6.14). As in PSII, absorption of light leads to the production of two charged species, in this case $P700^+$ and A_0^-. A_0^- is a very strong reducing agent with a redox potential of approximately −1.0 V, which is well above that needed to reduce $NADP^+$ (redox potential of −0.32 V). The positive charge of $P700^+$ pigment is neutralized by an incoming electron donated by plastocyanin, as noted earlier.

The initial separation of charge in PSI is stabilized by passage of the electron from A_0^- through several cofactors beginning with a type of quinone called *phylloquinone* (designated A_1) and then three iron-sulfur clusters (designated F_X, F_B, and F_A) (steps 2–4, Figure 6.15). The oxidation of P700 to $P700^+$ occurs at the luminal side of the membrane. As indicated in Figure 6.15, the electron that is lost to the primary acceptor passes through PSI to the iron-sulfur centers bound at the stromal side of the membrane. The electron is subsequently transferred out of PSI to a small, water-soluble, iron-sulfur protein called *ferredoxin* (step 5, Figure 6.15) associated with the stromal surface of the membrane. The reduction of $NADP^+$ to form NADPH (step 6, Figure 6.15) is catalyzed by a large enzyme called ferredoxin-$NADP^+$ reductase, which contains an FAD prosthetic group capable of accepting and transferring two electrons (page 185). An individual ferredoxin
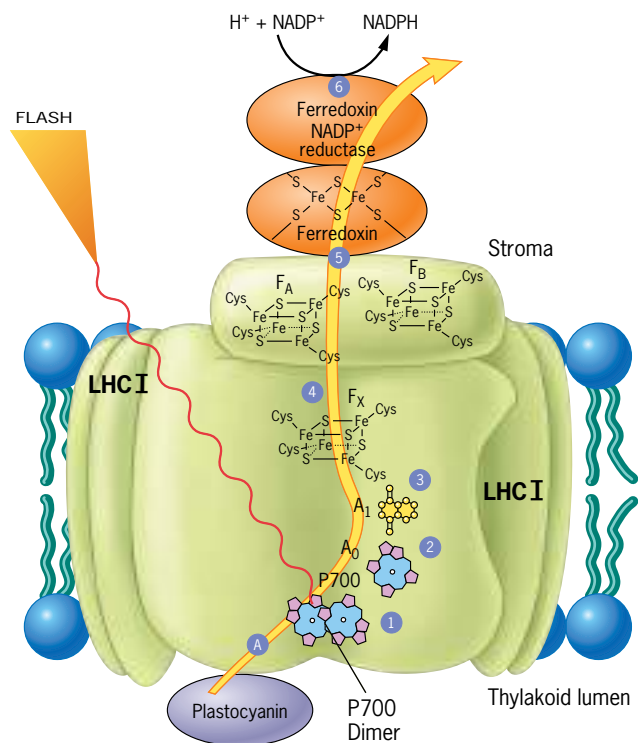
FIGURE 6.15 The functional organization of photosystem I. The path taken by electrons is indicated by the yellow arrow. Events begin with the absorption of light by an antenna pigment and transfer of the energy to a P700 chlorophyll at the PSI reaction center. Absorption of energy by P700 causes the excitation of an electron and its transfer (step 1) to A_0, which is the primary electron acceptor of PSI. The electron subsequently passes to A_1 (step 2) and then to an iron-sulfur center named F_X (step 3). From F_X the electron is transferred (step 4) through two more iron-sulfur centers (F_A and F_B), which are bound by a peripheral protein at the stromal side of the membrane. The electron is finally transferred to ferredoxin, a small iron-sulfur protein (step 5) that is external to the PSI complex. When two different ferredoxin molecules have accepted an electron, they act together to reduce a molecule of $NADP^+$ to NADPH (step 6). The electron-deficient reaction-center pigment ($P700^+$) is reduced by an electron donated by plastocyanin (step A).

molecule can donate only one electron, so that two ferredoxins act together in the reduction:

$$2 \text{ Ferredoxin}_{red} + H^+ + NADP^+ \xrightarrow{\text{Ferredoxin}-NADP^+ \text{ Reductase}} 2 \text{ Ferredoxin}_{ox} + NADPH$$
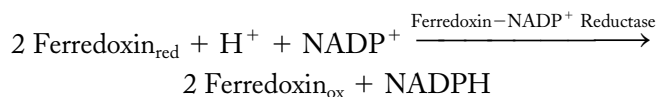
The removal of a proton from the stroma also adds to the proton gradient across the thylakoid membrane. We can write the overall reaction for PSI, based on the absorption of four photons as was done for PSII, as

$$4 \ e^- + 2 H^+_{stroma} + 2 NADP^+ \xrightarrow{\text{4 photons}} 2 \text{ NADPH}$$

Not all electrons passed to ferredoxin inevitably end up in NADPH; alternate routes can be taken depending on the particular organism and conditions. For example, electrons from PSI can be used to reduce various inorganic acceptors. These paths for electrons can lead to the eventual reduction of nitrate (NO_3^-) to ammonia (NH_3), or of sulfate (SO_4^{2-}) to sulfhydryl
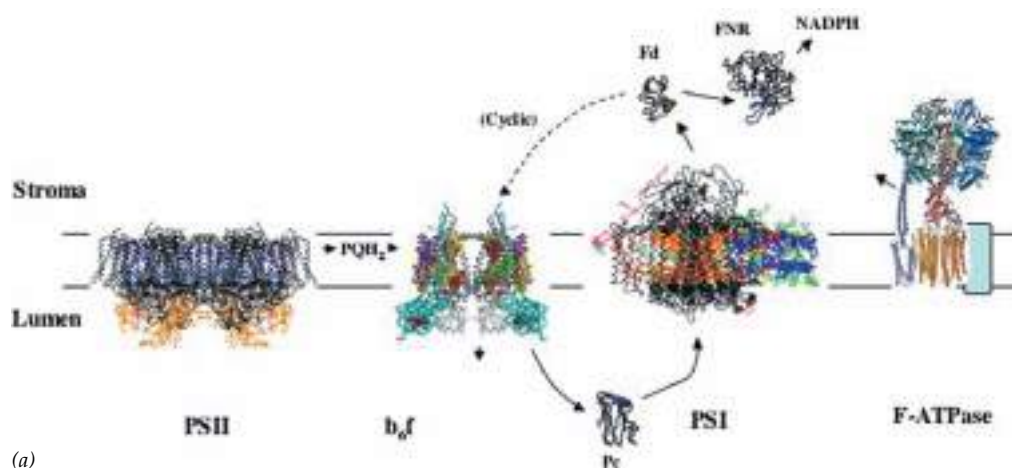
(a)
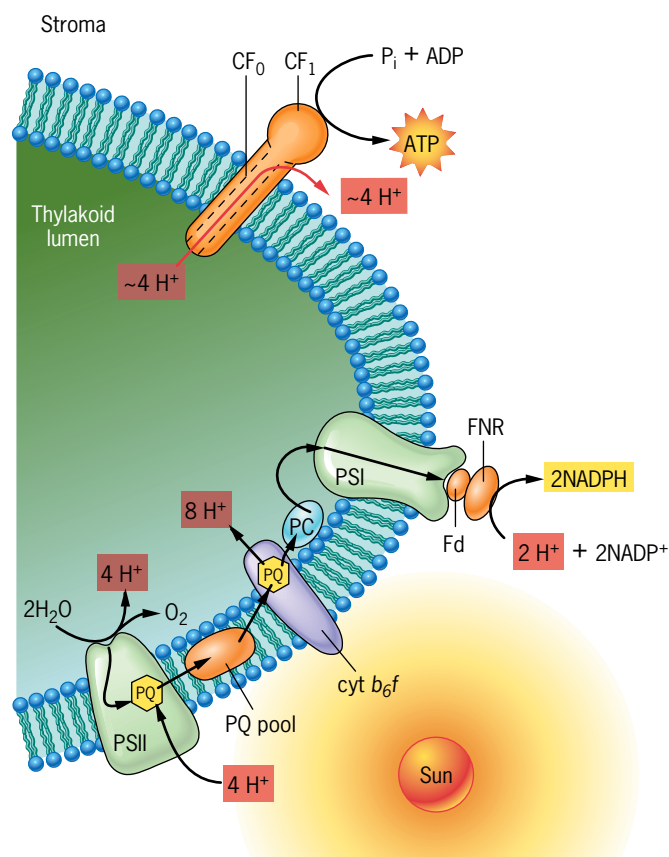
(b)

(—SH), which are key ingredients of biological molecules. Thus, the energy of sunlight is used not only to reduce the most oxidized carbon atoms (those in CO_2), but also to reduce highly oxidized forms of nitrogen and sulfur atoms.

An Overview of Photosynthetic Electron Transport   Figure 6.16*a* shows the molecular structure of the major components involved in the light-dependent reactions of the thylakoid membrane. Knowledge of these structures has proven invaluable in unlocking many of the mysteries of photosynthesis at the molecular level. If we look back over the entire process of electron transport that takes place during oxygenic photosynthesis (summarized in Figure 6.16), we can see that electrons travel from water to NADP$^+$ by the action

of two light-absorbing photosystems. Events occurring in PSII generate a strong oxidizing agent capable of producing O_2 from water, whereas events in PSI generate a strong reducing agent capable of producing NADPH from NADP$^+$. These two events lie at the opposite ends of redox chemistry in living organisms. As noted on page 216, the production of one molecule of O_2 requires the removal of four electrons from two molecules of water. The removal of four electrons from water requires the absorption of four photons, one for each electron. At the same time, the reduction of one molecule of NADP$^+$ requires the transfer of two electrons. Thus, hypothetically, if only one photosystem were able to transfer electrons from H_2O to NADP$^+$, four photons would be sufficient to produce two molecules of NADPH. Because two photosystems are utilized in the cell, that number is doubled to eight photons, four being utilized in PSII and four in PSI. In other words, a total of eight moles of photons must be absorbed by the cell to generate one mole of molecular oxygen and two moles of NADPH. Thus, if we add the reactions of PSII (page 216) and PSI (page 218), *ignoring the protons for a moment*, we arrive at an overall equation for the light reactions of

$$2\,H_2O + 2\,NADP^+ \xrightarrow{\text{8 photons}} 1\,O_2 + 2\,NADPH$$
$$\text{(overall light reaction)}$$
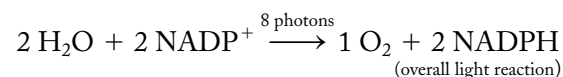
In addition, the light reactions of photosynthesis establish a proton gradient across the thylakoid membrane that leads to the

formation of ATP. The proton gradient forms as the result of the removal of H^+ from the stroma and the addition of H^+ to the thylakoid lumen. Contributions to the proton gradient (see Figure 6.16) arise from (1) the splitting of water in the lumen, (2) oxidation of plastoquinol (PQH_2) by cytochrome $b_6 f$, releasing protons into the lumen, and (3) reductions of $NADP^+$ and PQ, which remove protons from the stroma.

Killing Weeds by Inhibiting Electron Transport

The light reactions of photosynthesis employ a considerable number of electron carriers, which serve as targets for a variety of different plant-killing chemicals (herbicides). A number of common herbicides, including diuron, atrazine, and terbutryn, act by binding to a core protein of PSII. We saw on page 216 how absorption of light by PSII leads to production of a PQH_2 molecule that is subsequently released from the Q_B site of PSII and replaced by a PQ from the pool. The herbicides listed above act by binding to the open Q_B site after release of PQH_2, blocking electron transport through PSII. The herbicide paraquat has received attention in the news media because it is used to kill marijuana plants and because its residues are highly toxic to humans. Paraquat interferes with PSI function by competing with ferredoxin for electrons from the PSI reaction center. Electrons that are diverted to paraquat are subsequently transferred to O_2, generating highly reactive oxygen radicals (page 34) that damage the chloroplasts and kill the plant. Paraquat destroys human tissue by generating oxygen radicals using electrons diverted from complex I of the respiratory chain (page 189).

REVIEW ?

1. What is the relationship between the energy content of a photon and the wavelength of light? How does the wavelength of light determine whether or not it will stimulate photosynthesis? How do the absorbance properties of photosynthetic pigments determine the direction in which energy is transferred within a photosynthetic unit?

2. What is the role in photosynthesis of the light-harvesting antenna pigments?

3. Describe the sequence of events following the absorption of a photon of light by the reaction-center pigment of photosystem II. Describe the comparable events in photosystem I. How are the two photosystems linked to one another?

4. Describe the difference in the redox potentials of the reaction-center pigments of the two photosystems.

5. Describe the process by which water is split during photolysis. How many photons must be absorbed by PSII for this to occur?

6.5 PHOTOPHOSPHORYLATION

The light-dependent reactions discussed in previous pages provide the energy required to reduce CO_2 to carbohydrate. The conversion of one mole of CO_2 to one mole of carbohy-

drate (CH_2O) requires the input of three moles of ATP and two moles of NADPH (see Figure 6.19). We have seen how plant cells produce the NADPH required for manufacture of carbohydrate; let us now examine how these same cells produce the required ATP.

The machinery for ATP synthesis in a chloroplast is virtually identical to that of a mitochondrion or a plasma membrane of an aerobic bacterium. As in those cases, the ATP synthase (Figure 6.16) consists of a head (called CF_1 in chloroplasts), which contains the catalytic site of the enzyme, and a base (CF_0), which spans the membrane and mediates proton movement. The two parts are connected by a rotary stalk. The CF_1 heads project outward into the stroma in keeping with the orientation of the proton gradient, which has its higher concentration within the thylakoid lumen (Figure 6.16). Thus, protons move from higher concentration in the lumen through the CF_0 base of the ATP synthase and into the stroma, thereby driving phosphorylation of ADP, as described in Chapter 5 for the mitochondrion.

Measurements made during periods of maximal ATP synthesis suggest that differences in H^+ concentrations of 1000- to 2000-fold exist across the thylakoid membranes, corresponding to a pH gradient (ΔpH) of more than three units. The movement of protons into the lumen during electron transport is neutralized by movement of other ions, so that a significant membrane potential is not built up. Thus, unlike in mitochondria where the proton-motive force is expressed primarily as an electrochemical potential (page 191), the proton-motive force (Δp) acting in the chloroplasts is largely, if not exclusively, due to a gradient of pH.

Noncyclic Versus Cyclic Photophosphorylation

The formation of ATP during the process of oxygenic photosynthesis is called **noncyclic photophosphorylation** because electrons move in a linear (i.e., noncyclic) path from H_2O to $NADP^+$ (Figure 6.16). During the 1950s, Daniel Arnon of the University of California, Berkeley, discovered that isolated chloroplasts were not only capable of synthesizing ATP from ADP but could do so even in the absence of added CO_2 or $NADP^+$. These experiments indicated that chloroplasts had a means for ATP formation that did not require most of the photosynthetic reactions that would have led to oxygen production, CO_2 fixation, or $NADP^+$ reduction. All that was necessary was illumination, chloroplasts, ADP, and P_i. The process Arnon had discovered was later called **cyclic photophosphorylation** and is a process that is carried out by PSI independent of PSII. Despite the fact that it was discovered over 50 years ago, cyclic photophosphorylation is not well understood. Recent studies suggest that there are two overlapping pathways for cyclic electron transport, one of which is outlined in Figure 6.17. Cyclic electron transport begins with the absorption of a quantum of light by PSI and transfer of a high-energy electron to the primary acceptor. In the pathway depicted in Figure 6.17, the electron is passed along to ferredoxin, as is always the case, but rather than being transferred to $NADP^+$, the electron is routed back to the electron-deficient reaction center (as shown in Figure 6.17) to complete the cycle. During the flow
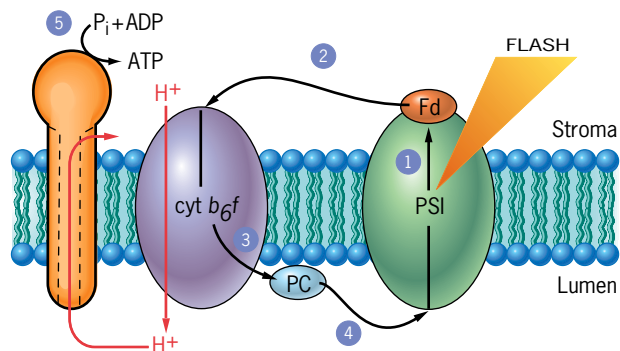
FIGURE 6.17 Simplified scheme for cyclic photophosphorylation. Absorption of light by PSI excites an electron, which is transferred to ferredoxin (step 1) and on to cytochrome b_6f (step 2), plastocyanin (step 3), and back to P700$^+$ (step 4). In the process, protons are translocated by cytochrome b_6f to form a gradient utilized for ATP synthesis (step 5). Another cyclic pathway for electron transport that involves movement of electrons from PSI through NADPH to cytochrome b_6f is not shown.

of an electron around this course, sufficient free energy is released to translocate protons (estimated at two H$^+$/e^-) across the membrane by the cytochrome b_6f complex and to build a proton gradient capable of driving ATP synthesis. Cyclic photophosphorylation is thought to provide additional ATP required for carbohydrate synthesis (see Figure 6.19) and also for other ATP-requiring activities in the chloroplast (e.g., molecular chaperone involvement in protein import). Inhibition of cyclic photophosphorylation leads to impaired development and growth of higher plants.

Now that we have seen how the light reactions of photosynthesis lead to the production of ATP and NADPH, which are the energy stores required for the synthesis of carbohydrates, we can turn to the reactions that lead to carbohydrate formation.

REVIEW ?

1. Which steps in the light-dependent reactions are responsible for generating an electrochemical proton gradient across the thylakoid membrane?

2. To what degree is this gradient reflected in a pH gradient versus a voltage?

3. How does the proton gradient lead to the formation of ATP?

6.6 CARBON DIOXIDE FIXATION AND THE SYNTHESIS OF CARBOHYDRATE

After World War II, Melvin Calvin of the University of California, Berkeley, along with colleagues Andrew Benson and James Bassham, began what was to be a decade-long study of the enzymatic reactions by which carbon dioxide was assimilated into the organic molecules of the cell. Armed with the newly available, long-lived radioactive isotope of carbon

(^{14}C) and a new technique, two-dimensional paper chromatography, they began the task of identifying all the labeled molecules produced when cells took up [^{14}C]O$_2$. The studies began with plant leaves but soon shifted to a simpler system, the green alga *Chlorella*. Algal cultures were grown in closed chambers in the presence of unlabeled CO$_2$, after which radioactive CO$_2$ was introduced by injection into the culture medium. After the desired period of incubation with the labeled CO$_2$, the algal suspension was drained into a container of hot alcohol, which had the combined effects of killing the cells immediately, stopping enzyme activity, and extracting soluble molecules. Extracts of the cells were then placed as a spot on chromatographic paper and subjected to two-dimensional chromatography. To locate the radioactive compounds at the end of the procedure, a piece of X-ray film was pressed against the chromatogram, and the plates were kept in the dark to expose the film. After photographic development, identification of radiolabeled compounds on the autoradiogram was made by comparison to known standards and by chemical analysis of the original spots. Let us consider some of their findings.

Carbohydrate Synthesis in C$_3$ Plants

Labeled CO$_2$ was converted to reduced organic compounds very rapidly. If the incubation period was very short (up to a few seconds), one radioactive spot on the chromatogram predominated (Figure 6.18). The compound that formed the spot was determined to be 3-phosphoglycerate (PGA), one of the intermediates of glycolysis. Calvin's group initially suspected that CO$_2$ was being covalently linked (or *fixed*) to a two-carbon compound to form the three-carbon PGA molecule. Because the first intermediate to be identified was a three-carbon molecule, plants that utilize this pathway to fix atmospheric CO$_2$ are referred to as **C$_3$ plants**.

After considerable investigation, it became apparent that the initial acceptor was not a two-carbon compound, but a five-carbon compound, ribulose 1,5-bisphosphate (RuBP), which condensed with CO$_2$ to produce a six-carbon molecule. This six-carbon compound rapidly fragmented into two mol-

FIGURE 6.18 Chromatogram showing the results of an experiment in which algal cells were incubated for 5 seconds in [^{14}C]O$_2$ prior to immersion in alcohol. One spot, which corresponds to 3-phosphoglycerate (labeled PGA), contains most of the radioactivity. (COURTESY OF JAMES BASSHAM AND MELVIN CALVIN.)

ecules of PGA, one of which contained the recently added carbon atom. Both the condensation of RuBP and the splitting of the six-carbon product (Figure 6.19a) are carried out in the stroma by a large multisubunit enzyme, *ribulose bisphosphate carboxylase*, which is known familiarly as *Rubisco*. As the enzyme responsible for converting inorganic carbon into useful biological molecules, Rubisco is one of the key proteins in the biosphere. As it happens, Rubisco is capable of fixing only about three molecules of CO_2 per second, which may be the worst turnover number of any known enzyme (see Table 3.3). To compensate for this inefficiency, as much as half of the protein of leaves is present as Rubisco. In fact, Rubisco constitutes the most abundant protein on Earth, accounting for 5–10 kg for every human.
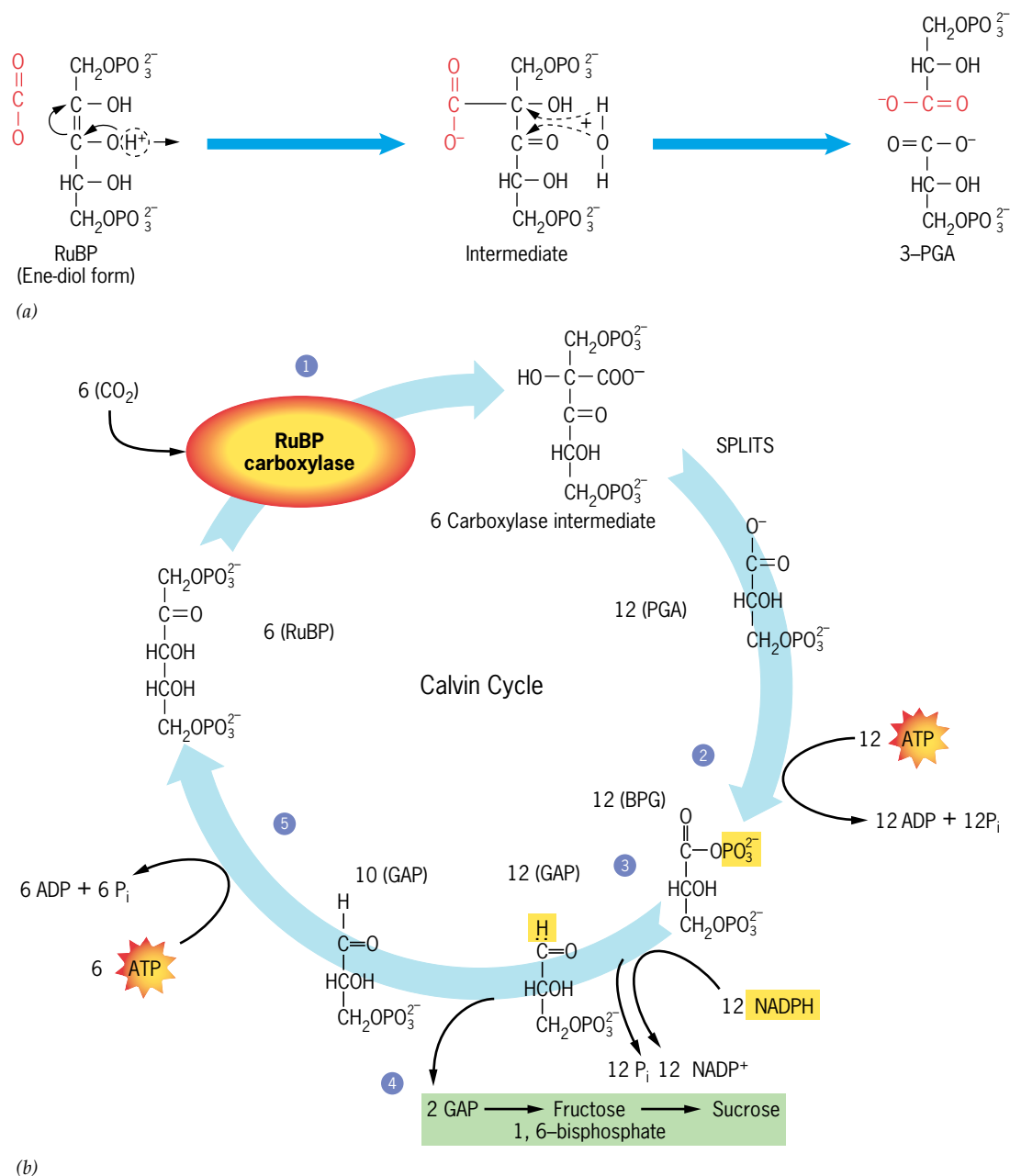
FIGURE 6.19 Converting CO_2 into carbohydrate. (*a*) The reaction catalyzed by ribulose bisphosphate carboxylase (Rubisco) in which CO_2 is fixed by linkage to RuBP. The product rapidly splits into two molecules of 3-phosphoglycerate (PGA). (*b*) An abbreviated version of the Calvin cycle showing the fate of 6 molecules of CO_2 that are fixed by combination with 6 molecules of RuBP. (Numerous reactions have been deleted). The fixation of CO_2 is indicated in step 1. In step 2, the 12 PGA molecules are phosphorylated via ATP hydrolysis to form 12 1,3-bisphosphoglycerate (BPG) molecules, which are reduced in step 3 by electrons provided by NADPH to form 12 molecules of glyceraldehyde 3-phosphate (GAP). Here, 2 of the GAPs are drained away (step 4) to be used in the synthesis of sucrose in the cytosol, which can be considered the product of the light-independent reactions. The other 10 molecules are converted into 6 molecules of RuBP (step 5), which can act as the acceptor for 6 more molecules of CO_2. The regeneration of 6 RuBPs requires the hydrolysis of 6 molecules of ATP. The NADPH and ATP used in the Calvin cycle represent the two high-energy products of the light-dependent reactions.

As the structure of the various intermediates was determined, together with the positions of the labeled carbon atoms, it became apparent that the pathway for conversion of CO_2 into carbohydrate was cyclic and complex. This pathway has become known as the **Calvin cycle (or Calvin-Benson cycle)**, and it occurs in cyanobacteria and all eukaryotic photosynthetic cells. A simplified version of the Calvin cycle is shown in Figure 6.19b. The cycle comprises three main parts: (1) carboxylation of RuBP to form PGA; (2) reduction of PGA to the level of a sugar (CH_2O) by formation of glyceraldehyde 3-phosphate (GAP) using the NADPH and ATP produced in the light-dependent reactions; and (3) the regeneration of RuBP, which also requires ATP. It can be seen in Figure 6.19b that for every 6 molecules of CO_2 fixed, 12 molecules of GAP are produced. (GAP is the spot marked triose phosphate on the chromatogram of Figure 6.18.) The atoms in 10 of these three-carbon GAP molecules are rearranged to regenerate 6 molecules of the five-carbon CO_2 acceptor, RuBP (Figure 6.19b). The other 2 GAP molecules can be considered as product. These GAP molecules can be exported to the cytosol in exchange for phosphate ions (see Figure 6.20) and used to synthesize the disaccharide sucrose. Alternatively, GAP can remain in the chloroplast where it is converted to starch. An overview of the entire process of photosynthesis, including the light reactions (light absorption, oxidation of water, reduction of $NADP^+$, translocation of protons), phosphorylation of ADP, the Calvin cycle, and the synthesis of starch or sucrose, is shown in Figure 6.20.

Sucrose molecules that are formed in the cytosol from the GAPs of the Calvin cycle are transported out of the leaf cells and into the phloem, where they are carried to the various nonphotosynthetic organs of the plant. Just as glucose serves as the source of energy and organic building blocks in most animals, sucrose serves an analogous role in most plants. Starch, on the other hand, is stored within chloroplasts as granules (see Figure 2.17b). Just as stored glycogen provides animals with readily available glucose in times of need, the starch stored in a plant's leaves provides it with sugars at night when the light reactions are not possible. It is evident from the reactions of Figure 6.19b that the synthesis of carbohydrate is an expensive activity. The conversion of 6 molecules of CO_2 to 1 six-carbon sugar molecule and the regeneration of RuBP requires 12 molecules of NADPH and 18 molecules of ATP. This large energy expenditure reflects the fact that CO_2 is the most highly oxidized form in which carbon can occur.

Redox Control   Through the course of this book, we will discuss a diverse array of mechanisms used by cells to control the activity of proteins. One of these mechanisms, known as *redox control*, is turning up more and more often as a regulator of basic cellular processes, including protein folding, transcription, and translation. We will consider it here because it is best understood as a regulator of chloroplast metabolism. Several key enzymes of the Calvin cycle are only active in the light when ATP and NADPH are being produced by photosynthesis. This light-dependent control of chloroplast enzymes is exercised by a small protein called thioredoxin that can occur in either a reduced or oxidized form. As noted
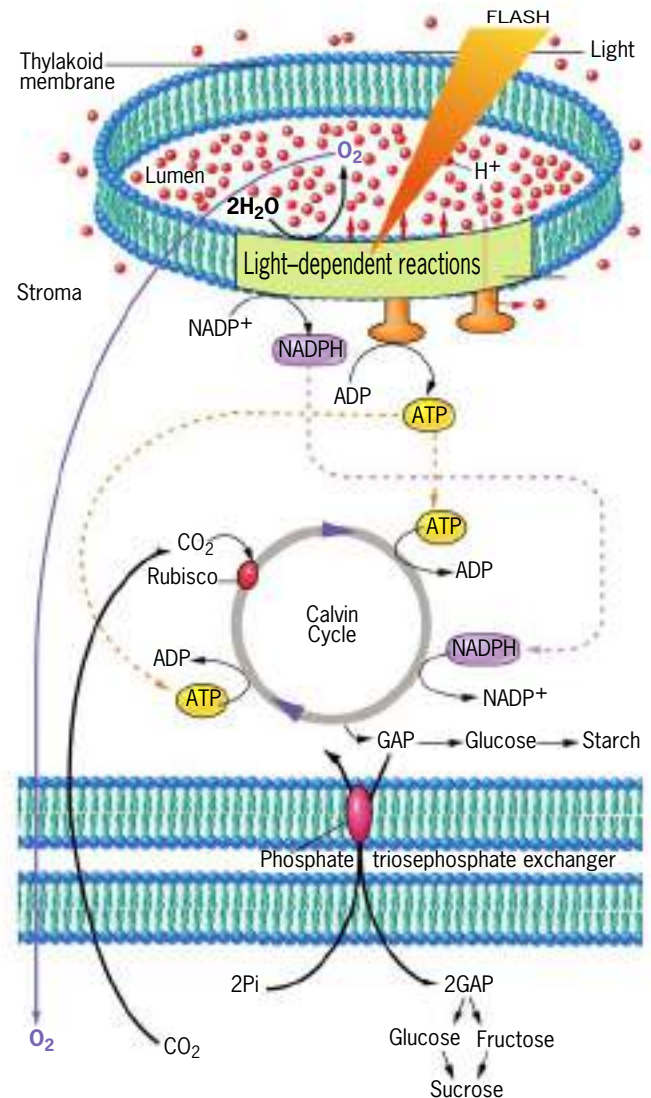
FIGURE 6.20 **An overview of the various stages of photosynthesis.**

on page 218, not all of the electrons that pass through ferredoxin are used to reduce $NADP^+$. In fact, some of these electrons are transferred to thioredoxin. Once it has accepted a pair of electrons, thioredoxin reduces certain disulfide bridges (—S-S—) into sulfhydryl groups (—SH) in selected Calvin cycle enzymes (Figure 6.21). This covalent change in protein structure activates these enzymes, promoting the synthesis of carbohydrates in the chloroplast. In the dark, when photosynthesis has ceased, thioredoxin is no longer reduced by ferredoxin, and the Calvin cycle enzymes revert to an oxidized (—S-S—) state in which they are inactive. It follows from these findings that reference to the reactions of the Calvin cycle as "dark reactions" is a misnomer.

Photorespiration   One of the spots that appeared on the chromatograms from Calvin's early work on algal cells was identified as the compound glycolate, which was ignored (and correctly so) in formulating the Calvin cycle of Figure 6.19. While glycolate is not part of the Calvin cycle, it is a product
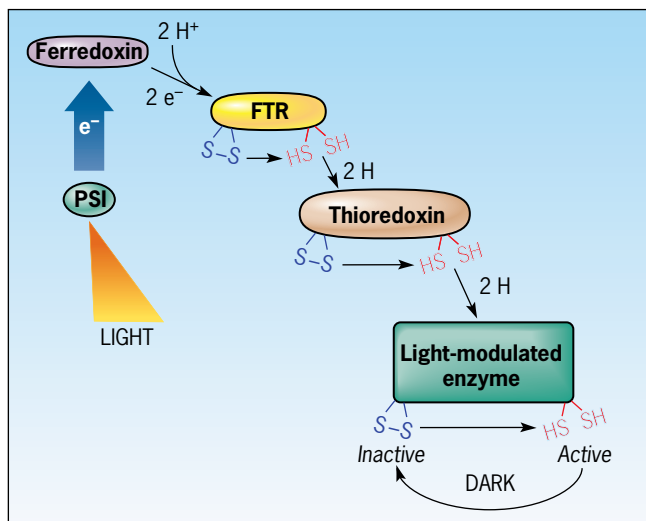
FIGURE 6.21 Redox control of the Calvin cycle. In the light, ferredoxin is reduced, and a fraction of these electrons are transferred to the small protein thioredoxin, which reduces the disulfide groups of certain Calvin cycle enzymes, maintaining them in an active state. In the dark, electron flow to thioredoxin ceases, the sulfhydryl groups of the regu-lated enzymes become oxidized to the disulfide state, and the enzymes are inactivated.

of a reaction catalyzed by Rubisco. Approximately 20 years af-ter it was found that Rubisco catalyzed the attachment of CO_2 to RuBP, it was discovered that Rubisco also catalyzed a second reaction in which O_2 is attached to RuBP to produce 2-phosphoglycolate (along with PGA) (Figure 6.22). Phos-

phoglycolate is subsequently converted to glycolate by an en-zyme in the stroma. Glycolate produced in the chloroplast is transferred to the peroxisome and eventually leads to the re-lease of CO_2 as described later (see Figure 6.23).

Because it involves the uptake of O_2 and the release of CO_2, this series of reactions is called **photorespiration**. Be-cause photorespiration leads to the release of previously fixed CO_2 molecules, it is a waste of the plant's energy. In fact, pho-torespiration can account for the loss of up to 50 percent of newly fixed carbon dioxide by crop plants growing under con-ditions of high light intensity. Thus, as you might expect, a concerted effort has been underway for decades to breed plants that are less likely to engage in photorespiration. So far, these efforts have met with virtually no success.

Studies of the enzymatic activity of purified Rubisco show that the enzyme exhibits only a modest preference for CO_2 as a substrate over O_2. The reason for this relative lack of specificity is that neither CO_2 nor O_2 binds to the active site of the enzyme (see Figure 3.11a). Rather, the enzyme binds RuBP, which adopts the enediol form shown in Figure 6.22. This form of RuBP can then be attacked by either CO_2 or O_2. Viewed in this way, photorespiration appears to be an un-avoidable consequence of the catalytic properties of Rubisco, an enzyme that is presumed to have evolved at a time when O_2 levels in the atmosphere were virtually nonexistent. Under today's atmospheric conditions, O_2 and CO_2 compete with one another, and the predominant direction of the Rubisco-catalyzed reaction is determined by the ratio of CO_2/O_2 that is available to the enzyme. When plants are grown in closed environments containing elevated levels of CO_2, they are ca-pable of much more rapid growth by virtue of their elevated

FIGURE 6.22 The reactions of photorespiration. Rubisco can catalyze two different reactions with RuBP as a substrate (shown in the ene-diol state within the plane; see Figure 6.19a). If the RuBP reacts with O_2 (step 1b), the reaction produces an oxygenase intermediate (step 2b) that breaks down into 3-PGA and 2-phosphoglycolate (step 3b). The subsequent reactions of phosphoglycolate are shown in Figure 6.23. The eventual outcome of these reactions is the release of CO_2, a molecule that the cell had previously expended energy to fix. In contrast, if the RuBP molecule reacts with CO_2 (step 1a), the reac-tion produces a carboxylase intermediate (step 2a) that breaks down into two molecules of PGA (step 3a), which continue through the Calvin cycle.
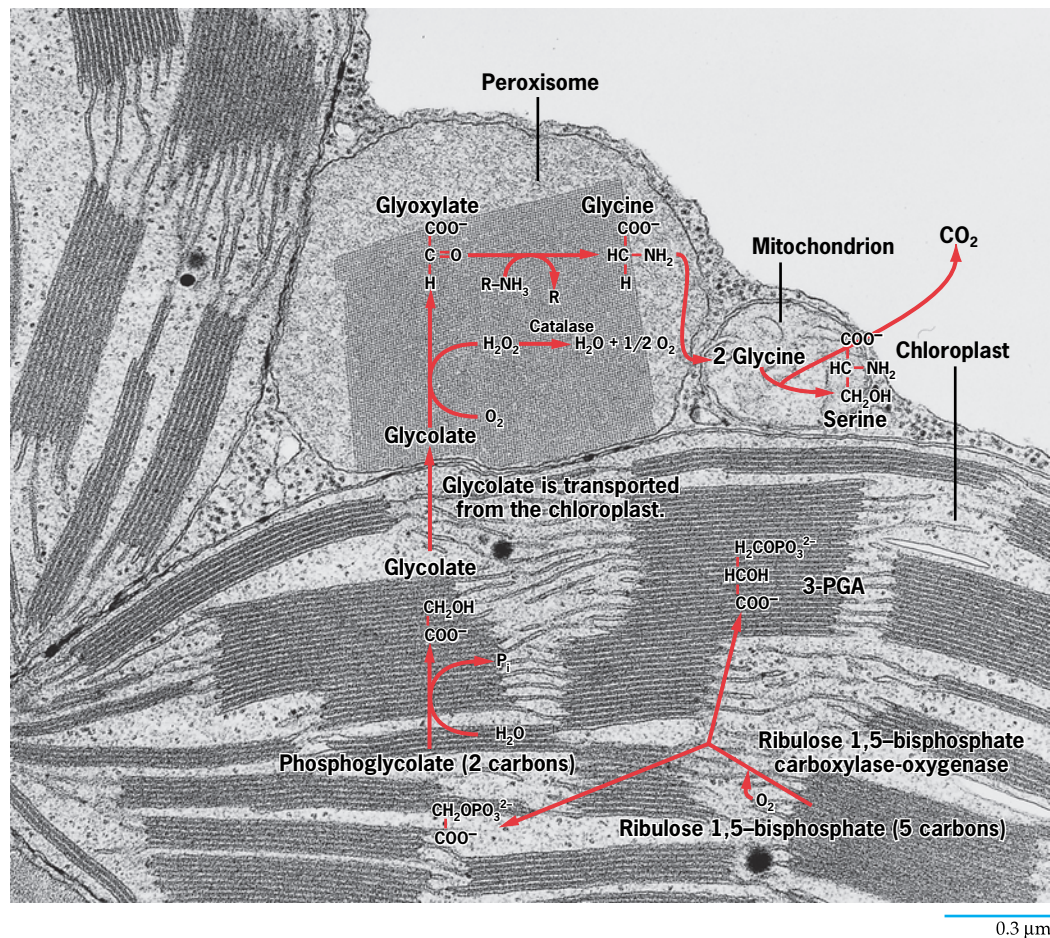
FIGURE 6.23 The cellular basis of photorespiration. Electron micrograph of a portion of a leaf mesophyll cell of a tobacco plant showing a peroxisome (identified by its crystalline core) pressed against a pair of chloroplasts and close to a mitochondrion. The reactions of photorespiration that occur in each of these organelles are described in the text and shown superimposed on the organelles in which they occur. This series of reactions is referred to as the C_2 cycle. The last steps of the cycle—the conversion of serine to glycerate in the peroxisome and then to 3-PGA in the chloroplast—are not shown. (MICROGRAPH COURTESY OF SUE ELLEN FREDERICK AND ELDON H. NEWCOMB.)

rates of CO_2 fixation. It is suggested that the rise in atmospheric CO_2 levels over the past century (from about 270 parts per million in 1870 to 390 ppm in 2009) is responsible for as much as 10 percent of the increased crop yield that has occurred during this period. This rise in atmospheric CO_2 concentration is also thought to be responsible for global warming. Even a small rise in Earth's temperature could have major effects on global conditions, such as a rise in the level of the seas and a spread of the planet's deserts.

Peroxisomes and Photorespiration Peroxisomes are cytoplasmic organelles whose role in oxidative metabolism was discussed in Section 5.6 of the last chapter. Studies on the peroxisomes of leaf cells have revealed a striking example of interdependence among different organelles, a feature that is often lost in studies with isolated cellular structures. The electron micrograph in Figure 6.23 shows a peroxisome of a leaf cell closely apposed to the surfaces of two adjacent chloroplasts and in close proximity to a nearby mitochondrion. This arrangement is not coincidental but reflects an underlying biochemical relationship whereby the products of one organelle serve as substrates in another organelle. The reactions that take place in the different organelles are superimposed on the micrograph in Figure 6.23 and summarized below.

It was noted above that photorespiration begins when RuBP reacts with O_2 to form 3-PGA and a two-carbon compound, phosphoglycolate (Figure 6.22). Once formed, phosphoglycolate is converted to glycolate, which is shuttled out of the chloroplast into a peroxisome. In the peroxisome, the enzyme glycolate oxidase converts glycolate to glyoxylate, which may then be converted to glycine and transferred to a mitochondrion. In the mitochondrion, two molecules of glycine (a two-carbon molecule) are converted to one molecule of serine (a three-carbon molecule), with the accompanying release of one molecule of CO_2. Thus, for every two molecules of phosphoglycolate produced by Rubisco, one carbon atom that had previously been fixed is lost back to the atmosphere. The serine produced in the mitochondrion may be shuttled back to the peroxisome and converted to glycerate, which can be transported to the chloroplast and utilized in carbohydrate synthesis by means of the formation of 3-PGA.

Two groups of plants, called C_4 and CAM plants, have overcome the negative effects of photorespiration by evolving metabolic mechanisms that increase the CO_2/O_2 ratio to which the Rubisco enzyme molecules are exposed.

Carbohydrate Synthesis in C_4 Plants

In 1965, Hugo Kortschak reported that when sugarcane was supplied with $[^{14}C]O_2$, radioactivity first appeared in organic compounds containing four-carbon skeletons rather than the three-carbon PGA molecule found in other types of plants. Further analysis revealed that these four-carbon compounds (predominantly oxaloacetate and malate) resulted from combination of CO_2 with phosphoenolpyruvate (PEP) and thus represented a second mechanism of fixation of atmospheric carbon dioxide (Figure 6.24). The enzyme responsible for linkage of CO_2 to PEP was named *phosphoenolpyruvate carboxylase*, and it catalyzes the first step of the **C_4 pathway**. Plants that utilize this pathway are referred to as **C_4 plants** and are represented primarily by tropical grasses. Before considering the fate of this newly fixed carbon atom, it is useful to examine the likely reasons an alternate pathway of CO_2 fixation has evolved.

When a C_3 plant is placed in a closed chamber and its photosynthetic activity monitored, it is found that, once the plant reduces CO_2 levels in the chamber to approximately 50 parts per million (ppm), the rate of CO_2 release by photorespiration equals the rate of CO_2 fixation by photosynthe-sis, so that net production of carbohydrate stops. In contrast, a plant that utilizes the C_4 pathway continues the net synthesis of carbohydrate until CO_2 levels have dropped to 1 to 2 ppm. C_4 plants have this ability because PEP carboxylase continues to operate at much lower CO_2 levels than does Rubisco and is not inhibited by O_2. But what is the value to a plant of being able to fix CO_2 at such low levels when the atmosphere invariably contains CO_2 at levels well above 300 ppm?

The value of the C_4 pathway becomes apparent when C_4 plants are placed in a dry, hot environment similar to that in which many of them live. The most serious problem faced by plants living in a hot, dry climate is loss of water through openings, called stomata, at the surfaces of their leaves (Figure 6.2). Although open stomata lead to loss of water, they also provide the channel through which CO_2 enters the leaves. C_4 plants are adapted to hot, arid environments because they are capable of closing their stomata to prevent water loss, yet they are still capable of maintaining sufficient CO_2 uptake to fuel their photosynthetic activity at a maximal rate. This is the reason crabgrass, a C_4 plant, tends to take over a lawn, displacing the domestic C_3 grasses originally planted. Sugarcane, corn, and sorghum are the most important crop plants that utilize the C_4 pathway. Because most C_4 plants do not do nearly as well in cooler temperatures, their distribution is limited at higher latitudes.

When the fate of the CO_2 fixed by means of the C_4 pathway is followed, it is found that the CO_2 group is soon released, only to be recaptured by Rubisco and converted to
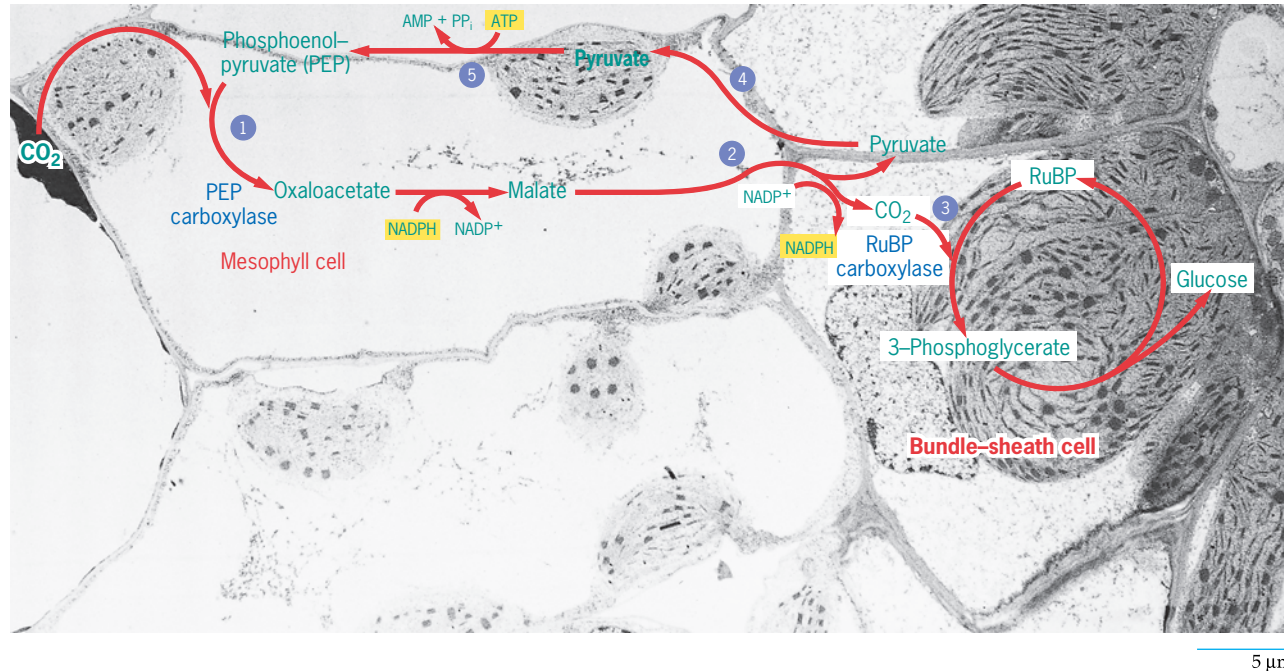
FIGURE 6.24 Structure and function in C_4 plants. Electron micrograph of a transverse section through the leaf of a C_4 plant showing the spatial relationship between the mesophyll and bundle-sheath cells. Superimposed on the micrograph are the reactions of CO_2 fixation that occur in each type of cell. In step 1, CO_2 is joined to PEP by the enzyme PEP carboxylase in a mesophyll cell that is located close to the leaf exterior. The four-carbon malate formed by the reaction is transported to the more centrally located bundle-sheath cell (step 2), where the CO_2 is released. The CO_2 becomes highly concentrated in the bundle-sheath cell, favoring the fixation of CO_2 by Rubisco to form 3-PGA (step 3), which can be circulated through the Calvin cycle. The pyruvate formed when CO_2 is released is shipped back to the mesophyll cell (step 4), where it is converted to PEP. Although the process requires the hydrolysis of ATP (step 5), the high CO_2/O_2 in the bundle-sheath cell minimizes the rate of photorespiration. (ELECTRON MICROGRAPH COURTESY OF S. CRAIG.)

metabolic intermediates of the C_3 pathway (Figure 6.24). The basis for this seemingly paradoxical metabolism becomes evident when the leaf anatomy of a C_4 plant is examined. Unlike those of C_3 plants, leaves of C_4 plants contain two concentric cylinders of cells. An outer cylinder is made of *mesophyll cells* and an inner cylinder of *bundle-sheath cells* (Figure 6.24). The fixation of CO_2 to PEP occurs in the outer mesophyll cells. PEP carboxylase activity can continue, even when the leaf stomata are almost entirely closed and the CO_2 level in the cells is very low. Once formed, the C_4 products are transported through plasmodesmata in the adjoining cell wall (see Figure 7.35) and into the thick-walled bundle-sheath cells, which are sealed off from atmospheric gases. Once in the bundle-sheath cells, the previously fixed CO_2 can be split from the C_4 carrier, producing a high CO_2 level in these interior cells—a level suitable for fixation by Rubisco. Carbon dioxide levels in the bundle-sheath cells may be 100 times that of the mesophyll. Thus, the C_4 pathway provides a mechanism to drive CO_2 fixation by the less efficient C_3 pathway by "pumping" CO_2 into the bundle sheath. Once CO_2 is split from the four-carbon compound, the pyruvate formed returns to the mesophyll cells to be recharged as PEP (Figure 6.24). In addition to saving water, plants that utilize the C_4 pathway are able to generate high CO_2/O_2 ratios in the local Rubisco environment, thereby favoring the process of CO_2 fixation over that of photorespiration. In fact, attempts to demonstrate photorespiration in the intact leaves of C_4 plants usually fail.

Before leaving the subject of C_4 photosynthesis, it can be noted that a number of plant science laboratories are attempting to introduce parts of the C_4 photosynthetic machinery into C_3 plants to boost crop productivity. For example, the PEP carboxylase gene from maize (a C_4 plant) has been introduced into rice (a C_3 plant) in hopes of creating a CO_2-concentrating mechanism within the rice mesophyll cells where Rubisco is housed. Some of the field studies on these genetically engineered plants have been encouraging, but it is too soon to tell whether any observed increases in photosynthetic efficiency are due to increased levels of CO_2 assimilation, decreased levels of photorespiration, increased resistance to stress, or some other mechanism.

Carbohydrate Synthesis in CAM Plants

Many desert plants, such as cacti, have another biochemical adaptation that allows them to survive in very hot, dry habitats. These plants are called **CAM plants**, and they utilize PEP carboxylase in fixing atmospheric CO_2 just like C_4 plants.[3] Unlike C_4 plants, however, CAM species carry out light-dependent reactions and CO_2 fixation at different times of the day, rather than in different cells of the leaf. Whereas C_3 and C_4 plants open their leaf stomata and fix CO_2 during the daytime, CAM plants keep their stomata closed during the hot, dry daylight hours. Then, at night, when the rate of water vapor loss is greatly reduced, they open their stomata and fix CO_2 by means of PEP carboxylase. As more and more carbon dioxide is fixed in the mesophyll cells during the night, the malate that is generated is transported into the cell's central vacuole. The presence of malate (in the form of malic acid within the acidic vacuole) is evident by the sour "morning taste" of the plants. During the daylight hours, the stomata close, and the malic acid is moved into the cytoplasm. There it gives up its CO_2, which can be fixed by Rubisco under conditions of low O_2 concentration that exist when the stomata are closed. Carbohydrates are then synthesized using energy from ATP and NADPH generated by the light-dependent reactions.

REVIEW ?

1. Describe the basic plan of the Calvin cycle, indicating the reactions that require energy input. Why is it described as a cycle? Why does energy have to be expended in this type of pathway? What are the eventual products of the pathway?
2. Describe the major structural and biochemical differences between C_3 plants and C_4 plants. How do these differences affect the ability of these two types of plants to grow in hot, dry climates?

---

[3]CAM is an acronym for *crassulacean acid metabolism*, named for plants of the family Crassulaceae in which it was first discovered.

SYNOPSIS

The earliest life forms are presumed to have been heterotrophs that depended on organic molecules formed abiotically; these forms were eventually succeeded by autotrophs—organisms that were capable of surviving on CO_2 as their principal carbon source. The first autotrophs are thought to have carried out nonoxygenic photosynthesis in which compounds such as H_2S were oxidized as a source of electrons. The evolution of oxygenic photosynthesis, in which water is oxidized and O_2 is released, allowed cyanobacteria to exploit a wide array of habitats and set the stage for aerobic respiration. (p. 206)

Chloroplasts are large membrane-bound organelles that have evolved from a photosynthetic prokaryote. Chloroplasts are bounded by a double membrane made porous by the inclusion of porins in the outer membrane. Thylakoids are flattened membranous sacs arranged in orderly stacks, or grana. The thylakoids are surrounded by a fluid stroma containing DNA, ribosomes, and the machinery required for gene expression. (p. 208)

The light-dependent reactions of photosynthesis begin with the absorption of photons of light by photosynthetic pigments, an event that pushes electrons to outer orbitals from which they can be transferred to an electron acceptor. The major light-absorbing pigments of plants are chlorophylls and carotenoids. Each chlorophyll molecule consists of a Mg^{2+}-containing porphyrin ring that acts in light absorption and a hydrocarbon tail (a phytol) that keeps the pigment embedded in the bilayer. Chlorophyll absorbs most strongly in the blue and red region of the visible spectrum and least strongly in the green. Carotenoids absorb most strongly in the blue and green region and least strongly in the red and orange. The action

spectrum of photosynthesis, which shows the wavelengths of light that are able to stimulate photosynthesis, follows the absorption spectrum of the pigments quite closely. Photosynthetic pigments are organized into functional units in which only one molecule—the reaction-center chlorophyll—transfers electrons to an electron acceptor. The bulk of the pigment molecules form a light-harvesting antenna that traps photons of differing wavelengths and transfers the energy of excitation to the pigment molecule at the reaction center. (p. 211)

The transfer of a pair of electrons from H_2O to $NADP^+$ under conditions operating in the chloroplast utilizes the input of approximately 2 V of energy, which is delivered by the combined action of two separate photosystems. Photosystem II (PSII) boosts electrons from an energy level below that of water to a midway point, while photosystem I (PSI) raises electrons to the highest energy level, above $NADP^+$. As photons are absorbed by each photosystem, energy is passed to their respective reaction-center pigments (P680 for PSII and P700 for PSI). The energy absorbed by the reaction-center chlorophylls serves to boost an electron to an outer orbital where it is transferred to a primary acceptor, producing a positively charged pigment ($P680^+$ and $P700^+$). (p. 213)

The path of noncyclic electron flow from water to PSII and then to PSI and $NADP^+$ is denoted in the form of a Z. The first leg of the Z is from H_2O to PSII. Most of the antenna pigments that collect light for PSII reside within a separate complex, called LHCII. Energy flows from LHCII to the PSII reaction center, where an electron is transferred from P680 to a primary acceptor, which is a chlorophyll-like pheophytin molecule. This transfer of electrons generates a strong oxidizing agent ($P680^+$) and a weak reducing agent ($Pheo^-$). The separation of charge in PSII is stabilized by moving the oppositely charged molecules farther apart; this is accomplished as the electron is transferred from pheophytin to the quinone PQ_A and then PQ_B. The successive absorption of two photons by PSII leads to the transfer of two electrons to PQ_B to form PQ_B^{2-}, which then picks up two protons from the stroma, forming PQH_2 (reduced plastoquinone). The reduced plastoquinone leaves the reaction center and is replaced by an oxidized plastoquinone that is capable of receiving additional electrons. As each electron is transferred from P680 to the primary acceptor and on to PQ_B, the positively charged reaction-center pigment ($P680^+$) is neutralized by an electron from a cluster containing four manganese ions and a calcium ion. As each electron is transferred to P680, the Mn-Ca-containing protein accumulates an oxidizing equivalent. Once the cluster has accumulated four oxidizing equivalents, it is capable of removing four electrons from water, a reaction that generates O_2 and delivers four H^+ to the thylakoid lumen, which contributes to the proton gradient. (p. 215)

Electrons from reduced plastoquinone are transferred to the multiprotein complex cytochrome b_6f, while the protons are released into the thylakoid lumen, thus contributing to the proton gradient. Electrons from cytochrome b_6f are passed to plastocyanin located on the luminal side of the thylakoid membrane and on to $P700^+$, the reaction-center pigment of PSI that had lost an electron following absorption of a photon. As each photon is absorbed by P700, the electron is transferred to a primary acceptor, A_0, and then through several iron-sulfur centers of the PSI reaction center to ferredoxin. Electrons are transferred from ferredoxin to $NADP^+$ forming NADPH, which requires a proton from the stroma, contributing to the proton gradient. In summary, noncyclic electron flow results in the oxidation of H_2O to O_2, with the transfer of electrons

to $NADP^+$, forming NADPH, and the establishment of a H^+ gradient across the membrane. (p. 217)

The proton gradient established during the light-dependent reactions provides the energy required for the formation of ATP in the chloroplast, a process called photophosphorylation. The machinery for ATP synthesis in the chloroplast is virtually identical to that of the mitochondrion; the ATP synthase consists of a CF_1 headpiece that projects into the stroma and a CF_0 base that is embedded in the thylakoid membrane. Protons drive the formation of ATP as they move from higher concentration in the thylakoid lumen through the CF_0 base and into the stroma, thereby dissipating the H^+ gradient. ATP synthesis can occur in the absence of the oxidation of H_2O through a process of cyclic photophosphorylation that does not involve PSII. Light is absorbed by P700 of PSI, passed to ferredoxin, and returned to the electron-deficient PSI reaction center through cytochrome b_6f. As electrons pass around the cyclic pathway, protons are translocated into the thylakoid lumen and subsequently drive the formation of ATP. (p. 220)

During the light-independent reactions, the chemical energy stored in NADPH and ATP is used in the synthesis of carbohydrates from CO_2. CO_2 is converted to carbohydrate by the C_3 pathway (or Calvin cycle) in which CO_2 is fixed by RuBP carboxylase (Rubisco) to a five-carbon compound, RuBP, forming an unstable six-carbon intermediate, which splits into two molecules of 3-phosphoglyceric acid. NADPH and ATP are used to convert PGA molecules to glyceraldehyde phosphate (GAP). For every 6 molecules of CO_2 fixed, 2 molecules of GAP can be directed toward the formation of sucrose or starch, while the remaining 10 molecules of GAP can be used to regenerate RuBP for additional rounds of CO_2 fixation. (p. 221)

Rubisco can also catalyze a reaction in which O_2, rather than CO_2, is covalently joined to RuBP. This process, which is called photorespiration, leads to the formation of compounds that are metabolized in reactions that lead to the loss of CO_2. Because photorespiration involves the uptake of O_2 and the release of CO_2, it represents a waste of the plant's energy. The rate of photorespiration versus CO_2 fixation depends on the CO_2/O_2 ratio faced by Rubisco. C_4 and CAM plants possess mechanisms that increase this ratio. (p. 224)

C_4 and CAM plants possess an additional CO_2-fixing enzyme, called PEP carboxylase, that is able to operate at very low CO_2 concentrations. C_4 plants possess a unique leaf structure containing an outer cylinder of mesophyll cells and an inner cylinder of bundle sheath cells that are sealed off from atmospheric gases. PEP carboxylase operates in the mesophyll, where CO_2 is fixed to the three-carbon compound phosphoenolpyruvate (PEP) to form a four-carbon acid, which is transported to the bundle sheath, where it is decarboxylated. The CO_2 released in the bundle sheath accumulates to high concentration, favoring CO_2 fixation to RuBP and the formation of PGA and GAP by means of the Calvin cycle. CAM plants carry out the light-dependent and light-independent reactions at different times of the day. CAM plants keep their stomata closed during hot, dry daylight hours, thereby avoiding water loss. Then, during the night, they open their stomata and fix CO_2 by means of PEP carboxylase. The malic acid produced is stored in the vacuole until daylight hours, when the compound is moved back into the chloroplast. There it gives up its CO_2, which can be fixed by Rubisco under conditions of low O_2 concentration and converted to carbohydrate using energy from the ATP and NADPH generated by the light-dependent reactions. (p. 226)

ANALYTIC QUESTIONS

1. Which of the two photosystems operates at the most negative redox potential? Which generates the strongest reducing agent? Which must absorb four photons during each round of noncyclic photophosphorylation?

2. Which type of plant (C_3, C_4, or CAM) might you expect to do best if exposed to continuous daylight under hot, dry conditions? Why?

3. Of the following substances, PQH_2, reduced cytochrome b_6, reduced ferredoxin, $NADP^+$, NADPH, O_2, H_2O, which is the strongest reducing agent? Which is the strongest oxidizing agent? Which has the greatest affinity for electrons? Which has the most energetic electrons?

4. Suppose you were to add the uncoupler dinitrophenol (DNP, page 191) to a preparation of chloroplasts carrying out photosynthesis. Which of the following activities would you expect to be affected? (1) absorption of light, (2) cyclic photophosphorylation, (3) electron transport between PSII and PSI, (4) noncyclic photophosphorylation, (5) synthesis of PGA, (6) $NADP^+$ reduction.

5. Calculate the proton-motive force that would be formed across a thylakoid membrane that maintained a 10,000-fold difference in $[H^+]$ and no electric potential difference. (The equation for proton-motive force is on page 191.)

6. Under what conditions would you expect a plant to be engaged most heavily in cyclic photophosphorylation?

7. Contrast the change in pH of the medium that occurs when isolated chloroplasts carry out photosynthesis compared to isolated mitochondria carrying out aerobic respiration.

8. It was noted in the previous chapter that the majority of the proton-motive force in mitochondria is expressed as a voltage. In contrast, the proton-motive force generated during photosynthesis is almost exclusively expressed as a pH gradient. How can you explain these differences?

9. Contrast the roles of PSI and PSII in generating the electrochemical gradient across the thylakoid membrane.

10. Would you agree that a C_3 plant has to expend more energy per molecule of CO_2 converted to carbohydrate than a C_4 plant? Why or why not?

11. In photosynthesis, the capture of light energy results in the release and subsequent transfer of electrons. From what molecules are the electrons originally derived? In what molecules do these electrons eventually reside?

12. How many ATP and NADPH molecules are required in the C_3 pathway to make one six-carbon sugar? If the synthesis of a molecule of ATP were to require four protons, would you expect that these relative requirements for ATP and NADPH could be met by noncyclic photophosphorylation in the absence of cyclic photophosphorylation?

13. If pheophytin and A_0 (a chlorophyll *a* molecule) are the primary electron acceptors of PSII and PSI, respectively, what are the primary electron donors of each photosystem?

14. Contrast the roles of three different metal atoms in the light-dependent reactions of photosynthesis.

15. Would you expect the greenhouse effect (an increase in the CO_2 content of the atmosphere) to have a greater effect on C_4 plants or C_3 plants? Why?

16. Would you expect the water content of the atmosphere to be an important factor in the success of a C_3 plant? Why?

17. It has been suggested that low CO_2 levels play a key role in keeping O_2 levels stable at 21 percent. How is it possible that CO_2 levels could affect O_2 levels in the atmosphere?

18. Suppose CO_2 levels were to rise to 600 ppm, which in fact they probably were about 300 million years ago. What effect would you suppose this might have regarding competition between C_3 and C_4 plants?

19. Suppose you were to place a C_3 plant under hot, dry conditions and provide it with radioactively labeled $^{18}O_2$. In what compounds would you find this label incorporated?

7
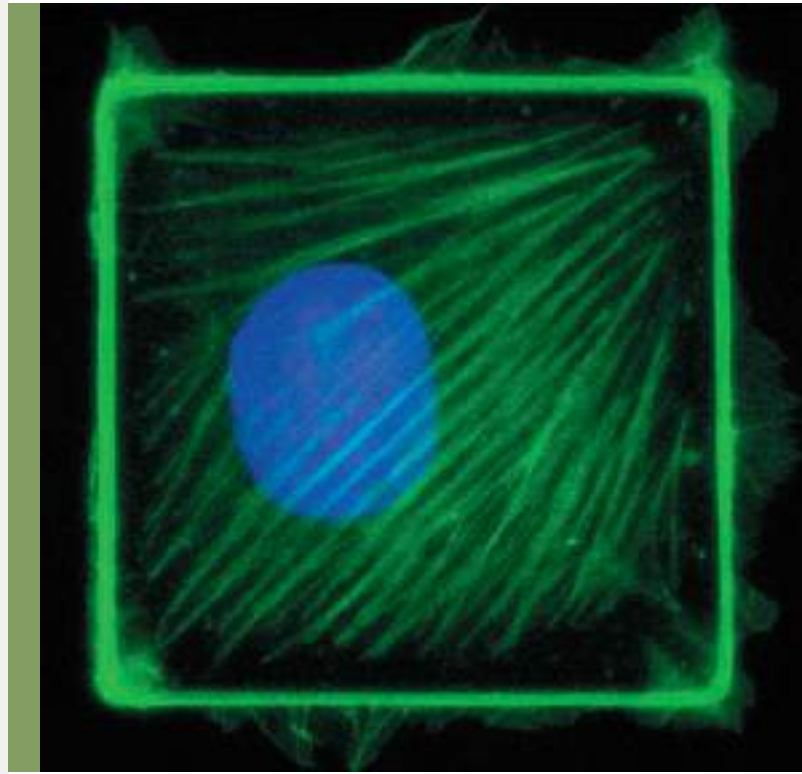
Interactions Between Cells and Their Environment

Although the plasma membrane constitutes the boundary between a living cell and its nonliving environment, materials present outside the plasma membrane play an important role in the life of a cell. Most cells in a multicellular plant or animal are organized into clearly defined tissues in which the component cells maintain a defined relationship with one another and with the extracellular materials that lie between the cells. Even those cells that have no fixed relationship within a solid tissue, such as white blood cells that patrol the body, must interact in highly specific ways with other cells and with extracellular materials with which they come in contact. These interactions regulate such diverse activities as cell migration, cell growth, and cell differentiation, and they determine the three-dimensional organization of tissues and organs that emerges during embryonic development.

We will focus in this chapter on the extracellular environment of cells and the types of interactions in which cells engage. Figure 7.1 shows a diagrammatic section of human skin and provides an overview of several of the subjects to be discussed in the chapter. The outer layer of skin (the epidermis) is a type of **epithelial tissue**. Like other epithelia, which line the spaces within the body, the epidermis consists largely of closely packed cells attached to one another and to an underlying noncellular layer by specialized contacts (depicted in the top inset of Figure 7.1). These contacts provide

Fluorescence micrograph of an endothelial cell—a type of cell that lines the inner surface of blood vessels. The cell is square because it has spread itself over a tiny square-shaped patch of an adhesive protein called fibronectin that was applied to a culture dish. The cell appears to be mounted in a green frame because it was treated with a green fluorescent antibody that binds to the cytoplasmic protein actin, a component of the cytoskeleton. (Reprinted from Christopher S. Chen, Clifford Brangwynne, and Donald E. Ingber, Trends Cell Biol. 9:283, 1999; Copyright 1999, with permission from Elsevier Science.)
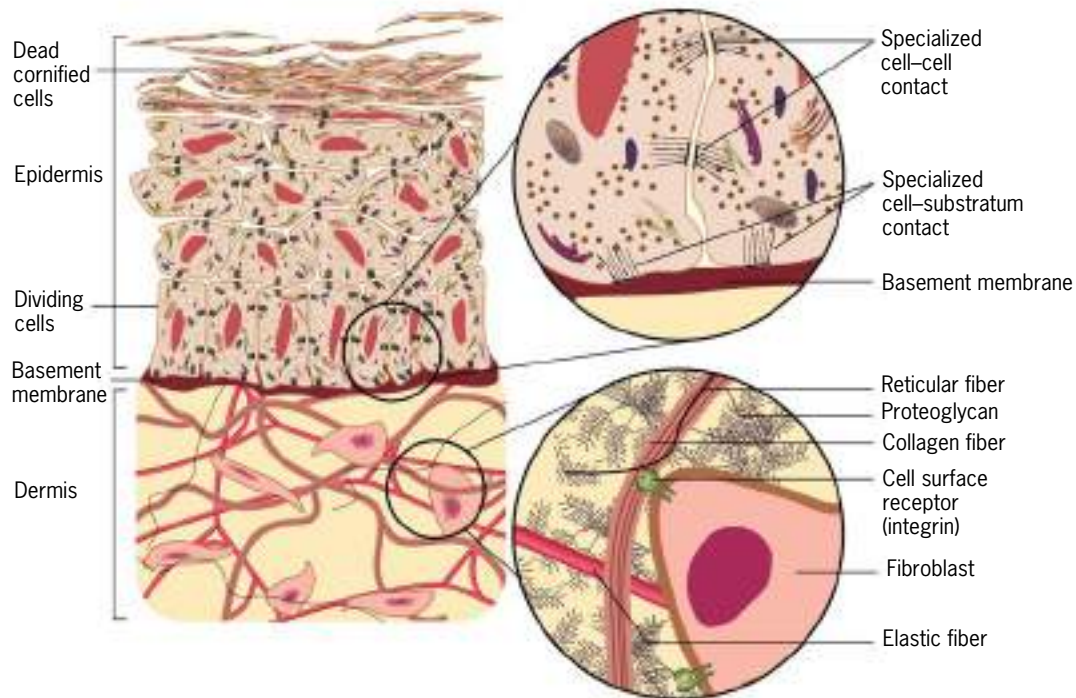
FIGURE 7.1 An overview of how cells are organized into tissues and how they interact with one another and with their extracellular environment. In this schematic diagram of a section through human skin, the cells of the epidermis are seen to adhere to one another by specialized contacts. The basal layer of epidermal cells also adheres to an underlying, noncellular layer (the basement membrane). The dermis consists largely of extracellular elements that interact with each other and with the surfaces of scattered cells (primarily fibroblasts). The cells contain receptors that interact with extracellular materials and transmit signals to the cell interior.

a mechanism for cells to adhere to and communicate with one another. In contrast, the deeper layer of the skin (the dermis) is a type of **connective tissue**. Like other connective tissues, such as those that form a tendon or cartilage, the dermis consists largely of extracellular material, including a variety of distinct fibers that interact with each other in specific ways. A closer look at one of the scattered cells (fibroblasts) of the dermis shows that the outer surface of the plasma membrane contains receptors that mediate interactions between the cell and the components of its evironment (depicted in the lower inset of Figure 7.1). These cell-surface receptors interact not only with the external surroundings, but are connected at their internal end to various cytoplasmic proteins. Receptors with this type of dual attachment are well suited to transmit messages between a cell and its environment. ■

7.1 THE EXTRACELLULAR SPACE

If we begin at the plasma membrane and move outward, we can examine the types of extracellular elements that surround various types of cells. It was noted in Chapter 4 that virtually all integral membrane proteins, as well as certain membrane lipids, bear chains of sugars (oligosaccharides) of variable length that project outward from the plasma membrane (see Figure 4.4c). These carbohydrate projections form part of a layer closely applied to the outer surface of the plasma membrane called the **glycocalyx** (or *cell coat*) (Figure 7.2a). This extracellular material is very prominent in some types of cells, such as the epithelial cells that line the mammalian digestive
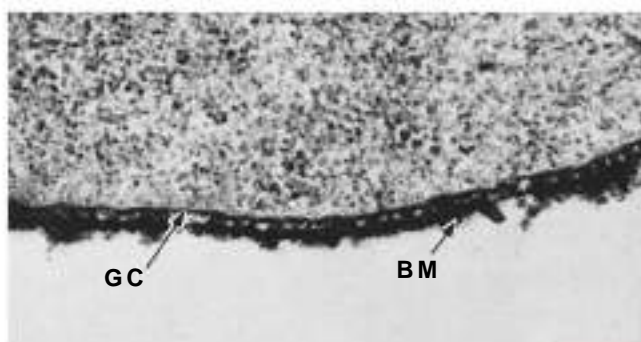
(a)
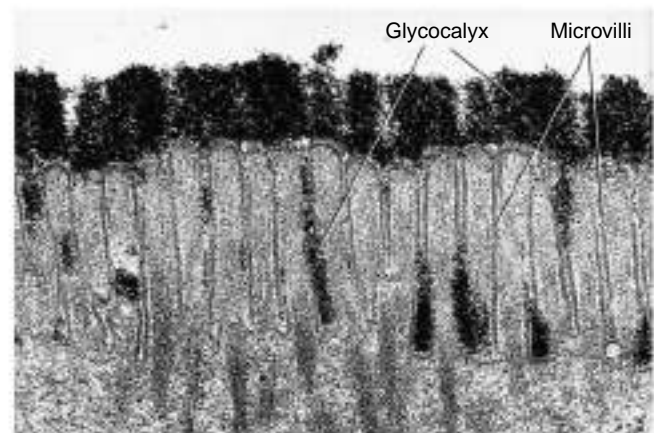
(b)

FIGURE 7.2 The glycocalyx. (a) The basal surface of an ectodermal cell of an early chick embryo. Two distinct structures closely applied to the outer cell surface can be distinguished: an inner glycocalyx (GC) and an outer basement membrane (BM). (b) Electron micrograph of the apical surface of an epithelial cell from the lining of the intestine, showing the extensive glycocalyx, which has been stained with the iron-containing protein ferritin. (A: FROM A. MARTINEZ-PALOMO, INT. REV. CYTOL. 29:64, 1970; B: FROM S. ITO AND D. W. FAWCETT/PHOTO RESEARCHERS.)

tract (Figure 7.2*b*). The glycocalyx is thought to mediate cell–cell and cell–substratum interactions, provide mechanical protection to cells, serve as a barrier to particles moving toward the plasma membrane, and bind important regulatory factors that act on the cell surface.

The Extracellular Matrix

Many types of animal cells are surrounded by an **extracellular matrix (ECM)**—an organized network of extracellular materials that is present beyond the immediate vicinity of the plasma membrane (Figure 7.3). The ECM is more than an inert packing material or a nonspecific glue that holds cells
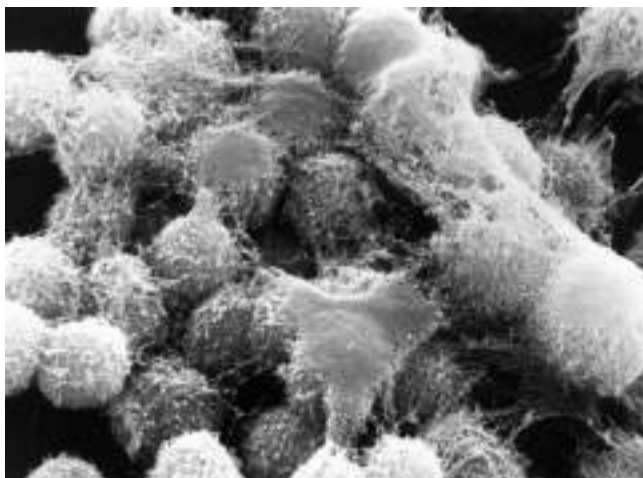
(a)
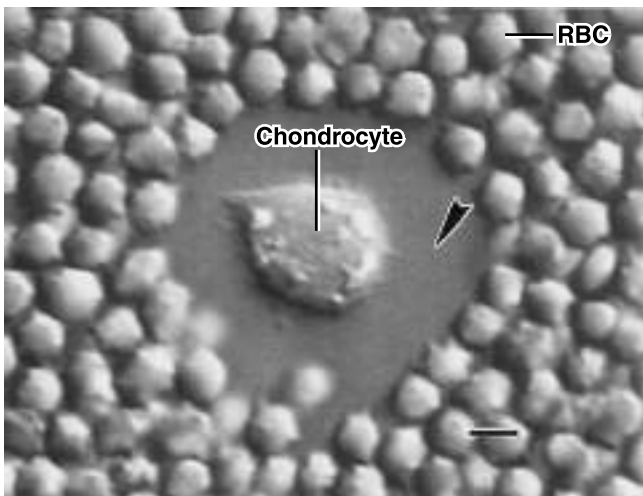
(b)

FIGURE 7.3 **The extracellular matrix (ECM) of cartilage cells.** (*a*) Scanning electron micrograph of a portion of a colony of cartilage cells (chondrocytes) showing the extracellular materials secreted by the cells. (*b*) The ECM of a single chondrocyte has been made visible by adding a suspension of red blood cells (RBCs). The thickness of the ECM is evident by the clear space (arrowhead) that is not penetrated by the RBCs. The bar represents 10 μm. (A: MICHAEL SOLURSH AND GERALD KARP; B: COURTESY OF GRETA M. LEE, BRIAN JOHNSTON, KEN JACOBSON, AND BRUCE CATERSON.)

together; it often plays a key regulatory role in determining the shape and activities of the cell. For example, enzymatic digestion of the ECM that surrounds cultured cartilage cells or mammary gland epithelial cells causes a marked decrease in the synthetic and secretory activities of the cells. Addition of extracellular matrix materials back to the culture dish can restore the differentiated state of the cells and their ability to manufacture their usual cell products (see Figure 7.29).

One of the best defined extracellular matrices is the **basement membrane** (or basal lamina), a continuous sheet 50 to 200 nm thick that (1) surrounds nerve fibers, muscles, and fat cells, (2) underlies the basal surface of epithelial tissues, such as the epidermis of the skin (Figures 7.1 and 7.4*a*), or the lining of the digestive and respiratory tracts, and (3) underlies the inner endothelial lining of blood vessels. Basement membranes provide mechanical support for the attached cells, generate signals that maintain cell survival, serve as a substratum for cell migration, separate adjacent tissues within an organ, and act as a barrier to the passage of macromolecules. In this latter capacity, basement membranes of the capillaries play an important role in preventing the passage of proteins out of the blood and into the tissues. This is particularly important in the kidney, where the blood is filtered under high pressure through a double-layered basement membrane that separates the capillaries of the glomerulus from the wall of the kidney tubules (Figure 7.4*b*). Kidney failure in long-term diabetics may result from an abnormal thickening of the basement membranes surrounding the glomeruli. Basement membranes also serve as a barrier to invasion of tissues by cancer cells. The molecular organization of basement membranes is discussed later (see Figure 7.12).

Even though the extracellular matrix may take diverse forms in different tissues and organisms, it tends to be composed of similar macromolecules. Unlike most proteins present inside of cells, which are compact, globular molecules, those of the extracellular space are typically extended, *fibrous* species. These proteins are secreted into the extracellular space where they are capable of self-assembling into an interconnected three-dimensional network, which is depicted in Figure 7.5 and discussed in the following sections. Among their diverse functions, the proteins of the ECM serve as trail markers, scaffolds, girders, wire, and glue. As noted throughout the following discussion, alterations in the amino acid sequence of extracellular proteins can lead to serious disorders. We will begin with one of the most important, and ubiquitous, molecules of the ECM, the glycoprotein collagen.

Collagen   **Collagens** comprise a family of fibrous glycoproteins that are present only in extracellular matrices. Collagens are found throughout the animal kingdom and are noted for their high tensile strength, that is, their resistance to pulling forces. It is estimated that a collagen fiber 1 mm in diameter is capable of suspending a weight of 10 kg (22 lb) without breaking. Collagen is the single most abundant protein in the human body (constituting more than 25 percent of all protein), a fact that reflects the widespread occurrence of extracellular materials.

(a)

10 μm

(b)
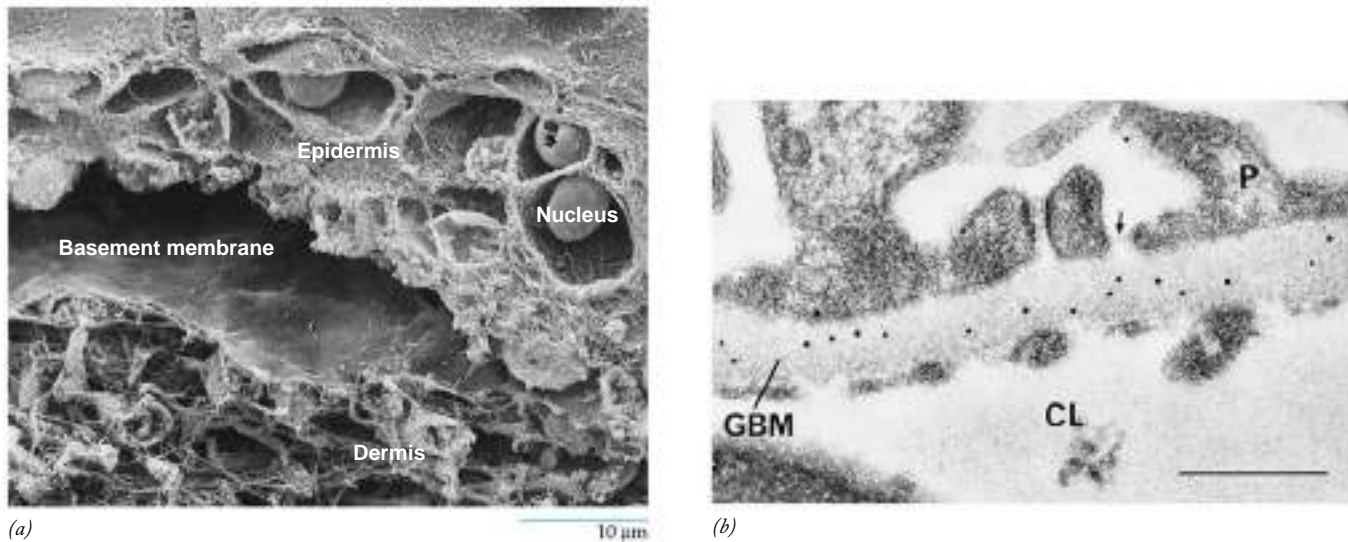
FIGURE 7.4 The basement membrane (basal lamina). (*a*) Scanning electron micrograph of human skin. The epidermis has pulled away from part of the basement membrane, which is visible beneath the epidermal cells. (*b*) An unusually thick basement membrane is formed between the blood vessels of the glomerulus and the proximal end of the renal tubules of the kidney. This extracellular layer plays an important role in filtering the fluid that is forced out of the capillaries into the renal tubules during urine formation. The black dots within the glomerular basement membrane (GBM) are gold particles attached to antibodies that are bound to type IV collagen molecules in the basement membrane (CL, capillary lumen; P, podocyte of tubule). The bar represents 0.5 μm. (A: COURTESY OF KAREN HOLBROOK; B: FROM MICHAEL DESJARDINS AND M. BENDAYAN, J. CELL BIOL. 113:695, 1991; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

Collagen is produced primarily by fibroblasts, the cells found in various types of connective tissues, and also by smooth muscle cells and epithelial cells. To date, 27 distinct types of human collagen have been identified. Each collagen type is restricted to particular locations within the body, but two or more different types are often present together in the same ECM. Additional functional complexity is provided by mixing several collagen types within the same fiber. These
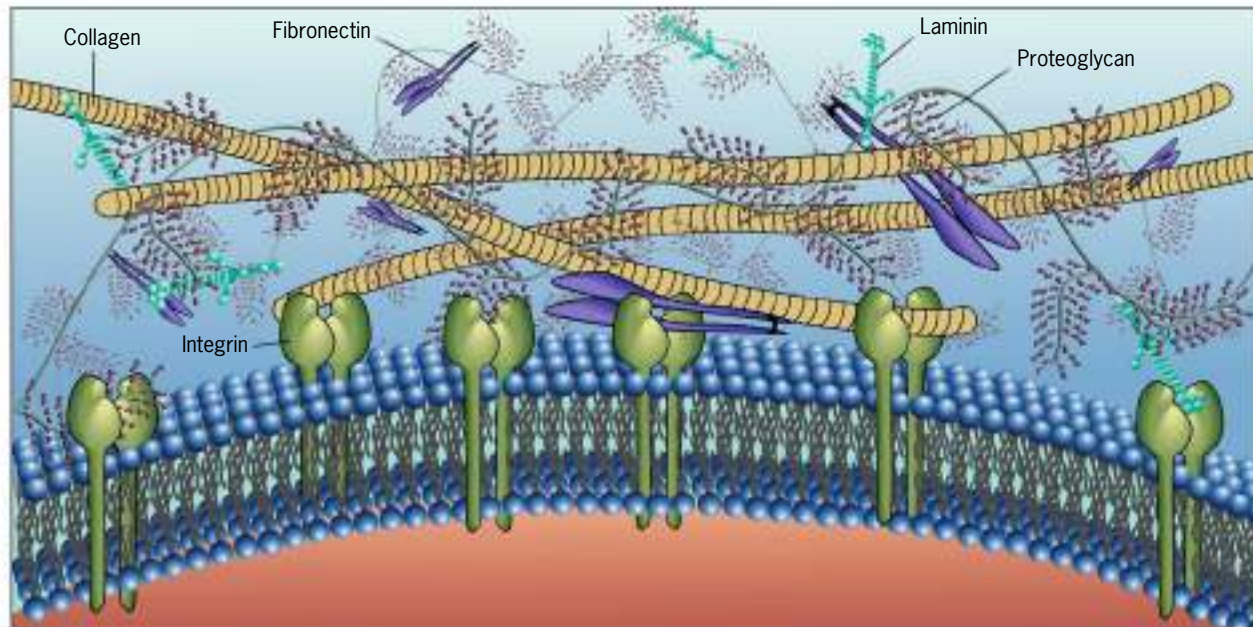
FIGURE 7.5 An overview of the macromolecular organization of the extracellular matrix. The proteins and polysaccharides shown in this illustration will be discussed in the following sections. The proteins depicted (fibronectin, collagen, and laminin) contain binding sites for one another, as well as binding sites for receptors (integrins) that are located at the cell surface. The proteoglycans are huge protein–polysaccharide complexes that occupy much of the volume of the extracellular space.

"heterotypic" fibers are the biological equivalent of a metal alloy. It is likely that different structural and mechanical properties result from different mixtures of collagens in fibers. Although there are many differences among the members of the collagen family, all share at least two important structural features. (1) All collagen molecules are trimers consisting of three polypeptide chains, called α chains. (2) Along at least part of their length, the three polypeptide chains of a collagen molecule are wound around each other to form a unique, rod-like triple helix (Figure 7.6a).

The α chains of collagen molecules contain large amounts of proline, and many of the proline (and lysine) residues are hydroxylated following synthesis of the polypeptide. The hydroxylated amino acids are important in maintaining the stability of the triple helix by forming hydrogen bonds between component chains. Failure to hydroxylate collagen chains has serious consequences for the structure and function of connective tissues. This is evident from the symptoms of scurvy, a disease that results from a deficiency of vitamin C (ascorbic acid) and is characterized by inflamed gums and tooth loss, poor wound healing, brittle bones, and the weakening of the lining of blood vessels, causing internal bleeding. Ascorbic acid is required as a coenzyme by the enzymes that add hydroxyl groups to the lysine and proline amino acids of collagen.

A number of the collagens, including types I, II, and III, are described as *fibrillar collagens* because they assemble into rigid, cable-like fibrils, which in turn become packaged into thicker fibers that are typically large enough to be seen in the light microscope. The side-by-side packaging of the rows of collagen I molecules within a collagen fibril is depicted in Figure 7.6b,c. The fibrils are strengthened by covalent cross-links between lysine and hydroxylysine residues on adjacent collagen molecules. This cross-linking process continues through life and may contribute to the decreased elasticity of skin and increased brittleness of bones among the elderly. Recent studies that probe the surface of collagen fibrils suggest that they are composed of strands that are twisted around the axis of the fibril to form a structure that is more like a nanosized rope than a cable (Figure 7.6d).

Among the various components of the ECM, collagen molecules provide the insoluble framework that determines many of the mechanical properties of the matrix. In fact, the properties of a particular tissue can often be correlated with the three-dimensional organization of its collagen molecules. For example, tendons, which connect muscles to bones, must resist tremendous pulling forces during times of muscular contraction. Tendons contain an ECM in which the collagen fibrils are aligned parallel to the long axis of the tendon and thus parallel to the direction of the pulling forces. The cornea is also a remarkable tissue; it must serve as a durable, protective layer at the surface of the eyeball but must also be transparent so that light can pass through the lens to the

(a)

(b)

(c) 0.3 μm

(d) 250 nm

FIGURE 7.6 The structure of collagen I. This figure depicts several levels of organization of a fibrillar collagen. (*a*) The collagen molecule (or monomer) is a triple helix composed of three helical α chains. Some types of collagen contain three identical α chains and thus are homotrimers, while others are heterotrimers containing two or three different chains. Each collagen I molecule is 295 nm in length. (*b*) Collagen I molecules become aligned in rows in which the molecules in one row are staggered relative to those in the neighboring row. A bundle of collagen I molecules, such as that shown here, form a collagen fibril. The staggered arrangement of the molecules produces bands (horizontal black lines in the illustration) across the fibril that repeat with a periodicity of 67 nm (which is equal to the length of the gap between molecules plus the overlap). (*c*) An electron micrograph of human collagen fibrils observed after metal shadowing (see Figure 18.15). The banding pattern of the fibrils is evident. (*d*) An atomic force micrograph showing the surface of a collagen fibril that suggests that its subcomponents are twisted spirally around the fibril axis like a rope. The banding pattern remains evident. The arrows point to places where the fibril is slightly unwound, as would occur if you were to twist a rope in the opposite direction it is wound. (C: COURTESY OF JEROME GROSS AND FRANCIS O. SCHMITT; D: FROM LAURENT BOZEC ET AL., BIOPHYS. J. 92:71, 2007.)

200 nm
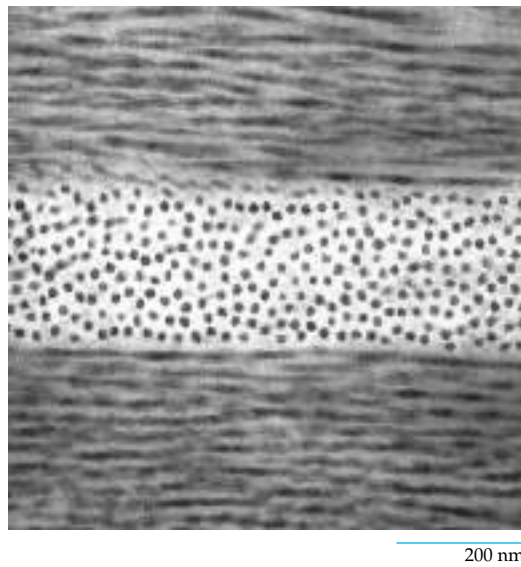
FIGURE 7.7 **The corneal stroma** consists largely of layers of collagen fibrils of uniform diameter and spacing. The molecules of alternate layers are arranged at right angles to one another, resembling the structure of plywood. (REPRINTED FROM NIGEL J. FULLWOOD, STRUCTURE 12:169, 2004; COPYRIGHT 2004, WITH PERMISSION FROM ELSEVIER SCIENCE.)

50 nm

FIGURE 7.8 **The type IV collagen network of the basement membrane.** Electron micrograph of a basement membrane from human amniotic tissue that had been extracted with a series of salt solutions to remove noncollagenous materials. The treatment leaves behind an extensive, branching, polygonal network of strands that form an irregular lattice. Evidence indicates that this lattice consists of type IV collagen molecules covalently linked to one another in a complex three-dimensional array. A model of the basement membrane scaffold is shown in Figure 7.12. (B: FROM PETER D. YURCHENCO AND GEORGE C. RUBEN, J. CELL BIOL. 105:2561, 1987; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

retina. The thick, middle layer of the cornea is the stroma, which contains relatively short collagen fibrils that are organized into distinct layers. The layered structure of the stroma is similar to that of plywood: the fibrils of each layer are parallel to other fibrils in the layer but nearly perpendicular to the fibrils in the layer on either side (Figure 7.7). This plywood-like structure provides strength to this delicate tissue, while the uniformity of size and ordered packing of the fibrils minimize the scattering of incoming light rays, thereby promoting the tissue's transparency.

Given their abundance and widespread distribution, it is not surprising that abnormalities in fibrillar collagen formation can lead to serious disorders. Burns or traumatic injuries to internal organs can lead to an accumulation of scar tissue, which consists largely of fibrillar collagen. Mutations in genes encoding type I collagen can produce *osteogenesis imperfecta*, a potentially lethal condition characterized by extremely fragile bones, thin skin, and weak tendons. Mutations in genes encoding type II collagen alter the properties of cartilage tissue, causing dwarfism and skeletal deformities. Mutations in a number of other collagen genes can lead to a variety of distinct but related defects in collagen matrix structure (called *Ehler-Danlos syndromes*). Persons with one of these syndromes have hyperflexible joints and highly extensible skin.

Not all collagens form fibrils. One of the nonfibrillar collagens is type IV, whose distribution is restricted to basement membranes. Basement membranes are thin, supportive sheets, and the type IV collagen molecules are organized into a network that provides mechanical support and serves as a lattice for the deposition of other extracellular materials (see Figure 7.12). Unlike type I collagen, which consists of a long, un-

interrupted triple helix, the type IV collagen trimer contains nonhelical segments interspersed along the molecule and globular domains at each end. The nonhelical segments make the molecule flexible, while the globular ends serve as sites of interaction between molecules that give the complex its lattice-like character (Figure 7.8). Mutations in type IV collagen genes have been identified in patients with *Alport syndrome*, an inherited kidney disease in which the glomerular basement membrane (Figure 7.4b) is disrupted.

Proteoglycans In addition to collagen, basement membranes and other extracellular matrices typically contain large amounts of a distinctive type of protein–polysaccharide complex called a **proteoglycan**. A proteoglycan (Figure 7.9a) consists of a core protein molecule (shown in black in Figure 7.9a) to which chains of *glycosaminoglycans* (*GAGs*) are covalently attached (shown in red in the figure). Each glycosaminoglycan chain is composed of a repeating disaccharide; that is, it has the structure -A-B-A-B-A-, where A and B represent two different sugars. GAGs are highly acidic due to the presence of both sulfate and carboxyl groups attached to the sugar rings (Figure 7.9b). Proteoglycans of the extracellular matrix may be assembled into gigantic complexes by linkage of their core
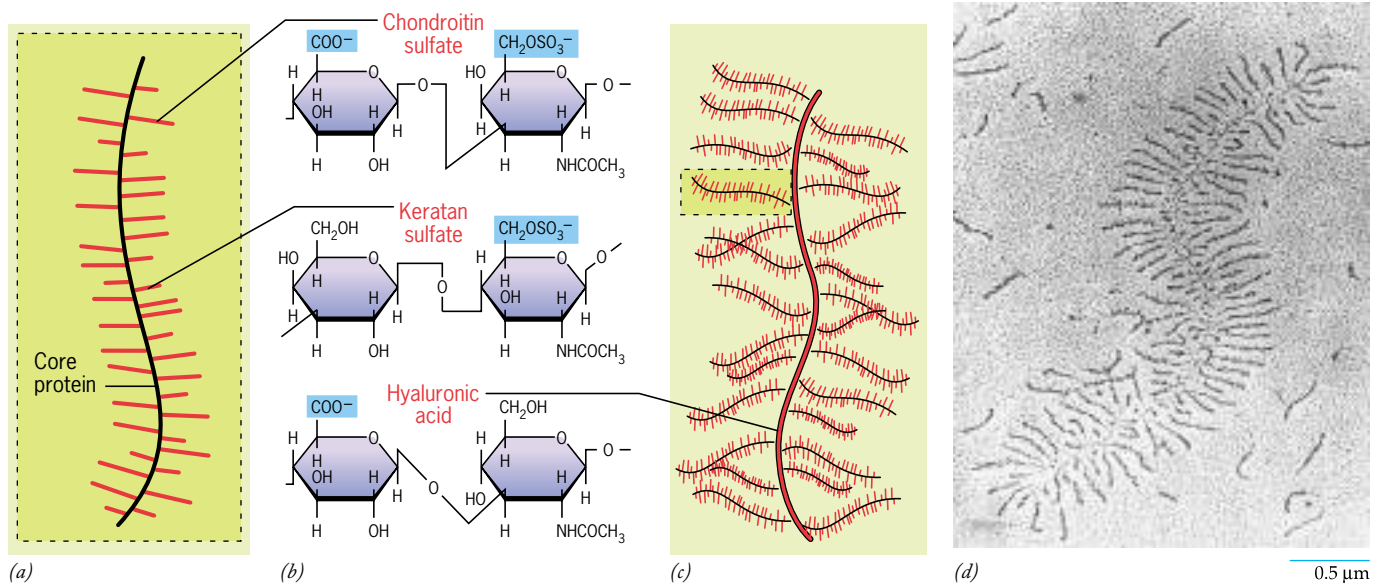
FIGURE 7.9 The structure of a cartilage-type proteoglycan complex. (*a*) Schematic representation of a single proteoglycan consisting of a core protein to which a large number of glycosaminoglycan chains (GAGs, shown in red) are attached. A proteoglycan from cartilage matrix (e.g., aggrecan) may contain about 30 keratan sulfate and 100 chondroitin sulfate chains. Proteoglycans found in basement membranes (e.g., perlecan and agrin) have only a few GAG chains attached to the core protein. (*b*) The structures of the repeating disaccharides that make up each of the GAGs shown in this figure. All GAGs bear large numbers of negative charges (indicated by the blue shading). (*c*) In the cartilage matrix, individual proteoglycans are linked to a nonsulfated GAG, called hyaluronic acid (or hyaluronan), to form a giant complex with a molecular mass of about 3 million daltons. The box indicates one of the proteoglycans of the type shown in part *a*. (*d*) Electron micrograph of a proteoglycan complex, comparable to that illustrated in part *c*, that was isolated from cartilage matrix. (D: COURTESY OF LAWRENCE C. ROSENBERG.)

proteins to a molecule of *hyaluronic acid*, a nonsulfated GAG (Figure 7.9*c*). The microscopic appearance of one of these complexes, which can occupy a volume equivalent to that of a bacterial cell, is shown in Figure 7.9*d*.

Because of the negative charges borne on the sulfated GAGs, proteoglycans bind huge numbers of cations, which in turn bind large numbers of water molecules. As a result, proteoglycans form a porous, hydrated gel that fills the extracellular space like packing material (Figure 7.5) and resists crushing (compression) forces. This property complements that of the adjacent collagen molecules, which resist pulling forces and provide a scaffold for the proteoglycans. Together, collagens and proteoglycans give cartilage and other extracellular matrices strength and resistance to deformation. The extracellular matrix of bone is also composed of collagen and proteoglycans, but it becomes hardened by impregnation with calcium phosphate salts.

Fibronectin The term *matrix* implies a structure made up of a network of interacting components. The term has proven very apt for the extracellular matrix, which contains a number of proteins, in addition to collagen and proteoglycans, that interact with one another in highly specific ways (Figure 7.5). *Fibronectin*, like several other proteins to be discussed in this chapter, consists of a linear array of distinct "building blocks" that gives each polypeptide a modular construction (Figure 7.10*a*). Each fibronectin polypeptide is constructed from a sequence of approximately 30 independently folding Fn modules, two of which are depicted in the top inset of Figure 7.10*a*. While Fn-type modules were first discovered in fibronectin, they are found as part of many other proteins, ranging from blood clotting factors to membrane receptors (see Figure 7.22). As discussed on page 58, the presence of shared segments among diverse proteins strongly suggests that many present-day genes have arisen during evolution by the fusion of parts of separate ancestral genes. In fibronectin, the 30 or so structural modules combine to form five or six larger functional domains, illustrated by the colored cylinders in Figure 7.10*a*. Each of the two polypeptide chains that make up a fibronectin molecule contains

1. Binding sites for numerous components of the ECM, such as collagens, proteoglycans, and other fibronectin molecules. These binding sites facilitate interactions that link these diverse molecules into a stable, interconnected network (Figure 7.5).

2. Binding sites for receptors on the cell surface. These binding sites hold the ECM in a stable attachment to the cell (Figure 7.5). The importance of a fibronectin-containing substratum for cell attachment is illustrated in the chapter-opening micrograph on page 230. The cultured endothelial cell shown in this photo has adopted a shape unlike any it would have in the body because it has spread itself over the available surface provided by a square coat of fibronectin.

The importance of fibronectin and other extracellular proteins is particularly evident during embryonic develop-

(a)

(b)

(c)
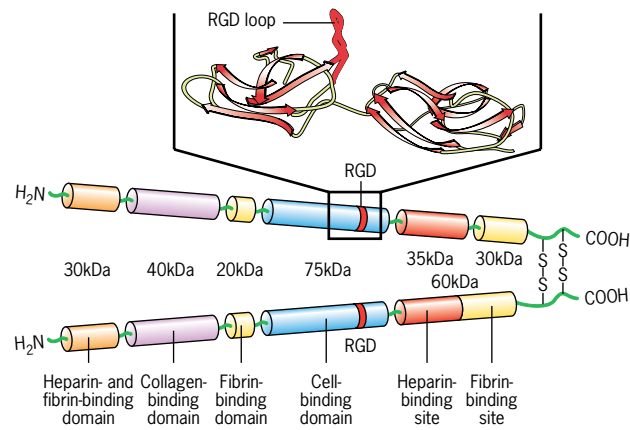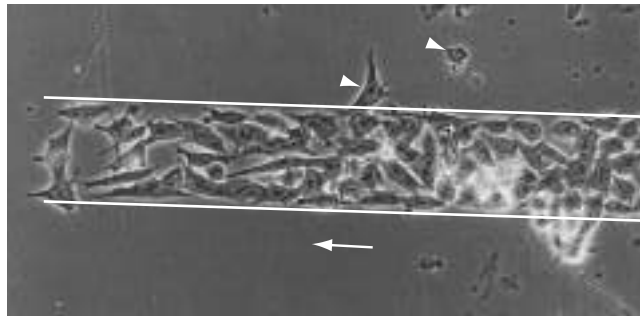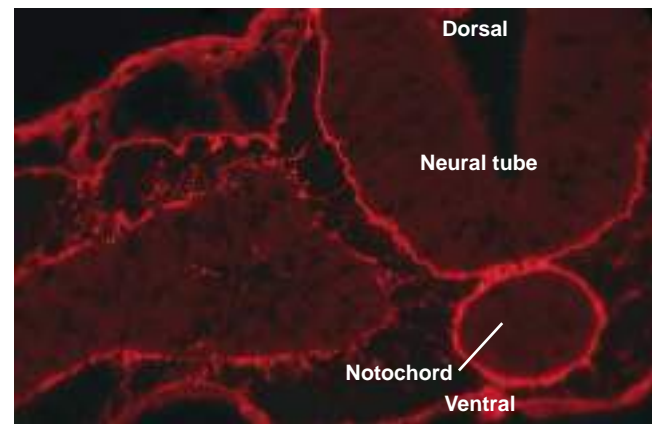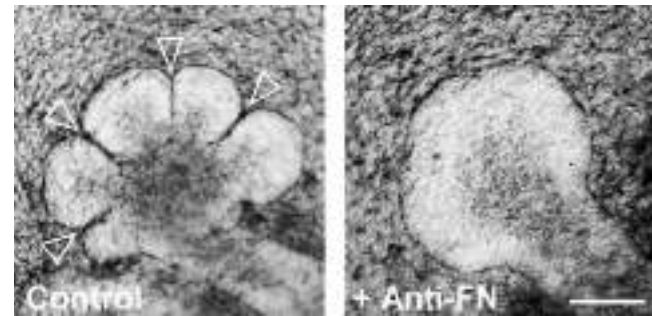
(d)   *(e)*

FIGURE 7.10 Structure of fibronectin and its importance during embryonic development. (*a*) A human fibronectin molecule consists of two similar, but nonidentical, polypeptides joined by a pair of disulfide bonds located near the C-termini. Each polypeptide is composed of a linear series of distinct modules, which are organized into several larger functional units, illustrated by the colored cylinders in this figure. Each of these functional units contains one or more binding sites for either a specific component of the ECM or for the surface of cells. Some of these binding activities are indicated by the labels. The cell-binding site of the polypeptide that contains the sequence arg-gly-asp, or RGD, is indicated. As discussed later in the chapter, this sequence binds specifically to a particular class of integral plasma membrane proteins (integrins) that are involved in cell attachment and signal transduction. The inset shows two of the 30 or so repeating Fn modules that comprise the polypeptide; the RGD sequence forms a loop of the polypeptide, protruding from one module. (*b*) A section through a young chick embryo that has been treated with fluorescent antibodies against fibronectin. Fibronectin is present as fibrils in the basement membranes (dark red sites) that lie beneath the embryonic epithelia and provide a substratum over which cells migrate. (*c*) In this micrograph, neural crest cells are migrating out from a portion of the developing chick nervous system (beyond the edge of the photo) onto a glass culture dish that contains strips of fibronectin-coated surface alternating with strips of bare glass. The boundary of a fibronectin-coated region is indicated by the white lines. It is evident that the cells remain exclusively on the fibronectin-coated surface. Cells that reach the glass substratum (arrowheads) tend to round up and lose their migratory capabilities. The arrow indicates the direction of migration. (*d*, *e*) The role of fibronectin in the formation of the embryonic salivary gland. The micrograph in *d* shows a mouse embryonic salivary gland that has been grown for 10 hours in culture. The gland is seen to be divided into separate buds by a series of clefts (open triangles). The gland shown in *e* has been grown for the same period of time in the presence of an anti-fibronectin antibody, which has completely inhibited cleft formation. Scale bar equals 100 μm. (B: COURTESY OF JAMES W. LASH; C: FROM GIOVANNI LEVI, JEAN-LOUP DUBAND, AND JEAN PAUL THIERY, INT. REV. CYTOL. 123:213, 1990; D,E: REPRINTED FROM TAKAYOSHI SAKAI, MELINDA LARSEN, & KENNETH M. YAMADA, NATURE 423:877, 2003; COPYRIGHT 2003 MACMILLAN MAGAZINES LTD.)

ment. Development is characterized by waves of cell migration during which different cells follow different routes from one part of the embryo to another (Figure 7.11*a*). Migrating cells are guided by proteins, such as fibronectin, that are contained within the molecular landscape through which they pass. For example, the cells of the neural crest, which migrate out of the developing nervous system into virtually all parts of the embryo, traverse pathways rich in fibrils composed of interconnected fibronectin molecules (Figure 7.10*b*). The importance of fibronectin in neural crest cell migration is readily revealed on a culture dish as shown in Figure 7.10*c*. A number of or-

gans of the body, including the salivary gland, kidney, and lung are formed by a process of branching in which the epithelial layer becomes divided by a series of clefts (Figure 7.10*d*). The importance of fibronectin in the formation of these clefts is illustrated in Figure 7.10*e*, which shows a salivary gland that has been incubated with antibodies that bind to fibronectin. Cleft formation and branching is entirely abolished as a result of the inactivation of fibronectin molecules.

Laminin   *Laminins* are a family of extracellular glycoproteins that consist of three different polypeptide chains linked by

(a)
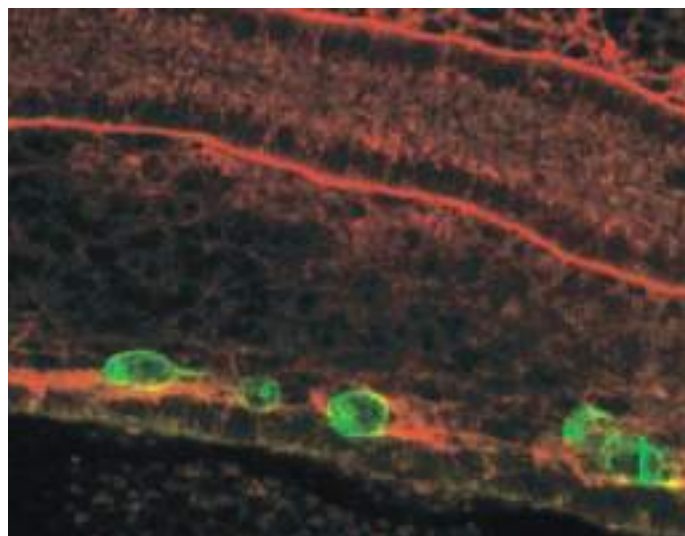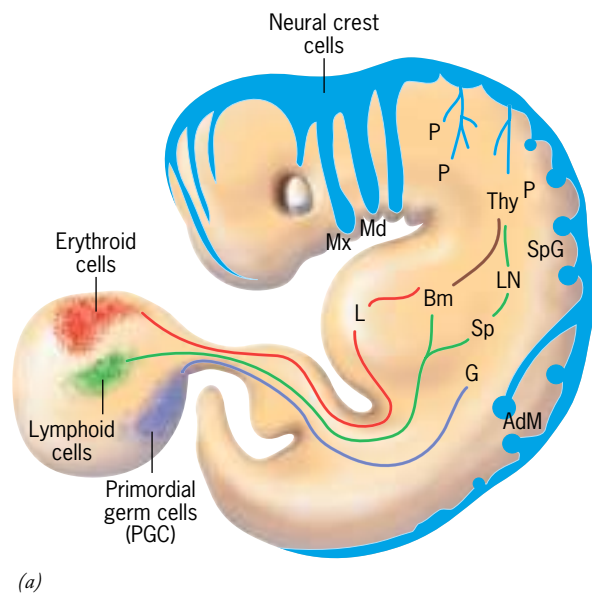
(b)　　20 µm

FIGURE 7.11　The role of cell migration during embryonic development. (*a*) A summary of some of the cellular traffic occurring during mammalian development. The most extensive movements are conducted by the neural crest cells (shown in blue), which migrate out of the neural plate in the dorsal midline of the embryo and give rise to all the pigment cells of the skin (P), the sympathetic ganglia (SpG), adrenal medulla (AdM), and the cartilage of the embryonic skull (Mx, Md for maxillary and mandibular arches). The primordial germ cells (PGC) migrate from the yolk sac to the site of gonad (G) formation within the embryo. The progenitors of lymphoid cells are transported to the liver (L), bone marrow (Bm), thymus (Thy), lymph nodes (LN), and spleen (Sp). (*Note:* The "pathways" shown here connect the original sites of cells with their destinations; they do not accurately depict the actual routes of the cells.) (*b*) Micrograph of a section of a portion of the hind gut of a 10-day mouse embryo. The primordial germ cells (green) are seen to be migrating along the dorsal mesentery on their way to the developing gonad. The tissue has been stained with antibodies against the protein laminin (red), which is seen to be concentrated in the surface over which the cells are migrating. (A: AFTER AARON A. MOSCONA AND R. E. HAUSMAN, IN CELL AND TISSUE INTERACTIONS, J. W. LASH AND M. M. BURGER, EDS, RAVEN PRESS, 1977; B: FROM MARTIN I. GARCIA-CASTRO, ROBERT ANDERSON, JANET HEASMAN, AND CHRISTOPHER WYLIE, J. CELL BIOL. 138:475, 1997; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

disulfide bonds and organized into a molecule resembling a cross with three short arms and one long arm (Figure 7.5). At least 15 different laminins have been identified. Like fibronectin, extracellular laminins can greatly influence a cell's potential for migration, growth, and differentiation. For example, laminins play a critical role in the migration of primordial germ cells (Figure 7.11*a*). These cells arise in the yolk sac, which is located outside the embryo itself, and then migrate by way of the bloodstream and embryonic tissues to the developing gonad, where they eventually give rise to sperm or eggs. During their migration, the primordial germ cells traverse surfaces that are particularly rich in laminin (Figure 7.11*b*). Studies indicate that the primordial germ cells possess a cell-surface protein that adheres strongly to one of the subunits of the laminin molecule. The influence of laminin on nerve outgrowth is shown in Figure 9.48*a*.

In addition to binding tightly to cell-surface receptors, laminins can bind to other laminin molecules, to proteoglycans, and to other components of basement membranes (Figure 7.5). The laminin and type IV collagen molecules of basement membranes are thought to form separate, but interconnected, networks, as depicted in Figure 7.12. These interwoven networks give basement membranes both strength and flexibility. In fact, basement membranes fail to form in mouse embryos that are unable to synthesize laminin, causing the death of the embryo around the time of implantation.

Dynamic Properties　The micrographs and diagrams presented in the first section of this chapter portray the ECM as a static structure. In actual fact, however, the ECM can exhibit dynamic properties, both in space and over time. Spatially, for example, ECM fibrils can be seen to stretch several times their normal length as they are pulled on by cells and contract when tension is relieved. Temporally, the components of an ECM are subject to continual degradation and reconstruction. These processes serve to renew the matrix and to allow it to be remodeled during embryonic development or following tissue injury. Even the calcified matrix of our bones, which we think of as a stable, inert structure, is subject to continual restoration.

The degradation of extracellular materials, along with cell-surface proteins, is accomplished largely by a family of zinc-containing enzymes called **matrix metalloproteinases (MMPs)** that are either secreted into the extracellular space or anchored to the plasma membrane. As a group, MMPs can digest nearly all of the diverse ECM components, although individual family members are limited as to the types of extracellular molecules they can attack. The physiological
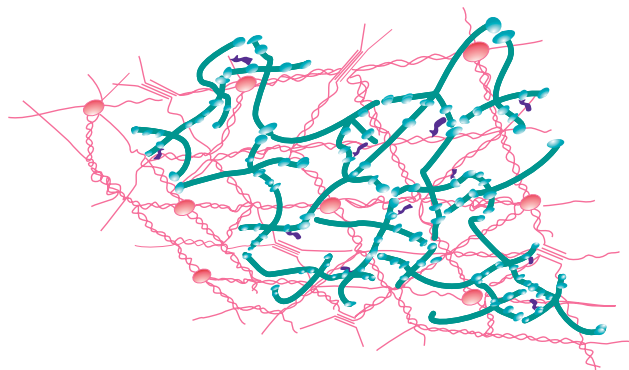
FIGURE 7.12 A model of the basement membrane scaffold. Basement membranes contain two network-forming molecules, collagen IV (pink), which was illustrated in Figure 7.8, and laminin (green), which is indicated by the thickened cross-shaped molecules. The collagen and laminin networks are connected by entactin molecules (purple). (AFTER PETER D. YURCHENCO, YI-SHAN CHENG, AND HOLLY COLOGNATO, J. CELL BIOL. 117:1132, 1992; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

roles of MMPs are not well understood, but they are thought to be involved in tissue remodeling, embryonic cell migration, wound healing, and the formation of blood vessels. As might be expected from enzymes whose normal function is to destroy extracellular materials, the excessive or inappropriate activity of MMPs is likely to cause disease. In fact, MMPs have been implicated in a number of pathological conditions, including arthritis, hepatitis, atherosclerosis, tooth and gum disease, and tumor progression (page 248). At least three inherited skeletal disorders have been traced to mutations in MMP genes.

REVIEW ?

1. Distinguish between the glycocalyx, a basement membrane, and the extracellular matrix of cartilage tissue.
2. Contrast the roles of collagen and proteoglycans in the extracellular space. How do fibronectin and laminin contribute to embryonic development?
3. List a few of the functions of the extracellular matrix in animal tissues.

7.2 INTERACTIONS OF CELLS WITH EXTRACELLULAR MATERIALS

It was noted in the previous discussion that components of the ECM, such as fibronectin, laminin, proteoglycans, and collagen, are capable of binding to receptors situated on the cell surface (as in Figure 7.5). The most important family of receptors that attach cells to their extracellular microenvironment is the integrins.

Integrins

Integrins, which are found only in animals, are a family of membrane proteins that play a key role in integrating the extracellular and intracellular environments. On one side of the plasma membrane, integrins bind to a remarkably diverse array of molecules (ligands) that are present in the extracellular environment. On the intracellular side of the membrane, integrins interact either directly or indirectly with dozens of different proteins to influence the course of events within the cell. Integrins are composed of two membrane-spanning polypeptide chains, an α chain and a β chain, that are noncovalently linked. Eighteen different α subunits and eight different β subunits have been identified. Although more than a hundred possible pairings of α and β subunits could theoretically occur, only about two dozen different integrins have been identified on the surfaces of cells, each with a specific distribution within the body. Whereas many subunits occur in only a single integrin heterodimer, the β_1 subunit is found in 12 different integrins and the α_v subunit in 5 of them (see Table 7.1). Most cells possess a variety of different integrins, and conversely, most integrins are present on a variety of different cell types.

Electron microscopic observations of integrin molecules beginning in the late 1980s suggested that the two subunits are oriented so as to form a globular extracellular head connected to the membrane by a pair of elongated "legs" (as depicted in Figure 7.5). The legs of each subunit extend through the lipid bilayer as a single transmembrane helix and end in a small cytoplasmic domain of about 20 to 70 amino acids.[1] Each integrin can bind a number of divalent cations, including Ca^{2+}, Mg^{2+}, and Mn^{2+}, although the effect of each ion on the structure and ligand-binding capabilities of the protein remains unclear. The first X-ray crystallographic structure of the extracellular portion of an integrin was published in 2001 and displayed a highly unexpected feature. Rather than "standing upright" as predicted, the integrin $\alpha_v\beta_3$ was bent dramatically at the "knees" so that the head faces the outer surface of the plasma membrane (Figure 7.13a). To understand the significance of this bent structure, we need to explore the properties of integrins.

Many integrins, including the one shown in Figure 7.13, can exist on the surface of a cell in an inactive conformation. These integrins can be activated rapidly by events within the cell that alter the conformation of the cytoplasmic domains of the integrin's subunits. Changes in conformation of the cytoplasmic domains are propagated through the molecule, increasing the integrin's affinity for an extracellular ligand. For example, the aggregation of platelets during blood clotting (see Figure 7.15) occurs only after the cytoplasmic activation of $\alpha_{IIb}\beta_3$ integrins, which increases their affinity for fibrinogen. This type of alteration in integrin affinity triggered by changes occurring inside the cell is referred to as "inside-out"

[1]One exception to this molecular architecture is the β_4 chain, which has an extra thousand or so amino acids as part of its cytoplasmic domain. This huge addition makes β_4 integrins capable of extending much more deeply into the cytoplasm (see Figure 7.19)

(a)

(b)
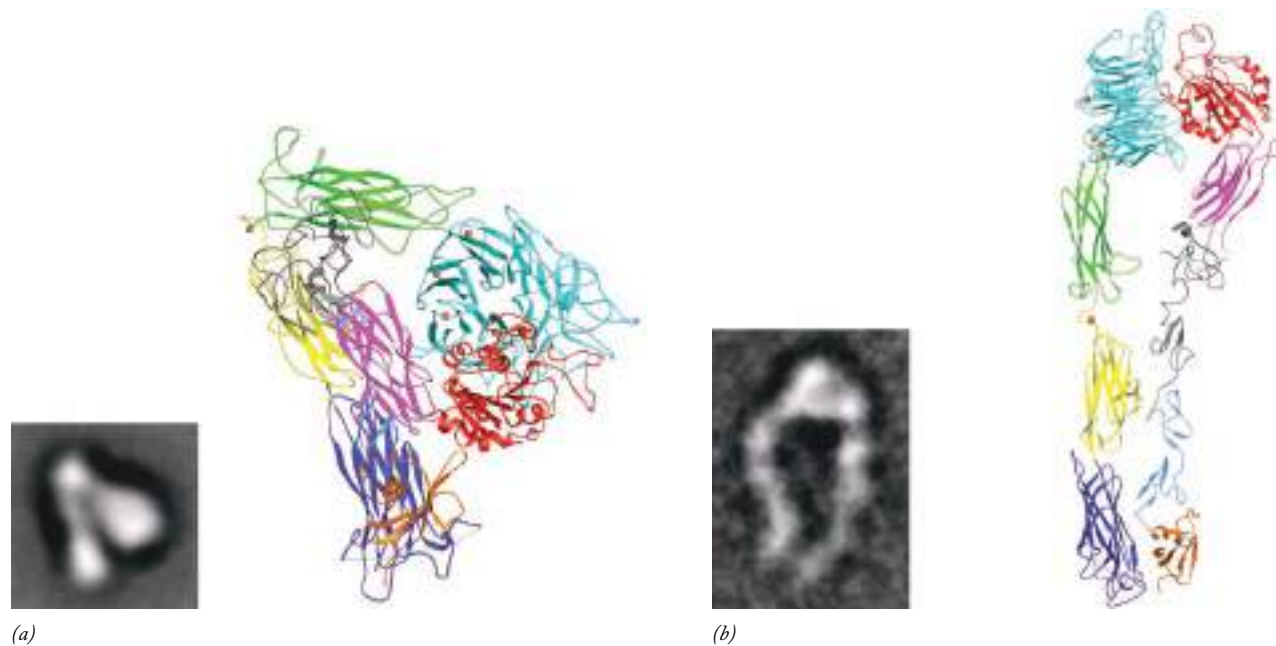
FIGURE 7.13 Integrin conformations. (*a*) Ribbon drawing of the extracellular domains of an integrin ($\alpha_v\beta_3$) in the "bent" conformation, as determined by X-ray crystallography, and a corresponding electron micrograph of the same portion of an integrin molecule. Studies suggest that this is an inactive conformation of the integrin. (*b*) Ribbon drawing and electron micrograph of the same integrin in the "upright" conformation, which is thought to represent the active (i.e., ligand-binding) structure. Divalent metal ions (Ca^{2+}, Mg^{2+}, and/or Mn^{2+}) are shown as orange spheres. (*Note:* The relationship between integrin conformation and ligand-binding activity has been a focus of controversy. Conformations that are intermediate between those shown here may also exhibit ligand-binding, as discussed in *J. Cell Sci.* 122:165, 2009.) (ELECTRON MICROGRAPHS FROM JUNICHI TAKAGI ET AL., CELL 110:601, 2002; BY PERMISSION OF CELL PRESS; RIBBON DRAWINGS FROM M. AMIN ARNAOUT ET AL., CURR. OPIN. CELL BIOL. 14:643, 2002.)

signaling. In the absence of the inside-out signal, the integrin remains in the inactive state, which protects the body against the formation of an inappropriate blood clot.

A body of evidence strongly suggests that the bent conformation of an integrin shown in Figure 7.13*a* corresponds to the inactive state of the protein, one that is incapable of binding a ligand. In fact, when an $\alpha_v\beta_3$ integrin containing a bound ligand is analyzed, the integrin no longer exhibits the bent structure but is present instead in an upright conformation as illustrated in Figure 7.13*b*. The ligand is bound to the head of the integrin in a cleft where the α and β subunits come together (as in Figure 7.14). If the bent and upright structures represent the inactive and active states of an integrin, respectively, then it is important to consider what type of stimulus is capable of triggering such a remarkable transformation in protein conformation.

Studies suggest that the ligand-binding ability of the extracellular domains of an integrin is dependent on the spatial arrangement of the α and β cytoplasmic tails present on the inner side of the membrane. The cytoplasmic domains of integrins bind a wide array of proteins; one of these proteins, called talin, causes separation of the α and β subunits (Figure 7.14). Mutations in talin that block its interaction with β-integrin subunits also prevent activation of integrins and adhesion to the ECM. Separation of the cytoplasmic ends of the integrins by talin is thought to send a change in conforma-

tion through the legs of the integrin, causing the molecule to assume an upright position in which the head of the protein is capable of interacting specifically with its appropriate ligand. An increase in affinity of individual integrins is often followed by clustering of the activated integrins, which greatly enhances the overall strength of the interaction between the cell surface and its extracellular ligands.

Integrins have been implicated in two major types of activities: adhesion of cells to their substratum (or to other cells) and transmission of signals from the external environment to the cell interior, a phenomenon known as "outside-in" signaling. The binding of the extracellular domain of an integrin to a ligand, such as fibronectin or collagen, can induce a conformational change at the opposite, cytoplasmic end of the integrin. Changes at the cytoplasmic end can, in turn, alter the way the integrin interacts with nearby cytoplasmic proteins, modifying their activity. Thus, as integrins bind to an extracellular ligand, they can trigger the activation of cytoplasmic protein kinases, such as FAK and Src (see Figure 7.17*c*). These kinases can then phosphorylate other proteins, initiating a chain reaction. In some cases, the chain reaction leads all the way to the nucleus, where a specific group of genes may be activated.

Outside-in signals transmitted by integrins (and other cell-surface molecules) can influence many aspects of cell behavior, including differentiation, motility, growth, and even

the survival of the cell. The influence of integrins on cell survival is best illustrated by comparing normal and malignant cells. Most malignant cells are capable of growing while suspended in a liquid culture medium. Normal cells, in contrast, can only grow and divide if they are cultured on a solid substratum; if they are placed in suspension cultures, they die. Normal cells are thought to die in suspension culture because their integrins are not able to interact with extracellular substrates and, as a result, are not able to transmit life-saving signals to the interior of the cell. When cells become malignant, their survival no longer depends on integrin binding. The role of integrins in transmitting signals is a very active area of research and will be explored further on page 253.

Table 7.1 lists a number of known integrins and the key extracellular ligands to which they bind. Linkage between integrins and these ligands mediates adhesion between cells and their environment. Because individual cells may express a variety of different integrins on their cell surface, such cells are capable of binding to a variety of different extracellular components (Figure 7.5). Despite the apparent overlap seen in Table 7.1, most integrins do appear to have unique functions, as knockout mice (Section 18.18) that lack different integrin subunits exhibit distinct phenotypes. For example, α_8 knock-

outs show kidney defects, α_4 knockouts exhibit heart defects, and α_5 knockouts display vascular defects.

Many of the extracellular proteins that bind to integrins do so because they contain the amino acid sequence arginine-glycine-aspartic acid (or, in abbreviated amino acid nomenclature, RGD). This tripeptide is present in the cell-binding sites of proteoglycans, fibronectin, laminin, and various other extracellular proteins. The cell-binding domain of fibronectin, with its extended RGD-containing loop, is shown in Figure 7.10a.

The discovery of the importance of the RGD sequence has opened the door to new avenues for treatment of medical conditions that involve receptor-ligand interactions. When the wall of a blood vessel is injured, the damaged region is sealed by the controlled aggregation of blood platelets, which are nonnucleated cells that circulate in the blood. When it occurs at an inappropriate time or place, the aggregation of platelets can form a potentially dangerous blood clot (*thrombus*) that can block the flow of blood to major organs and is one of the leading causes of heart attack and stroke. The aggregation of platelets requires the interaction of a platelet-specific integrin ($\alpha_{IIb}\beta_3$) with soluble RGD-containing blood proteins, such as fibrinogen and von Willebrand factor, which act as linkers that hold the platelets together (Figure 7.15, top). Experiments with animals had indicated that RGD-containing peptides can inhibit blood clot formation by preventing the platelet integrin from binding to blood proteins (Figure 7.15, bottom). These findings led to the design of a new class of antithrombotic agents (Aggrastat and Integrelin) that resemble the RGD structure but bind selectively to the platelet integrin. A specific antibody (ReoPro) directed against the RGD binding site of $\alpha_{IIb}\beta_3$ integrins can also
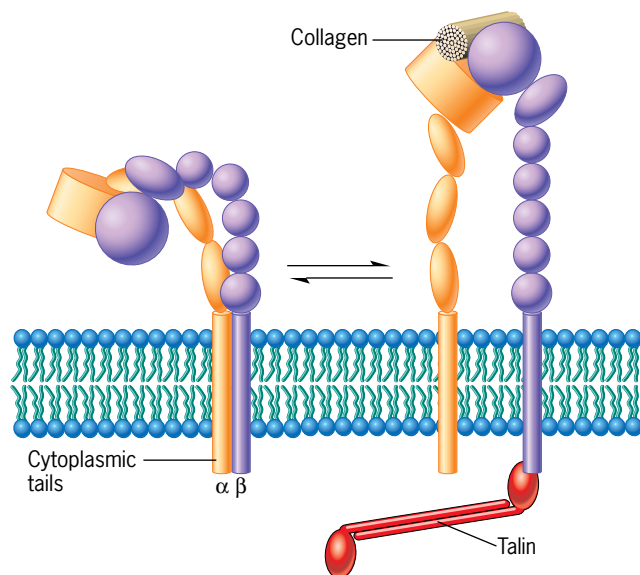
FIGURE 7.14 A model of integrin activation. Schematic representation of a heterodimeric integrin molecule in the bent, inactive conformation (left) and the upright, active conformation (right). The switch in conformation is triggered by the binding of a protein, in this case talin, to the small cytoplasmic domain of the β subunit. Binding of talin induces a separation of the two subunits and conversion to the active conformation. Activated integrins typically become clustered as the result of interactions of their cytoplasmic domains with the underlying cytoskeleton as indicated in Figure 7.17c. The domain structure of each subunit seen in the ribbon drawings of Figure 7.13 is shown here using globular-shaped segments. The extracellular ligand, in this case a collagen fiber, binds between the two subunits in the head region of the activated integrin dimer.

TABLE 7.1 Classification of Integrin Receptors Based on Recognition of RGD Sequences

RGD Recognition		Non-RGD Recognition	
Integrin receptor	Key ligands	Integrin receptor	Key ligands
$\alpha_3\beta_1$	Fibronectin	$\alpha_1\beta_1$	Collagen
$\alpha_5\beta_1$	Fibronectin	$\alpha_2\beta_1$	Collagen
$\alpha_v\beta_1$	Fibronectin		Laminin
		$\alpha_3\beta_1$	Collagen
$\alpha_{IIb}\beta_3$	Fibronectin		Laminin
	von Willebrand factor	$\alpha_4\beta_1$	Fibronectin
	Vitronectin		VCAM
	Fibrinogen	$\alpha_6\beta_1$	Laminin
$\alpha_v\beta_3$	Fibronectin	$\alpha_L\beta_2$	ICAM-1
	von Willebrand factor		ICAM-2
	Vitronectin	$\alpha_M\beta_2$	Fibrinogen
$\alpha_v\beta_5$	Vitronectin		ICAM-1
$\alpha_v\beta_6$	Fibronectin		

Source: S. E. D'Souza, M. H. Ginsberg, and E. F. Plow, *Trends Biochem. Sci.* 16:249, 1991.
See *Annu. Rev. Neurosci.* 30:463, 2007 for a complete list of all integrins and their ligands.
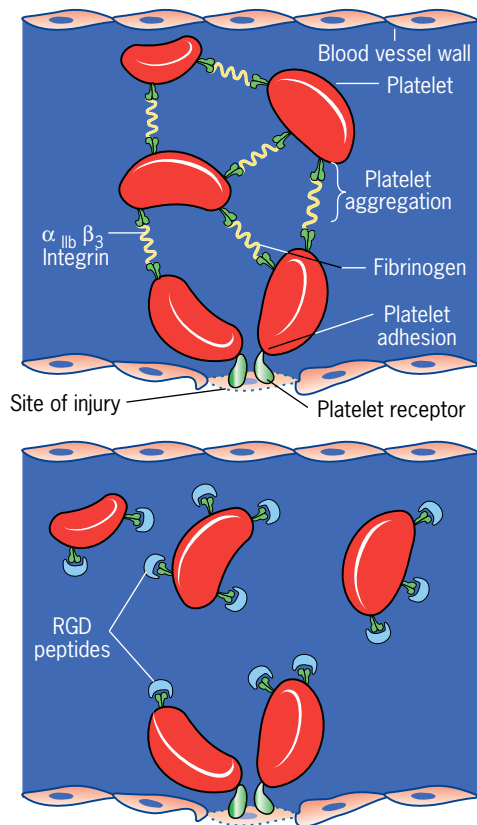
FIGURE 7.15 **Blood clots form when platelets adhere to one another through fibrinogen bridges that bind to the platelet-integrins.** The presence of synthetic RGD peptides can inhibit blood clot formation by competing with fibrinogen molecules for the RGD-binding sites on platelet $\alpha_{IIb}\beta_3$ integrins (also known as GPIIb/IIIa). Nonpeptide RGD analogues and anti-integrin antibodies can act in a similar way to prevent clot formation in high-risk patients.

prevent blood clots in certain patients undergoing high-risk vascular surgeries. A number of compounds that target integrins involved in other diseases are currently in clinical trials.

The cytoplasmic domains of integrins contain binding sites for a variety of cytoplasmic proteins, including several that act as adaptors that link the integrin to actin filaments of the cytoskeleton (see Figure 7.17). The role of integrins in making the connection between the ECM and the cytoskeleton is best seen in two specialized structures: focal adhesions and hemidesmosomes.

Focal Adhesions and Hemidesmosomes: Anchoring Cells to Their Substratum

It is much easier to study the interaction of cells with the bottom of a culture dish than with an extracellular matrix inside an animal. Consequently, much of our knowledge of cell–matrix interactions has been obtained from the study of cells adhering to various substrates in vitro. The stages in the attachment of a cultured cell to its substratum are shown in Figure 7.16. At first, the cell has a rounded morphology, as is

(a)   2.5 μm

(b)   2.5 μm

(c)   2.5 μm
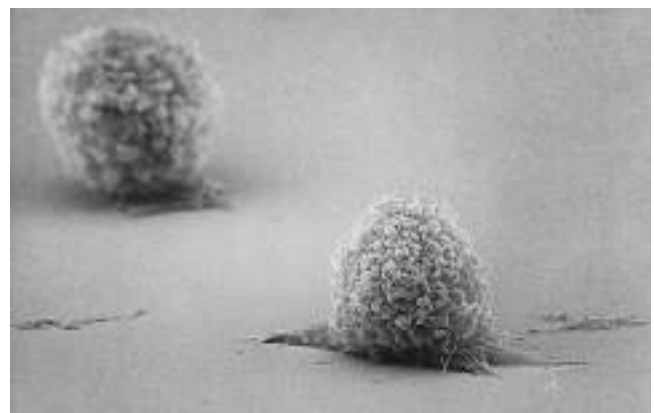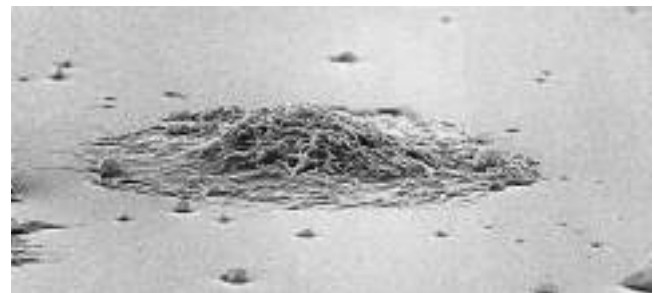
(d)   2.5 μm

FIGURE 7.16 **Steps in the process of cell spreading.** Scanning electron micrographs showing the morphology of mouse fibroblasts at successive times during attachment and spreading on glass coverslips. Cells were fixed after (a) 30 minutes, (b) 60 minutes, (c) 2 hours, and (d) 24 hours of attachment. (FROM J. J. ROSEN AND L. A. CULP, EXP. CELL RES. 107:141, 1977.)

generally true of animal cells suspended in aqueous medium. Once the cell makes contact with the substratum, it sends out projections that form increasingly stable attachments. Over time, the cell flattens and spreads itself out on the substratum.

When fibroblasts or epithelial cells spread onto the bottom of a culture dish, the lower surface of the cell is not pressed uniformly against the substratum. Instead, the cell is anchored to the surface of the dish only at scattered, discrete sites, called **focal adhesions** (Figure 7.17a). Focal adhesions are dynamic structures that can be rapidly disassembled if the adherent cell is stimulated to move or enter mitosis. The plasma membrane in the region of a focal adhesion contains large clusters of integrins—often $\alpha_v\beta_3$, the integrin whose structure is shown in Figure 7.13. The cytoplasmic domains of the integrins are con-

nected by various adaptors to actin filaments of the cytoskeleton (Figure 7.17b,c). Focal adhesions may act as a type of sensory structure, collecting information about the physical and chemical properties of the extracellular environment and transmitting that information to the cell interior, which may lead to changes in cell adhesion, proliferation, or survival. Focal adhesions have also been implicated in cell locomotion, during which the integrins develop transient interactions with extracellular materials.

Focal adhesions are capable of creating or responding to mechanical forces, which might be expected from a structure that contains actin and myosin, two of the cell's major contractile proteins. Figure 7.18 shows a cultured fibroblast attached to a gelated surface that can be deformed by local forces. The original surface contained a uniform grid pattern that has been dis-
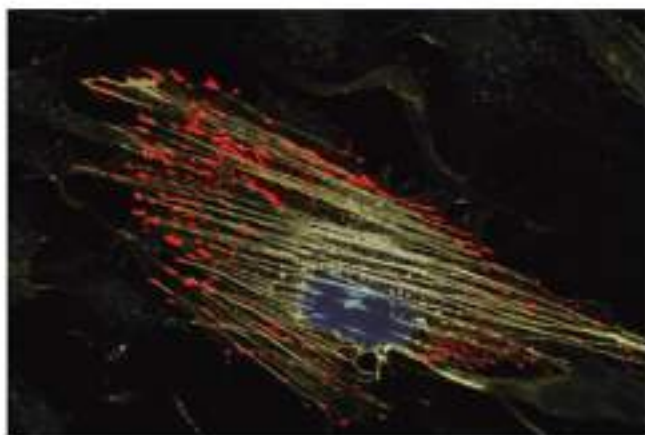
(a)
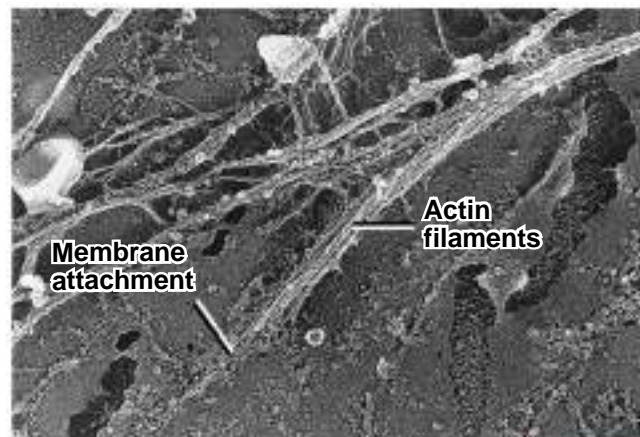
(b)

FIGURE 7.17 Focal adhesions are sites where cells adhere to their substratum and send signals to the cell interior. (a) This cultured cell has been stained with fluorescent antibodies to reveal the locations of actin filaments (gray-green) and integrins (red). The integrins are localized in small patches that correspond to the sites of focal adhesions. (b) The cytoplasmic surface of a focal adhesion of a cultured amphibian cell is shown here after the inner surface of the membrane was processed for quick-freeze, deep-etch analysis. Bundles of microfilaments are seen to associate with the inner surface of the membrane in the region of a focal adhesion. (c) Schematic drawing of a focal adhesion showing the interactions of integrin molecules with other proteins on both sides of the lipid bilayer. The binding of extracellular ligands, such as collagen and fibronectin, is thought to induce conformational changes in the cytoplasmic domains of the integrins that cause the integrins to become linked to actin filaments of the cytoskeleton. Linkages with the cytoskeleton lead, in turn, to the clustering of integrins at the cell surface. Linkages with the cytoskeleton are mediated by various actin-binding proteins, such as talin and α-actinin, that bind to the β subunit of the integrin. The cytoplasmic domains of integrins are also associated with protein kinases, such as FAK (focal adhesion kinase) and Src. The attachment of the integrin to an extracellular ligand can activate these protein kinases and start a chain reaction that transmits signals throughout the cell. The association of myosin molecules with the actin filaments can generate traction forces that are transmitted to sites of cell–substrate attachment. (A: FROM MARGO LAKONISHOK AND CHRIS DOE, SCI. AM. P. 68, MAY 1997; B: FROM STEVEN J. SAMUELSSON,
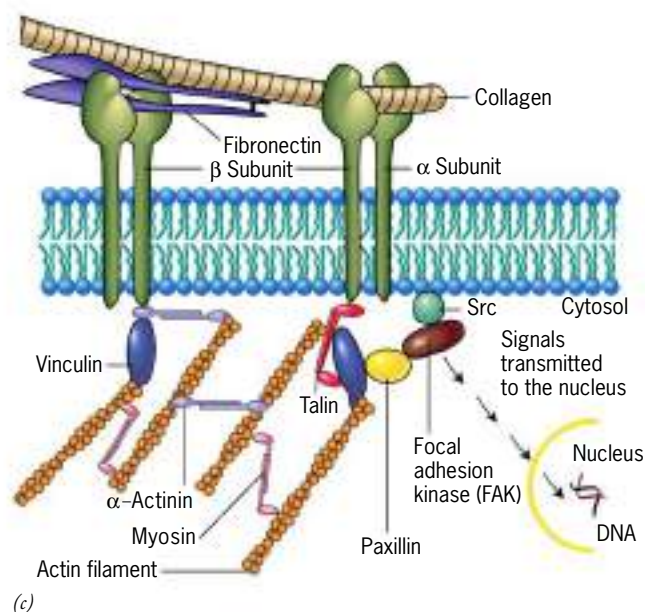
(c)

PAUL J. LUTHER, DAVID W. PUMPLIN, AND ROBERT J. BLOCH, J. CELL BIOL. 122:487, 1993; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)
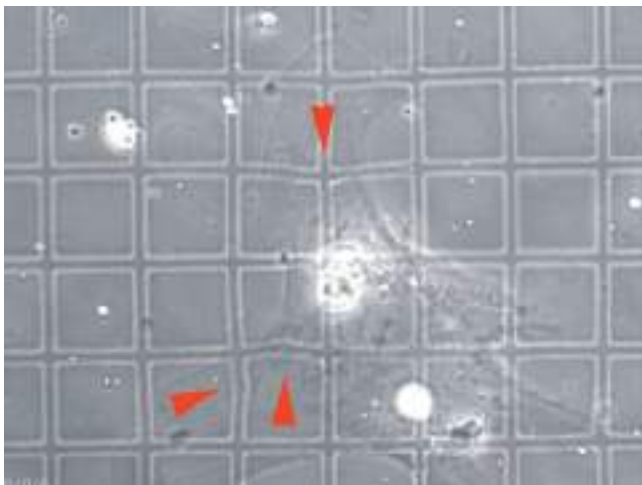
FIGURE 7.18 Experimental demonstration of forces exerted by focal adhesions. In this approach, fibroblasts are plated on a deformable surface that contains a visible, uniform grid pattern. Traction forces generated by focal adhesions can be monitored by observing deformations (arrowheads) in the grid pattern of the substratum to which the cells are adhering. The generation of force can be correlated with the location of fluorescently labeled focal adhesions (see Figure 9.73). (REPRINTED FROM N. Q. BALABAN ET AL., NATURE CELL BIOL. 3:468, 2001, COURTESY OF BENJAMIN GEIGER; COPYRIGHT 2001, MACMILLAN MAGAZINES LTD.)

torted by traction (gripping/pulling) forces generated by focal adhesions at the undersurface of the cell. Acting in the opposite direction, mechanical forces applied to the surfaces of cells can be converted by focal adhesions into cytoplasmic signals. In one study, for example, cells were allowed to bind to beads that had been covered with a coating of fibronectin. When the membrane-bound beads were pulled by an optical tweezer (page 139), the mechanical stimulus was transmitted into the cell interior where it generated a wave of Src kinase activation. Figure 7.17c illustrates how the pull of a fibronectin or collagen molecule could activate protein kinases, such as FAK or Src. Activation of these protein kinases can, in turn, transmit signals throughout the cell, including the cell nucleus, where they can promote changes in gene expression. Within the body, activation of Src protein kinases can dramatically alter cell behavior. The importance of the physical properties of a cell's environment in influencing cellular behavior is illustrated by a study in which mesenchymal stem cells (MSCs) derived from adult bone marrow were grown on substrates of varying elasticity (or stiffness). When these MSCs were grown on a soft, pliable substratum, such as might be encountered by a cell within the developing brain, the MSCs differentiated into nerve cells. When grown on a substratum of greater stiffness, these same cells differentiated into muscle cells. Finally, when grown on even stiffer substrates, such as those that might be home to cells growing in a skeletal tissue such as cartilage or bone, the MSCs differentiated into osteoblasts, which give rise to bone cells.

   Focal adhesions are most commonly seen in cells grown in vitro, although similar types of adhesive contacts are found in certain tissues, such as muscle and tendon. Within the
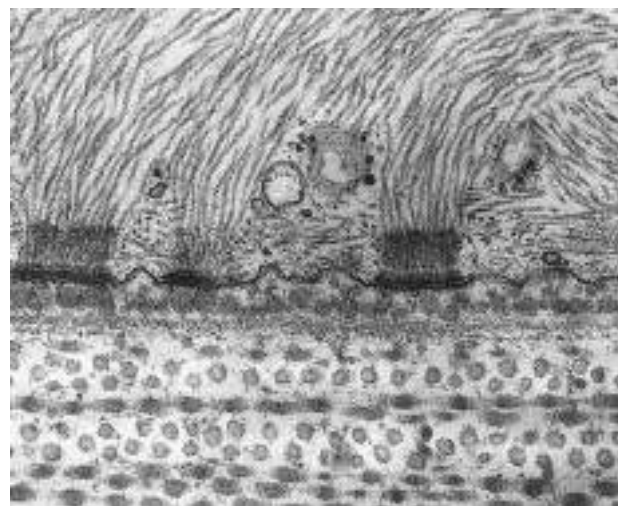
(a)   0.3 μm
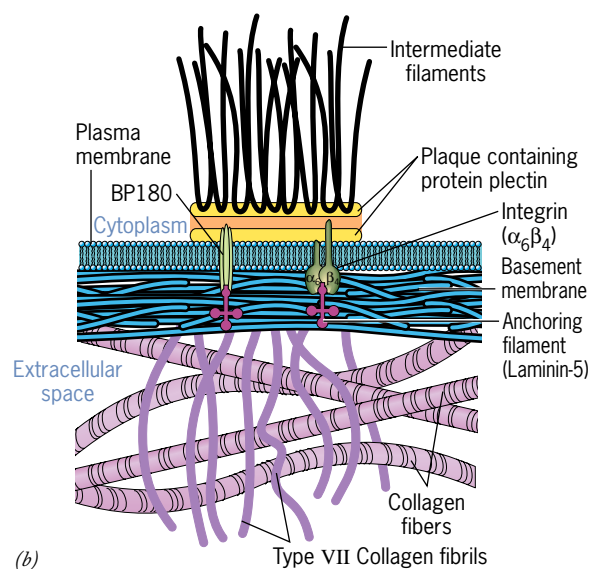
(b)

FIGURE 7.19 Hemidesmosomes are differentiated sites at the basal surfaces of epithelial cells where the cells are attached to the underlying basement membrane. (*a*) Electron micrograph of several hemidesmosomes showing the dense plaque on the inner surface of the plasma membrane and the intermediate filaments projecting into the cytoplasm. (*b*) Schematic diagram showing the major components of a hemidesmosome connecting the epidermis to the underlying dermis. The $\alpha_6\beta_4$ integrin molecules of the epidermal cells are linked to cytoplasmic intermediate filaments by a protein called plectin that is present in the dark-staining plaque and to the basement membrane by anchoring filaments of a particular type of laminin. A second transmembrane protein (BP180) is also present in hemidesmosomes. The collagen fibers are part of the underlying dermis. (A: FROM DOUGLAS E. KELLY, J. CELL BIOL. 28:51, 1966; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

body, the tightest attachment between a cell and its extracellular matrix is seen at the basal surface of epithelial cells where the cells are anchored to the underlying basement membrane by a specialized adhesive structure called a **hemidesmosome** (Figures 7.1 and 7.19). Hemidesmosomes contain a dense

plaque on the inner surface of the plasma membrane with filaments coursing outward into the cytoplasm. Unlike the filaments of focal adhesions, which consist of actin, the filaments of the hemidesmosome are thicker and consist of the protein keratin. Keratin-containing filaments are classified as intermediate filaments, which serve primarily in a supportive function (as discussed in detail in Section 9.4). The keratin filaments of the hemidesmosome are linked to the extracellular matrix by membrane-spanning integrins, including $\alpha_6\beta_4$ (Figure 7.19*b*). Like their counterparts in focal adhesions, these integrins also transmit signals from the ECM that affect the shape and activities of the attached epithelial cells.

The importance of hemidesmosomes is revealed by a rare disease, *bullous pemphigoid*, in which individuals produce antibodies that bind to proteins (the *bullous pemphigoid antigens*) present in these adhesive structures. Diseases caused by production of antibodies directed against one's own tissues (i.e., autoantibodies) are called autoimmune disorders and are responsible for a wide variety of conditions. In this case, the presence of autoantibodies causes the lower layer of the epidermis to lose attachment to the underlying basement membrane (and thus to the underlying connective tissue layer of the dermis). The leakage of fluid into the space beneath the epidermis results in severe blistering of the skin. A similar inherited blistering disease, *epidermolysis bullosa*, can occur in patients with genetic alterations in any one of a number of hemidesmosomal proteins, including the α_6 or β_4 integrin subunit, collagen VII, or laminin-5.

REVIEW ?

1. How are integrins able to link the cell surface with materials that make up the ECM? How do the inactive and active structures of integrins differ from one another structurally and functionally? What is the significance of the presence of an RGD motif in an integrin ligand?
2. How can a cell-surface protein be involved in both cell adhesion and transmembrane signal transduction?
3. What is the difference between inside-out and outside-in signaling? What is the importance of each to cell activities?
4. List two characteristics that distinguish hemidesmosomes from focal adhesions.

7.3 INTERACTIONS OF CELLS WITH OTHER CELLS

Examination of a thin section through a major organ of an animal reveals a complex architecture involving a variety of different types of cells. Little is known about the mechanisms responsible for generating the complex three-dimensional cellular arrangements found within developing organs. It is presumed that this process depends heavily on *selective* interactions between cells of the same type, as well as between cells of a different type. It is evident that cells can recognize the surfaces of other cells, interacting with some and ignoring others.

It is very difficult to study cellular interactions that occur within the inaccessible organs of a developing embryo as tissue formation takes place. Early attempts to learn about cell–cell recognition and adhesion were carried out by removing a developing organ from a chick or amphibian embryo, dissociating the organ's tissues to form a suspension of single cells, and determining the ability of the cells to reaggregate in culture. In experiments in which cells from two different developing organs were dissociated and mixed together, the cells would initially aggregate to form a mixed clump. Over time, however, the cells would move around within the aggregate and eventually "sort themselves out" so that each cell adhered only to cells of the same type (Figure 7.20). Once separated into a homogeneous cluster, these embryonic cells would often differentiate into many of the structures they would have formed within an intact embryo.

Little was known about the nature of the molecules that mediated cell–cell adhesion until the development of techniques for purifying integral membrane proteins and, more recently, for the isolation and cloning of genes that encode these proteins. Dozens of different proteins involved in cell adhesion have now been identified. Different arrays of these molecules on the surfaces of different types of cells are thought to be responsible for the specific interactions between cells within complex tissues. Four distinct families of integral membrane proteins play a major role in mediating cell–cell adhesion: (1) selectins, (2) certain members of the immunoglobulin superfamily (IgSF), (3) certain members of the integrin family, and (4) cadherins.

Selectins

During the 1960s, it was discovered that lymphocytes that had been removed from peripheral lymph nodes, radioactively labeled, and injected back into the body would return to the sites from which they were originally derived—a phenomenon called *lymphocyte homing*. It was subsequently found that homing could be studied in vitro by allowing lymphocytes to adhere to frozen sections of lymphoid organs. Under these experimental conditions, the lymphocytes would selectively adhere to the endothelial lining of the venules (the smallest veins) of peripheral lymph nodes. Binding of lymphocytes to venules could be blocked by antibodies that bound to a specific glycoprotein on the lymphocyte surface. This lymphocyte glycoprotein was named LEU-CAM1 and later L-selectin.

Selectins comprise a family of integral membrane glycoproteins that recognize and bind to a particular arrangement of sugars in the oligosaccharides that project from the surfaces of other cells (page 126). The name of this class of cell-surface receptors is derived from the word "lectin," a general term for a compound that binds to specific carbohydrate groups. Selectins possess a small cytoplasmic domain, a single membrane-spanning domain, and a large extracellular seg-

FIGURE 7.20 Experimental demonstration of cell–cell recognition. When cells from different parts of an embryo are dissociated and then intermixed, the cells initially aggregate and then sort out by associating with other cells of the same type. The results of two such classic experiments are shown here. (*a*) In this experiment, two regions of an early amphibian embryo (the ectoderm and mesoderm) were dissociated into single cells and combined. At first the cells form a mixed aggregate, but eventually they sort out. The ectodermal cells (shown in red) move to the outer surface of the aggregate, which is where they would be located in the embryo, and the mesodermal cells (shown in purple) move to the interior, the position they would occupy in the embryo. Both types of cells then differentiate into the types of structures they would normally give rise to. (*b*) Light micrograph showing the results of an experiment in which precartilage cells from a chick limb are mixed with chick heart ventricle cells. The two types of cells have sorted themselves out of the mixed aggregate with the heart cells forming a layer on the outside of the precartilage cells. It is proposed that the precartilage cells collect in the center of the aggregate because the cells adhere to one another more strongly than do the cells from the heart. (This and other models are discussed in *Nat. Cell Biol.* 10:375, 2008.) (A: FROM P. L. TOWNES AND JOHANNES HOLTFRETER, J. EXP. ZOOL. 128:53, 1955; B: FROM MALCOLM S. STEINBERG, J. EXP. ZOOL. 173:411, 1970.)

Ectoderm + Mesoderm

(*a*)

Heart cells
Precartilage cells

100 µm

(*b*)
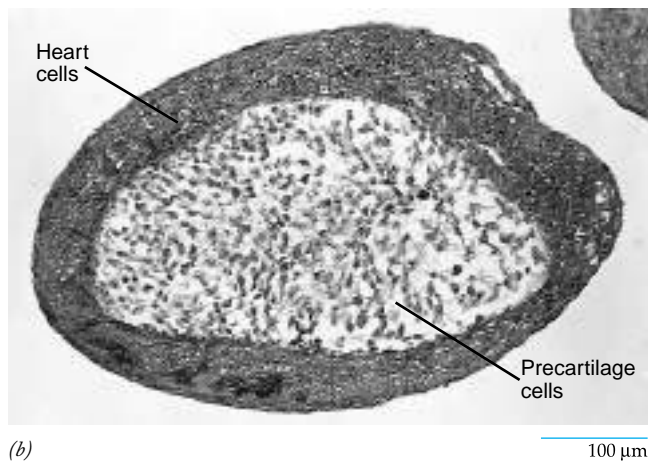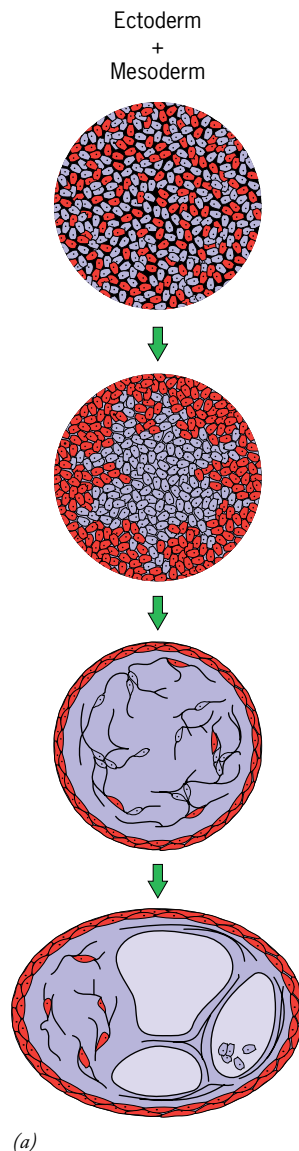
ment that consists of a number of separate modules, including an outermost domain that acts as the lectin (Figure 7.21). There are three known selectins: E-selectin, present on endothelial cells; P-selectin, present on platelets and endothelial cells; and L-selectin, present on leukocytes (white blood cells). All three selectins recognize a particular grouping of sugars (Figure 7.21*c*) that is found at the ends of the carbohydrate chains of certain complex glycoproteins. Binding of selectins to their carbohydrate ligands requires calcium. As a group, selectins mediate transient interactions between circulating leukocytes and vessel walls at sites of inflammation and clotting. Capturing a leukocyte is a challenging task because these cells are flowing through the bloodstream at a considerable rate of speed. Selectins are well-suited for this function because the bonds that they form with their ligands become stronger when the interaction is placed under mechanical stress, as occurs when the leukocyte is being pulled away from a given site on the vessel wall. The role of selectins in inflammation is discussed further in the Human Perspective on page 247.

L-Selectin

E-Selectin

P-Selectin

Lectin-like domain
EGF-like domain
Structural domain

(*a*)

Extracellular

Glycoprotein

(*b*)

Ligand

SiA — α2→3 — Gal
β1→4
SO₄ — GlcNAc — Remainder of oligosaccharide
α1→3
Fuc

(*c*)
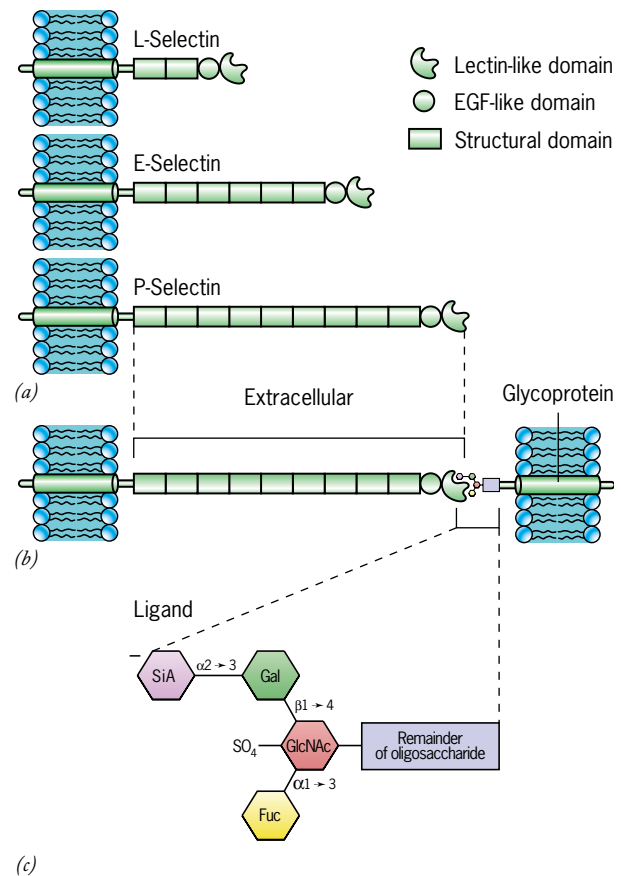
FIGURE 7.21 Selectins. Schematic drawing showing the three types of known selectins (*a*). All of them recognize and bind to a similar carbohydrate ligand at the ends of oligosaccharide chains on glycoproteins, such as the one shown in (*b*). (*c*) The detailed structure of the carbohydrate ligand. The terminal fucose and sialic acid moieties are particularly important in selectin recognition, and the *N*-acetylglucosamine moiety is often sulfated.

The Role of Cell Adhesion in Inflammation and Metastasis

Inflammation is one of the primary responses to infection. If a part of the body becomes contaminated by bacteria, as might occur following a puncture wound to the skin, the site of injury becomes a magnet for a variety of white blood cells. White blood cells (leukocytes) that would normally remain in the bloodstream are stimulated to traverse the endothelial layer that lines the smallest veins (venules) in the region and enter the tissue. Once in the tissue, the leukocytes move in response to chemical signals toward the invading microorganisms, which they ingest.[a] Although inflammation is a protective response, it also produces adverse side effects, such as fever, swelling due to fluid accumulation, redness, and pain.

Inflammation can also be triggered inappropriately. For example, damage to the tissues of the heart or brain can occur when blood flow to these organs is blocked during a heart attack or stroke. When blood flow to the organ is restored, circulating leukocytes may attack the damaged tissue, causing a condition known as *reperfusion damage*. An overzealous inflammatory response can also lead to asthma, toxic shock syndrome, and respiratory distress syndrome. A great deal of research has focused on questions related to these conditions: How are leukocytes recruited to sites of inflammation? How are they able to stop flowing through the bloodstream and adhere to vessel walls? How do they penetrate the walls of the vessels? How can some of the negative side effects of inflammation be blocked without interfering with the beneficial aspects of the response? Answers to questions about inflammation have focused on three types of cell-adhesion molecules: selectins, integrins, and IgSF proteins.

A chain of events that has been proposed to occur during acute inflammation is shown in Figure 1. The first step begins as the walls of the venules become activated in response to chemical signals from nearby damaged tissue. The endothelial cells that line these venules become more adhesive to circulating neutrophils, a type of phagocytic leukocyte that carries out a rapid, nonspecific attack on invading pathogens. This change in adhesion is mediated by a temporary display of P- and E-selectins on the surfaces of the activated endothelial cells in the damaged area (step 2, Figure 1). When neutrophils encounter the selectins, they form transient adhesions that dramati-

cally slow their movement through the vessel. At this stage, the neutrophils can be seen to "roll" slowly along the wall of the vessel. A number of biotechnology companies are attempting to develop anti-inflammatory drugs that act by interfering with binding of ligands to E- and P-selectins. Anti-selectin antibodies block neutrophil rolling on selectin-coated surfaces in vitro and suppress inflammation and reperfusion damage in animals. A similar type of blocking effect has been attained using synthetic carbohydrates (e.g., efomycines) that bind to E- and P-selectin, thereby competing with carbohydrate ligands on the surfaces of the neutrophils.

As the neutrophils interact with the inflamed venule endothelium, interactions take place between other molecules on the surfaces of the two types of cells. One of the molecules displayed on the surface of the endothelial cells is a phospholipid called *platelet activating factor* (*PAF*). PAF binds to a receptor on the surface of the neutrophil sending a signal into the neutrophil that leads to an increase in the binding activity of certain integrins (e.g., $\alpha_L\beta_2$ and $\alpha_4\beta_1$) already situated on the neutrophil surface (step 3, Figure 1). This is an example of the type of inside-out signal that is illustrated in Figure 7.14. The activated integrins then bind with high affinity to IgSF molecules (e.g., ICAM-1 and VCAM-1) on the surface of the endothelial cells, causing the neutrophils to stop their rolling and adhere firmly to the wall of the vessel (step 4). The bound neutrophils then change their shape and squeeze between adjacent endothelial cells into the damaged tissue (step 5). Invading neutrophils appear capable of disassembling the adherens junctions (page 251) that form the major barrier between cells of the vessel wall. This cascade of events, which involves several different types of cell-adhesion molecules, ensures that attachment of blood cells to the walls of blood vessels and subsequent penetration will occur only at sites where leukocyte invasion is required.

The importance of integrins in the inflammatory response is demonstrated by a rare disease called *leukocyte adhesion deficiency* (*LAD*). Persons with this disease are unable to produce the β_2 subunit as part of a number of leukocyte integrins. As a result, the leukocytes of these individuals lack the ability to adhere to the endothelial layer of venules, a step required for their exit from the bloodstream. These patients suffer from repeated life-threatening bacterial infections. The disease is best treated by bone marrow

---

[a]A remarkable film of a leukocyte "chasing" a bacterium can be found on the Web by searching the keywords "neutrophil crawling."

1	2	3	4	5
Endothelial activation	Neutrophil trapping	Neutrophil activation	Neutrophil adhesion	Neutrophil invasion

♀ – ICAM   ♂ – Selectin   ● – Platelet activating factor   ▮ – Integrin (unactivated)   ▮▮ – Integrin (activated)   Invasion
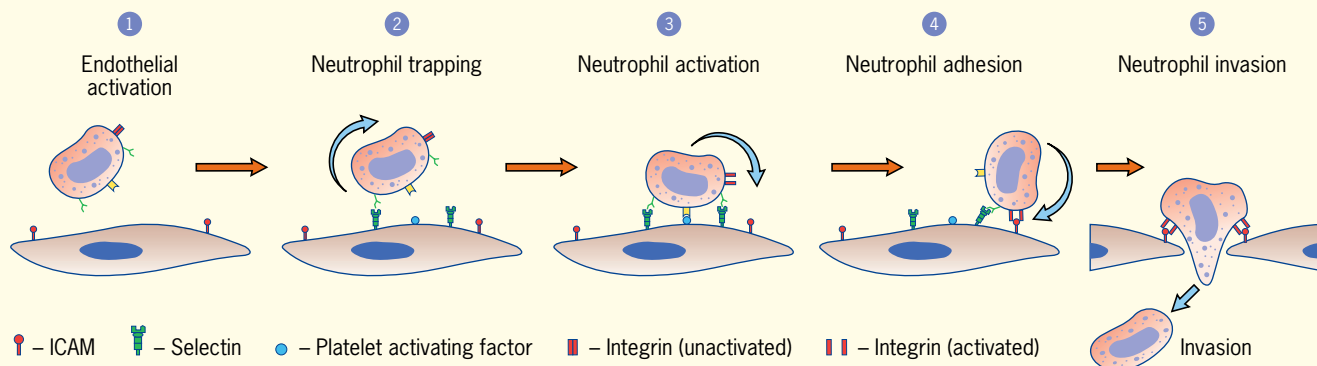
FIGURE 1 **Steps in the movement of neutrophils from the bloodstream during inflammation.** The steps are described in the text.

transplantation, which provides the patient with stem cells capable of forming normal leukocytes. Administration of antibodies against the β_2 subunit can mimic the effects of LAD, blocking the movement of neutrophils and other leukocytes out of blood vessels. Such antibodies might prove useful in preventing inflammatory responses associated with diseases such as asthma and rheumatoid arthritis or with reperfusion.

Cancer is a disease in which cells escape from the body's normal growth control mechanisms and proliferate in an unregulated manner. Were the malignant cells to remain in a single mass, as often occurs in some types of skin cancer or thyroid cancer, most cancers would be readily cured by surgical removal of the diseased tissue. Most malignant tumors, however, spawn cells that are capable of leaving the primary tumor mass and entering the bloodstream or lymphatic channels, thereby initiating the growth of secondary tumors in other parts of the body (Figure 2). The spread of a tumor within the body is called **metastasis** and is the reason cancer is such a devastating disease. Metastatic cells (cancer cells that are able to initiate the formation of secondary tumors) are thought to have special cell-surface properties that are not shared by most other cells in the tumor. For example:

1. Metastatic cells must be less adhesive than other cells to break free of the tumor mass.
2. They must be able to penetrate numerous barriers, such as the extracellular matrices of surrounding connective tissue and the basement membranes that underlie the epithelium and line the blood vessels that carry them to distant sites (Figure 2).
3. They must be able to invade normal tissues if they are to form secondary colonies.

The mechanisms used by cancer cells to penetrate extracellular matrices are poorly understood because such events have been virtually impossible to study within the tissues of a living animal. It is thought that movement through basement membranes is accomplished largely by ECM-digesting enzymes, most notably the matrix metalloproteinases (MMPs) discussed on page 238. These enzymes are presumed to degrade the proteins and proteoglycans that stand in the way of cancer cell migration. In addition, cleavage of certain proteins of the ECM by MMPs produces active protein fragments that act back on the cancer cells to stimulate their growth and invasive character. Because of their apparent roles in the development of malignant tumors, MMPs became a prominent target of the pharmaceutical industry. Once it was demonstrated that synthetic MMP inhibitors were able to reduce metastasis in mice, a number of clinical trials of these drugs were conducted on patients with a variety of advanced, inoperable cancers. Unfortunately, these inhibitors have shown little promise in stopping late-stage tumor progression and, in some cases, have led to joint damage. To date, the only FDA-approved MMP inhibitor (Periostat) is used to treat periodontal disease.

Changes in the numbers and types of various cell-adhesion molecules—and thus the ability of cells to adhere to other cells or to extracellular matrices—have also been implicated in promoting metastasis. The major studies in this area have focused on E-cadherin, which is the predominant cell-adhesion molecule of the adherens junctions that hold epithelial cells in a cohesive sheet. Page 250
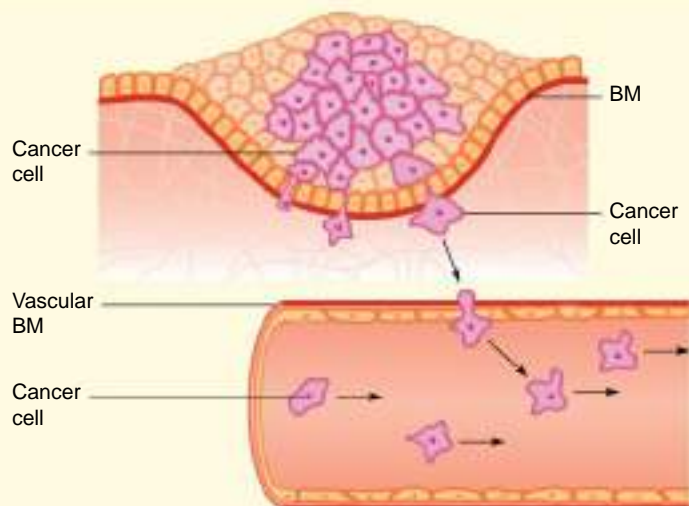
FIGURE 2 **Steps leading to the metastatic spread of an epithelial cancer (a carcinoma).** A fraction of the cells of the primary tumor lose their adhesiveness to other tumor cells and gain the capability to penetrate the basement membrane (BM) barrier that underlies the epithelial tissue. These cells, which have assumed a mesenchymal-like appearance, migrate through the surrounding stromal tissue and cross the BM of a blood or lymph vessel, thereby entering the general circulation. The cells are carried to other tissues, where they migrate back across the BM of the vessel and enter a tissue in which they possess the potential to form secondary tumors. Only a very small percentage of tumor cells that are released from a primary tumor manage to overcome these numerous hurdles, but those that do pose a threat to the life of the host. (FROM R. G. ROWE AND S. J. WEISS, TRENDS CELL BIOL. 18:562, 2008; COPYRIGHT 2008, WITH PERMISSION FROM ELSEVIER SCIENCE.)

describes how the loss of E-cadherin from epithelial cells during certain events of embryonic development is associated with conversion of the cells into a less adhesive, more motile, mesenchymal phenotype. A remarkably similar epithelial-mesenchymal transition occurs during the growth and development of a tumor as malignant cells separate from the primary tumor mass and invade the adjacent normal tissue (Figure 2). This is an important step in the process of metastasis. Surveys of a variety of epithelial cell tumors (e.g., breast, prostate, and colon cancers) confirm that these malignant cells have greatly reduced levels of E-cadherin; the lower the level of expression of E-cadherin, the greater the cell's metastatic potential. Conversely, when malignant cells are forced to express extra copies of the E-cadherin gene, the cells become much less capable of causing tumors when injected into host animals. The presence of E-cadherin is thought to favor the adhesion of cells to one another and suppress the dispersal of tumor cells to distant sites. E-cadherin may also inhibit the signaling pathways within the cell that lead to tissue invasion and metastasis. The importance of E-cadherin is evident from a study of a family of native New Zealanders that had lost 25 members to stomach cancer over a 30-year period. Analysis of the DNA of family members has revealed that susceptible individuals carry mutations in the gene encoding E-cadherin.

The Immunoglobulin Superfamily

Elucidation of the structure of blood-borne antibody molecules in the 1960s was one of the milestones in our understanding of the immune response. Antibodies, which are a type of protein called an immunoglobulin (or Ig), were found to consist of polypeptide chains composed of a number of similar domains. Each of these Ig domains, as they are called, is composed of 70 to 110 amino acids organized into a tightly folded structure, as shown in the inset of Figure 7.22. The human genome encodes 765 distinct Ig domains, making it the most abundant domain in human proteins. Taken as a group, these proteins are members of the **immunoglobulin superfamily**, or **IgSF**. Most members of the IgSF are involved in various aspects of immune function, but some of these proteins mediate calcium-*independent* cell–cell adhesion. In fact, the discovery of Ig-like domains in cell-adhesion receptors in invertebrates—animals that lack a classic immune system—suggests that Ig-like proteins originally evolved as cell-adhesion mediators and only secondarily took on their functions as effectors of the vertebrate immune system.

Most IgSF cell-adhesion molecules mediate the specific interactions of lymphocytes with cells required for an immune response (e.g., macrophages, other lymphocytes, and target cells). However, some IgSF members, such as VCAM (vascular cell-adhesion molecule), NCAM (neural cell-adhesion molecule), and L1, mediate adhesion between nonimmune cells. NCAM and L1, for example, play important roles in nerve outgrowth, synapse formation, and other events during the development of the nervous system. Like fibronectin and many other proteins involved in cell adhesion, the IgSF cell-adhesion molecules have a modular construction (Figure 7.22) and are composed of individual domains of similar structure to the domains in other proteins.

The importance of L1 in neural development has been revealed in several ways. In humans, mutations in the *L1* gene can have devastating consequences. In extreme cases, babies are born with a fatal condition of hydrocephalus ("water on the brain"). Children with less severe mutations typically exhibit mental retardation and difficulty in controlling limb movements (spasticity). Autopsies on patients that have died of an L1-deficiency disease reveal a remarkable condition: they are often missing two large nerve tracts, one that runs between the two halves of the brain and the other that runs between the brain and the spinal cord. The absence of such nerve tracts suggests that L1 is involved in the directed growth of axons within the embryonic nervous system.

Various types of proteins serve as ligands for IgSF cell-surface molecules. As described earlier, most integrins facilitate adhesion of cells to their substratum, but a few integrins mediate cell–cell adhesion by binding to proteins on other cells. For example, the integrin $\alpha_4\beta_1$ on the surface of leukocytes binds to VCAM, an IgSF protein on the endothelial lining of certain blood vessels.

FIGURE 7.22 **L1 is a cell-adhesion molecule of the immunoglobulin (Ig) superfamily.** A model of cell–cell adhesion resulting from the specific interactions of the immunoglobulin (Ig) domains of two L1 molecules projecting from the surfaces of neighboring cells. Each L1 molecule contains a small cytoplasmic domain, a transmembrane segment, several segments that resemble one type of module found in fibronectin, and six Ig domains situated at the N-terminal portion of the molecule. The inset shows the structure of the two N-terminal Ig domains of VCAM, an IgSF molecule on the surface of endothelial cells. The Ig domains of VCAM and L1 have a similar three-dimensional structure consisting of two β sheets packed face-to-face. (INSET REPRINTED WITH PERMISSION FROM E. YVONNE JONES ET AL., NATURE 373:540, 1995; COPYRIGHT 1995, MACMILLAN MAGAZINES LTD.)

Cadherins

The **cadherins** are a large family of glycoproteins that mediate Ca^{2+}-dependent cell–cell adhesion and transmit signals from the ECM to the cytoplasm. Cadherins typically join cells of similar type to one another and do so predominantly by binding to the same cadherin present on the surface of the neighboring cell. This property of cadherins was first demonstrated by genetically engineering cells that were normally nonadhesive to express one of a variety of different cadherins. The cells were then mixed in various combinations and their interactions monitored. It was found that cells expressing one species of cadherin preferentially adhered to other cells expressing the same cadherin.

Like selectins and IgSF molecules, cadherins have a modular construction. The best studied cadherins are E-cadherin (epithelial), N-cadherin (neural), and P-cadherin (placental). These "classical" cadherins, as they are called, contain a relatively large extracellular segment consisting of five tandem domains of similar size and structure, a single transmembrane
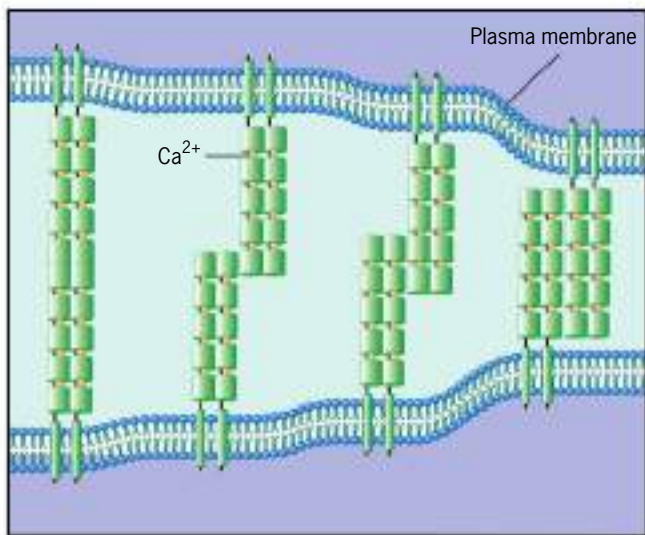
FIGURE 7.23 **Cadherins and cell adhesion.** Schematic representation of two cells adhering to one another as the result of interactions between similar types of cadherins projecting from the plasma membrane of each cell. Calcium ions (shown as small yellow spheres) are situated between the successive domains of the cadherin molecule where they play a critical role in maintaining the rigidity of the extracellular portion of the protein. This illustration shows several alternate models by which cadherins from opposing cells might interact. Different types of studies have suggested different degrees of overlap (interdigitation) between the extracellular domains of molecules from opposing cells (see *Annu. Rev. Cell Dev. Biol.* 23:237, 2007, for discussion). For consistency, subsequent figures will depict cadherins with a single domain overlap.

segment, and a small cytoplasmic domain (Figure 7.23). The cytoplasmic domain is often associated with members of the *catenin* family of cytosolic proteins, which have a dual role: they tether the cadherins to the cytoskeleton (see Figure 7.26), and they transmit signals to the cytoplasm and nucleus.

A number of models of cadherin adhesion are depicted in Figure 7.23. Structural studies indicate that cadherins from the same cell-surface associate laterally to form parallel dimers. These studies also shed light on the role of calcium, which has been known for decades to be essential for cell–cell adhesion. As indicated in Figure 7.23, calcium ions form bridges between successive domains of a given molecule. These calcium ions maintain the extracellular portion of each cadherin in a rigid conformation required for cell adhesion. Adhesion between cells results from the interaction between the extracellular domains of cadherins from opposing cells to form a "cell-adhesion zipper." Considerable controversy has arisen over the degree to which cadherins from opposing cells overlap with one another, which is the reason that several alternate configurations are shown in Figure 7.23. Different types of cells bearing different cadherins likely engage in different types of interactions, so that more than one (or all) of the configurations depicted in Figure 7.23 may occur within an organism. Just as cadherin interdigitation can be compared to a zipper, cadherin clusters can be compared to Velcro; the greater the number of interacting cadherins in a cluster, the greater the strength of adhesion between apposing cells.

Cadherin-mediated adhesion is thought to be responsible for the ability of like cells to "sort out" of mixed aggregates, as was illustrated in Figure 7.20. In fact, cadherins may be the single most important factor in molding cells into cohesive tissues in the embryo and holding them together in the adult. As discussed in the Human Perspective, the loss of cadherin function may play a key role in the spread of malignant tumors.

Embryonic development is characterized by change: change in gene expression, change in cell shape, change in cell motility, change in cell adhesion, and so forth. Cadherins are thought to mediate many of the dynamic changes in adhesive contacts that are required to construct the tissues and organs of an embryo, which is known as the process of *morphogenesis*. For example, a number of morphogenetic events during embryonic development involve a group of cells changing from an epithelium (tightly adherent, polarized cell layer) to a mesenchyme (solitary, nonadhesive, nonpolarized, migratory cells), or vice versa. The **epithelial-mesenchymal transition** (or **EMT**) is illustrated by the formation of the mesoderm during gastrulation in a chick or mammalian embryo. Typically, these cells break away from a cohesive epithelial layer (called the epiblast) at the dorsal surface of the early embryo and wander into the interior regions as mesenchymal cells (Figure 7.24*a,b*). These mesenchymal cells will eventually give rise to mesodermal tissues such as blood, muscle, and bone. The cells of the epiblast display E-cadherins on their surface, which is presumed to promote their close association with one another. Prior to their separation from the epiblast, the future mesodermal cells stop expressing E-cadherin, which is thought to promote their release from the epithelium and transformation into mesenchymal cells (Figure 7.24*a*). At a later stage of development, another major event, the formation of the primitive nervous system, is also characterized by changes in cadherin expression. Following gastrulation, the dorsal surface of the embryo is covered by a single-celled epithelial layer, which will become the ectodermal tissues of the animal (including the skin and nervous system). At this stage, the cells in the central region of this layer stop expressing E-cadherin and begin expressing N-cadherin (Figure 7.24*c*). In subsequent stages, the epithelial cells expressing N-cadherin separate from their neighbors on either side and roll into the neural tube, which will go on to become the animal's brain and spinal cord. It is thought that cadherins (and other cell-adhesion molecules) play a key role in these events by changing the adhesive properties of cells.

Whereas cadherins are typically distributed diffusely all along the surfaces of two adherent cells, they also participate in the formation of specialized intercellular junctions, which are the subjects of the following section.

Adherens Junctions and Desmosomes: Anchoring Cells to Other Cells

The cells of certain tissues, particularly epithelia and cardiac muscle, are notoriously difficult to separate from one another because they are held together tightly by specialized calcium-dependent adhesive junctions. There are two main types of cell–cell adhesive junctions: adherens junctions and desmosomes. In addition to adhesive junctions, epithelial cells often contain other types of cell junctions that are also located along

(a)

(b)

Neural tube

(c)
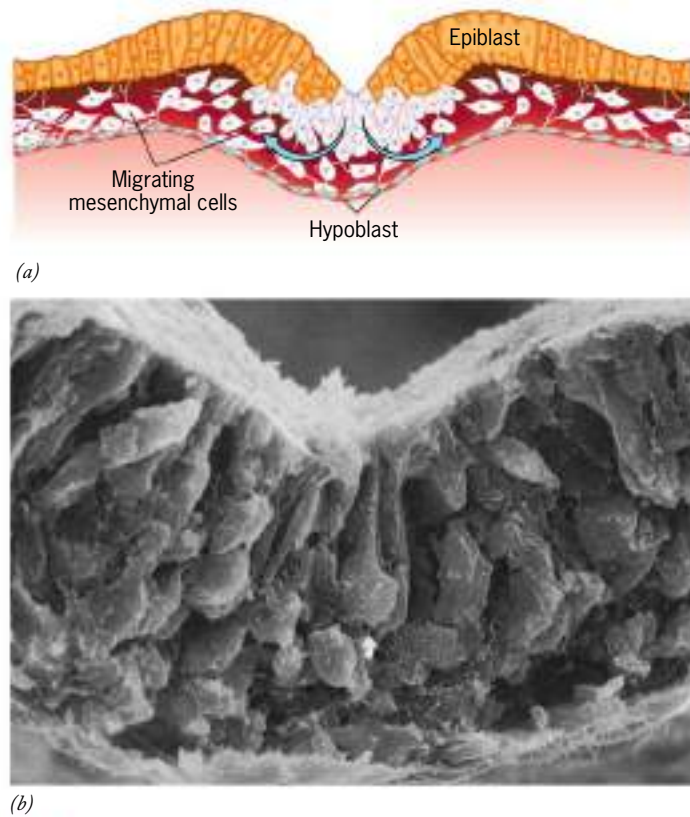
FIGURE 7.24 Cadherins and morphogenesis. (*a*) During gastrulation, cells in the upper layer of the embryo (the epiblast) move toward a groove in the center of the embryo, sink into the groove, and then migrate laterally as mesenchymal cells in the space beneath the epiblast. This epithelial-mesenchymal transition is marked by a loss of expression of E-cadherin that is characteristic of epithelial cells. Cells expressing E-cadherin are depicted in orange. (*b*) Scanning electron micrograph of a chick embryo during gastrulation that has been fractured to reveal the cells undergoing the epithelial-mesenchymal transition (arrow) depicted in part *a*. (*c*) This sequence of drawings depicts the development of the neural tube, which is an epithelial layer that forms by separation from the overlying layer of dorsal ectoderm. In the top drawing, the epithelial cells are expressing E-cadherin. In the lower drawings, the cells of the neural tube stop expressing E-cadherin (orange) and instead express N-cadherin (blue). (B: COURTESY OF MICHAEL SOLURSH AND JEAN PAUL REVEL.)
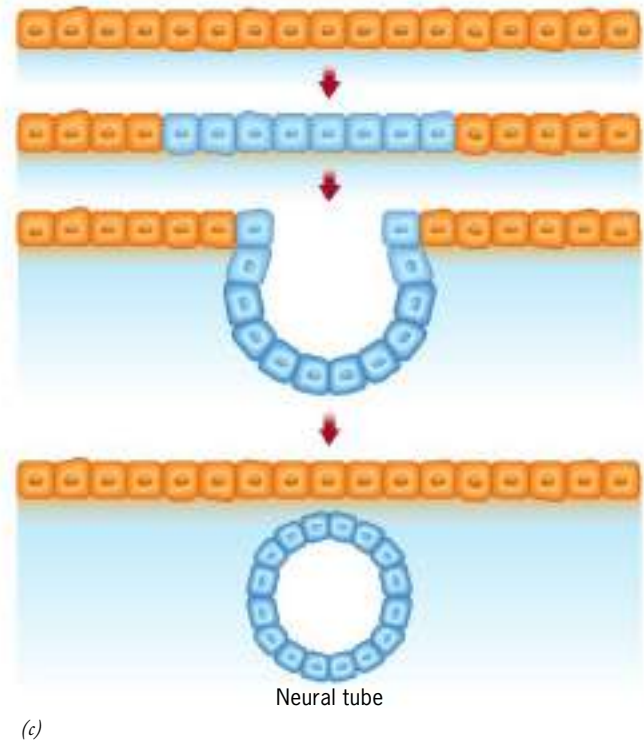
their lateral surfaces near the apical lumen (Figure 7.25). When these junctions are arranged in a specific array, this assortment of surface specializations is called a *junctional complex*. The structures and functions of the two adhesive junctions of the complex are described in the following paragraphs, whereas discussion of the other types of epithelial junctions (tight junctions and gap junctions) is taken up later in the chapter.

Adherens junctions are found in a variety of sites within the body. They are particularly common in epithelia, such as the lining of the intestine, where they occur as "belts" (or *zonulae adherens*) that encircle each of the cells near its apical surface, binding that cell to its surrounding neighbors (Figure 7.25*a*). In an adherens junction, cells are held together by calcium-dependent linkages formed between the extracellular domains of cadherin molecules that bridge the 30-nm gap between neighboring cells (Figure 7.26). As Figure 7.26 illustrates, the cytoplasmic domain of these cadherins is linked by α- and β-catenins to a variety of cytoplasmic proteins, including actin filaments of the cytoskeleton. Thus, like the integrins of a focal adhesion, the cadherin clusters of an adherens junction (1) connect the external environment to the actin cytoskeleton and (2) provide a pathway for signals to be transmitted from the cell exterior to the cytoplasm.

To give one example, the adherens junctions situated between endothelial cells that line the walls of blood vessels transmit signals that ensure the survival of the cells. Mice that are lacking an endothelial cell cadherin are unable to transmit these survival signals, and these animals die during embryonic development as a result of the death of the cells lining the vessel walls.

Desmosomes (or *maculae adherens*) are disk-shaped adhesive junctions approximately 1 μm in diameter (Figure 7.27*a*) that are found in a variety of tissues. Desmosomes are particularly numerous in tissues that are subjected to mechanical stress, such as cardiac muscle and the epithelial layers of the skin and uterine cervix. Like adherens junctions, desmosomes contain cadherins that link the two cells across a narrow extracellular gap. The cadherins of desmosomes have a different domain structure from the classical cadherins found in adherens junctions and are referred to as *desmogleins* and *desmocollins* (Figure 7.27*b*). Dense cytoplasmic plaques on the inner surface of the plasma membranes serve as sites of anchorage for looping intermediate filaments similar to those of hemidesmosomes (Figure 7.19). The three-dimensional network of ropelike intermediate filaments provides structural continuity and tensile strength to the entire sheet of cells. Intermediate filaments are linked to the
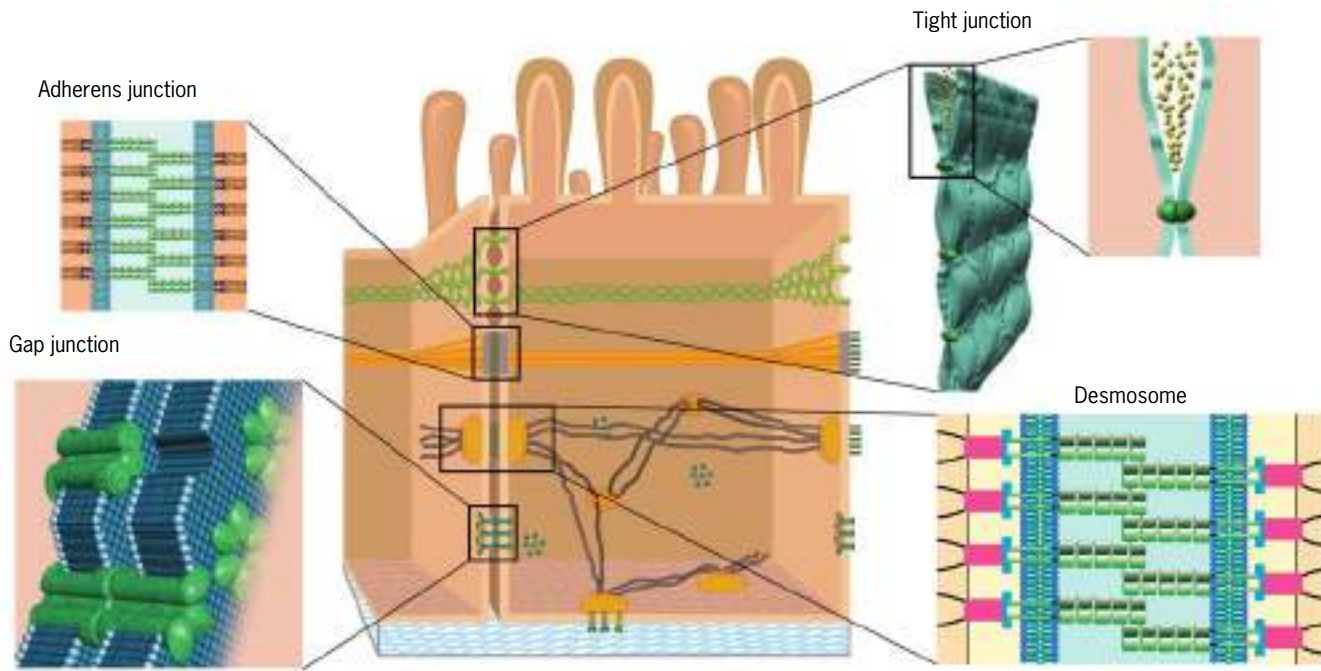
(a)
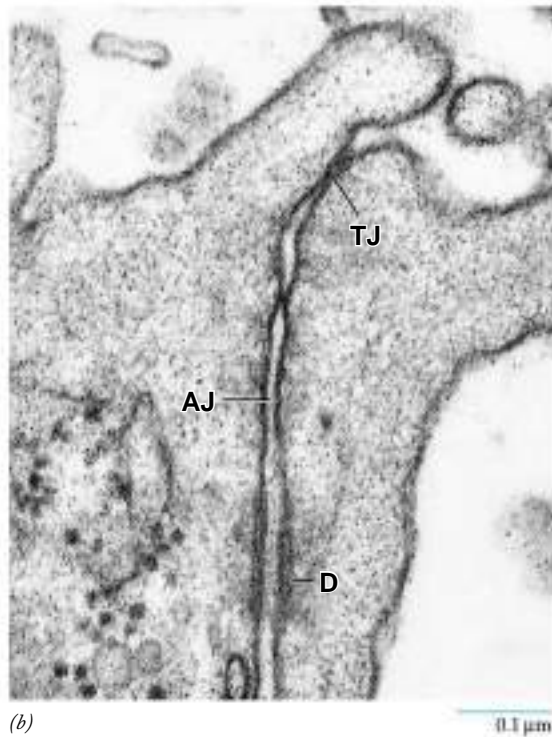
(b)

FIGURE 7.25 An intercellular junctional complex. (*a*) Schematic diagram showing the junctional complex on the lateral surfaces of a simple columnar epithelial cell. The complex consists of a tight junction (zonula occludens), an adherens junction (zonula adherens), and a desmosome (macula adherens). Other desmosomes and gap junctions are located deeper along the lateral surfaces of the cells. Adherens junctions and tight junctions encircle the cell, whereas desmosomes and gap junctions are restricted to a particular site between adjacent cells. Hemidesmosomes are shown at the basal cell surface. (*b*) Electron micrograph of a junctional complex between two rat airway epithelial cells (TJ, tight junction; AJ, adherens junction; D, desmosome). (B: FROM EVELINE E. SCHNEEBERGER AND ROBERT D. LYNCH, AM. J. PHYSIOL. 262:L648, 1992.)

FIGURE 7.26 Schematic model of the molecular architecture of an adherens junction. The cytoplasmic domain of each cadherin molecule is connected to the actin filaments of the cytoskeleton by linking proteins, including β-catenin, α-catenin, and various actin-binding proteins. One of these actin-binding proteins is formin, which participates in the polymerization of the actin filaments. Actin filament assembly at the junction may be regulated by α-catenin. β-catenin is also a key element in the Wnt signaling pathway, which transmits signals from the cell surface to the cell nucleus. Disassembly of adherens junctions may release β-catenin to participate in this pathway, leading to the activation of gene expression. Another member of the catenin family, p120 catenin, binds to a site on the cytoplasmic domain of the cadherin. p120 catenin may regulate the adhesive strength of the junction and serve as a component of intracellular signaling pathways. Numerous other proteins found in these junctions are not indicated.

(b)
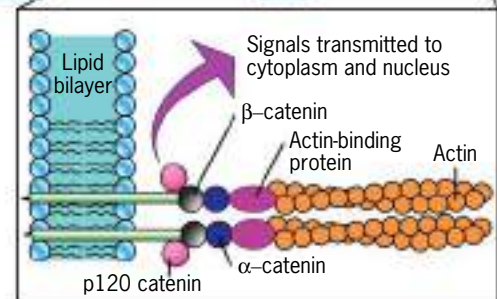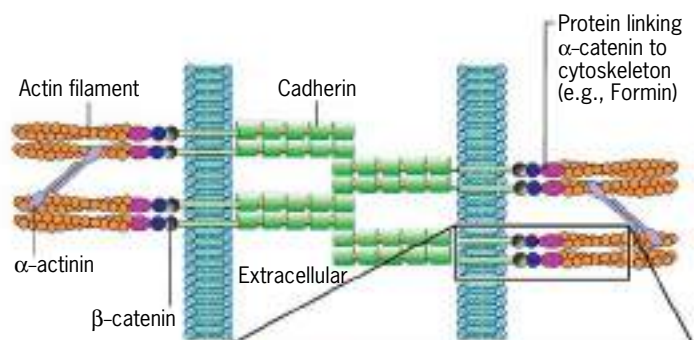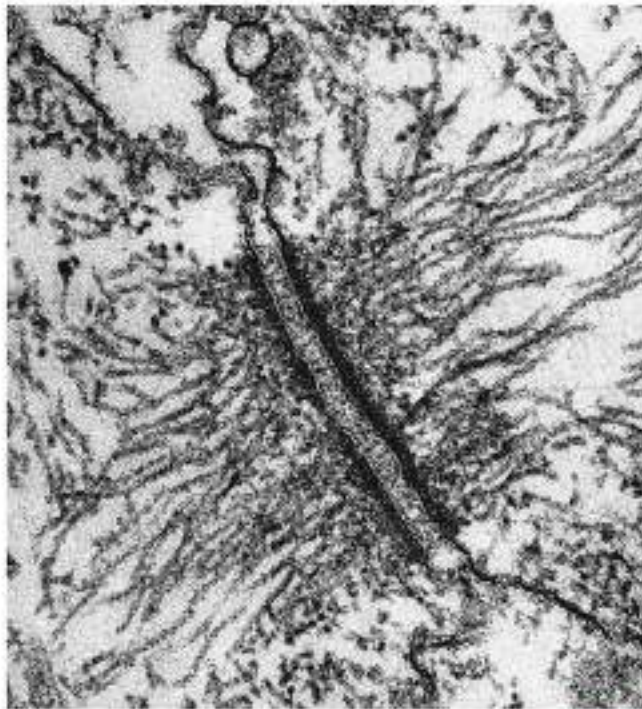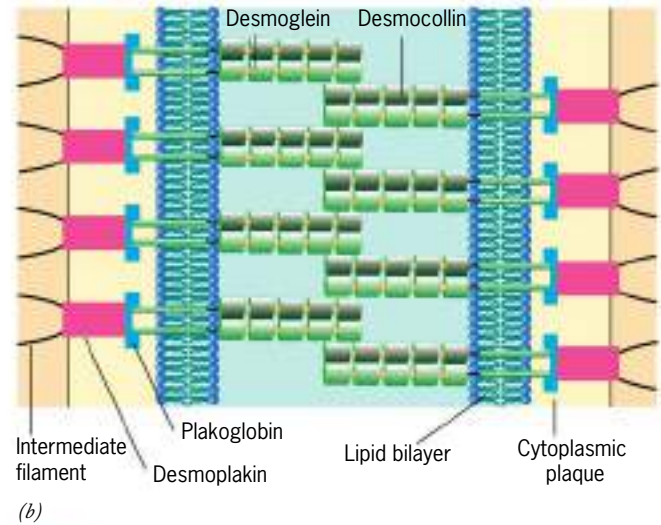
FIGURE 7.27 **The structure of a desmosome.** (*a*) Electron micrograph of a desmosome from newt epidermis. (*b*) Schematic model of the molecular architecture of a desmosome. (A: FROM DOUGLAS E. KELLY, J. CELL BIOL. 28:51, 1966; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

(a)

cytoplasmic domains of the desmosomal cadherins by additional proteins, as depicted in Figure 7.27*b*. The importance of cadherins in maintaining the structural integrity of an epithelium is illustrated by an autoimmune disease (*pemphigus vulgaris*) in which antibodies are produced against one of the desmogleins. The disease is characterized by a loss of epidermal cell–cell adhesion and severe blistering of the skin.

The Role of Cell-Adhesion Receptors in Transmembrane Signaling

Figure 7.28 provides a summary of some of the points that have been made in this chapter. The drawing depicts each of the four types of cell-adhesion molecules discussed and their interactions with both extracellular and cytoplasmic materials. One of the roles of integral membrane proteins is to transfer informa-
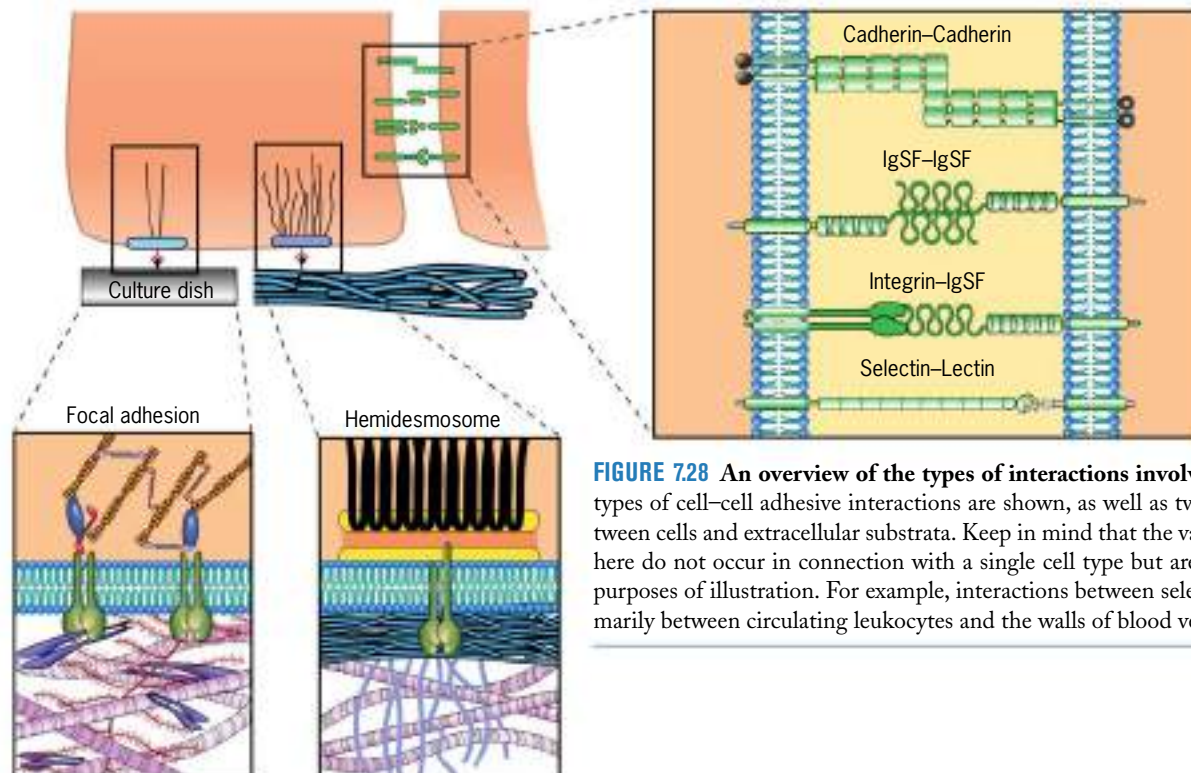
FIGURE 7.28 **An overview of the types of interactions involving the cell surface.** Four types of cell–cell adhesive interactions are shown, as well as two types of interactions between cells and extracellular substrata. Keep in mind that the various interactions depicted here do not occur in connection with a single cell type but are shown in this manner for purposes of illustration. For example, interactions between selectins and lectins occur primarily between circulating leukocytes and the walls of blood vessels.

tion across the plasma membrane, a process known as **transmembrane signaling**. While the subject will be explored in detail in Chapter 15, it can be noted that all four types of cell-adhesion molecules illustrated in Figure 7.28 have the potential to carry out this function. For example, integrins and cadherins can transmit signals from the extracellular environment to the cytoplasm by means of linkages with the cytoskeleton and with cytosolic regulatory molecules, such as protein kinases and G proteins. Protein kinases activate (or inhibit) their target proteins through phosphorylation, whereas G proteins activate (or inhibit) their protein targets through physical interaction (see Figure 15.19b). The engagement of an integrin with its ligand can induce a variety of responses within a cell, including changes in cytoplasmic pH or Ca^{2+} concentration, protein phosphorylation, and gene expression. These changes, in turn, can alter a cell's growth potential, migratory activity, state of differentiation, or survival. This type of phenomenon is

(a)
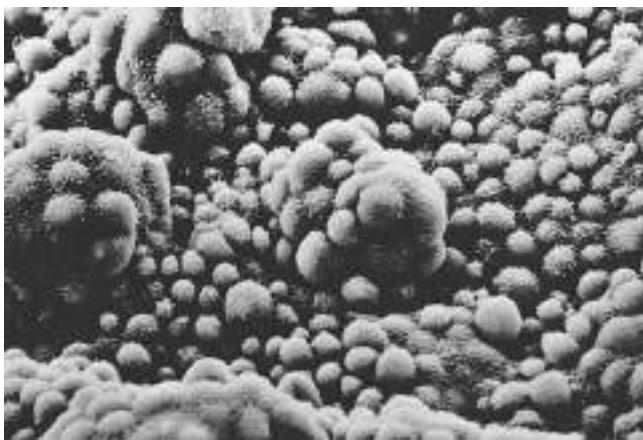
(b)

FIGURE 7.29 The role of extracellular proteins in maintaining the differentiated state of cells. (*a*) These mammary gland epithelial cells were removed from a mouse and cultured in the absence of an extracellular matrix. Unlike normal differentiated mammary cells, these cells are flattened and are not engaged in the synthesis of milk proteins. (*b*) When extracellular matrix molecules are added back to the culture, the cells regain their differentiated appearance and synthesize milk proteins. (COURTESY OF JOANNE EMERMAN.)

illustrated by the mammary gland epithelial cells shown in Figure 7.29. When these cells are removed from a mammary gland and grown on a bare culture dish, they lose their ability to synthesize milk proteins and appear as flattened, undifferentiated cells (Figure 7.29a). When these same undifferentiated cells are cultured in the presence of certain extracellular molecules (e.g., laminin), they regain their differentiated appearance and become organized into milk-producing, gland-like structures (Figure 7.29b). Laminin is thought to stimulate mammary cells by binding to cell-surface integrins and activating kinases at the inner surface of the membrane (as in Figure 7.17c).

> ## REVIEW ?
>
> 1. Distinguish between a hemidesmosome and a desmosome; a desmosome and an adherens junction.
> 2. Which type(s) of cell junctions contain actin filaments? Which contain(s) intermediate filaments? Which contain(s) integrins? Which contain(s) cadherins?
> 3. How do cadherins, IgSF members, and selectins differ at the molecular level in the way they mediate cell–cell adhesion?

7.4 TIGHT JUNCTIONS: SEALING THE EXTRACELLULAR SPACE

A simple epithelium, like the lining of the intestine or lungs, is composed of a layer of cells that adhere tightly to one another to form a thin cellular sheet. Biologists have known for decades that when certain types of epithelia, such as frog skin or the wall of the urinary bladder, are mounted between two compartments containing different solute concentrations, very little diffusion of ions or solutes is observed across the wall of the epithelium from one compartment to the other. Given the impermeability of plasma membranes, it is not surprising that solutes cannot diffuse freely through the cells of an epithelial layer. But why aren't they able to pass between cells by way of the *paracellular pathway* (as in Figure 7.30a)? The reason became apparent during the 1960s with the discovery of specialized contacts, called **tight junctions** (or *zonulae occludens*), between neighboring epithelial cells.

Tight junctions (TJs) are located at the very apical end of the junctional complex between adjacent epithelial cells (see Figure 7.25). An electron micrograph of a section through a TJ that has been cut to include the plasma membranes of the adjacent cells is shown in Figure 7.30a. A higher magnification view showing the interaction between the membranes of a TJ is shown in the inset of Figure 7.30a. It is evident that the adjoining membranes make contact at intermittent points, rather than being fused over a large surface area. As indicated in Figure 7.30b, the points of cell–cell contact are sites where integral proteins of two adjacent membranes meet within the extracellular space.

Freeze fracture, which allows observation of the internal faces of a membrane (Figure 4.15), shows that the plasma

membranes of a TJ contain interconnected strands (Figure 7.30*c*) that run mostly parallel to one another and to the apical surface of the epithelium. The strands (or grooves in the opposite face of a fractured membrane) correspond to paired rows of aligned integral membrane proteins that are illustrated in the inset of Figure 7.30*b*. The integral proteins of TJs form continuous fibrils that completely encircle the cell like a gasket and make contact with neighboring cells on all sides (Figure 7.30*d*). As a result, TJs serve as a barrier to the free diffusion of water and solutes from the extracellular compartment on one side of an epithelial sheet to that on the other side. Tight junctions also serve as "fences" that help maintain

the polarized character of epithelial cells (see Figure 4.30). They accomplish this by blocking the diffusion of integral proteins between the apical domain of the plasma membrane and its lateral and basal domains. Like other sites of cell adhesion, tight junctions are also involved in signaling pathways that regulate numerous cellular processes.

Not all TJs exhibit the same permeability properties. Part of the explanation can be seen under the electron microscope: TJs with several parallel strands (such as that in Figure 7.30*c*) tend to form better seals than junctions with only one or a couple of strands. But there is more to the story than numbers of strands. Some TJs are permeable to *specific* ions or solutes to which other

(a)

(c)

0.2 μm

(b)

(d)
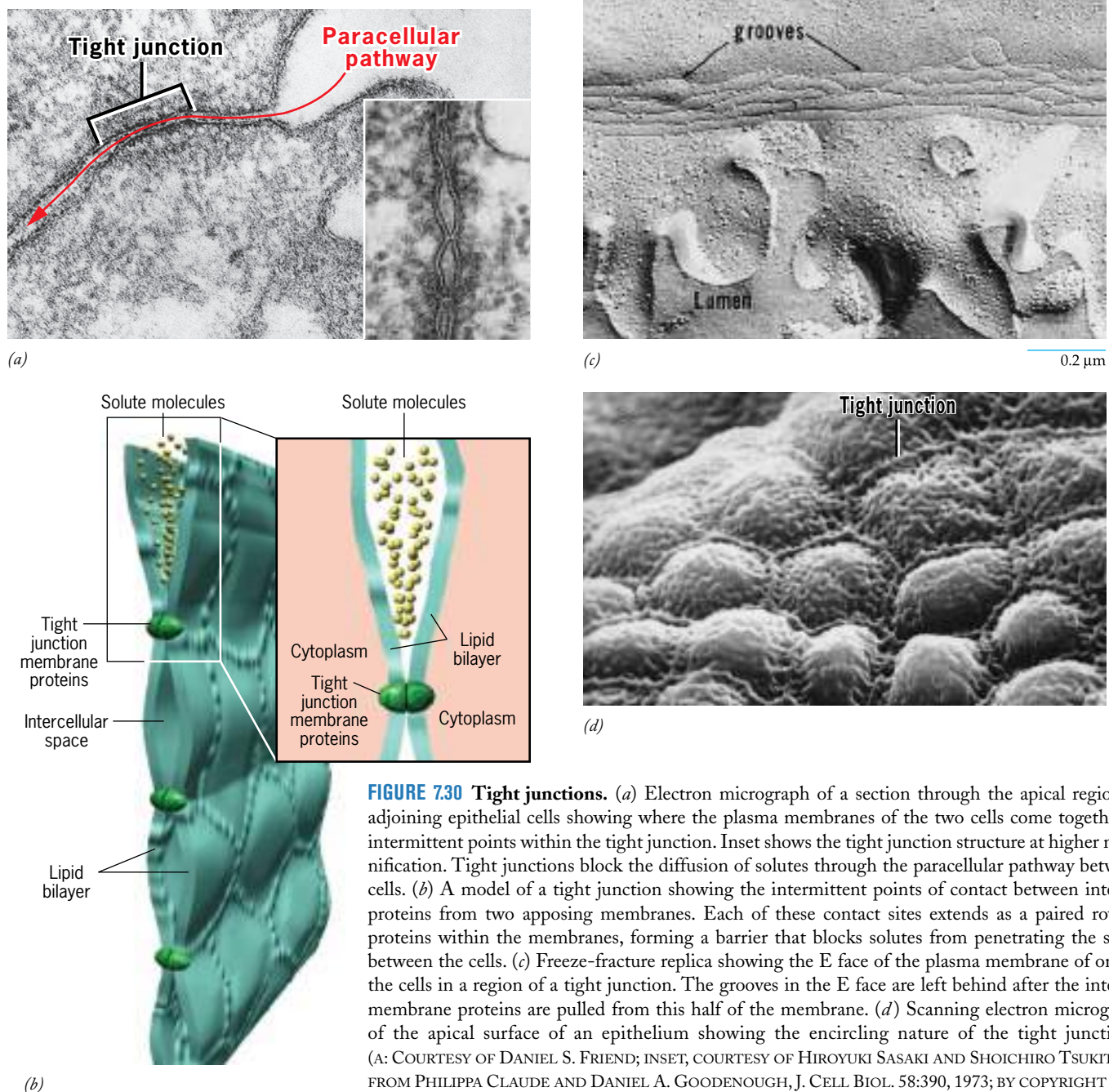
FIGURE 7.30 Tight junctions. (*a*) Electron micrograph of a section through the apical region of adjoining epithelial cells showing where the plasma membranes of the two cells come together at intermittent points within the tight junction. Inset shows the tight junction structure at higher magnification. Tight junctions block the diffusion of solutes through the paracellular pathway between cells. (*b*) A model of a tight junction showing the intermittent points of contact between integral proteins from two apposing membranes. Each of these contact sites extends as a paired row of proteins within the membranes, forming a barrier that blocks solutes from penetrating the space between the cells. (*c*) Freeze-fracture replica showing the E face of the plasma membrane of one of the cells in a region of a tight junction. The grooves in the E face are left behind after the integral membrane proteins are pulled from this half of the membrane. (*d*) Scanning electron micrograph of the apical surface of an epithelium showing the encircling nature of the tight junctions. (A: COURTESY OF DANIEL S. FRIEND; INSET, COURTESY OF HIROYUKI SASAKI AND SHOICHIRO TSUKITA; C: FROM PHILIPPA CLAUDE AND DANIEL A. GOODENOUGH, J. CELL BIOL. 58:390, 1973; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS; D: COURTESY OF D. TARIN.)

TJs are impermeable. Studies over the past decade have shed considerable light on the molecular basis of TJ permeability.

Up until 1998 it was thought that TJ strands were composed of a single protein, *occludin*. Then, it was found that cultured cells that lacked a gene for occludin, and thus could not produce the protein, were still able to form TJ strands of normal structure and function. Subsequent studies by Shoichiro Tsukita and his colleagues at the University of Kyoto led to the discovery of a family of proteins called *claudins* that form the major structural component of TJ strands. The electron micrograph of Figure 7.31 shows that occludin and claudin are present together within the linear fibrils of a TJ. At least 24 different claudins have been identified, and differences in the distribution of these proteins may explain selective differences in TJ permeability. For example, one small region of a human kidney tubule—a region known as the thick ascending limb (or TAL)—has TJs that are permeable to magnesium (Mg^{2+}) ions. It is thought that the loops of the claudin molecules that extend into the extracellular space form pores in the TAL that are selectively permeable to Mg^{2+} ions. Support for this concept has come from the finding that one specific member of the claudin family, claudin-16, is expressed primarily in the TAL. The importance of claudin-16 in kidney function was revealed in studies of patients suffering from a rare disease characterized by abnormally low Mg^{2+} levels in their blood. These patients were found to have mutations in both copies of their *claudin-16* gene. The Mg^{2+} level in their blood is low because tight junctions containing the abnormal claudin are impermeable to Mg^{2+}. As a result, this important ion fails to be reabsorbed from the tubule and is simply excreted in the urine.

Another important function of tight junctions came to light in 2002. It was thought for decades that the impermeability of mammalian skin to water was solely a property of the outer, cornified layer of the skin (see Figure 7.1), which contains tightly packed protein filaments and associated lipids. It was discovered, however, that mice lacking a gene for claudin-1 died shortly after birth as a result of dehydration. Further investigation revealed that the cells in one of the outer layers of *normal* epidermis are connected to one another by tight junctions. Animals lacking the gene for claudin-1 were unable to assemble watertight epidermal tight junctions and, as a result, suffered from uncontrolled water loss.

Tight junctions are also present between the endothelial cells that line the walls of capillaries. These junctions are particularly evident in the brain where they help form the *blood–brain barrier*, which prevents substances from passing from the bloodstream into the brain. Although small ions and even water molecules may not be able to penetrate the blood–brain barrier, cells of the immune system are able to pass across the endothelium through these junctions. These cells are thought to send a signal that opens the junction, allowing the cells to pass. While protecting the brain from unwanted solutes, the blood–brain barrier also prevents access of many drugs to the central nervous system. Consequently, a major goal of the pharmaceutical industry is to develop drugs that temporarily open the tight junctions of the brain to allow passage of therapeutic compounds.

REVIEW ?

1. What does freeze-fracture analysis tell you about the structure of a junction that can't be learned from examination of stained tissue sections?

2. How does the structure of a tight junction contribute to its function?

FIGURE 7.31 The molecular composition of tight junction strands. Electron micrograph of a freeze-fracture replica of cells that had been joined to one another by tight junctions. The fracture faces were incubated with two types of gold-labeled antibodies. The smaller gold particles (arrowheads) reveal the presence of claudin molecules, whereas the larger gold particles (arrows) indicate the presence of occludin. These experiments demonstrate that both proteins are present in the same tight junction strands. Bar equals 0.15 μm. The inset shows a possible arrangement of the two integral membrane proteins as they make contact in the intercellular space. Both the claudins (red) and occludin (brown) span the membrane four times. (MICROGRAPH FROM MIKIO FURUSE, HIROYUKI SASAKI, KAZUSHI FUJIMOTO, AND SHOICHIRO TSUKITA, J. CELL BIOLOGY 143:398, 1998; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

7.5 GAP JUNCTIONS AND PLASMODESMATA: MEDIATING INTERCELLULAR COMMUNICATION

Gap junctions are sites between animal cells that are specialized for intercellular communication. Electron micrographs reveal gap junctions to be sites where the plasma membranes of adjacent cells come very close to one another (within about 3 nm) but do not make direct contact. Instead, the cleft between the cells is spanned by very fine strands (Figure 7.32*a*) that are actually molecular "pipelines" that pass through the adjoining plasma membranes and open into the cytoplasm of the neighboring cells (Figure 7.32*b*).

Gap junctions have a simple molecular composition; they are composed entirely of an integral membrane protein called *connexin*. Connexins are organized into multisubunit complexes, called **connexons**, that completely span the membrane (Figure 7.32*b*). Each connexon is composed of six connexin subunits arranged in a ring around a central opening, or *annu-*

lus, that is approximately 1.5 nm in diameter at its extracellular surface (Figure 7.32*c*, left).

During the formation of gap junctions, the connexons in the plasma membranes of apposing cells become tightly linked to one another through extensive noncovalent interactions of the extracellular domains of the connexin subunits. Once aligned, connexons in apposing plasma membranes form complete intercellular channels connecting the cytoplasm of one cell with the cytoplasm of its neighbor (Figure 7.32*b*). Large numbers of connexons become clustered in specific regions of the membrane, forming gap-junction plaques that can be visualized when the membrane is split down the middle by freeze fracture (Figure 7.32*d*).

As discussed in the Experimental Pathways (see www. wiley.com/college/karp), gap junctions are sites of communi-

cation between the cytoplasms of adjacent cells. The existence of gap-junction intercellular communication (GJIC) is revealed through the passage of either ionic currents or low-molecular-weight dyes, such as fluorescein, from one cell to its neighbors (Figure 7.33). Mammalian gap junctions allow the diffusion of molecules having a molecular mass below approximately 1000 daltons. In contrast to the highly selective ion channels that connect a cell to the external medium (page 149), gap-junction channels are relatively nonselective. Just as ion channels can be open or closed, gap-junction channels are also thought to be gated. Channel closure is probably triggered primarily by phosphorylation of connexin subunits. Closure can also be triggered by changes in voltage or abnormally high Ca^{2+} concentrations (Figure 7.32*c*, right).

We saw in Chapter 4 how skeletal muscle cells are stimulated by chemicals released from the tips of nearby nerve cells. Stimulation of a cardiac or smooth muscle cell occurs by a very different process, one involving gap junctions. The contraction of the mammalian heart is stimulated by an electrical im-

(a)   0.15 μm

(b)

(c)

(d)
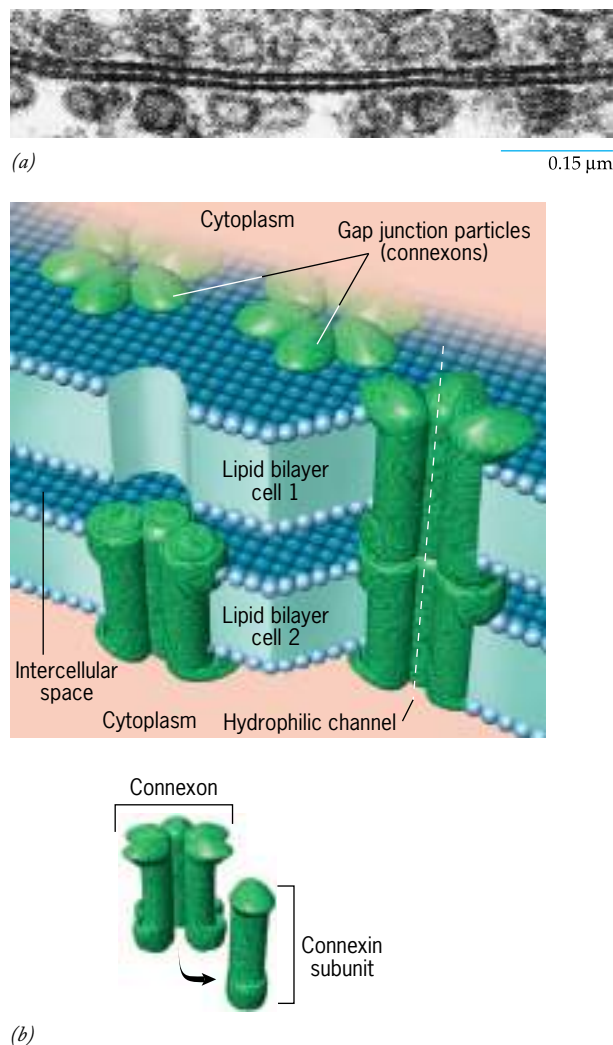
FIGURE 7.32 Gap junctions. (*a*) Electron micrograph of a section through a gap junction perpendicular to the plane of the two adjacent membranes. The "pipelines" between the two cells are seen as electron-dense beads on the apposed plasma membranes. (*b*) Schematic model of a gap junction showing the arrangement of six connexin subunits to form a connexon, which contains half of the channel that connects the cytoplasm of the two adjoining cells. Each connexin subunit is an integral protein with four transmembrane domains. (*c*) High-resolution images derived from atomic force microscopy of the extracellular surface of a single connexon in the open (left) and closed (right) conformations. Closure of the connexon was induced by exposure to elevated Ca^{2+} ion concentration. (*d*) Freeze-fracture replica of a gap junction plaque showing the large numbers of connexons and their high concentration. (A: FROM CAMILLO PERACCHIA AND ANGELA F. DULHUNTY, J. CELL BIOL. 70:419, 1976; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS; C: COURTESY OF GINA E. SOSINSKY; D: COURTESY OF DAVID ALBERTINI.)
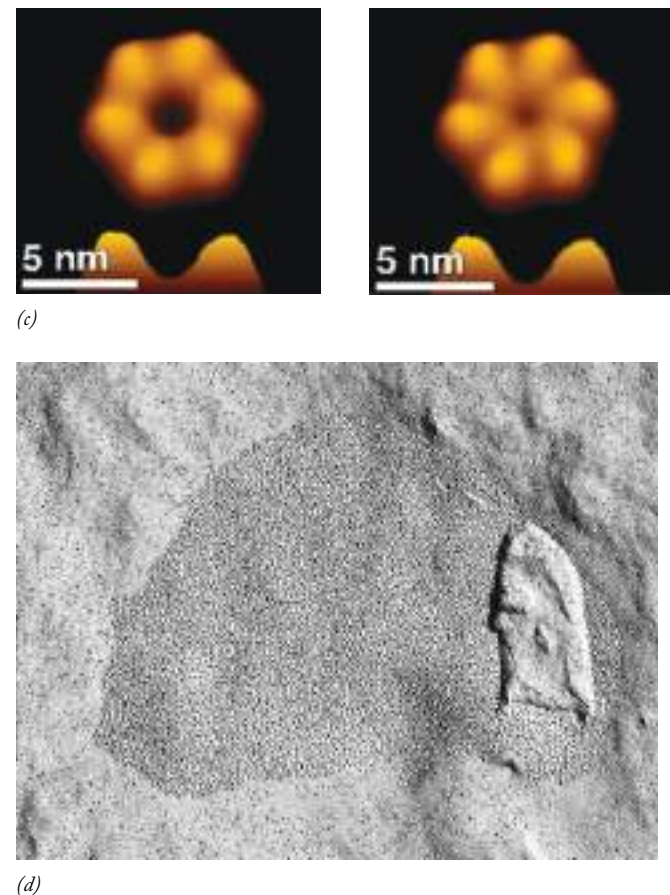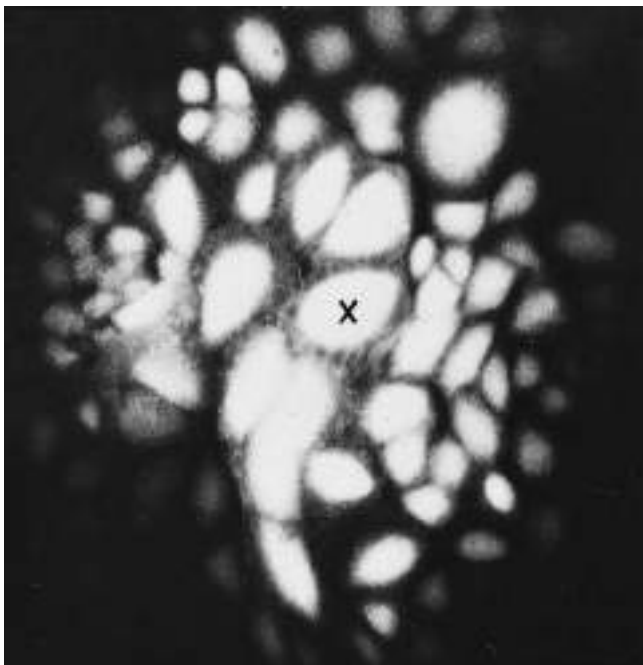
FIGURE 7.33 Results of an experiment demonstrating the passage of low-molecular-weight solutes through gap junctions. Micrograph showing the passage of fluorescein from one cell into which it was injected (X) to the surrounding cells. (FROM R. AZARNIA AND W. R. LOEWENSTEIN, J. MEMB. BIOL. 6:378, 1971; WITH PERMISSION FROM SPRINGER SCIENCE AND BUSINESS MEDIA.)

pulse generated in a small region of specialized heart muscle called the *sinoatrial node*, which acts as the heart's pacemaker. The impulse spreads rapidly as a current of ions flows through gap junctions from one cardiac muscle cell to its neighbors, causing the cells to contract in synchrony. Similarly, the flow of ions through gap junctions that interconnect smooth muscle cells in the wall of the esophagus or intestine leads to the generation of coordinated peristaltic waves that move down the length of the wall.[2]

Gap junctions can put a large number of cells of a tissue into intimate cytoplasmic contact. This has important physiologic consequences because a number of highly active regulatory substances, such as cyclic AMP and inositol phosphates (Chapter 15), are small enough to fit through gap-junction channels. As a result, gap junctions have the potential to integrate the activities of individual cells of a tissue into a functional unit. If, for example, only a few cells near a particular blood vessel happen to be stimulated by a hormone, the stimulus can be rapidly transmitted to all cells of the tissue. Gap junctions also allow cells to cooperate metabolically by sharing key metabolites, such as ATP, sugar phosphates, amino acids, and many coenzymes, which are small enough to pass through these intercellular channels. This is particularly important in tissues such as the lens, which are avascular (i.e., lack blood vessels).

---

[2]As discussed in the Experimental Pathways on the Web, gap junctions also occur between the presynaptic and postsynaptic membranes of adjacent nerve cells in certain parts of the brain, allowing nerve impulses to be transmitted directly from one neuron to another without requiring the release of chemical transmitters.

Connexins (Cx), the proteins of which gap junctions are constructed, are members of a multigene family. Approximately 20 different connexins with distinct tissue-specific distributions have been identified. Connexons composed of different connexins exhibit marked differences in conductance, permeability, and regulation. In some cases, connexons in neighboring cells that are composed of different connexins are able to dock and form functional channels, whereas in other cases, they are not. These compatibility differences may play important roles in either promoting or preventing communication between different types of cells in an organ. For example, connexons joining cardiac muscle cells are composed of the connexin Cx43, whereas connexons joining the cells that make up the heart's electrical conduction system are composed of Cx40. Because these two connexins form incompatible connexons, the two types of cells are electrically insulated from one another even though they are in physical contact. A number of inherited disorders have been associated with mutations in genes encoding connexins. Consequences of these disorders include deafness, blindness, cardiac arrhythmias, skin abnormalities, or nerve degeneration.

Over the past few years a new type of communication system has been discovered that consists of thin, highly elongated tubules capable of conducting cell-surface proteins, cytoplasmic vesicles, and calcium signals from one cell to another. To date, these *tunneling nanotubes*, as they are called, have been observed almost exclusively between cells growing in culture (as in Figure 7.34), so it remains to be seen whether they have a significant physiological role within the body.

Plasmodesmata

Unlike animals, whose cells make intimate contact with one another, plant cells are separated from one another by a substantial barrier—the cell wall. It is not surprising, therefore, that plants lack the cell-adhesion molecules we have been discussing in this chapter. Although plants lack the specialized junctions found in animal tissues, most plant cells are connected to one another by plasmodesmata. **Plasmodesmata** (singular, **plasmodesma**) are cytoplasmic channels that pass through the cell walls of adjacent cells. A simple (i.e., unbranched) plasmodesma is shown in Figure 7.35*a,b*. Plasmodesmata are lined by plasma membrane and usually contain a dense central structure, the *desmotubule*, derived from the smooth endoplasmic reticulum of the two cells. Like gap junctions between animal cells, plasmodesmata serve as sites of cell-to-cell communication, as substances pass through the annulus surrounding the desmotubule.

It was thought for many years that plasmodesmata were impermeable to molecules greater than about 1000 daltons (1 kDa). This conclusion was based on studies in which different-sized fluorescent dyes were injected into cells. More recent studies suggest that plasmodesmata allow much larger molecules (up to 50 kDa) to pass between cells, owing to the fact that the plasmodesmatal pore is capable of dilation. The first insight into this dynamic property was gained from studies in the 1980s on plant viruses that spread from one cell to another through plasmodesmata. It was found that the viruses encoded a *movement protein* that interacted with the wall of the plasmodes-
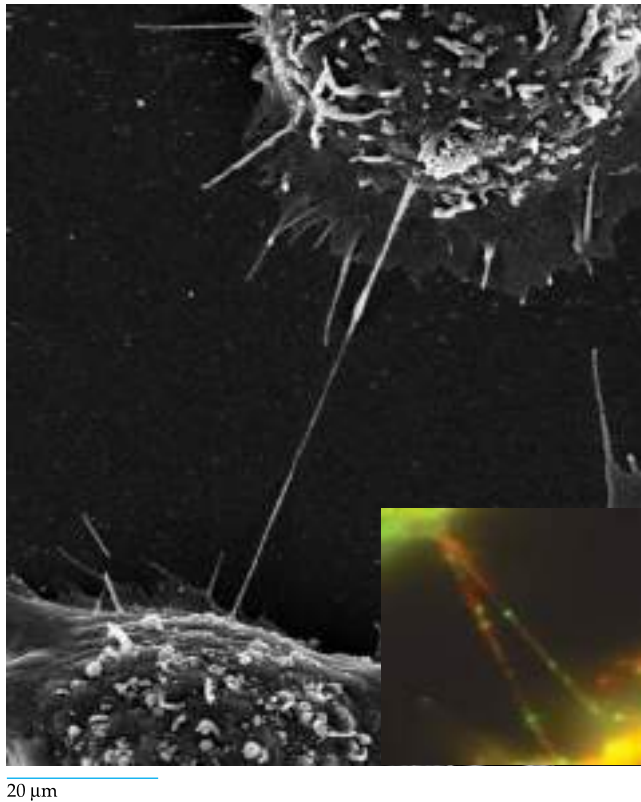
FIGURE 7.34 Tunneling nanotubes. Scanning electron micrograph showing two cultured neuroendocrine cells connected to one another by a thin tubular process capable of carrying materials between the cytoplasm of the neighboring cells. These processes, which are only about 100 nm in diameter, are supported by an internal actin "skeleton." The inset shows a number of fluorescently labeled vesicles caught in the act of movement between two cells. (REPRINTED WITH PERMISSION FROM AMIN RUSTOM ET AL., COURTESY OF HANS-HERMANN GERDES, SCIENCE 303:1007, 2004; COPYRIGHT 2004, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)
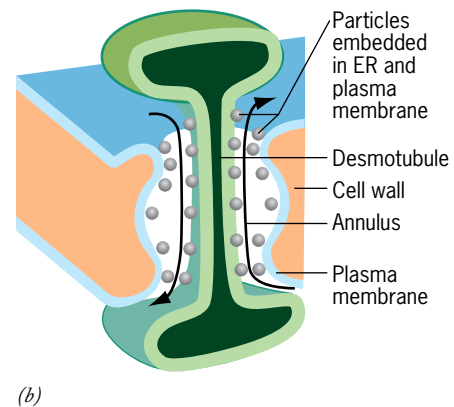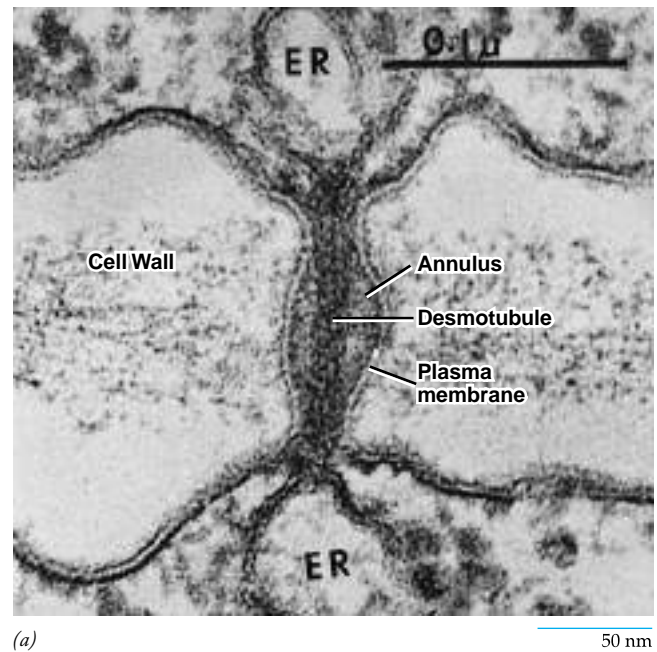
FIGURE 7.35 Plasmodesmata. (*a*) Electron micrograph of a section through a plasmodesma of a fern gametophyte. The desmotubule is seen to consist of membrane that is continuous with the endoplasmic reticulum (ER) of the cytoplasm on both sides of the plasma membrane. (*b*) Schematic drawing of a plasmodesma. The black arrows indicate the pathways taken by molecules as they pass through the annulus from cell to cell. (*c*) An example of the movement of a protein from one cell to another within a plant root. The smaller inset shows the localization of the fluorescently labeled messenger RNA molecules (green) that encode a protein called Shr. The mRNA is localized within the cells of the stele (Ste), which is thus the tissue in which this protein is synthesized. The larger photo shows the localization of the fluorescently labeled Shr protein (also green), which is present both within the stele cells where it is synthesized and the adjoining endodermal cells (End) into which it has passed by way of the connecting plasmodesmata. The transported protein is localized within the nuclei of the endodermal cells where it acts as a transcription factor. Bars: 50 μm and 25 μm (inset). (A: FROM LEWIS G. TILNEY, TODD J. COOKE, PATRICIA S. CONNELLY, AND MARY S. TILNEY, J. CELL BIOL. 112:740, 1991; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS; C: REPRINTED FROM KEIJI NAKAJIMA ET AL., COURTESY OF PHILIP N. BENFEY, NATURE 413:308, 2001; COPYRIGHT 2001 MACMILLAN MAGAZINES LTD.)
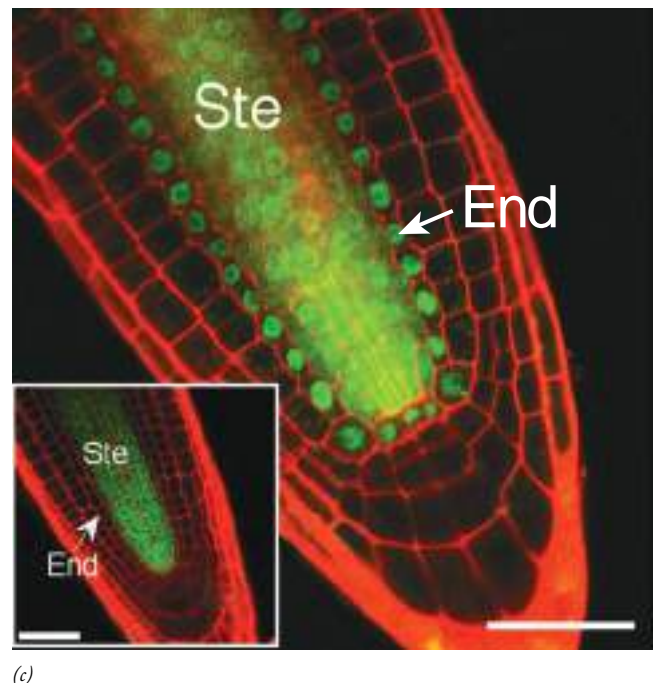
mata and increased the diameter of the pore. Subsequent studies revealed that plant cells produce their own movement proteins that regulate the flow of proteins and RNAs from cell to cell. Some of these macromolecules find their way into the vascular system of the plant, where they integrate plant-wide activities, such as the growth of new leaves and flowers or defense against pathogens. Figure 7.35*c* documents the movement of a protein (labeled with green fluorescence) from one type of plant tissue (the stele), where it was synthesized, to an adjoining tissue (the endodermis). The protein is seen to be concentrated in the spherical nuclei of the single layer of endodermal cells where it acts to stimulate gene transcription.

REVIEW ?

1. Compare the arrangement of integral membrane proteins in a tight junction versus a gap junction.
2. How are plasmodesmata and gap junctions similar? How are they dissimilar? Would you expect a moderately sized protein to pass through a plasmodesma?

7.6 CELL WALLS

Because a lipid–protein plasma membrane of less than 10-nm thickness can be expected to offer only minimal protection for a cell's contents, it is not surprising that "naked" cells are extremely fragile structures. The cells of nearly all organisms other than animals are enclosed in a protective outer envelope. Protozoa have a thickened outer coat, whereas bacteria, fungi, and plants have distinct **cell walls**. We will restrict our discussion to the cell walls of plants, which were the first cellular structures to be observed with a light microscope (page 2).

Plant cell walls perform numerous vital functions. As discussed on page 146, plant cells develop turgor pressure that pushes against their surrounding wall. As a result, the wall gives the enclosed cell its characteristic polyhedral shape (Figure 7.36*a*). In addition to providing support for individual cells, cell walls serve collectively as a type of "skeleton" for the entire plant. In fact, a tree without cell walls might resemble, in many respects, a human without bones. Cell walls also pro-

(a)

(b)   100 nm

(c)

FIGURE 7.36 The plant cell wall. (*a*) Electron micrograph of a plant cell surrounded by its cell wall. The middle lamella is a pectin-containing layer situated between adjacent cell walls. (*b*) Electron micrograph showing the cellulose microfibrils and hemicellulose cross-links of an onion cell wall after extraction of the nonfibrous pectin polymers. (*c*) Schematic diagram of a model of a generalized plant cell wall. (A: COURTESY OF W. GORDON WHALEY; B: FROM M. C. MCCANN, B. WELLS, AND K. ROBERTS, J. CELL SCI. 96:329, 1990; BY PERMISSION OF THE COMPANY OF BIOLOGISTS LTD.)

(a)

0.1 μm

5-nm microfibril

Lipid bilayer

Microtubule

Rosette

Cytoplasm

(b)

FIGURE 7.37 Synthesis of plant cell wall macromolecules. (*a*) Freeze-fracture replica of the membrane of an algal cell. The rosettes are thought to represent the cellulose-synthesizing enzyme (cellulose synthase) situated within the plasma membrane. (*b*) A model of cellulose fibril deposition. Each rosette is thought to form a single microfibril that associates laterally with the microfibrils from other rosettes to form a larger fiber. The entire array of rosettes might move laterally within the membrane as it is pushed by the elongating cellulose molecules. Studies suggest that the direction of movement of the membrane rosettes is determined by oriented microtubules present in the cortical cytoplasm beneath the plasma membrane (discussed in Chapter 9). (*c*) Electron micrograph of a Golgi complex of a peripheral root cap cell stained with antibodies against a polymer of galacturonic acid, one of the major components of pectin. This material, like hemicellulose, is assembled in the Golgi complex. The antibodies have been linked to gold particles to make them visible as dark granules. The bar represents 0.25 μm. (A: FROM T. H. GIDDINGS, JR., D. L. BROWER, AND L. A. STAEHELIN, J. CELL BIOL. 84:332, 1980; C: FROM MARGARET LYNCH AND L. A. STAEHELIN, J. CELL BIOL. 118:477, 1992; ALL BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

(c)
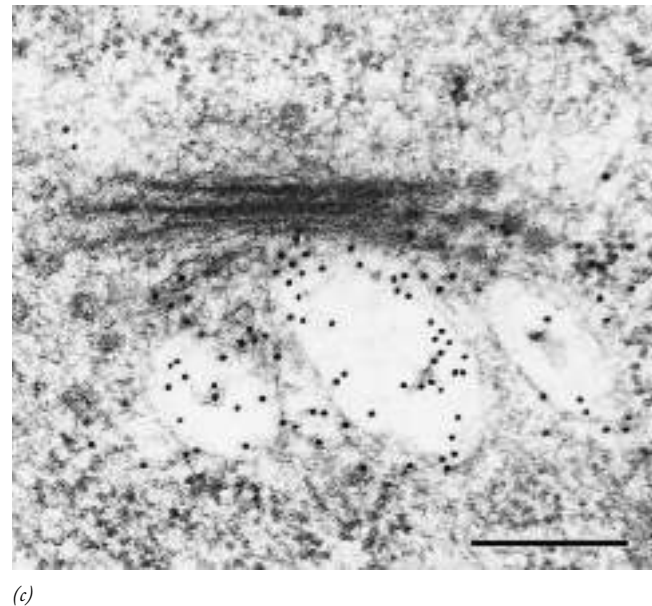
tect the cell against damage from mechanical abrasion and pathogens, and they mediate cell–cell interactions. Like the ECM at the surface of an animal cell, a plant cell wall is a source of signals that alter the activities of the cells that it contacts.

Plant cell walls are often likened to fabricated materials such as reinforced concrete or fiberglass because they contain a fibrous element embedded in a nonfibrous, gel-like matrix. Cellulose, whose structure was described on page 46, provides the fibrous component of the cell wall, and proteins and pectin (described below) provide the matrix. Cellulose molecules are organized into rod-like **microfibrils** (Figure 7.36*b,c*) that confer rigidity on the cell wall and provide resistance to tensile (pulling) forces. Each microfibril is about 5 nm in diameter and is typically composed of bundles of 36 cellulose molecules oriented parallel to one another and held together by hydrogen bonds. The walls of many plant cells are composed of layers in which the microfibrils of one layer are oriented at approximately 90° to those of adjacent layers (Figure 7.36*b*).

Cellulose molecules are polymerized at the cell surface. Glucose subunits are added to the end of a growing cellulose molecule by a multisubunit enzyme called *cellulose synthase*. The subunits of the enzyme are organized into a six-membered ring, or rosette, which is embedded within the plasma membrane (Figure 7.37*a,b*). In contrast, materials of the matrix are synthesized within the cytoplasm (Figure 7.37*c*) and carried to the cell surface in secretory vesicles. The matrix is highly complex, requiring hundreds of enzymes for its synthesis and degradation. The matrix of the cell wall is composed of three types of macromolecules (Figure 7.36*c*):

1. **Hemicelluloses** are branched polysaccharides whose backbone consists of one sugar, such as glucose, and side chains of other sugars, such as xylose. Hemicellulose molecules bind to the surfaces of cellulose microfibrils, cross-linking them into a resilient structural network.

2. **Pectins** are a heterogeneous class of negatively charged polysaccharides containing galacturonic acid. Like the

glycosaminoglycans of animal cell matrices, pectins hold water and thus form an extensive hydrated gel that fills in the spaces between the fibrous elements. When a plant is attacked by pathogens, fragments of pectins released from the wall trigger a defensive response by the plant cell. Purified pectin is used commercially to provide the gel-like consistency of jams and jellies.

3. **Proteins**, whose functions are not well understood, mediate dynamic activities. One class, the *expansins*, facilitate cell growth. These proteins cause localized relaxation of the cell wall, which allows the cell to elongate at that site in response to the turgor pressure generated within the cell. Cell wall-associated protein kinases span the plasma membrane and are thought to transmit signals from the cell wall to the cytoplasm.

The percentages of these various materials in cell walls are highly variable, depending on the type of plant, the type of cell, and the stage of the wall. Like the extracellular matrices of animal connective tissues, plant cell walls are dynamic structures that can be modified in response to changing environmental conditions.

Cell walls arise as a thin **cell plate** (described in Figure 14.38) that forms between the plasma membranes of newly formed daughter cells following cell division. The cell wall matures by the incorporation of additional materials that are assembled inside the cell and secreted into the extracellular space. In addition to providing mechanical support and protection from foreign agents, the cell wall of a young, undifferentiated plant cell must be able to grow in conjunction with the enormous growth of the cell it surrounds. The walls of growing cells are called **primary walls**, and they possess an extensibility that is lacking in the thicker **secondary walls** present around many mature plant cells. The transformation from primary to secondary cell wall occurs as the wall increases in cellulose content and, in most cases, incorporates a phenol-containing polymer called *lignin*. Lignin provides structural support and is the major component of wood. The lignin in the walls of water-conducting cells of the xylem provides the support required to move water through the plant.

REVIEW ?

1. Describe the components that make up a plant cell wall and the role of each in the wall's structure and function.

2. Distinguish between cellulose and hemicellulose, a cellulose molecule and a microfibril, a primary cell wall and a secondary cell wall.

SYNOPSIS

The extracellular space extends outward from the outer surface of the plasma membrane and contains a variety of secreted materials that influence cell behavior. Epithelial tissues rest on a basement membrane, which consists of a thin, interwoven network of extracellular materials. Various types of connective tissue, including tendons, cartilage, and the corneal stroma, contain an expansive extracellular matrix that gives the tissue its characteristic properties. (p. 231)

The major components of extracellular matrices include collagens, proteoglycans, and a variety of proteins, such as fibronectin, and laminin. Each of the proteins of the ECM has a modular construction composed of domains that contain binding sites for one another and for receptors on the cell surface. As a result, these various extracellular materials interact to form an interconnected network that is bound to the cell surface. Collagens are highly abundant, fibrous proteins that give the extracellular matrix the capability to resist pulling forces. Proteoglycans consist of protein and glycosaminoglycans and serve as an amorphous packing material that fills the extracellular space. (p. 232)

Integrins are cell-surface receptors that mediate interactions between cells and their substratum. Integrins are heterodimeric integral membrane proteins whose cytoplasmic domains interact with cytoskeletal components, and whose extracellular domains contain binding sites for various extracellular materials. Internal changes within a cell can send "inside-out" signals that activate the ligand-binding activity of integrins. This activation is associated with a dramatic change in structure of the integrin from a bent to an upright conformation. Conversely, binding of an extracellular ligand to an integrin may send "outside-in" signals that initiate changes in cellular activities. (p. 239)

Cells attach to their substratum by means of cell-surface specializations such as focal adhesions and hemidesmosomes. Focal adhesions are sites of attachment where the plasma membrane contains clusters of integrins linked to actin-containing microfilaments of the cytoskeleton. Hemidesmosomes are sites of attachment where the plasma membrane contains clusters of integrins that are connected to the basement membrane on their outer surface and indirectly to keratin-containing intermediate filaments on their inner surface. Both types of adhesions are also sites of potential cell signaling. (p. 242)

The adhesion of cells to other cells is mediated by several distinct families of integral membrane proteins, including selectins, integrins, cadherins, and members of the immunoglobulin superfamily (IgSF). Selectins bind to specific arrangements of carbohydrate groups projecting from the surfaces of other cells and mediate calcium-dependent, transient interactions between circulating leukocytes and the walls of blood vessels at sites of inflammation and clotting. Cell-adhesion molecules of the Ig superfamily mediate calcium-independent cell–cell adhesion. An IgSF protein on one cell may interact with an integrin or with the same or a different IgSF protein projecting from another cell. Cadherins mediate calcium-dependent cell–cell adhesion by binding to the same cadherin species on the opposing cell, thereby facilitating the formation of tissues composed of similar types of cells. (p. 245)

Strong adhesions between cells are facilitated by the formation of specialized adherens junctions and desmosomes. Adherens junctions encircle a cell near its apical surface, allowing contact between the cell and all of its immediate neighbors. The plasma membranes of adherens junctions contain clusters of cadherins that

are linked by means of intermediary proteins to the actin filaments of the cytoskeleton. Desmosomes occur as patches between cells and are characterized by dense cytoplasmic plaques on the inner surfaces of the membranes. Desmosomes are sites of concentration of cadherins, which are connected by means of intermediary proteins with intermediate filaments that loop through the cytoplasmic plaques. (p. 250)

Tight junctions are specialized sites of contact that block solutes from diffusing between the cells in an epithelium. A cross section through a tight junction shows the outer surfaces of adjacent cells come into direct contact at intermittent sites. Examination of the membranes by freeze fracture shows the sites to contain rows of aligned particles that form strands within the plasma membranes of the adjacent cells. (p. 254)

Gap junctions and plasmodesmata are specialized sites of communication between adjoining cells in animals and plants, respectively. The plasma membranes of a gap junction contain channels formed by a hexagonal array of connexin subunits forming a connexon. These channels connect the cytoplasm of one cell with the cytoplasm of the adjoining cell. The central channel of a connexon allows the direct diffusion of substances up to about 1000 daltons between cells. The passage of ionic currents through gap junctions plays a pivotal role in numerous physiologic processes, including the spread of excitation through the cardiac muscle tissue of the heart. Plasmodesmata are cylindrical cytoplasmic channels that extend between adjacent plant cells directly through the intervening cell wall. These channels normally allow the free passage of solute molecules of approximately 1000 daltons, but they can be dilated by specific proteins to allow passage of macromolecules. (p. 256)

Plant cells are surrounded by a complex cell wall composed of cellulose and a variety of secreted materials that provide mechanical support and protect the cell from potentially damaging influences. Cellulose microfibrils, which are composed of bundles of cellulose molecules that are constructed by enzymes residing within the plasma membrane, provide rigidity and tensile strength. Hemicellulose molecules cross-link the cellulose fibers, whereas pectins form an extensively cross-linked gel that fills the spaces between the fibrous elements of the cell wall. Like the extracellular matrices of animal cells, plant cell walls are dynamic structures that can be modified in response to changing environmental conditions. (p. 260)

ANALYTIC QUESTIONS

1. Cell adhesion can often be blocked in vitro by treating cells with specific agents. Which of the following substances would be expected to interfere with cell adhesion mediated by selectins? with cell adhesion mediated by L1 molecules? The substances are trypsin, which digests proteins; a peptide containing RGD; neuraminidase, which removes sialic acid from an oligosaccharide; collagenase, which digests collagen; hyaluronidase, which digests hyaluronic acid; EGTA, which binds Ca^{2+} ions in the medium.

2. What substance might you add to a culture dish to block the migration of neural crest cells? to block the adhesion of fibroblasts to the substratum?

3. Mice lacking a gene for fibronectin do not survive early embryonic development. Name two processes that might be disrupted in such embryos.

4. Suppose you found that molecule A, which had a molecular mass of 1500 daltons, was able to penetrate the channels of a gap junction, but molecule B, whose molecular mass was only 1200 daltons, was unable to diffuse between the same cells. How might these molecules be different so as to explain these results?

5. In what ways are the extracellular matrices of animals and the cell walls of plants similar in construction?

6. It was noted that two different autoimmune diseases, one producing antibodies against a component of hemidesmosomes and the other producing antibodies against a component of desmosomes, both cause severe blistering of the skin. Why do you think these two conditions have such similar symptoms?

7. Which of the various types of molecules that mediate cell adhesion is most likely to be responsible for the type of sorting out demonstrated by the cells of Figure 7.20? Why? How could you test your conclusion?

8. Follicle-stimulating hormone (FSH) is a pituitary hormone that acts on the follicle cells of the ovary to trigger the synthesis of cyclic AMP, which stimulates various metabolic changes. FSH normally has no effect on cardiac muscle cells. However, when ovarian follicle cells and cardiac muscle cells are grown together in a mixed culture, a number of the cardiac muscle cells are seen to contract following addition of FSH to the medium. How might this observation be explained?

9. Why do you think animal cells are able to survive without the types of cell walls that are found in nearly every other group of organisms?

10. Why would lowering the temperature of the medium in which cells were growing be expected to affect the ability of the cells to form gap junctions with one another?

11. Certain of the cell junctions occur as encircling belts, while others occur as discrete patches. How do these two types of structural arrangements correlate with the functions of the respective junctions?

12. Propose some mechanism that might explain how tobacco mosaic virus was able to alter the permeability of a plasmodesma. How might you be able to test your proposal?

13. One type of vertebrate cell that is thought to lack integrins is the erythrocyte (red blood cell). Does this surprise you? Why or why not?

8

Cytoplasmic Membrane Systems: Structure, Function, and Membrane Trafficking

U nder the light microscope, the cytoplasm of living cells appears relatively devoid of structure. Yet, even before the beginning of the twentieth century, examination of stained sections of animal tissues hinted at the existence of an extensive membrane network within the cytoplasm. It wasn't until the development of the electron microscope in the 1940s, however, that biologists began to appreciate the diverse array of membrane-bound structures present in the cytoplasm of most eukaryotic cells. These early electron microscopists saw membrane-bound vesicles of varying diameter containing material of different electron density; long channels bounded by membranes that radiate through the cytoplasm to form an interconnected network of canals; and stacks of flattened membrane-bound sacs, called *cisternae.*

It became evident from these early electron microscopic studies and the biochemical investigations that followed that the cytoplasm of eukaryotic cells was subdivided into a variety of distinct compartments bounded by membrane barriers. As more types of cells were examined, it became apparent that these membranous compartments in the cytoplasm formed different organelles that could be identified in

Colorized scanning electron micrograph of a human neutrophil, a type of white blood cell, ingesting a number of bacteria by the process of phagocytosis. Neutrophils are essential components of our innate immune response against pathogens. (FROM SCOTT D. KOBAYASHI ET AL., COVER OF PNAS, VOL. 100, #19, 2003, COURTESY OF FRANK R. DELEO.)

diverse cells from yeast to multicellular plants and animals. The extent to which the cytoplasm of a eukaryotic cell is occupied by membranous structures is illustrated in the electron micrograph of a maize root cell shown in Figure 8.1. As we will see in the following pages, each of these organelles contains a particular complement of proteins and is specialized for particular types of activities. Thus, just as a house or restaurant is divided into specialized rooms where different activities can take place independent of one another, the cytoplasm of a cell is divided into specialized membranous compartments for analogous reasons. Keep in mind, as you examine the micrographs in this chapter, that these cytoplasmic organelles may appear as stable structures, like the rooms of a house or restaurant, but in fact they are dynamic compartments that are in continual flux.

In the present chapter, we will examine the structure and functions of the endoplasmic reticulum, Golgi complex, endosomes, lysosomes, and vacuoles. Taken together, these organelles form an **endomembrane system** in which the individual components function as part of a coordinated unit. (Mitochondria and chloroplasts are not part of this interconnected system and were the subjects of Chapters 5 and 6. Current evidence suggests that peroxisomes, which were also discussed in Chapter 6, have a dual origin. The basic elements of the boundary membrane are thought to arise from the endoplasmic reticulum, but many of the membrane proteins and the soluble internal proteins are taken up from the cytosol, as described in Section 8.9.) ■

3 μm
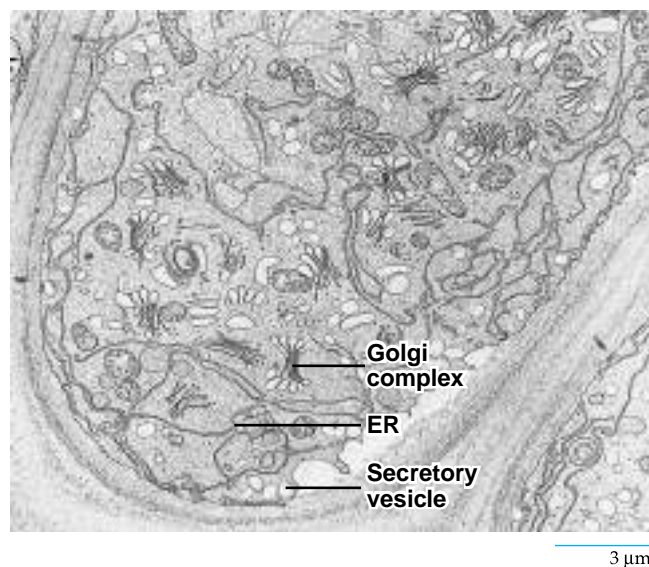
FIGURE 8.1 Membrane-bound compartments of the cytoplasm. The cytoplasm of this root cap cell of a maize plant contains an array of membrane-bound organelles whose structure and function will be examined in this chapter. As is evident in this micrograph, the combined surface area of the cytoplasmic membranes is many times greater than that of the surrounding plasma membrane. (COURTESY OF HILTON H. MOLLENHAUER.)

8.1 AN OVERVIEW OF THE ENDOMEMBRANE SYSTEM

The organelles of the endomembrane system are part of a dynamic, integrated network in which materials are shuttled back and forth from one part of the cell to another. For the most part, materials are shuttled between organelles—from the Golgi complex to the plasma membrane, for example—in small, membrane-bounded **transport vesicles** that bud from a donor membrane compartment (Figure 8.2a).[1] Transport vesicles move through the cytoplasm in a directed manner, often pulled by motor proteins that operate on tracks formed by microtubules and microfilaments of the cytoskeleton (see Figure 9.1a). When they reach their destination, the vesicles fuse with the membrane of the acceptor compartment, which receives the vesicle's soluble cargo as well as its membranous wrapper (Figure 8.2a). Repeated cycles of budding and fusion shuttle a diverse array of materials along numerous pathways that traverse the cell.

Several distinct pathways through the cytoplasm have been identified and are illustrated in the overview shown in Figure 8.2b. A **biosynthetic pathway** can be discerned in which proteins are synthesized in the endoplasmic reticulum, modified during passage through the Golgi complex, and transported from the Golgi complex to various destinations, such as the plasma membrane, a lysosome, or the large vacuole of a plant cell. This route is also referred to as the **secretory pathway**, as many of the proteins synthesized in the endoplasmic reticulum (as well as complex polysaccharides synthesized in the Golgi complex, page 284) are destined to be discharged (**secreted**) from the cell. Secretory activities of cells can be divided into two types: constitutive and regulated (Figure 8.2b). During **constitutive secretion**, materials are transported in secretory vesicles from their sites of synthesis and discharged into the extracellular space in a continual manner. Most cells engage in constitutive secretion, a process that contributes not only to the formation of the extracellular matrix (Section 7.1), but to the formation of the plasma membrane itself. During **regulated secretion**, materials are stored as membrane-bound packages and discharged only in response to an appropriate stimulus. Regulated secretion occurs, for example, in endocrine cells that release hormones, in pancreatic acinar cells that release digestive enzymes, and in nerve cells that release neurotransmitters. In some of these cells, materials to be secreted are stored in large, densely packed, membrane-bound **secretory granules** (see Figure 8.3).

Proteins, lipids, and complex polysaccharides are transported through the cell along the biosynthetic or secretory pathway. We will focus in the first part of the chapter on the synthesis and transport of proteins, as summarized in Figure 8.2b. During the discussion, we will consider several distinct classes of proteins. These include soluble proteins that

---

[1]The term *vesicle* implies a spherical-shaped carrier. Cargo may also be transported in irregular or tubular-shaped membrane-bound carriers. For the sake of simplicity, we will generally refer to carriers as "vesicles," while keeping in mind they are not always spherical.
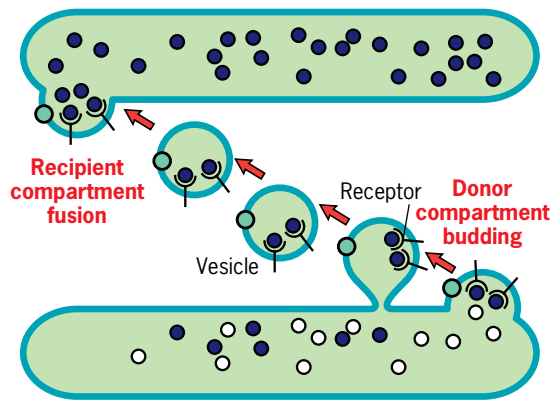
FIGURE 8.2 **An overview of the biosynthetic/secretory and endocytic pathways that unite endomembranes into a dynamic, interconnected network.** (*a*) Schematic diagram illustrating the process of vesicle transport by which materials are transported from a donor compartment to a recipient compartment. Vesicles form by membrane budding, during which membrane proteins of the donor membrane are incorporated into the vesicle membrane and soluble proteins in the donor compartment are bound to specific receptors. When the transport vesicle subsequently fuses to another membrane, the proteins of the vesicle membrane become part of the recipient membrane, and the soluble proteins become sequestered within the lumen of the recipient compartment. (*b*) Materials follow the biosynthetic (or secretory) pathway from the endoplasmic reticulum, through the Golgi complex, and out to various locations including lysosomes, endosomes, secretory vesicles, secretory granules, vacuoles, and the plasma membrane. Materials follow the endocytic pathway from the cell surface to the interior by way of endosomes and lysosomes, where they are generally degraded by lysosomal enzymes.
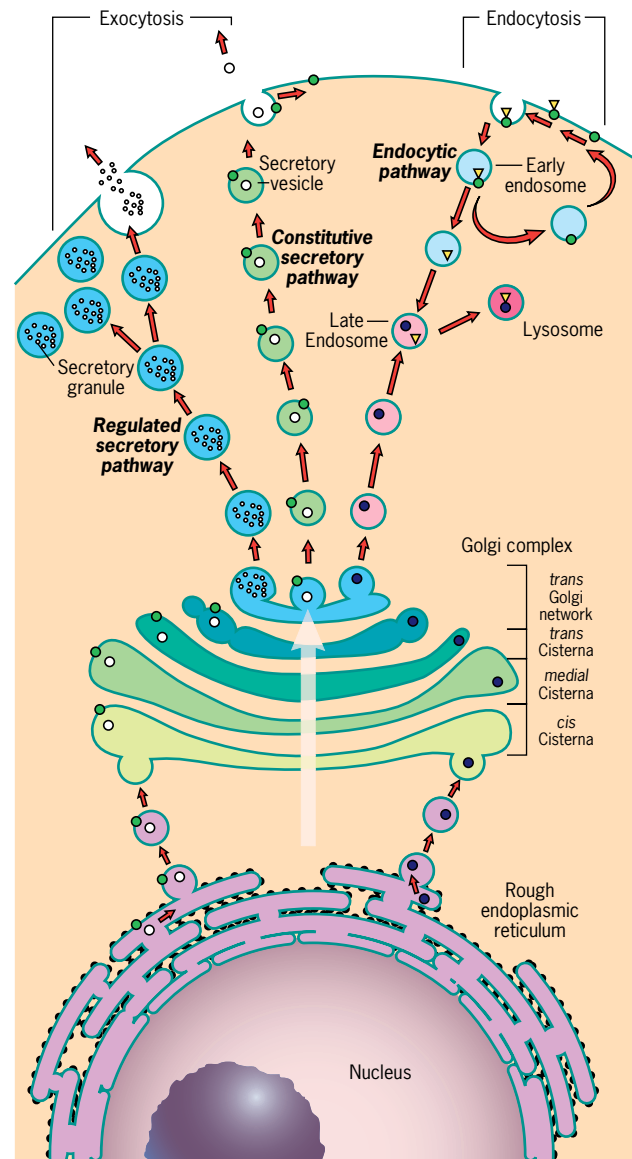
are discharged from the cell, integral proteins of the various membranes depicted in Figure 8.2*b*, and soluble proteins that reside within the various compartments enclosed by the endomembranes (e.g., lysosomal enzymes). Whereas materials move out of the cell by the secretory pathway, the endocytic pathway operates in the opposite direction. By following the **endocytic pathway**, materials move from the outer surface of the cell to compartments, such as endosomes and lysosomes, located within the cytoplasm (Figure 8.2*b*).

The movement of vesicles and their contents along the various pathways of a cell is analogous to the movement of trucks carrying different types of cargo along the various highways of a city. Both types of transport require defined *traffic patterns* to ensure that materials are accurately delivered to the appropriate sites. For example, protein trafficking within a salivary gland cell requires that the proteins of salivary mucus, which are synthesized in the endoplasmic reticulum, are specifically *targeted* for secretory granules, while lysosomal enzymes, which are also manufactured in the endoplasmic reticulum, are specifically targeted to a lysosome. Different organelles also contain different integral membrane proteins. Consequently, membrane proteins must also be targeted to particular organelles, such as a lysosome or Golgi cisterna. These various types of cargo—secreted proteins,

lysosomal enzymes, and membrane proteins—are routed to their appropriate cellular destinations by virtue of specific "addresses" or *sorting signals* that are encoded in the amino acid sequence of the proteins or in the attached oligosaccharides. The sorting signals are recognized by specific receptors that reside in the membranes or surface coats of budding vesicles, ensuring that the protein is transported to the appropriate destination. For the most part, the machinery responsible for driving this complex distribution system consists of soluble proteins that are *recruited* to specific membrane surfaces. During the course of this chapter we will try to understand why one protein is recruited, for example, to the endoplasmic reticulum whereas another protein might be recruited to a particular region of the Golgi complex.

Great advances have been made over the past three decades in mapping the traffic patterns that exist in eukaryotic cells, identifying the specific addresses and receptors that gov-

ern the flow of traffic, and dissecting the machinery that ensures that materials are delivered to the appropriate sites in the cell. These subjects will be discussed in detail in the following pages. Motor proteins and cytoskeletal elements, which play key roles in the movements of transport vesicles and other endomembranes, will be described in the following chapter. We will begin the study of endomembranes by discussing a few of the most important experimental approaches that have led to our current understanding of the subject.

8.2 A FEW APPROACHES TO THE STUDY OF ENDOMEMBRANES

Early studies with the electron microscope provided biologists with a detailed portrait of the structure of cells but gave them little insight into the functions of the components they were observing. Determining the functions of cytoplasmic organelles required the development of new techniques and the execution of innovative experiments. The experimental approaches described in the following sections have proven particularly useful in providing the foundation of knowledge on which current research on cytoplasmic organelles is based.

Insights Gained from Autoradiography

Among the hundreds of different cells in the body, the acinar cells of the pancreas have a particularly extensive endomembrane system. These cells function primarily in the synthesis and secretion of digestive enzymes. After secretion from the pancreas, these enzymes are shipped through ducts to the small intestine, where they degrade ingested food matter. Where within the pancreatic acinar cells are the secretory proteins synthesized, and how do they reach the surface of the cells where they are discharged? These questions are inherently difficult to answer because all of the steps in the process of secretion occur simultaneously within the cell. To follow the steps of a single cycle from start to finish, that is, from the synthesis of a secretory protein to its discharge from the cell, James Jamieson and George Palade of Rockefeller University utilized the technique of **autoradiography**.

As discussed in Chapter 18, autoradiography provides a means to visualize biochemical processes by allowing an investigator to determine the location of radioactively labeled materials within a cell. In this technique, tissue sections containing radioactive isotopes are covered with a thin layer of photographic emulsion, which is exposed by radiation emanating from radioisotopes within the tissue. Sites in the cells

containing radioactivity are revealed under the microscope by silver grains in the overlying emulsion (Figure 8.3).

To determine the sites where secretory proteins are synthesized, Palade and Jamieson incubated slices of pancreatic tissue in a solution containing radioactive amino acids for a brief period of time. During this period, labeled amino acids were taken up by the living cells and incorporated into the digestive enzymes as they were being synthesized on ribosomes. The tissues were quickly fixed, and the locations of proteins that had been synthesized during the brief incubation with labeled amino acids were determined autoradiographically. Using this approach, the endoplasmic reticulum was discovered to be the site of synthesis of secretory proteins (Figure 8.3a).

To determine the intracellular path followed by secretory proteins from their site of synthesis to their site of discharge, Palade and Jamieson carried out an additional experiment. After incubating the tissue for a brief period in radioactive amino acids, they washed the tissue free of excess isotope and transferred the tissue to a medium containing only unlabeled amino acids. An experiment of this type is called a *"pulse-chase."* The *pulse* refers to the brief incubation with radioactivity during which labeled amino acids are incorporated into protein. The *chase* refers to the period when the tissue is exposed to the unlabeled medium, a period during which additional proteins are synthesized using nonradioactive amino acids. The longer the chase, the farther the radioactive proteins manufactured during the pulse will have traveled from their site of synthesis within the cell. Using this approach, one can ideally follow the movements of newly synthesized molecules by observing a wave of radioactive material moving through the cytoplasmic organelles of cells from one location to the next until the process is complete. The results of these experiments—which first defined the biosynthetic (or secretory) pathway and tied a number of seemingly separate membranous compartments into an integrated functional unit—are summarized in Figure 8.3b–d.

Insights Gained from the Use of the Green Fluorescent Protein

The autoradiographic experiments described in the previous section require investigators to examine thin sections of different cells that have been fixed at various times after introduction of a radioactive label. An alternative technology allows researchers to follow the dynamic movements of specific proteins with their own eyes as they occur within a single living cell. This technology utilizes a gene isolated from a jellyfish that encodes a small protein, called the **green fluorescent protein (GFP)**, which emits a green fluorescent light. In this approach, DNA encoding GFP is fused to DNA encoding the protein to be studied and the resulting chimeric (i.e., composite) DNA is introduced into cells that can be observed under the microscope. Once inside a cell, the chimeric DNA expresses a chimeric protein consisting of GFP fused to the end of the protein to be studied. In most cases, the presence of GFP joined to the end of a protein has little or no effect on the movement or function of that protein.

Silver Grains

Rough ER

Secretory Granule
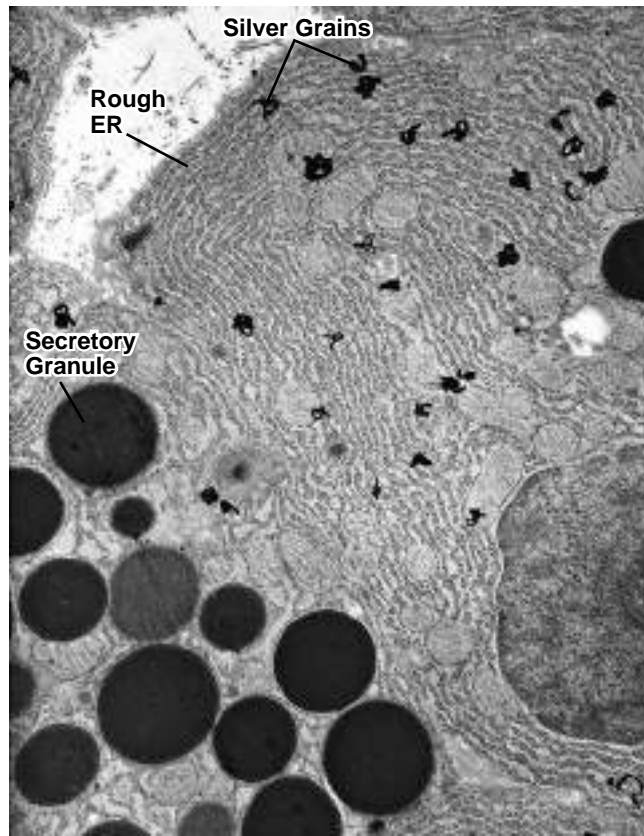
(a)

FIGURE 8.3 Autoradiography reveals the sites of synthesis and subsequent transport of secretory proteins. (*a*) Electron micrograph of a section of a pancreatic acinar cell that had been incubated for 3 minutes in radioactive amino acids and then immediately fixed and prepared for autoradiography (see Section 18.4 for discussion of the technique). The black silver grains that appear in the emulsion following development are localized over the endoplasmic reticulum. (*b–d*) Diagrams of a sequence of autoradiographs showing the movement of labeled secretory proteins (represented by the silver grains in red) through a pancreatic acinar cell. When the cell is pulse-labeled for 3 minutes and immediately fixed (as shown in *a*), radioactivity is localized in the endoplasmic reticulum (*b*). After a 3-minute pulse and 17-minute chase, radioactive label is concentrated in the Golgi complex and adjacent vesicles (*c*). After a 3-minute pulse and 117-minute chase, radioactivity is concentrated in the secretory granules and is beginning to be released into the pancreatic ducts (*d*). (A: Courtesy of James D. Jamieson and George Palade.)

Lumen

Granules

Golgi complex

Endoplasmic reticulum

Nucleus

Mitochondria

3 min

20 min

120 min
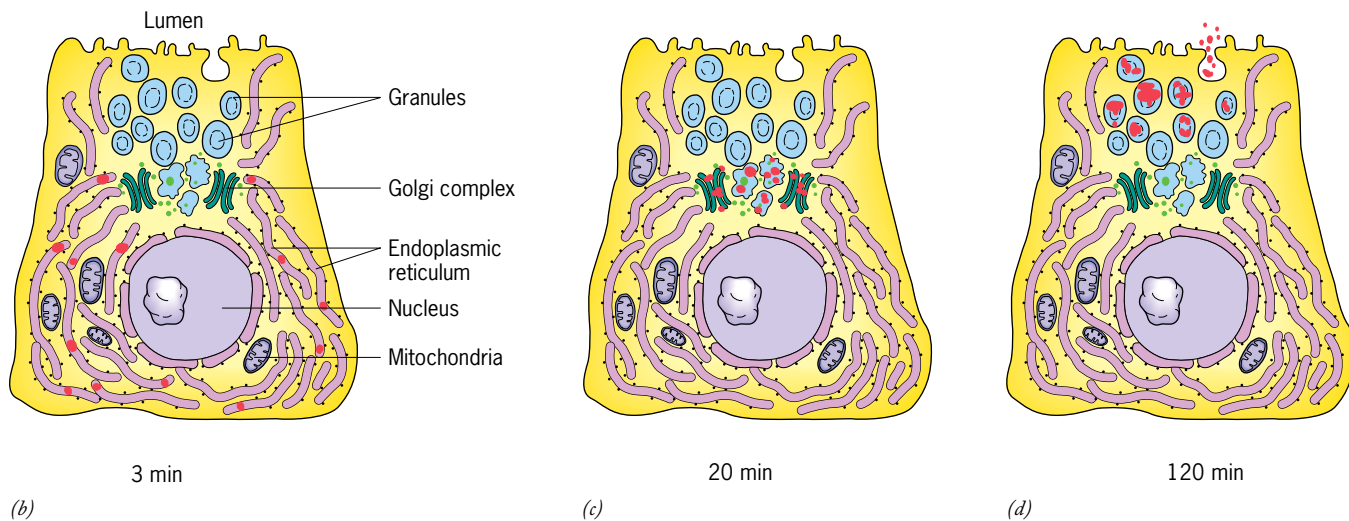
(b) *(c)* *(d)*

Figure 8.4 shows a pair of micrographs depicting cells that contain a GFP fusion protein. In this case, the cells were infected with a strain of the vesicular stomatitis virus (VSV) in which one of the viral genes (*VSVG*) is fused to the *GFP* gene. Viruses are useful in these types of studies because they turn infected cells into factories for the production of viral proteins, which are carried like any other protein cargo through the biosynthetic pathway. When a cell is infected with VSV, massive amounts of the VSVG protein are produced in the endoplasmic reticulum (ER). The VSVG molecules then traffic through the Golgi complex and are transported to the plasma membrane of the infected cell where they are incorporated into viral envelopes. As in a radioactive pulse-chase experiment, the use of a virus allows investigators to follow a relatively synchronous wave of protein movement, in this case represented by a wave of green fluorescence that begins soon after infection. Synchrony can be enhanced, as was done in the experiment depicted in Figure 8.4, by use of a virus with a mutant VSVG protein that is unable to leave the ER of infected cells that are grown at an elevated temperature (e.g., 40°C). When the temperature is lowered to 32°C, the fluorescent GFP-VSVG protein that had accumulated in the
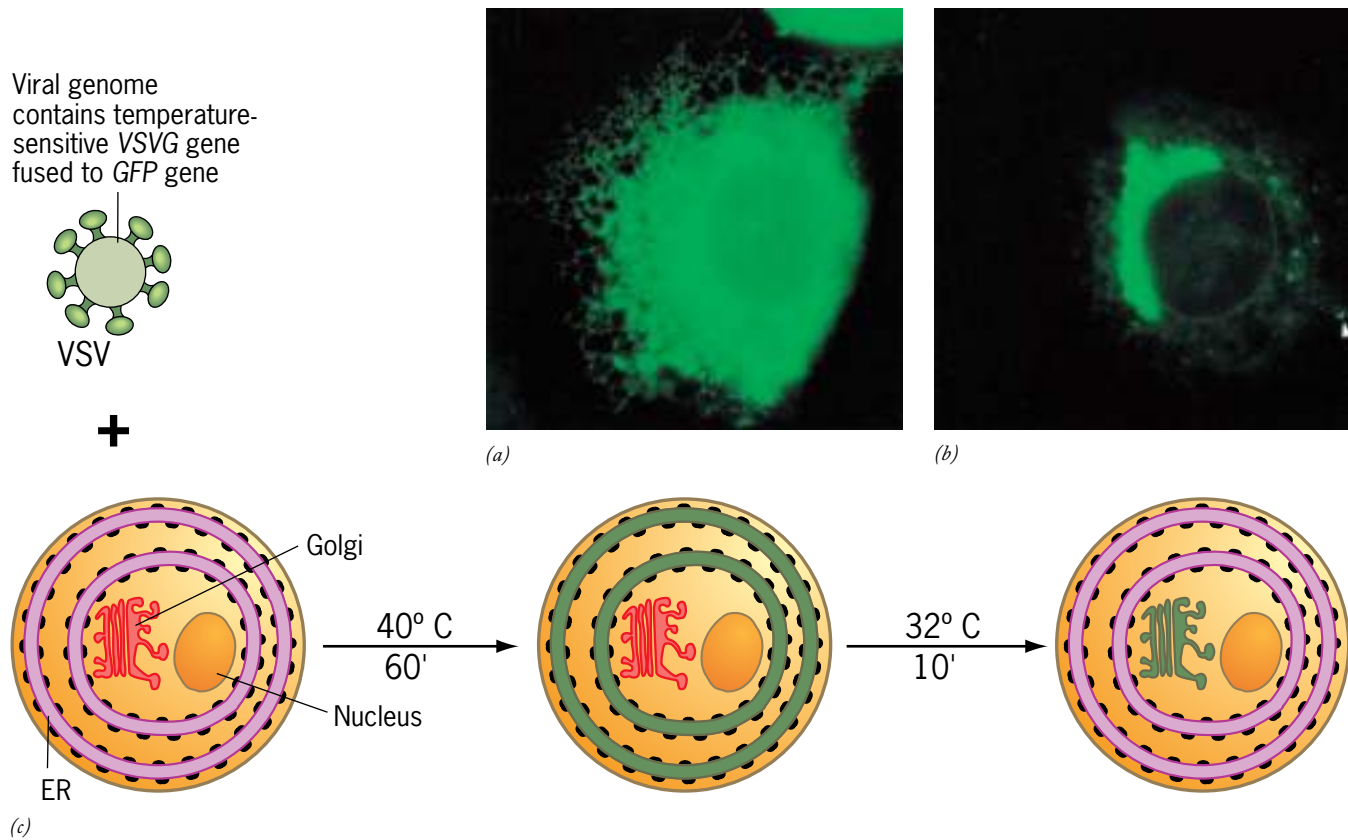
FIGURE 8.4 The use of green fluorescent protein (GFP) reveals the movement of proteins within a living cell. (*a*) Fluorescence micrograph of a live cultured mammalian cell that had been infected with the VSV virus at 40°C. This particular strain of the VSV virus contained a *VSVG* gene that (1) was fused to a gene for the fluorescent protein GFP and (2) contained a temperature-sensitive mutation that prevented the newly synthesized VSVG protein from leaving the ER when kept at 40°C. The green fluorescence in this micrograph is restricted to the ER. (*b*) Fluorescence micrograph of a live infected cell that was held at 40°C to allow the VSVG protein to accumulate in the ER and then incubated at 32°C for 10 minutes. The fluorescent VSVG protein has moved on to the region containing the Golgi complex. (*c*) Schematic drawing showing the retention of the mutant VSVG protein in the ER at 40°C and its synchronous movement to the Golgi complex within 10 minutes of incubation at the lower temperature. (A,B: FROM DANIEL S. CHAO ET AL., COURTESY OF RICHARD H. SCHELLER, J. CELL BIOL. 144:873, 1999; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

ER (Figure 8.4*a,c*) moves synchronously to the Golgi complex (Figure 8.4*b,c*), where various processing events occur, and then on to the plasma membrane. Mutants of this type that function normally at a reduced (permissive) temperature, but not at an elevated (restrictive) temperature are described as *temperature-sensitive mutants*. An experiment utilizing two different fluorescent probes is described on page 314.

Insights Gained from the Biochemical Analysis of Subcellular Fractions

Electron microscopy, autoradiography, and the use of GFP provide information on the structure and function of cellular organelles but fail to provide much insight into the molecular composition of these structures. Techniques to break up (**homogenize**) cells and isolate particular types of organelles were pioneered in the 1950s and 1960s by Albert Claude and Christian De Duve. When a cell is ruptured by homogenization, the cytoplasmic membranes become frag-

mented and the fractured edges of the membrane fragments fuse to form spherical vesicles less than 100 nm in diameter. Vesicles derived from different organelles (nucleus, mitochondrion, plasma membrane, endoplasmic reticulum, and so forth) have different properties, which allow them to be separated from one another, an approach that is called **subcellular fractionation**.

Membranous vesicles derived from the endomembrane system (primarily the endoplasmic reticulum and Golgi complex) form a heterogeneous collection of similar sized vesicles referred to as **microsomes**. A rapid (and crude) preparation of the microsomal fraction of a cell is depicted in Figure 8.5*a*. The microsomal fraction can be further fractionated into smooth and rough membrane fractions (Figure 8.5*b,c*) by the gradient techniques discussed in Section 18.6. Once isolated, the biochemical composition of various fractions can be determined. It was found, for example, using this approach, that vesicles derived from different parts of the Golgi complex contained enzymes that added different sugars to the end of a

FIGURE 8.5 **Isolation of a microsomal fraction by differential centrifugation.** (*a*) When a cell is broken by mechanical homogenization (step 1), the various membranous organelles become fragmented and form spherical membranous vesicles. Vesicles derived from different organelles can be separated by various techniques of centrifugation. In the procedure depicted here, the cell homogenate is first subjected to low-speed centrifugation to pellet the larger particles, such as nuclei and mitochondria, leaving the smaller vesicles (microsomes) in the supernatant (step 2). The microsomes can be removed from the supernatant by centrifugation at higher speeds for longer periods of time (step 3). A crude microsomal fraction of this type can be fractionated into different vesicle types in subsequent steps. (*b*) Electron micrograph of a smooth microsomal fraction in which the membranous vesicles lack ribosomes. (*c*) Electron micrograph of a rough microsomal fraction containing ribosome-studded membranes. (B,C: COURTESY OF J. A. HIGGINS AND R. J. BARNETT.)

growing carbohydrate chain of a glycoprotein or glycolipid. A specific enzyme could be isolated from the microsomal fraction and then used as an antigen to prepare antibodies against that enzyme. The antibodies could then be linked to materials, such as gold particles, that could be visualized in the electron microscope, and the localization of the enzyme in the membrane compartment could be determined. These studies detailed the role of the Golgi complex in the stepwise assembly of complex carbohydrates.

In the past few years, the identification of the proteins present in cell fractions has been taken to a new level using sophisticated proteomic technology. Once a particular organelle has been isolated, the proteins can be extracted, separated, and identified by mass spectrometry as discussed on page 69. Hundreds of proteins can be identified simultaneously, providing a comprehensive molecular portrait of any organelle that can be prepared in a relatively pure state. In one example of this technology, it was found that a simple phagosome (page 308) containing an ingested latex bead comprised more than 160 different proteins, many of which had never previously been identified or were not known to be involved in phagocytosis.

Insights Gained from the Use of Cell-Free Systems

Once techniques to fractionate membranous organelles were developed, researchers began to probe the capabilities of these crude subcellular preparations. They found that the isolated parts of a cell were capable of remarkable activities. These early **cell-free systems**—so called because they did not contain whole cells—provided a wealth of information about complex processes that were impossible to study using intact cells. During the 1960s, for example, George Palade, Philip Siekevitz, and their colleagues at Rockefeller University set out to learn more about the properties of the rough microsomal fraction (shown in Figure 8.5c), which is derived from the rough ER (page 273). They found that they could strip a rough microsomal preparation of its attached particles, and the isolated particles (i.e., ribosomes) were capable of synthe-

(*a*)

(*b*)   (*c*)   0.3 μm
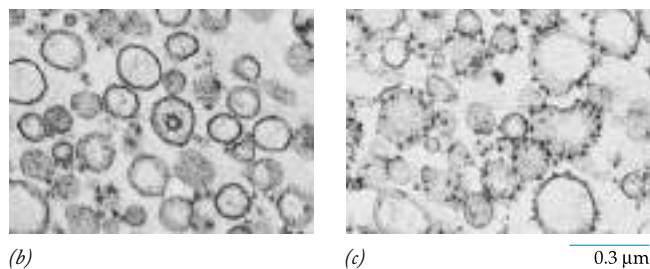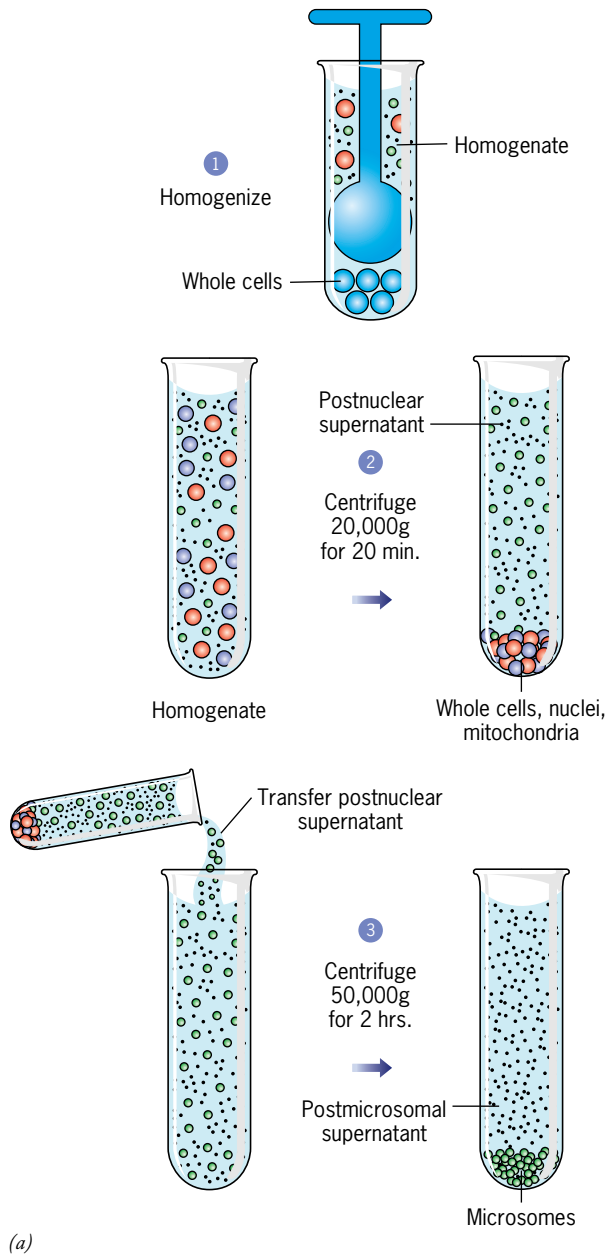
sizing proteins when provided with required ingredients from the cytosol. Under these conditions, the newly synthesized proteins were simply released by the ribosomes into the aqueous fluid of the test tube. When the same experiment was carried out using intact rough microsomes, the newly synthesized proteins were no longer released into the incubation medium
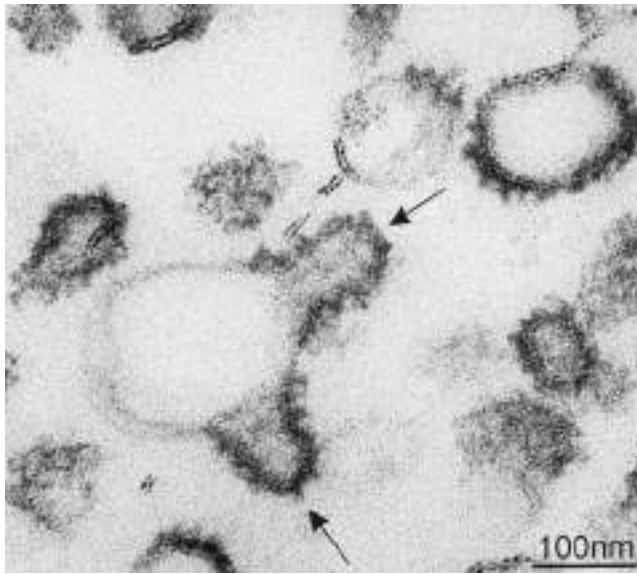
FIGURE 8.6 Formation of coated vesicles in a cell-free system. Electron micrograph of a liposome preparation that had been incubated with the components required to promote vesicle budding within the cell. The proteins in the medium have become attached to the surface of the liposomes and have induced the formation of protein-coated buds (arrows). (COURTESY OF LELIO ORCI AND RANDY SCHEKMAN.)

but were trapped within the lumen of the membranous vesicles. It was concluded from these studies that the microsomal membrane was not required for the incorporation of amino acids into proteins but for sequestering newly synthesized secretory proteins within the ER cisternal space.

Over the past few decades, researchers have used cell-free systems to identify the roles of many of the proteins involved in membrane trafficking. Figure 8.6 shows a liposome with vesicles budding from its surface (arrows). As discussed on page 124, liposomes are vesicles whose surface consists of an artificial bilayer that is created in the laboratory from purified phospholipids. The buds and vesicles seen in Figure 8.6 were produced after the preparation of liposomes was incubated with purified proteins that normally comprise coats on the cytosolic surface of transport vesicles within the cell. Without the added coat proteins, vesicle budding could not occur. Using this strategy in which cellular processes are *reconstituted* in vitro from purified components, researchers have been able to study the proteins that bind to the membrane to initiate vesicle formation, the proteins responsible for cargo selection, and the proteins that sever the vesicle from the donor membrane.

Insights Gained from the Study of Mutant Phenotypes

A mutant is an organism (or cultured cell) whose chromosomes contain one or more genes that encode abnormal proteins. When a protein encoded by a mutant gene is unable to carry out its normal function, the cell carrying the mutation exhibits a characteristic deficiency. Determining the precise nature of

the deficiency provides information on the function of the normal protein. Study of the genetic basis of secretion has been carried out largely on yeast cells, most notably by Randy Schekman and colleagues at the University of California, Berkeley. Yeasts are particularly amenable to genetic studies because they have a small number of genes compared to other types of eukaryotes; they are small, single-celled organisms easily grown in culture; and they can be grown as haploids for the majority of their life cycle. A mutation in a single gene in a haploid cell produces an observable effect because the cells lack a second copy of the gene that would mask the presence of the abnormal one.

In yeast, as in all eukaryotic cells, vesicles bud from the ER and travel to the Golgi complex, where they fuse with the Golgi cisternae (Figure 8.7a). To identify genes whose encoded protein is involved in this portion of the secretory pathway (i.e., *SEC* genes), researchers screen for mutant cells that exhibit an abnormal distribution of cytoplasmic membranes. An electron micrograph of a wild-type yeast cell is shown in Figure 8.7b. The cell depicted in Figure 8.7c has a mutation in a gene that encodes a protein involved in the formation of vesicles at the ER membrane (step 1, Figure 8.7a). In the absence of vesicle formation, mutant cells accumulate an expanded endoplasmic reticulum. In contrast, the cell depicted in Figure 8.7d carries a mutation in a gene that encodes a protein involved in vesicle fusion (step 2, Figure 8.7a). When this gene product is defective, the mutant cells accumulate an excess number of unfused vesicles. Researchers have isolated dozens of different mutants which, taken as a group, exhibit disruptions in virtually every step of the secretory pathway. The genes responsible for these defects have been cloned and sequenced and the proteins they encode have been isolated. Isolation of proteins from yeast has launched successful searches for homologous proteins (i.e., proteins with related sequence) in mammals.

One of the most important lessons learned from the use of all of these techniques is that the dynamic activities of the endomembrane system are highly conserved. Not only do yeast, plant, insect, and human cells carry out similar processes, but they do so with remarkably similar proteins. It is evident that the structural diversity of cells belies their underlying molecular similarities. In many cases, proteins from widely divergent species are interchangeable. For example, cell-free systems derived from mammalian cells can often utilize yeast proteins to facilitate vesicle transport. Conversely, yeast cells with genetic deficiencies that disrupt some phase of the biosynthetic pathway can often be "cured" by genetically engineering them to carry mammalian genes.

Over the past decade or so, researchers interested in searching for genes that affect a particular cellular process in plant or animal cells have taken advantage of a cellular phenomenon called *RNA interference* (*RNAi*). RNAi is a process in which cells produce small RNAs (called *siRNAs*) that bind to specific mRNAs and inhibit the translation of these mRNAs into proteins. This phenomenon and its use are discussed in detail in Sections 11.5 and 18.18. For the present purpose we will simply note that researchers can synthesize a

(a)

(b)   (c)   (d)

collection (library) of siRNAs that are capable of inhibiting the translation of virtually any mRNA that is produced by a genome. Each mRNA represents the expression of a specific gene; therefore, one can find which genes are involved in a particular process by determining which siRNAs interfere with that process. In the experiment depicted in Figure 8.8, researchers set out to identify genes that were involved in various steps of the secretory pathway, a goal similar to that of the investigators studying the yeast mutants shown in Figure 8.7. In this case, researchers used a strain of cultured *Drosophila* cells and attempted to identify genes that affected the localization of mannosidase II, an enzyme that is synthesized in the endoplasmic reticulum and moves via transport vesicles to the Golgi complex, where it takes up residence. Figure 8.8*a* shows a control cell that is synthesizing a GFP-labeled version of mannosidase II; the fluorescence becomes localized in the numerous Golgi complexes of the cell as would be expected. Figure 8.8*b* shows a cell that contains siRNA molecules that have caused a relocation of the GFP-mannosidase into the ER, which is seen to be fused with the Golgi complexes. This type of phenotype is typically caused by the absence of one of the proteins involved in the transport of the enzyme from the ER to the Golgi complex. Of the 130 different siRNAs that

were found to interfere in some way with the secretory pathway in this study, 31 of them generated a phenotype similar to that shown in Figure 8.8*b*. Included among these 31 siRNAs were numerous species that inhibited the expression of genes that were already known to be involved in the secretory pathway. In addition, the study identified other genes whose function had been unknown and are now presumed to be involved in these processes as well. Because it is easier to synthesize a small RNA than to generate an organism with a mutant gene, RNAi has become a common strategy to investigate the effect of a missing protein.

REVIEW ?

1. Describe the differences between an autoradiograph of a pancreas cell that had been incubated in labeled amino acids for 3 minutes and immediately fixed and one of a cell that had been labeled for 3 minutes, chased for 40 minutes, and then fixed.

2. What techniques or approaches might you use to learn which proteins are normally present in the endoplasmic reticulum?

(a) *(b)*
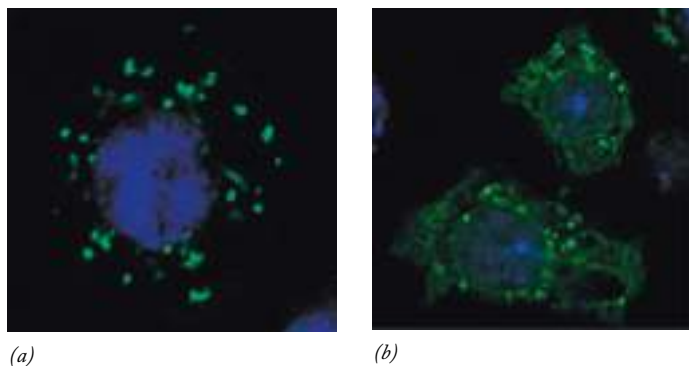
FIGURE 8.8 **Inhibition of gene expression with RNA interference.** (*a*) A control *Drosophila* S2 cultured cell expressing GFP-labeled mannosidase II. The fluorescent enzyme becomes localized in the Golgi complex after its synthesis in the ER. (*b*) A cell that has been genetically engineered to express a specific siRNA, which binds to a complementary mRNA and inhibits translation of the encoded protein. In this case, the siRNA has caused the fluorescent enzyme to remain in the ER, which has fused with the Golgi membranes. This phenotype suggests that the mRNA being affected by the siRNA encodes a protein involved in an early step of the secretory pathway during which the enzyme is synthesized in the ER and traffics to the Golgi complex. Among the genes that exhibit this phenotype when targeted by siRNAs are those encoding proteins of the COPI coat, Sar1, and Sec23. The functions of these proteins are discussed later in the chapter. (REPRINTED WITH PERMISSION FROM FREDERIC BARD ET AL., COURTESY OF VIVEK MALHOTRA, NATURE 439:604, 2006; © COPYRIGHT 2006, MACMILLAN MAGAZINES LTD.)

---

3. How does the isolation of a mutant yeast that accumulates vesicles provide information on the process of protein trafficking?
4. How can GFP be used to study membrane dynamics?

8.3 THE ENDOPLASMIC RETICULUM

The **endoplasmic reticulum (ER)** is divided into two subcompartments, the **rough endoplasmic reticulum (RER)** and the **smooth endoplasmic reticulum (SER)**. Both types of ER comprise a system of membranes that enclose a space, or lumen, that is separated from the surrounding cytosol. As will be evident in the following discussion, the composition of the **luminal** (or **cisternal**) **space** inside the ER membranes is quite different from that of the surrounding **cytosolic space**.

Fluorescently labeled proteins and lipids are capable of diffusing from one type of ER into the other, indicating that their membranes are continuous. In fact, the two types of ER share many of the same proteins and engage in certain common activities, such as the synthesis of certain lipids and cholesterol. At the same time, however, numerous proteins are found only in one or the other type of ER. As a result, the SER and RER have important structural and functional differences.

The rough ER is defined by the presence of ribosomes bound to its cytosolic surface, whereas the smooth ER lacks associated ribosomes. The RER is typically composed of a network of flattened sacs (**cisternae**), as shown in Figure 8.9. The RER is continuous with the outer membrane of the nuclear envelope, which also bears ribosomes on its cytosolic surface (Figure 8.2*b*). In contrast, the membranous elements of the SER are highly curved and tubular, forming an interconnecting system of pipelines curving through the cytoplasm. When cells are homogenized, the SER fragments into smooth-surfaced vesicles, whereas the RER fragments into rough-surfaced vesicles (Figure 8.5*b,c*).

Different types of cells contain markedly different ratios of the two types of ER, depending on the activities of the cell. For example, cells that secrete large amounts of proteins, such as the cells of the pancreas or salivary glands, have extensive regions of RER (Figure 8.9*b,c*). We will return to the function of the RER shortly, but first we will describe the activities of the SER.

The Smooth Endoplasmic Reticulum

The SER is extensively developed in a number of cell types, including those of skeletal muscle, kidney tubules, and steroid-producing endocrine glands (Figure 8.10). SER functions include:

- Synthesis of steroid hormones in the endocrine cells of the gonad and adrenal cortex.
- Detoxification in the liver of a wide variety of organic compounds, including barbiturates and ethanol, whose chronic use can lead to proliferation of the SER in liver cells. Detoxification is carried out by a system of oxygen-transferring enzymes (oxygenases), including the *cytochrome P450* family. These enzymes are noteworthy for their lack of substrate specificity, being able to oxidize thousands of different hydrophobic compounds and convert them into more hydrophilic, more readily excreted derivatives. The effects are not always positive. For example, the relatively harmless compound benzo[*a*]pyrene formed when meat is charred on a grill is converted into a potent carcinogen by the "detoxifying" enzymes of the SER. Cytochrome P450s metabolize many prescribed medications, and genetic variation in these enzymes among humans may explain differences from one person to the next in the effectiveness and side effects of many drugs.
- Sequestering calcium ions within the cytoplasm of cells. The regulated release of Ca^{2+} from the SER of skeletal and cardiac muscle cells (known as the *sarcoplasmic reticulum* in muscle cells) triggers contraction.

Functions of the Rough Endoplasmic Reticulum

Early investigations on the functions of the RER were carried out on cells that secrete large quantities of proteins, such as the acinar cells of the pancreas (Figure 8.3) or the mucus-secreting cells of the lining of the digestive tract (Figure 8.11). It is evident from the drawing and micrograph of Figure 8.11 that the organelles of these epithelial secretory cells are posi-

(a)

(b)

(c)   0.3 μm

(d)   10 μm
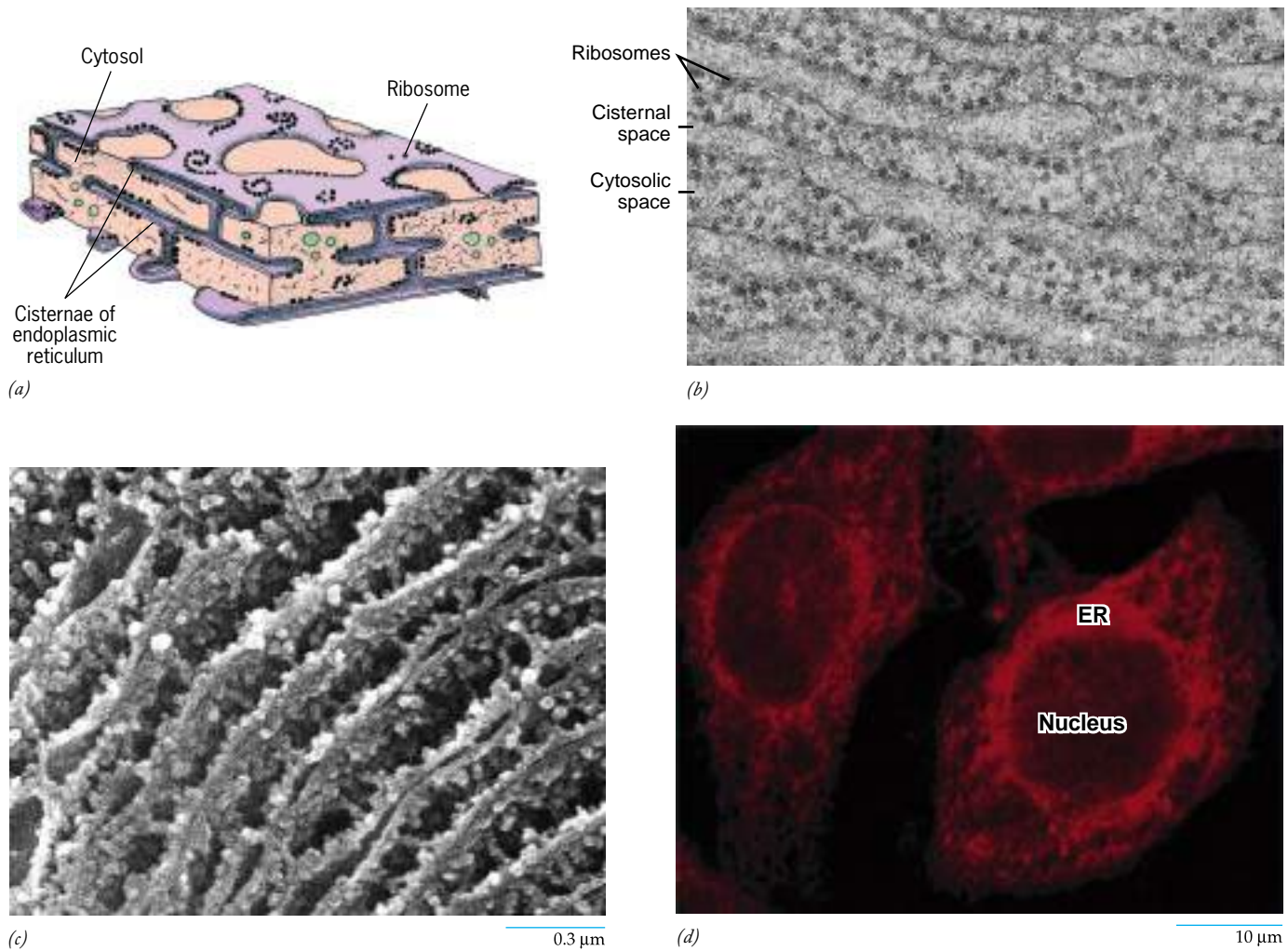
FIGURE 8.9 The rough endoplasmic reticulum (RER). (*a*) Schematic diagram showing the stacks of flattened cisternae that make up the rough ER. The cytosolic surface of the membrane contains bound ribosomes, which gives the cisternae their rough appearance. (*b*) Transmission electron micrograph of a portion of the rough ER of a pancreatic acinar cell. The division of the rough ER into a cisternal space (which is devoid of ribosomes) and a cytosolic space is evident. (*c*) Scanning electron micrograph of the rough ER in a pancreatic acinar cell. (*d*) Visualization of the rough ER in a whole cultured cell as revealed by immunofluorescent staining for the enzyme protein disulfide isomerase (PDI), an ER resident protein. (B: COURTESY OF S. ITO; C: FROM K. TANAKA, INT. REV. CYTOL. 68:101, 1980; D: FROM BRIAN STORRIE, RAINER PEPPERKOK, AND TOMMY NILSSON, TRENDS CELL BIOL. 10:388, 2000; COPYRIGHT 2000, WITH PERMISSION FROM ELSEVIER SCIENCE.)
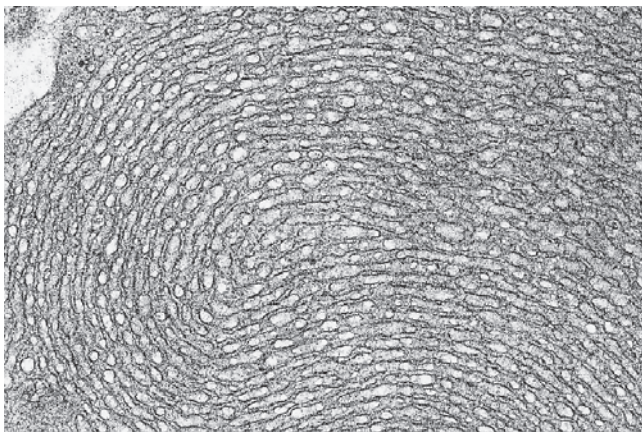
FIGURE 8.10 The smooth ER (SER). Electron micrograph of a Leydig cell from the testis showing the extensive smooth ER where steroid hormones are synthesized. (FROM DON FAWCETT/VISUALS UNLIMITED.)

tioned in the cell in such a way as to produce a distinct polarity from one end to the other. The nucleus and an extensive array of RER cisternae are located near the basal surface of the cell, which faces the blood supply. The Golgi complex is located in the central region of the cell. The apical surface of the cell faces a duct that carries the secreted proteins out of the organ. The cytoplasm at the apical end of the cell is filled with secretory granules whose contents are ready to be released into the duct upon arrival of the appropriate signal. The polarity of these glandular epithelial cells reflects the movement of secretory proteins through the cell from their site of synthesis to their site of discharge. The rough ER is the starting point of the biosynthetic pathway: it is the site of synthesis of the proteins, carbohydrate chains, and phospholipids that journey through the membranous compartments of the cell.

(a)

(b)

1.5 μm
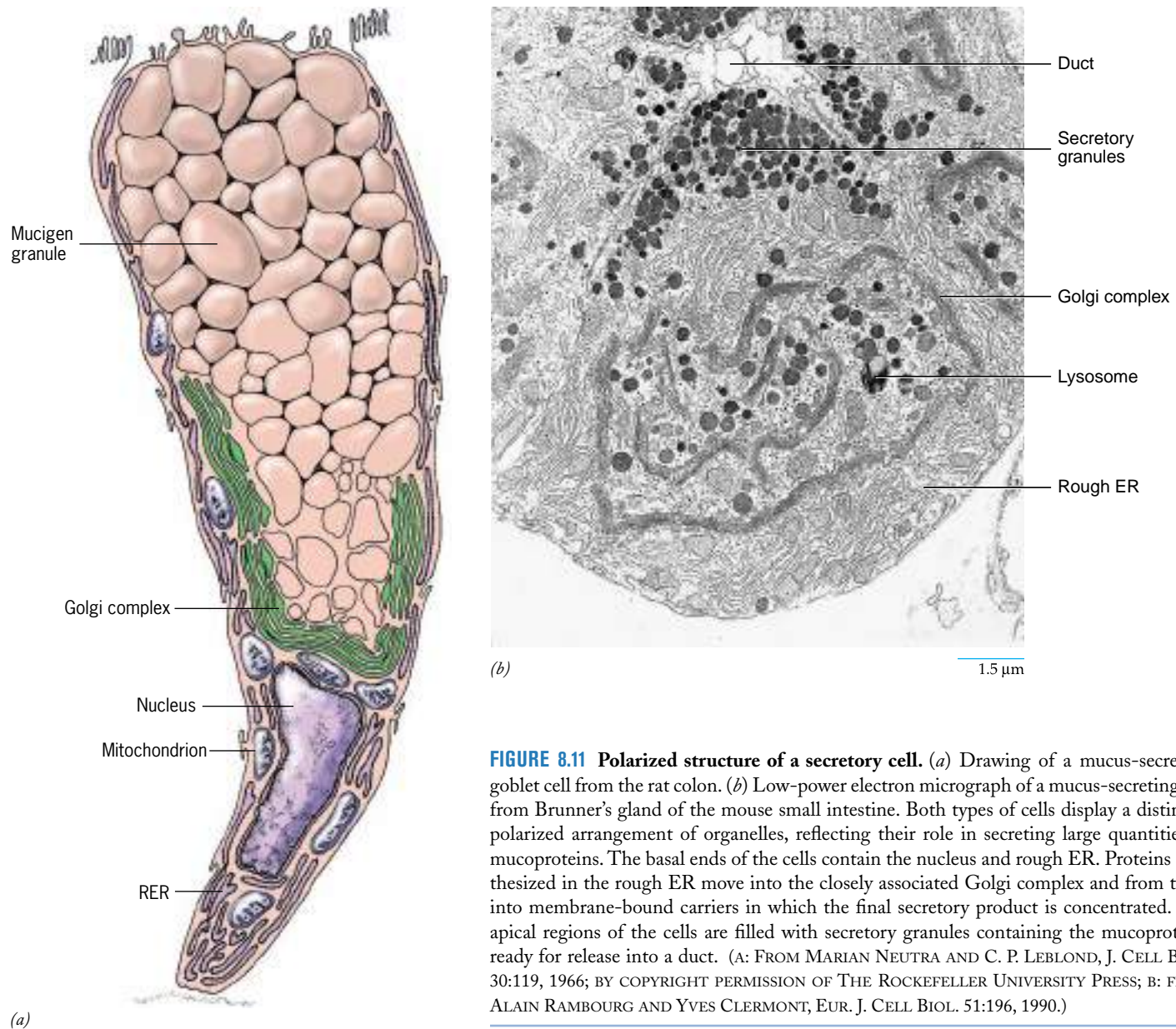
FIGURE 8.11 Polarized structure of a secretory cell. (*a*) Drawing of a mucus-secreting goblet cell from the rat colon. (*b*) Low-power electron micrograph of a mucus-secreting cell from Brunner's gland of the mouse small intestine. Both types of cells display a distinctly polarized arrangement of organelles, reflecting their role in secreting large quantities of mucoproteins. The basal ends of the cells contain the nucleus and rough ER. Proteins synthesized in the rough ER move into the closely associated Golgi complex and from there into membrane-bound carriers in which the final secretory product is concentrated. The apical regions of the cells are filled with secretory granules containing the mucoproteins ready for release into a duct. (A: FROM MARIAN NEUTRA AND C. P. LEBLOND, J. CELL BIOL. 30:119, 1966; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS; B: FROM ALAIN RAMBOURG AND YVES CLERMONT, EUR. J. CELL BIOL. 51:196, 1990.)

Synthesis of Proteins on Membrane-Bound versus Free Ribosomes

The discovery of the rough endoplasmic reticulum as the site of synthesis of secretory proteins in pancreatic acinar cells was described earlier (Figure 8.3). Similar results were found for other types of secretory cells, including intestinal goblet cells that secrete mucoproteins, endocrine cells that secrete polypeptide hormones, plasma cells that secrete antibodies, and liver cells that secrete blood serum proteins.

Further experiments revealed that polypeptides are synthesized at two distinct locales within the cell.

1. Certain polypeptides are synthesized on ribosomes attached to the cytosolic surface of the RER membranes. These include (a) secreted proteins, (b) integral membrane proteins, and (c) soluble proteins that reside within compartments of the endomembrane system, including the ER, Golgi complex, lysosomes, endosomes, vesicles, and plant vacuoles.

2. Other polypeptides are synthesized on "free" ribosomes, that is, on ribosomes that are not attached to the RER, and are subsequently released into the cytosol. This class includes (a) proteins destined to remain in the cytosol (such as the enzymes of glycolysis and the proteins of the cytoskeleton), (b) peripheral proteins of the cytosolic surface of membranes (such as spectrins and ankyrins that are only weakly associated with the plasma membrane's cytosolic surface), (c) proteins that are transported to the nucleus (Section 12.1), and (d) proteins to be incorporated into peroxisomes, chloroplasts, and mitochondria. Proteins in the latter two groups are synthesized to completion in the cytosol and

then imported *posttranslationally* into the appropriate organelle across its boundary membrane(s) (page 309).

What determines the location in a cell where a protein is synthesized? In the early 1970s, Günter Blobel, in collaboration with David Sabatini and Bernhard Dobberstein of Rockefeller University, first proposed, and then demonstrated, that the site of synthesis of a protein was determined by the sequence of amino acids in the N-terminal portion of the polypeptide, which is the first part to emerge from the ribosome during protein synthesis. They suggested the following:

1. Secretory proteins contain a **signal sequence** at their N-terminus that directs the emerging polypeptide and ribosome to the ER membrane.

2. The polypeptide moves into the cisternal space of the ER through a protein-lined, aqueous channel in the ER membrane. It was proposed that the polypeptide moves through the membrane as it is being synthesized, that is, *cotranslationally*.[2]

---

[2]It should be noted that protein transport across the ER membrane can also occur posttranslationally (i.e., after synthesis). In this process, the polypeptide is synthesized totally in the cytosol and then imported into the ER lumen through the same protein-conducting channels used in the cotranslational pathway. The posttranslational pathway is utilized much more heavily in yeast than in mammalian cells for import into the ER. In fact, yeast cells that are unable to engage in cotranslational transport into the ER are still able to survive, even though they grow more slowly than normal cells.

This proposal, known as the *signal hypothesis*, has been substantiated by a large body of experimental evidence. Even more importantly, Blobel's original concept that proteins contain built-in "address codes" has been shown to apply in principle to virtually all types of protein trafficking pathways throughout the cell.

Synthesis of Secretory, Lysosomal, or Plant Vacuolar Proteins on Membrane-Bound Ribosomes The steps that take place during the synthesis of a secretory, lysosomal, or plant vacuolar protein are shown in Figure 8.12. The synthesis of the polypeptide begins after a messenger RNA binds to a free ribosome, that is, one that is *not* attached to a cytoplasmic membrane. In fact, all ribosomes are thought to be

FIGURE 8.12 A schematic model of the synthesis of a secretory protein (or a lysosomal enzyme) on a membrane-bound ribosome of the RER. (*a*) Synthesis of the polypeptide begins on a free ribosome. As the signal sequence (shown in red) emerges from the ribosome, it binds to the SRP (step 1), which stops further translation until the SRP-ribosome-nascent chain complex can make contact with the ER membrane. The SRP-ribosome complex then collides with and binds to an SRP receptor (SR) situated within the ER membrane (step 2). Attachment of this complex to the SRP receptor is followed by release of the SRP and the association of the ribosome with a translocon of the ER membrane (step 3). These latter events are accompanied by the reciprocal hydrolyis of GTP molecules (not shown) bound to both the SRP and its receptor. In the model depicted here, the signal peptide then binds to the interior of the translocon, displacing the plug from the channel and allowing the remainder of the polypeptide to translocate through the membrane cotranslationally (step 4). After the nascent polypeptide passes into the lumen of the ER, the signal peptide is cleaved by a membrane protein (the signal peptidase, not shown), and the protein undergoes folding with the aid of ER chaperones, such as BiP. Studies suggest that translocons are organized into groups of two or four units rather than singly as shown here. (*b*) Cross-sectional view of the translocon channel from the side based on the X-ray crystal structure of an archaebacterial translocon. The hour-glass shape of the aqueous channel is evident. The helical plug that prevents the movement of solutes in the translocon's closed conformation and the ring of hydrophobic side chains (green) situated at the narrowest site within the channel are also shown. (B: FROM TOM A. RAPOPORT, NATURE 450:664, 2007; © COPYRIGHT 2007, MACMILLAN MAGAZINES LTD.)

identical; those employed in the synthesis of secretory, lyso-somal, or plant vacuolar proteins are taken from the same population (*pool*) as those used for production of proteins that remain in the cytosol. Polypeptides synthesized on membrane-bound ribosomes contain a signal sequence—which includes a stretch of 6–15 hydrophobic amino acid residues—that targets the *nascent* polypeptide to the ER membrane and leads to the compartmentalization of the polypeptide within the ER lumen. (A *nascent* polypeptide is one in the process of being synthesized.) Although the signal sequence is usually located at or near the N-terminus, it occupies an internal position in some polypeptides.

As it emerges from the ribosome, the hydrophobic signal sequence is recognized by a **signal recognition particle (SRP)**, which consists in mammalian cells of six distinct polypeptides and a small RNA molecule, called the 7S RNA. The SRP binds to both the signal sequence on the nascent polypeptide and the ribosome (step 1, Figure 8.12), temporarily arresting further synthesis of the polypeptide. The bound SRP serves as a tag that enables the entire complex (SRP-ribosome-nascent polypeptide) to bind specifically to the cytosolic surface of the ER membrane. Binding to the ER occurs through at least two distinct interactions: one between the SRP and the **SRP receptor** (step 2), and the other between the ribosome and the **translocon** (step 3). The translocon is a protein-lined channel embedded in the ER membrane through which the nascent polypeptide is able to move in its passage from the ribosome to the ER lumen.

One of the major feats in the field of membrane trafficking over the past few years has been the determination of the three-dimensional structure of a prokaryotic version of the translocon by X-ray crystallography. This effort revealed the presence within the translocon of a pore in the shape of an hourglass with a ring of six hydrophobic amino acids situated at its narrowest diameter. In the inactive (i.e., nontranslocating) state, which was the state in which the structure was crystallized, the opening in the pore ring is plugged by a short α helix. This plug is proposed to seal the channel, preventing the unwanted passage of calcium and other ions between the cytosol and the ER lumen.

Once the SRP-ribosome-nascent chain complex binds to the ER membrane (step 2, Figure 8.12), the SRP is released from its ER receptor, the ribosome becomes attached to the cytosolic end of the translocon, and the signal sequence on the nascent polypeptide is inserted into the narrow aqueous channel of the translocon (step 3). It is proposed that contact of the signal sequence with the interior of the translocon leads to displacement of the plug and opening of the passageway. The growing polypeptide is then translocated through the hydrophobic pore ring and into the ER lumen (step 4). Because the pore ring observed in the crystal structure has a diameter (5-8Å) that is considerably smaller than that of a helical polypeptide chain, it is presumed that the pore expands as the nascent chain traverses the channel. (Expansion is possible because the residues that make up the ring are situated on different helices of the translocon protein.) Upon termination of translation and passage of the completed polypeptide through the translocon, the membrane-bound ribosome is released

from the ER membrane and the helical plug is reinserted into the translocon channel.

Several of the steps involved in the synthesis and trafficking of secretory proteins are regulated by the binding or hydrolysis of GTP. As will be discussed at length in Chapter 15 and elsewhere in this chapter, **GTP-binding proteins** (or **G proteins**) play key regulatory roles in many different cellular processes.[3] G proteins can be present in at least two alternate conformations, one containing a bound GTP molecule and the other a bound GDP molecule. GTP-bound and GDP-bound versions of a G protein have different conformations and thus have different abilities to bind other proteins. Because of this difference in binding properties, G proteins act like "molecular switches"; the GTP-bound protein typically turns the process on, and hydrolysis of the bound GTP turns it off. Among the components depicted in Figure 8.12, both the SRP and the SRP receptor are G proteins that interact with one another in their GTP-bound states. The hydrolysis of GTP bound to these two proteins occurs between steps 2 and 3 and triggers the release of the signal sequence by the SRP and its insertion into the translocon.

Processing of Newly Synthesized Proteins in the Endoplasmic Reticulum   As it enters the RER cisterna, a nascent polypeptide is acted on by a variety of enzymes located within either the membrane or the lumen of the RER. The N-terminal portion containing the signal peptide is removed from most nascent polypeptides by a proteolytic enzyme, the **signal peptidase**. Carbohydrates are added to the nascent protein by the enzyme *oligosaccharyltransferase* (discussed on page 280). Both the signal peptidase and oligosaccharyltransferase are integral membrane proteins that reside in close proximity to the translocon and act on nascent proteins as they enter the ER lumen.

The RER is a major protein processing plant. To meet its obligations, the RER lumen is packed with molecular chaperones that recognize and bind to unfolded or misfolded proteins and give them the opportunity to attain their correct (native) three-dimensional structure (page 282). The ER lumen also contains a number of protein-processing enzymes, such as *protein disulfide isomerase* (*PDI*). Proteins enter the ER lumen with their cysteine residues in the reduced (—SH) state, but they leave the compartment with many of these residues joined to one another as oxidized disulfides (—SS—) (page 52). The formation (and rearrangement) of disulfide bonds is catalyzed by PDI. Disulfide bonds play an important role in maintaining the stability of proteins that are present at the extracellular surface of the plasma membrane or secreted into the extracellular space.

The endoplasmic reticulum is ideally constructed for its role as a port of entry for the biosynthetic pathway of the cell. Its membrane provides a large surface area to which many ribosomes can attach (an estimated 13 million per liver cell). The lumen of the ER cisternae provides a local environment

---

[3]GTP proteins generally require accessory proteins to carry out their function. The roles of these proteins are discussed in Chapter 15 and illustrated in Figure 15.19. They will not be considered in the present chapter, even though they are involved in these activities.

that favors the folding and assembly of proteins and a compartment in which secretory, lysosomal, and plant cell vacuolar proteins can be segregated from other newly synthesized proteins. The segregation of newly synthesized proteins in the ER cisternae removes them from the cytosol and allows them to be modified and dispatched toward their ultimate destination, whether outside the cell or within one of the cytoplasm's membranous organelles.

Synthesis of Integral Membrane Proteins on Membrane-Bound Ribosomes   Integral membrane proteins—other than those of mitochondria and chloroplasts—are also synthesized on membrane-bound ribosomes of the ER. These membrane proteins are translocated into the ER membrane as they are synthesized (i.e., cotranslationally) using the same machinery described for the synthesis of secretory and lysosomal proteins (Figure 8.12). Unlike soluble secretory and lysosomal proteins, however, which pass entirely through the ER membrane during translocation, integral proteins contain one or more hydrophobic transmembrane segments (page 130) that are shunted directly from the channel of the translocon into the lipid bilayer. How can such a transfer take place? The X-ray crystallographic studies of the translocon described above showed the translocon to have a clam-shaped conformation with a groove or seam along one side of the wall where the channel might open and close. As a polypeptide passes through the translocon, it is proposed that this lateral "gate" in the channel continually opens and closes, which gives each segment of the nascent polypeptide an opportunity to partition itself according to its solubility properties into either the aqueous compartment within the translocon channel or the surrounding hydrophobic core of the lipid bilayer. Those seg-

ments of the nascent polypeptide that are sufficiently hydrophobic will spontaneously "dissolve" into the lipid bilayer and ultimately become transmembrane segments of an integral membrane protein. This concept has received strong support from an in vitro study in which translocons were given the opportunity to translocate custom-designed nascent proteins that contained test segments of varying hydrophobicity. The more hydrophobic the test segment, the greater the likelihood it would pass through the wall of the translocon and become integrated as a transmembrane segment of the bilayer.

Figure 8.13 shows the synthesis of a pair of integral membrane proteins containing a single transmembrane segment. Single-spanning membrane proteins can have an orientation with their N-terminus facing either the cytosol or the lumen of the ER (and eventually the extracellular space). As noted on page 130, the most common determinant of membrane protein alignment is the presence of positively charged amino acid residues flanking the cytosolic end of a transmembrane segment (see Figure 4.18). During the synthesis of membrane proteins, the inner lining of the translocon is thought to orient the nascent polypeptide, as indicated in Figure 8.13, so that the more positive end faces the cytosol. In multispanning proteins (as shown in Figure 4.32d), sequential transmembrane segments typically have opposite orientations. For these proteins, their arrangement within the membrane is determined by the direction in which the first transmembrane segment is inserted. Once that has been determined, every other transmembrane segment has to be rotated 180° before it can exit the translocon. Studies performed with purified components in cell-free systems suggest that a translocon, by itself, is capable of properly orienting transmembrane segments. It would appear that the translocon is more than a sim-
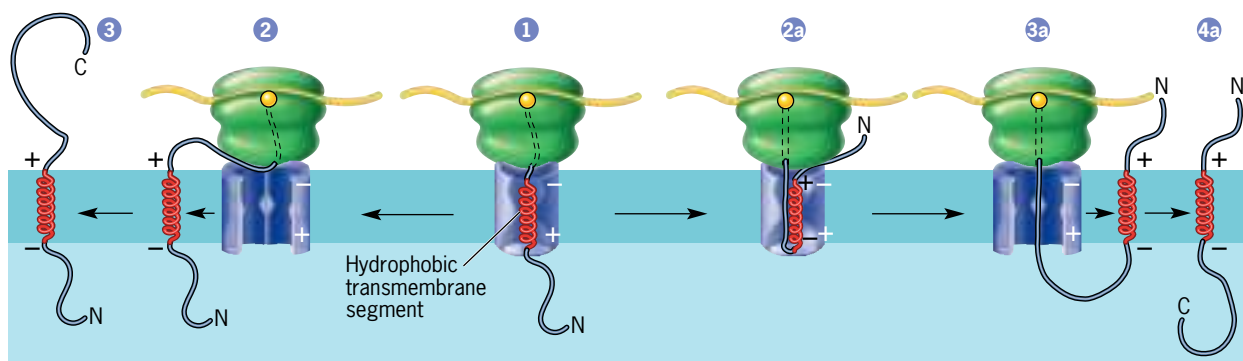
FIGURE 8.13 A schematic model for the synthesis of an integral membrane protein that contains a single transmembrane segment near the N-terminus of the nascent polypeptide. The SRP and the various components of the membrane that were shown in Figure 8.12 are also involved in the synthesis of integral proteins, but they have been omitted for simplicity. The nascent polypeptide enters the translocon just as if it were a secretory protein (step 1). However, the entry of the hydrophobic transmembrane sequence into the pore blocks further translocation of the nascent polypeptide through the channel. Steps 2–3 show the synthesis of a transmembrane protein whose N-terminus is in the lumen of the ER and C-terminus is in the cytosol. In step 2, the translocon has

opened laterally and expelled the transmembrane segment into the bilayer. Step 3 shows the final disposition of the protein. Steps 2a–4a show the synthesis of a transmembrane protein whose C-terminus is in the lumen and N-terminus is in the cytosol. In step 2a, the translocon has reoriented the transmembrane segment, in keeping with its reversed positively and negatively charged flanks. In step 3a, the translocon has opened laterally and expelled the transmembrane segment into the bilayer. Step 4a shows the final disposition of the protein. White-colored + and − signs indicate the proposed charge displayed by the inner lining of the translocon.

ple passageway through the ER membrane; it is a complex machine capable of recognizing various signal sequences and performing complex mechanical activities.

Membrane Biosynthesis in the ER   Membranes do not arise *de novo*, that is, as new entities from pools of protein and lipid that mix together. Instead, membranes arise from preexisting membranes. Membranes grow as newly synthesized proteins and lipids are inserted into existing membranes in the ER. As will be apparent in the following discussion, membrane components move from the ER to virtually every other compartment in the cell. As the membrane moves from one compartment to the next, its proteins and lipids are modified by enzymes that reside in the cell's various organelles. These modifications contribute to giving each membrane compartment a unique composition and distinct identity.

Keep in mind that cellular membranes are asymmetric: the two phospholipid layers (leaflets) of a membrane have different compositions (page 125). This asymmetry is established initially in the endoplasmic reticulum. Asymmetry is maintained as membrane carriers bud from one compartment and

FIGURE 8.14   Maintenance of membrane asymmetry. As each protein is synthesized in the rough ER, it becomes inserted into the lipid bilayer in a predictable orientation determined by its amino acid sequence. This orientation is maintained throughout its travels in the endomembrane system, as illustrated in this figure. The carbohydrate chains, which are first added in the ER, provide a convenient way to assess membrane sidedness because they are always present on the cisternal side of the cytoplasmic membranes, which becomes the exoplasmic side of the plasma membrane following the fusion of vesicles with the plasma membrane.

fuse to the next. As a result, domains situated at the cytosolic surface of the ER membrane can be identified on the cytosolic surface of transport vesicles, the cytosolic surface of Golgi cisternae, and the internal (cytoplasmic) surface of the plasma membrane (Figure 8.14). Similarly, domains situated at the luminal surface of the ER membrane maintain their orientation and are found at the external (exoplasmic) surface of the plasma membrane. In fact, in many ways, including its high calcium concentration and abundance of proteins with disulfide bonds and carbohydrate chains, the lumen of the ER (as well as other compartments of the secretory pathway) is a lot like the extracellular space.

Synthesis of Membrane Lipids   Most membrane lipids are synthesized entirely within the endoplasmic reticulum. The major exceptions are (1) sphingomyelin and glycolipids, whose synthesis begins in the ER and is completed in the Golgi complex, and (2) some of the unique lipids of the mitochondrial and chloroplast membranes, which are synthesized by enzymes that reside in those membranes. The enzymes involved in the synthesis of phospholipids are themselves integral proteins of the ER membrane with their active sites facing the cytosol. Newly synthesized phospholipids are inserted into the half of the bilayer facing the cytosol. Some of these lipid molecules are later flipped into the opposite leaflet through the action of enzymes called *flippases*. Lipids are carried from the ER to the Golgi complex and plasma membrane as part of the bilayer that makes up the walls of transport vesicles.

The membranes of different organelles have markedly different lipid composition (Figure 8.15*a*), which indicates that changes take place as membrane flows through the cell. Several factors may contribute to such changes (Figure 8.15*b*):

1. Most membranous organelles contain enzymes that modify lipids already present within a membrane, converting one type of phospholipid (e.g., phosphatidylserine) to another (e.g., phosphatidylcholine) (step 1, Figure 8.15*b*).

2. When vesicles bud from a compartment (as in Figure 8.2*a*), some types of phospholipids may be preferentially included within the membrane of the forming vesicle, while other types may be left behind (step 2, Figure 8.15*b*).

3. Cells contain **phospholipid-transfer proteins** that can bind and transport phospholipids *through the aqueous cytosol* from one membrane compartment to another (step 3, Figure 8.15*b*). These enzymes facilitate the movement of specific phospholipids from the ER to other organelles. This is especially important in delivering lipid to mitochondria and chloroplasts, which are not part of the normal flow of membrane along the biosynthetic pathway.

Glycosylation in the Rough Endoplasmic Reticulum Nearly all of the proteins produced on membrane-bound ribosomes—whether integral components of a membrane, soluble lysosomal or vacuolar enzymes, or parts of the extracellular matrix—become glycoproteins. Carbohydrate groups have key roles in the function of many glycoproteins, particularly as binding sites in their interactions with other

FIGURE 8.15 **Modifying the lipid composition of membranes.** (*a*) Histogram indicating percentage of each of three phospholipids (phosphatidylcholine, phosphatidylserine, and sphingomyelin) in three different cellular membranes (ER, Golgi complex, and plasma membrane). The percentage of each lipid changes gradually as membrane flows from the ER to the Golgi to the plasma membrane. (*b*) Schematic diagram showing three distinct mechanisms that might explain how the phospholipid composition of one membrane in the endomembrane system can be different from another membrane in the system, even though the membranous compartments are spatially and temporally continuous. (1) The head groups of phospholipids of the bilayer are modified enzymatically; (2) the membrane of a forming vesicle contains a different phospholipid composition from the membrane it buds from; (3) phospholipids can be removed from one membrane and inserted into another membrane by phospholipid-transfer proteins.

(a)

ER = Endoplasmic reticulum
GC = Golgi complex
PM = Erythrocyte plasma membrane

PC = Phosphatidylcholine
PS = Phosphatidylserine
SM = Sphingomyelin

(b)
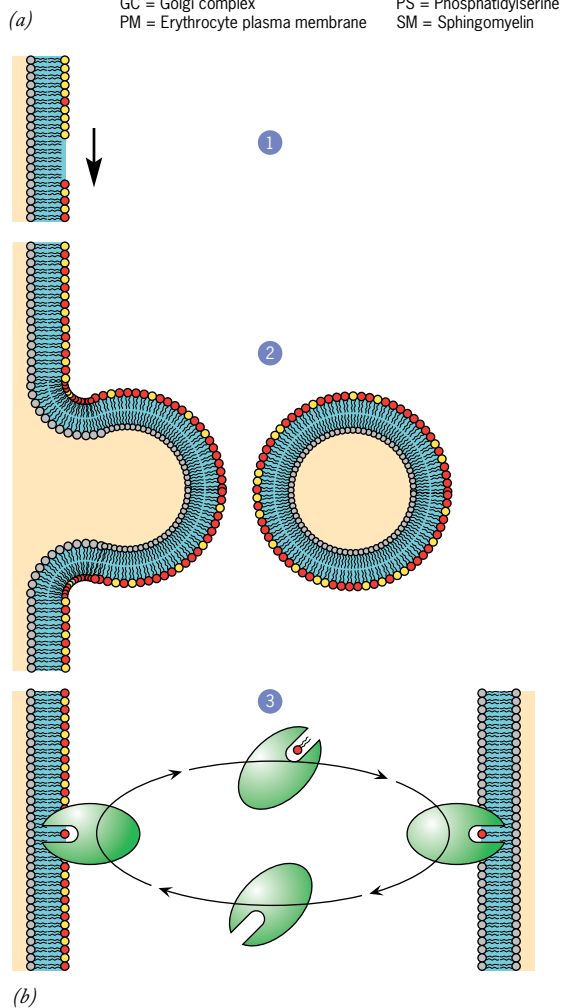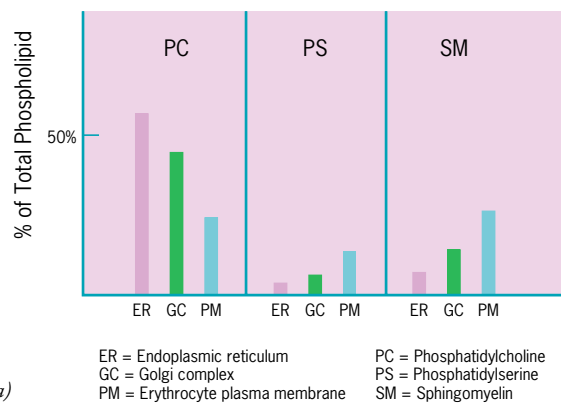
macromolecules, as occurs during many cellular processes. They also aid in the proper folding of the protein to which they are attached. The sequences of sugars that comprise the oligosaccharides of glycoproteins are highly specific; if the oligosaccharides are isolated from a purified protein, their sequence is consistent and predictable. How is the order of sugars in oligosaccharides achieved?

The addition of sugars to an oligosaccharide chain is catalyzed by a family of membrane-bound enzymes called **glycosyltransferases**. Each of these enzymes transfers a specific monosaccharide from a nucleotide sugar, such as GDP-mannose or UDP-*N*-acetylglucosamine (Figure 8.16), to the growing end of the carbohydrate chain. The sequence in which sugars are transferred during assembly of an oligosaccharide depends on the sequence of action of glycosyltransferases that participate in the process. This in turn depends on the location of specific enzymes within the various membranes of the secretory pathway. Thus the arrangement of sugars in the oligosaccharide chains of a glycoprotein depends on the spatial localization of particular enzymes in the assembly line.

The initial steps in the assembly of *N*-linked oligosaccharides (as opposed to *O*-linked oligosaccharides; see Figure 4.11) of both soluble proteins and integral membrane proteins are shown in Figure 8.16. The basal, or core, segment of each carbohydrate chain is not assembled on the protein itself but put together independently on a lipid carrier and then transferred, as a block, to specific asparagine residues of the polypeptide. This lipid carrier, which is named **dolichol phosphate**, is embedded in the ER membrane. Sugars are added to the dolichol phosphate molecule one at a time by membrane-bound glycosyltransferases, beginning with step 1 of Figure 8.16. This part of the glycosylation process is essentially invariant; in mammalian cells, it begins with the transfer of *N*-acetylglucosamine 1–phosphate, followed by the transfer of another *N*-acetylglucosamine, then nine mannose and three glucose units in the precise pattern indicated in Figure 8.16. This preassembled block of 14 sugars is then transferred by the ER enzyme oligosaccharyltransferase from dolichol phosphate to certain asparagines in the nascent polypeptide

(step 10, Figure 8.16) as the polypeptide is being translocated into the ER lumen.

Mutations that lead to the total absence of *N*-glycosylation cause the death of embryos prior to implantation. However, mutations that lead to partial disruption of the glycosylation pathway in the ER are responsible for serious inherited disorders affecting nearly every organ system. These diseases are called Congenital Diseases of Glycosylation, or CDGs and they are usually identified through blood tests that detect abnormal glycosylation of serum proteins. One of these diseases, CDG1b, can be managed through a remarkably
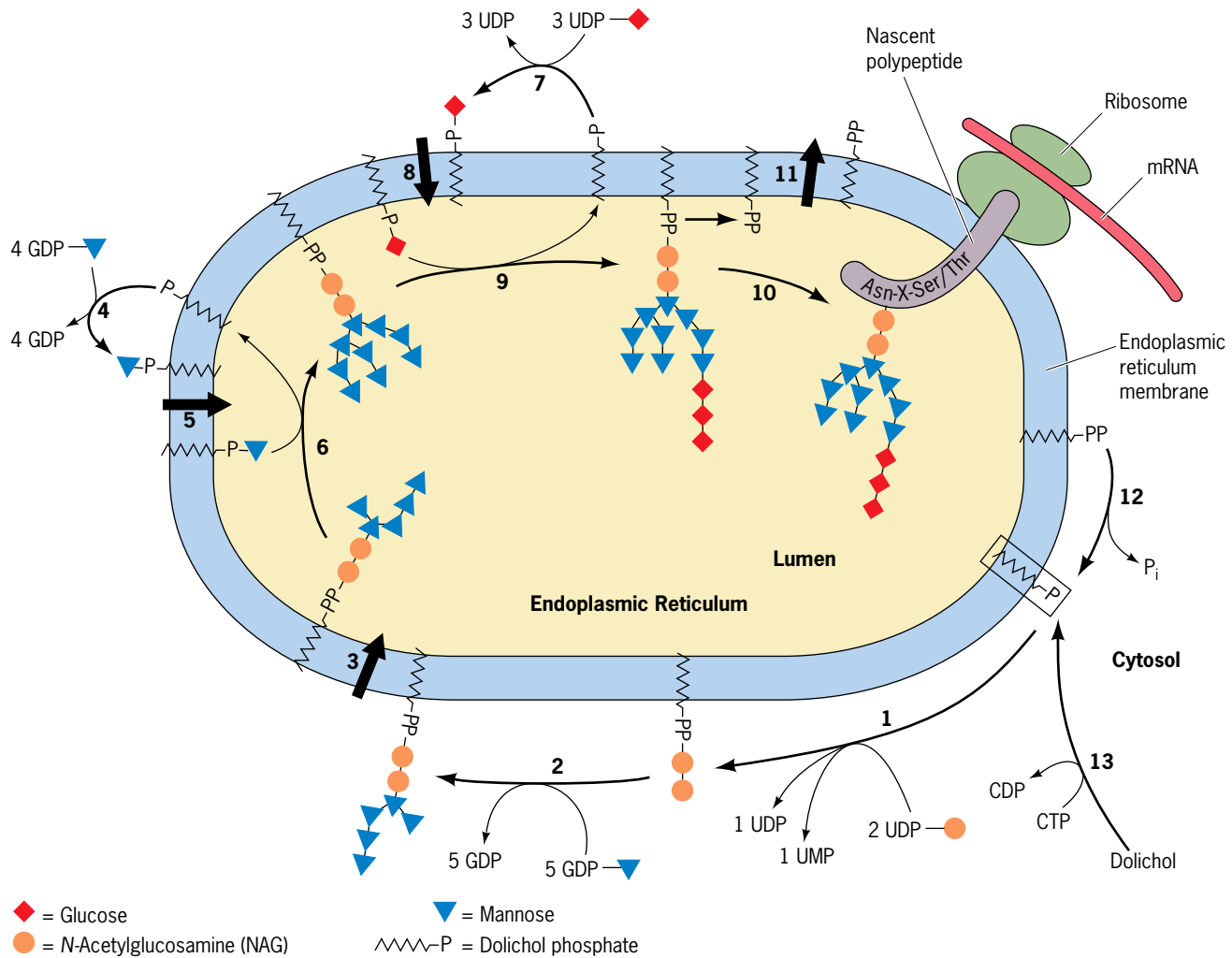
FIGURE 8.16 Steps in the synthesis of the core portion of an *N*-linked oligosaccharide in the rough ER. The first seven sugars (five mannose and two NAG residues) are transferred one at a time to the dolichol-PP on the cytosolic side of the ER membrane (steps 1 and 2). At this stage, the dolichol with its attached oligosaccharide is then flipped across the membrane (step 3), and the remaining sugars (four mannose and three glucose residues) are attached on the luminal side of the membrane. These latter sugars are attached one at a time on the cytosolic side of the membrane to the end of a dolichol phosphate molecule (as in steps 4 and 7), which then flips across the membrane (steps 5 and 8) and donates its sugar to the growing end of the oligosaccharide chain (steps 6 and 9). Once the oligosaccharide is completely assembled, it is transferred enzymatically to an asparagine residue of the nascent polypeptide (step 10). The dolichol-PP is flipped back across the membrane (step 11) and is ready to begin accepting sugars again (steps 12 and 13). (FROM D. VOET AND J. G. VOET, BIOCHEMISTRY, 2D ED., COPYRIGHT 1995; JOHN WILEY & SONS, INC. REPRINTED BY PERMISSION OF JOHN WILEY & SONS, INC.)

simple treatment. CDG1b results from the deficiency of the enzyme phosphomannose isomerase, which catalyzes the conversion of fructose 6-phosphate to mannose 6-phosphate, a crucial reaction in the pathway that makes mannose available for incorporation into oligosaccharides. The disease can be managed by providing patients with oral supplements of mannose. The treatment was first tested in a boy who was dying from uncontrolled gastrointestinal bleeding, one of the usual complications of the disease. Within months of taking mannose supplements the child was living a normal life.

Shortly after it is transferred to the nascent polypeptide, the oligosaccharide chain undergoes a gradual process of modification. This modification begins in the ER with the enzymatic removal of two of the three terminal glucose residues (step 1, Figure 8.17). This sets the stage for an important event in the life of a newly synthesized glycoprotein in which it is screened by a system of **quality control** that determines whether or not it is fit to move on to the next compartment of the biosynthetic pathway. To begin this screening process, each glycoprotein, which at this stage contains a single remaining glucose, binds to an ER chaperone (calnexin or calreticulin) (step 2). Removal of the remaining glucose by glucosidase II leads to release of the glycoprotein from the chaperone (step 3). If a glycoprotein at this stage has not completed its folding or has misfolded, it is recognized by a conformation-sensing enzyme (called GT) that adds a single glucose residue back to
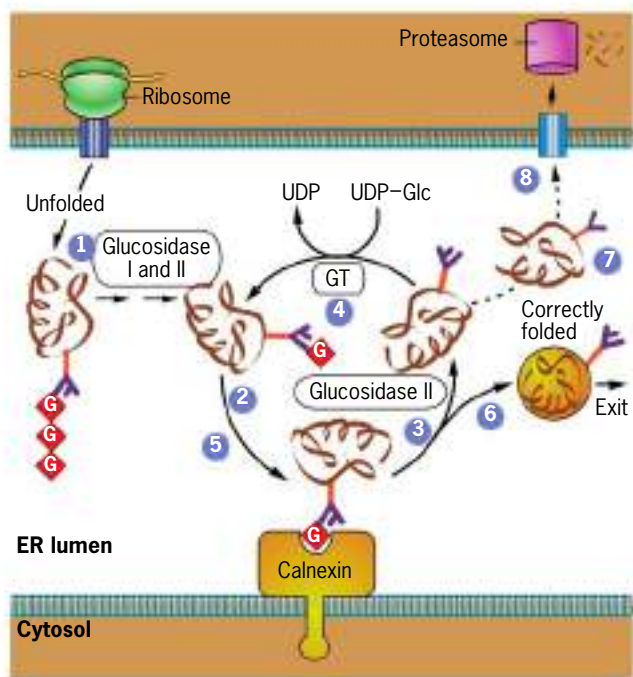
FIGURE 8.17 Quality control: ensuring that misfolded proteins do not proceed forward. Based on this proposed mechanism, misfolded proteins are recognized by a glucosyltransferase (GT) which adds a glucose to the end of the oligosaccharide chains. Glycoproteins containing monoglucosylated oligosaccharides are recognized by the membrane-bound chaperone calnexin and given an opportunity to achieve their correctly folded (native) state. If that does not occur after repeated attempts, the protein is translocated to the cytosol and destroyed. The steps are described in the text. A soluble chaperone (calreticulin) participates in this same quality-control pathway. (AFTER L. ELLGAARD ET AL. SCIENCE 286:1884, 1999; COPYRIGHT 1999, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

one of the mannose residues at the exposed end of the recently trimmed oligosaccharide (step 4). GT recognizes incompletely folded or misfolded proteins because they display exposed hydrophobic residues that are absent from properly folded proteins. Once the glucose residue has been added, the "tagged" glycoprotein is recognized by the same ER chaperones, which give the protein another chance to fold properly (step 5). After a period of time with the chaperone, the added glucose residue is removed and the conformation-sensing enzyme checks it again to see if it has achieved its proper three-dimensional structure. If it is still partially unfolded or misfolded, another glucose residue is added and the process is repeated until, eventually, the glycoprotein has either folded correctly and continues on its way (step 6), or remains misfolded and is destroyed. Studies suggest that the "decision" to destroy the defective protein is governed by a slow-acting enzyme in the ER that trims a mannose residue from an exposed end of the oligosaccharide of a protein that has been in the ER for an extended period. Once one or more of these mannose residues have been removed (step 7), the protein can no longer be recycled and, instead, is sentenced to degradation (step 8).

We will pick up the story of protein glycosylation again on page 284, where the oligosaccharide that is assembled in the ER is enlarged as it passes through the Golgi complex on its journey through the biosynthetic pathway.

Mechanisms that Ensure the Destruction of Misfolded Proteins   We have just seen how proteins that fail to fold properly are detected by ER enzymes. It was a surprise to discover that misfolded proteins are not destroyed in the ER, but instead are transported into the cytosol by a process of "dislocation." It remains unclear whether misfolded proteins are dislocated back into the cytosol through the translocons that brought them into the ER or by way of a separate dislocation channel of uncertain identity. Once in the cytosol, the oligosaccharide chains are removed, and the misfolded proteins are destroyed in proteasomes, which are protein-degrading machines whose structure and function are discussed in Section 12.7. This process, known as *ER-associated degradation* (*ERAD*), ensures that aberrant proteins are not transported to other parts of the cell, but it can have negative consequences. In most patients with cystic fibrosis, the plasma membrane of epithelial cells is lacking the abnormal protein encoded by the cystic fibrosis gene (page 156). In these cases, the mutant protein is destroyed by the quality control process of the ER and thus fails to reach the cell surface.

Under certain circumstances, misfolded proteins can be generated in the ER at a rate faster than they can be exported to the cytoplasm. The accumulation of misfolded proteins, which is potentially lethal to cells, triggers a comprehensive "plan of action" within the cell known as the **unfolded protein response (UPR)**. The ER contains protein sensors that monitor the concentration of unfolded or misfolded proteins in the ER lumen. According to the prevailing model outlined in Figure 8.18, the sensors are normally kept in an inactive state by molecular chaperones, particularly BiP. If circumstances should lead to an accumulation of misfolded proteins, the BiP molecules in the ER lumen are called into service as chaperones for the misfolded proteins, rendering them incapable of inhibiting the sensors. Activation of the sensors leads to a multitude of signals that are transmitted into both the nucleus and cytosol and result in

- the expression of hundreds of different genes whose encoded proteins have the potential to alleviate stressful conditions within the ER. These include genes that encode (1) ER-based molecular chaperones that can help misfolded proteins reach the native state, (2) proteins involved in the transport of the proteins out of the ER, and (3) proteins involved in the selective destruction of abnormal proteins as discussed above.

- phosphorylation of a key protein (eIFα) required for protein synthesis. This modification inhibits protein synthesis and decreases the flow of newly synthesized proteins into the ER. This gives the cell an opportunity to remove those proteins that are already present in the ER lumen.

Interestingly, the UPR is more than a cell-survival mechanism; it also includes the activation of a pathway that leads to the death of the cell. It is presumed that the UPR provides a mechanism to relieve itself of the stressful conditions. If these corrective measures are not successful, the cell-death pathway is triggered and the cell is destroyed.

FIGURE 8.18 A model of the mammalian unfolded protein response (UPR). The ER contains transmembrane proteins that function as sensors of events that occur within the ER lumen. Under normal conditions, these sensors are present in an inactive state as the result of their association with chaperones, particularly BiP (step 1). If the number of unfolded or misfolded proteins should increase to a high level, the chaperones are recruited to aid in protein folding, which leaves the sensors in their unbound, activated state and capable of initiating a UPR. At least three distinct UPR pathways have been identified in mammalian cells, each activated by a different protein sensor. Two of these pathways are depicted in this illustration. In one of these pathways, the release of the inhibitory BiP protein leads to the dimerization of a sensor (called PERK) (step 2). In its dimeric state, PERK beomes an activated protein kinase that phosphorylates a protein (eIF2α) that is required for the initiation of protein synthesis (step 3). This translation factor is inactive in the phosphorylated state, which stops the cell from synthesizing additional proteins in the ER (step 4), giving the cell more time to process those proteins already present in the ER lumen. In the second pathway depicted here, release of the inhibitory BiP protein allows the sensor (called ATF6) to move on to the Golgi complex where the cytosolic domain of the protein is cleaved away from its transmembrane domain (step 2a). The cytosolic portion of the sensor diffuses through the cytosol (step 3a) and into the nucleus (step 4a), where it stimulates the expression of genes whose encoded proteins can alleviate the stress in the ER (step 5a). These include chaperones, coat proteins that form on transport vesicles, and proteins of the quality-control machinery.
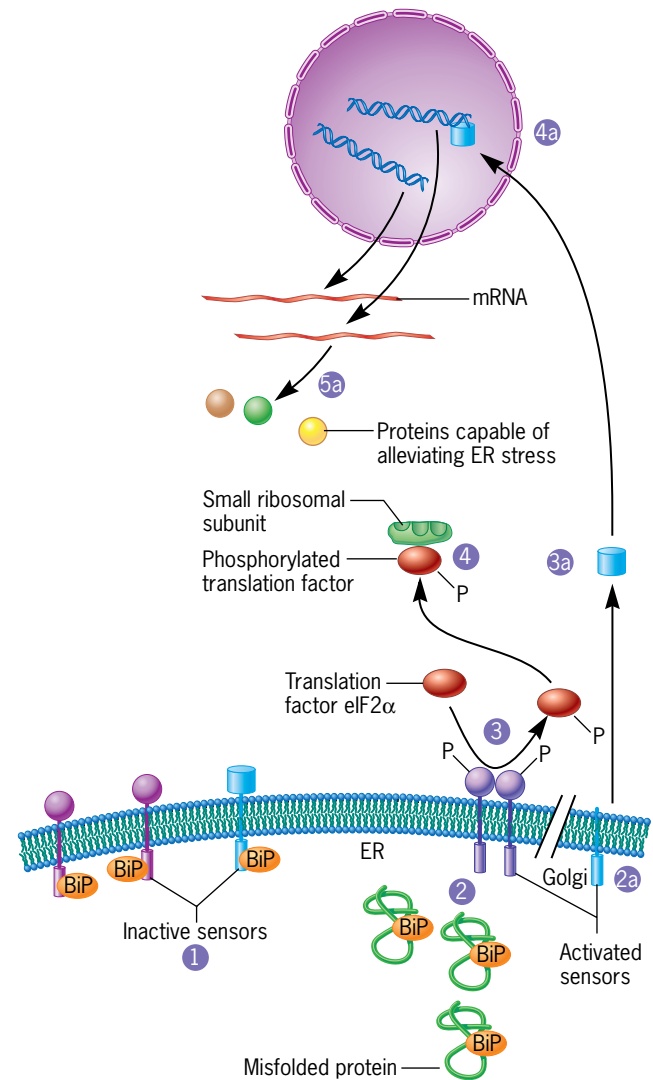
From the ER to the Golgi Complex: The First Step in Vesicular Transport

The exit sites of the RER cisternae are devoid of ribosomes and serve as places where the first transport vesicles in the biosynthetic pathway are formed. The trip from the ER toward the Golgi complex can be followed visually in living cells by tagging secretory proteins with the green fluorescent protein (GFP) as described on page 267. Using this technique it is found that soon after they bud from the ER membrane, transport vesicles fuse with one another to form larger vesicles and interconnected tubules in the region between the ER and Golgi complex. This region has been named the *ERGIC* (*e*ndoplasmic *r*eticulum *G*olgi *i*ntermediate *c*ompartment), and the vesicular-tubular carriers that form there are called *VTCs*

(see Figure 8.25*a*). Once formed, the VTCs move farther away from the ER toward the Golgi complex. The movement of two of these vesicular-tubular membranous carriers from the ERGIC to the Golgi complex is shown in Figure 8.19. Movement of VTCs occurs along tracks composed of microtubules.

FIGURE 8.19 Visualizing membrane traffic with the use of a fluorescent tag. This series of photographs shows a small portion of a living mammalian cell that has been infected with the vesicular stomatitis virus (VSV) containing a *VSVG–GFP* chimeric gene (page 268). Once it is synthesized in the RER, the fusion protein emits a green fluorescence, which can be followed as the protein moves through the cell. In the series of photographs shown here, two vesicular-tubular carriers (VTCs) (arrows) containing the fluorescent protein have budded from the ER and are moving toward the Golgi complex (GC). The series of events depicted here took place over a period of 13 seconds. Bar represents 6 mm. (REPRINTED WITH PERMISSION FROM JOHN F. PRESLEY ET AL., NATURE 389:82, 1997; © COPYRIGHT 1997, MACMILLAN MAGAZINES LTD.)

8.4 THE GOLGI COMPLEX

In the latter years of the nineteenth century, an Italian biologist, Camillo Golgi, was inventing new types of staining procedures that might reveal the organization of nerve cells within the central nervous system. In 1898, Golgi applied a metallic stain to nerve cells from the cerebellum and discovered a darkly stained reticular network located near the cell nucleus. This network, which was later identified in other cell types and named the **Golgi complex**, helped earn its discoverer the Nobel Prize in 1906. The Golgi complex remained a center of controversy for decades between those who believed that the organelle existed in living cells and those who believed it was an *artifact*, that is, an artificial structure formed during preparation for microscopy. It wasn't until the Golgi complex was clearly identified in unfixed, freeze-fractured cells (see Figure 18.17) that its existence was verified beyond reasonable doubt.

The Golgi complex has a characteristic morphology consisting primarily of flattened, disklike, membranous cisternae with dilated rims and associated vesicles and tubules (Figure 8.20a). The cisternae, whose diameters are typically 0.5 to 1.0 μm, are arranged in an orderly stack, much like a stack of pancakes, and are curved so as to resemble a shallow bowl (Figure 8.20b).[4] Typically, a Golgi stack contains fewer than eight cisternae. An individual cell may contain from a few to several thousand distinct stacks, depending on the cell type. The Golgi stacks in mammalian cells are interconnected by membranous tubules to form a single, large ribbon-like complex situated adjacent to the cell's nucleus (Figure 8.20c). A closer look at an individual cisterna suggests that vesicles bud from a peripheral tubular domain of each cisterna (Figure 8.20d). As discussed later, many of these vesicles contain a distinct protein coat that is visible in Figure 8.20d.

The Golgi complex is divided into several functionally distinct compartments arranged along an axis from the *cis* or entry face closest to the ER to the *trans* or exit face at the opposite end of the stack (Figures 8.20a,b). The *cis*-most face of the organelle is composed of an interconnected network of tubules referred to as the ***cis* Golgi network (CGN)**. The CGN is thought to function primarily as a sorting station that distinguishes between proteins to be shipped back to the ER (page 291) and those that are allowed to proceed to the next Golgi station. The bulk of the Golgi complex consists of a series of large, flattened cisternae, which are divided into ***cis***, ***medial***, and ***trans* cisternae** (Figure 8.20a). The *trans*-most face of the organelle contains a distinct network of tubules and vesicles called the ***trans* Golgi network (TGN)**. The TGN is a sorting station where proteins are segregated into different types of vesicles heading either to the plasma membrane or to various intracellular destinations. The membranous elements of the Golgi complex are thought to be supported mechanically by a peripheral membrane skeleton or scaffold composed of a variety of proteins, including members of the spectrin, ankyrin, and actin families—proteins that are also present as part of the plasma membrane skeleton (page 142). The Golgi scaffold may be physically linked with motor proteins that direct the movement of vesicles and tubules entering and exiting the Golgi complex. A separate group of fibrous proteins are thought to form a Golgi "matrix," that plays a key role in the disassembly and reassembly of the Golgi complex during mitosis.

Figure 8.21 provides visual evidence that the Golgi complex is not uniform in composition from one end to the other. Differences in composition of the membrane compartments from the *cis* to the *trans* face reflect the fact that the Golgi complex is primarily a "processing plant." Newly synthesized membrane proteins, as well as secretory and lysosomal proteins, leave the ER and enter the Golgi complex at its *cis* face and then pass across the stack to the *trans* face. As they progress along the stack, proteins that were originally synthesized in the rough ER are sequentially modified in specific ways. In the best-studied Golgi activity, a protein's carbohydrates are modified by a series of stepwise enzymatic reactions, as discussed in the following section.

Glycosylation in the Golgi Complex

The Golgi complex plays a key role in the assembly of the carbohydrate component of glycoproteins and glycolipids. When we left the topic of synthesis of *N*-linked carbohydrate chains on page 281, the glucose residues had just been removed from the ends of the core oligosaccharide. As newly synthesized soluble and membrane glycoproteins pass through the *cis* and *medial* cisternae of the Golgi stack, most of the mannose residues are also removed from the core oligosaccharides, and other sugars are added sequentially by various glycosyltransferases.

In the Golgi complex, as in the RER, the sequence in which sugars are incorporated into oligosaccharides is determined by the spatial arrangement of the specific glycosyltrans-

---

[4]A single Golgi stack in plant cells is sometimes called a *dictyosome*.

(a)

(b)

(c)

(d)
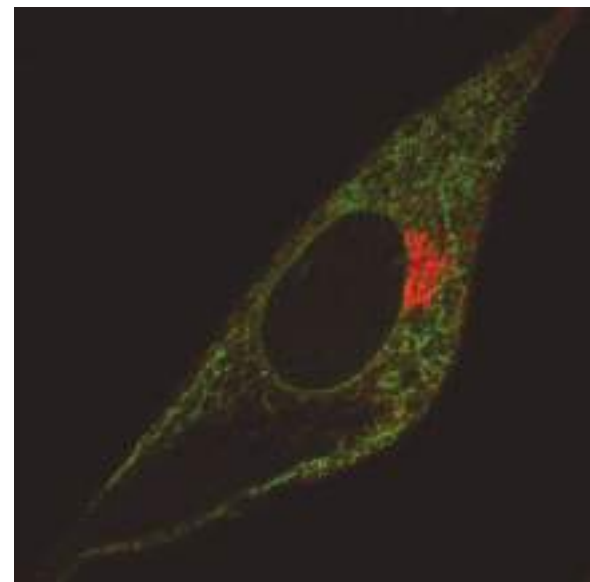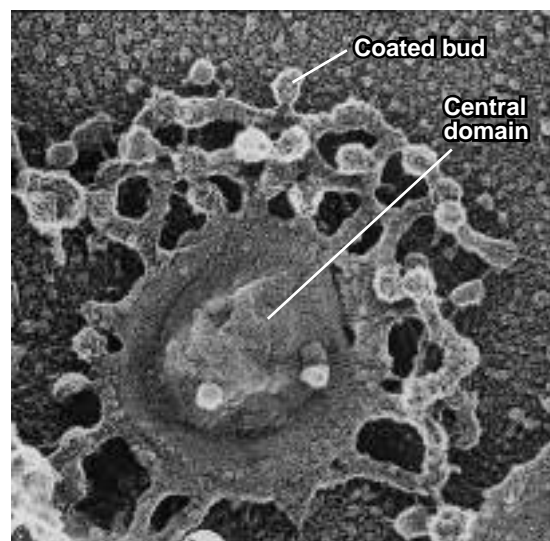
FIGURE 8.20 **The Golgi complex.** (*a*) Schematic model of a portion of a Golgi complex from an epithelial cell of the male rat reproductive tract. The elements of the *cis* and *trans* compartments are often discontinuous and appear as tubular networks. (*b*) Electron micrograph of a portion of a tobacco root cap cell showing the *cis* to *trans* polarity of the Golgi stack. (*c*) Fluorescence micrograph of a cultured mammalian cell. The position of the Golgi complex is revealed by the red fluorescence, which marks the localization of antibodies to a COPI coat protein. (*d*) Electron micrograph of a single isolated Golgi cisterna showing two distinct domains, a concave central domain and an irregular peripheral domain. The peripheral domain consists of a tubular network from which protein-coated buds are being pinched off. (A: FROM A. RAMBOURG AND Y. CLERMONT, EUR. J. CELL BIOL. 51:195, 1990; B: COURTESY OF THOMAS H. GIDDINGS; C: FROM ANDREI V. NIKONOV ET AL., J. CELL BIOL. 158:500, 2002; COURTESY OF GERT KREIBICH, BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS; FROM PEGGY J. WEIDMAN AND JOHN HEUSER, TRENDS CELL BIOL. 5:303, 1995; COPYRIGHT 1995, WITH PERMISSION FROM ELSEVIER SCIENCE.)

ferases that come into contact with the newly synthesized protein as it moves through the Golgi stack. The enzyme sialyltransferase, for example, which places a sialic acid at the terminal position of the chain in animal cells, is localized in the *trans* end of the Golgi stack, as would be expected if newly synthesized glycoproteins were continually moving toward this part of the organelle. In contrast to the glycosylation events that occur in the ER, which assemble a single core

(a)   (b)   (c)   0.15 μm
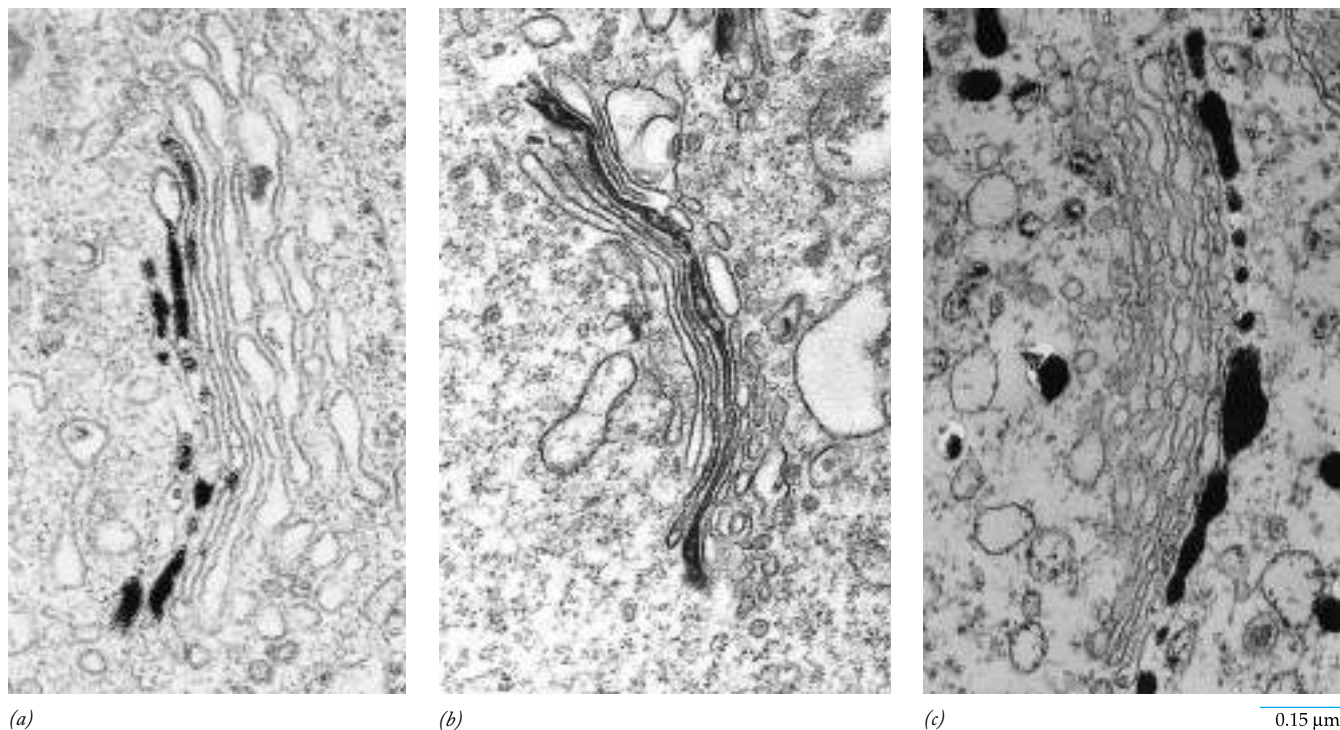
FIGURE 8.21 Regional differences in membrane composition across the Golgi stack. (*a*) Reduced osmium tetroxide preferentially impregnates the *cis* cisternae of the Golgi complex. (*b*) The enzyme mannosidase II, which is involved in trimming the mannose residues from the core oligosaccharide as described in the text, is preferentially localized in the *medial* cisternae. (*c*) The enzyme nucleoside diphosphatase, which splits dinucleotides (e.g., UDP) after they have donated their sugar, is preferentially localized in the *trans* cisternae. (A,C: FROM ROBERT S. DECKER, J. CELL BIOL. 61:603, 1974; B: FROM ANGEL VELASCO ET AL., J. CELL BIOL. 122:41, 1993; ALL BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

oligosaccharide, the glycosylation steps in the Golgi complex can be quite varied, producing carbohydrate domains of remarkable sequence diversity. One of many possible glycosylation pathways is shown in Figure 8.22. Unlike the *N*-linked oligosaccharides, whose synthesis begins in the ER, those at-tached to proteins by *O*-linkages (Figure 4.11) are assembled entirely within the Golgi complex.

The Golgi complex is also the site of synthesis of most of a cell's complex polysaccharides, including the glycosamino-glycan chains of the proteoglycan shown in Figure 7.9*a* and
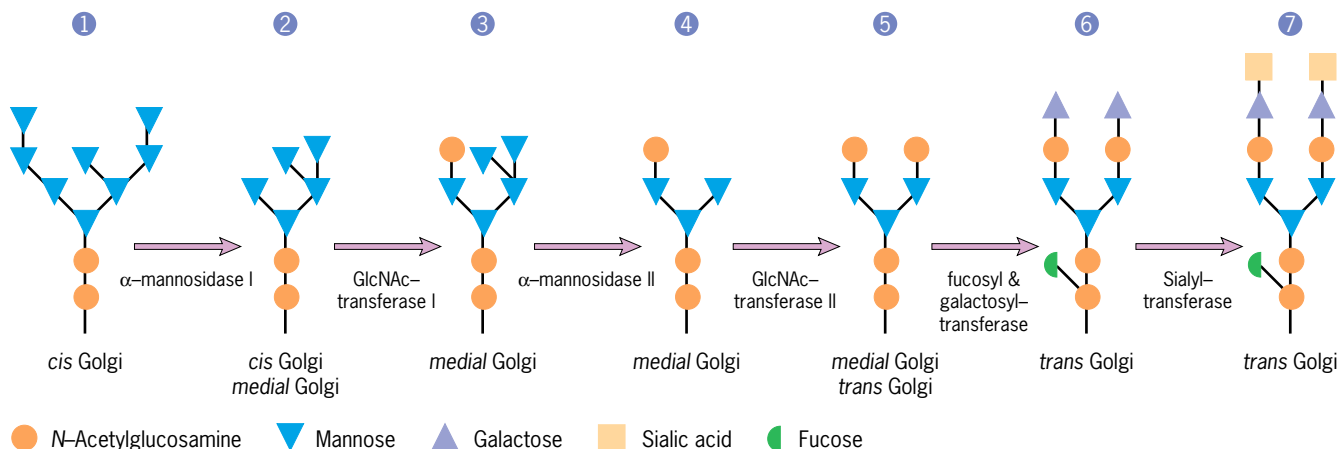
FIGURE 8.22 Steps in the glycosylation of a typical mammalian *N*-linked oligosaccharide in the Golgi complex. Following the removal of the three glucose residues, various mannose residues are subsequently removed, while a variety of sugars (*N*-acetylglucosamine, galactose, fucose, and sialic acid) are added to the oligosaccharide by specific glycosyl-transferases. These enzymes are integral membrane proteins whose ac-tive sites face the lumen of the Golgi cisternae. This is only one of numerous glycosylation pathways.

the pectins and hemicellulose found in the cell walls of plants (see Figure 7.37*c*).

The Movement of Materials through the Golgi Complex

That materials move through the various compartments of the Golgi complex has long been established; however, two contrasting views of the way this occurs have dominated the field for years. Up until the mid-1980s, it was generally accepted that Golgi cisternae were transient structures. It was supposed that Golgi cisternae formed at the *cis* face of the stack by fusion of membranous carriers from the ER and ERGIC and that each cisterna physically moved from the *cis* to the *trans* end of the stack, changing in composition as it progressed. This is known as the *cisternal maturation model* because, according to the model, each cisterna "matures" into the next cisterna along the stack.

From the mid-1980s to the mid-1990s, the maturation model of Golgi movement was largely abandoned and replaced by an alternate model, which proposed that the cisternae of a Golgi stack remain in place as stable compartments.

In this latter model, which is known as the *vesicular transport model*, cargo (i.e., secretory, lysosomal, and membrane proteins) is shuttled through the Golgi stack, from the CGN to the TGN, in vesicles that bud from one membrane compartment and fuse with a neighboring compartment farther along the stack. The vesicular transport model is illustrated in Figure 8.23*a*, and its acceptance was based largely on the following observations:

1. Each of the various Golgi cisternae of a stack has a distinct population of resident enzymes (Figure 8.21). How could the various cisternae have such different properties if each cisterna was giving rise to the next one in line, as suggested by the cisternal maturation model?

2. Large numbers of vesicles can be seen in electron micrographs to bud from the rims of Golgi cisternae. In 1983, James Rothman and his colleagues at Stanford University demonstrated, using cell-free preparations of Golgi membranes (page 280), that transport vesicles were capable of budding from one Golgi cisterna and fusing with another Golgi cisterna in vitro. This landmark experiment formed the basis for a hypothesis suggesting that inside the cell,

(a) Vesicular transport model   (b) Cisternal maturation model   (c)   0.2 μm

(d)
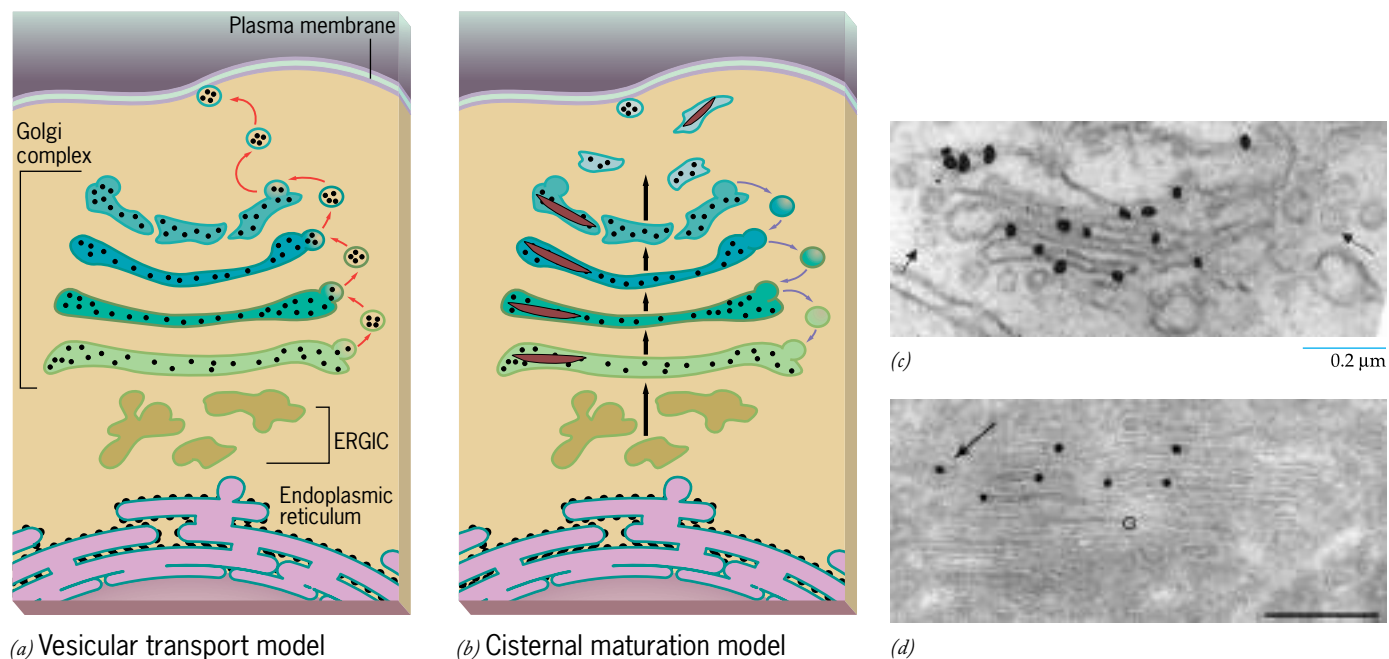
FIGURE 8.23 The dynamics of transport through the Golgi complex. (*a*) In the vesicular transport model, cargo (black dots) is carried in an anterograde direction by transport vesicles, while the cisternae themselves remain as stable elements. (*b*) In the cisternal maturation model, the cisternae progress gradually from a *cis* to a *trans* position and then disperse at the TGN. Transport vesicles carry resident Golgi enzymes (indicated by the colored vesicles) in a retrograde direction. The red lens-shaped objects represent large cargo materials, such as procollagen complexes of fibroblasts. (*c*) Electron micrograph of an area of Golgi complex in a thin frozen section of a cell that had been infected with vesicular stomatitis virus (VSV). The black dots are nanosized gold particles bound by means of antibodies to VSVG protein, an anterograde cargo molecule. The cargo is restricted to the cisternae and does not appear in nearby vesicles (arrows). (*d*) Electron micrograph of similar nature to that of *c* but, in this case, the gold particles are not bound to cargo, but to mannosidase II, a resident enzyme of the *medial* Golgi cisternae. The enzyme appears in both a vesicle (arrow) and cisternae. The labeled vesicle is presumably carrying the enzyme in a retrograde direction, which compensates for the anterograde movement of the enzyme as the result of cisternal maturation. Bar, 0.2 μm. (A third model for intra-Golgi transport is discussed in *Cell* 133:951, 2008.) (C: FROM ALEXANDER A. MIRONOV ET AL., COURTESY OF ALBERTO LUINI, J. CELL BIOL. 155:1234, 2001; D: FROM JOSE A. MARTINEZ-MENÁRGUEZ ET AL., COURTESY OF JUDITH KLUMPERMAN, J. CELL BIOL. 155:1214, 2001; BOTH BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

cargo-bearing vesicles budded from *cis* cisternae and fused with cisternae situated at a more *trans* position in the stack.

Although both models of Golgi function continue to have their proponents, the consensus of opinion has shifted back to the cisternal maturation model. Several of the major reasons for this shift can be noted:

■ The cisternal maturation model envisions a highly dynamic Golgi complex in which the major elements of the organelle, the cisternae, are continually being formed at the *cis* face and dispersed at the *trans* face. According to this view, the very existence of the Golgi complex itself depends on the continual influx of transport carriers from the ER and ERGIC. As predicted by the cisternal maturation model, when the formation of transport carriers from the ER is blocked either by treatment of cells with specific drugs or the use of temperature-sensitive mutants (page 268), the Golgi complex simply disappears. When the drugs are removed or the mutant cells are returned to the permissive temperature, the Golgi complex rapidly reassembles as ER–to–Golgi transport is renewed.

■ Certain materials that are produced in the endoplasmic reticulum and travel through the Golgi complex can be shown to remain within the Golgi cisternae and never appear within Golgi-associated transport vesicles. For example, studies on fibroblasts indicate that large complexes of procollagen molecules (the precursors of extracellular collagen) move from the *cis* cisternae to the *trans* cisternae without ever leaving the cisternal lumen.

■ It was assumed until the mid-1990s that transport vesicles always moved in a "forward" (*anterograde*) direction, that is, from a *cis* origin to a more *trans* destination. But a large body of evidence has indicated that vesicles can move in a "backward" (*retrograde*) direction, that is, from a *trans* donor membrane to a *cis* acceptor membrane.

■ Studies on live budding yeast cells containing fluorescently labeled Golgi proteins have shown directly that the composition of an individual Golgi cisterna can change over time—from one that contains early (*cis*) Golgi resident proteins to one that contains late (*trans*) Golgi resident proteins. The results of this experiment are shown in the opening micrograph of Chapter 18 and they are not compatible with the vesicular transport model. Whether or not these results on yeast can be extrapolated to a mammalian Golgi complex, which has a more complex, stacked structure, remains to be determined.

A current version of the cisternal maturation model is depicted in Figure 8.23*b*. Unlike the original versions of the cisternal maturation model, the version shown in Figure 8.23*b* acknowledges a role for transport vesicles, which have been clearly shown to bud from Golgi membranes. In this model, however, these transport vesicles do not shuttle cargo in an anterograde direction, but instead carry resident Golgi enzymes in a retrograde direction. This model of intra-Golgi transport is supported by electron micrographs of the type illustrated in Figure 8.23*c,d*. These micrographs depict ultrathin sections of cultured mammalian cells that were cut from a frozen block. In both

cases, the frozen sections were treated with antibodies that were linked to gold particles prior to examination in the electron microscope. Figure 8.23*c* shows a section through a Golgi complex after treatment with gold-labeled antibodies that bind to a cargo protein, in this case the viral protein VSVG (page 268). The VSVG molecules are present within the cisternae, but are absent from the nearby vesicles (arrows), indicating that cargo is carried in an anterograde direction within maturing cisternae but not within small transport vesicles. Figure 8.23*d* shows a section through a Golgi complex after treatment with gold-labeled antibodies that bind to a Golgi resident protein, in this case the processing enzyme mannosidase II. Unlike the VSVG cargo protein, mannosidase II molecules are found in both the cisternae and associated vesicles (arrow), which strongly supports the proposal that these vesicles are utilized to carry Golgi resident enzymes in a retrograde direction.

The cisternal maturation model depicted in Figure 8.23*b* explains how different Golgi cisternae in a stack can have a unique identity. An enzyme such as mannosidase II, for example, which removes mannose residues from oligosaccharides and is largely restricted to the *medial* cisternae (Figure 8.21), can be recycled backward in transport vesicles as each cisterna moves toward the *trans* end of the stack. It should be noted that a number of prominent researchers continue to argue, based on other experimental results, that cargo can be carried by transport vesicles between Golgi cisternae in an anterograde direction. Thus, the matter remains to be settled.

REVIEW   ?

1. Describe the steps that occur as a soluble secretory protein, such as a digestive enzyme in a pancreas cell, moves from the RER to the *cis* Golgi cisterna; from the *cis* cisterna to the TGN.

2. What is the role of dolichol phosphate in the synthesis of membrane glycoproteins? How is the sequence of sugars attached to the protein determined?

3. How does the process of glycosylation in the Golgi complex compare to that in the RER?

4. How can the cisternal maturation model of Golgi activity be reconciled with the presence of transport vesicles in the Golgi region?

8.5 TYPES OF VESICLE TRANSPORT AND THEIR FUNCTIONS

The biosynthetic pathway of a eukaryotic cell consists of a series of distinct membrane-bound organelles that function in the synthesis, modification, and delivery of soluble and membrane proteins to their appropriate destination in the cell. As was illustrated in Figure 8.2*a*, materials are carried between compartments by vesicles (or other types of membrane-bound carriers) that bud from donor membranes and fuse with acceptor membranes. If one scrutinizes electron micrographs for vesicles caught in the act of budding, one finds that most of

(a) *(b)* 20 nm
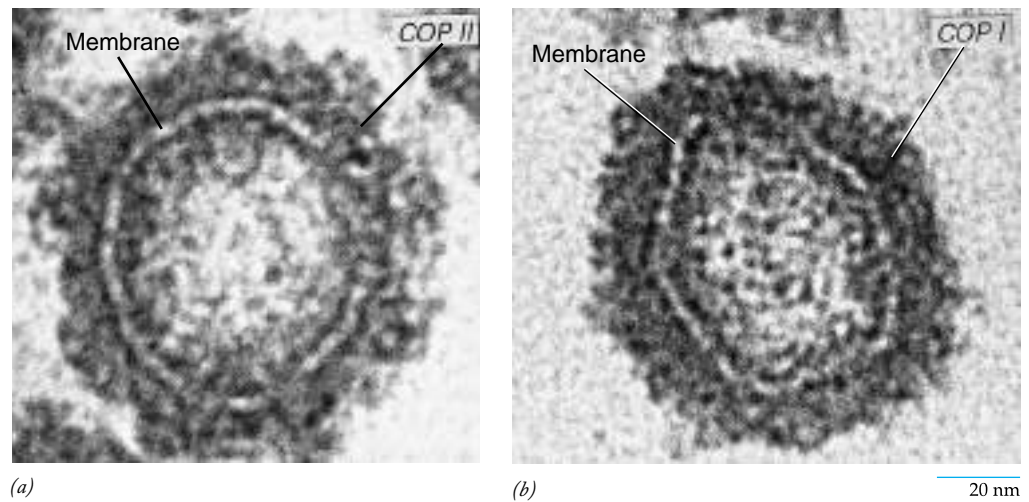
these membranous buds are covered on their cytosolic surface by a "fuzzy," electron-dense layer. Further analysis reveals that the dark-staining layer consists of a protein coat formed from soluble proteins that assemble on the cytosolic surface of the donor membrane at sites where budding takes place. Assembly is initiated by the activation of a small G protein that is specifically recruited to the site. Each coated bud pinches off to form a **coated vesicle** such as those shown in Figure 8.24. Vesicles of similar size and structure can be formed in cell-free systems as illustrated in Figure 8.6. (The discovery of coated vesicles is discussed in the Experimental Pathways on page 312.)

Protein coats have at least two distinct functions: (1) they act as a mechanical device that causes the membrane to curve and form a budding vesicle, and (2) they provide a mechanism for selecting the components to be carried by the vesicle. Selected components include (*a*) cargo consisting of secretory, lysosomal, and membrane proteins to be transported and (*b*) the machinery required to target and dock the vesicle to the correct acceptor membrane (page 296). In the two best understood cases (see Figures 8.26 and 8.40), the vesicle coat is composed of two distinct protein layers: an outer cage or scaffolding that forms the framework for the coat and an inner layer of adaptors that serves primarily to bind the vesicle's cargo. As discussed below, the adaptors are able to select specific cargo molecules by virtue of their specific affinity for the cytosolic "tails" of integral proteins that reside in the donor membrane (see Figure 8.25*b*).

Several distinct classes of coated vesicles have been identified; they are distinguished by the proteins that make up their coat, their appearance in the electron microscope, and their role in cell trafficking. The three best studied coated vesicles are the following:

1. **COPII-coated vesicles** (Figure 8.24*a*) move materials from the ER "forward" to the ERGIC and Golgi complex. (Recall from page 283 that the ERGIC is the intermediate compartment situated between the ER and Golgi complex.) (COP is an acronym for coat proteins.)

2. **COPI-coated vesicles** (Figure 8.24*b*) move materials in a retrograde direction (1) from the ERGIC and Golgi stack "backward" toward the ER and (2) from *trans* Golgi cisternae "backward" to *cis* Golgi cisternae (see Figure 8.25*a*).

3. **Clathrin-coated vesicles** move materials from the TGN to endosomes, lysosomes, and plant vacuoles. They also move materials from the plasma membrane to cytoplasmic compartments along the endocytic pathway. They have also been implicated in trafficking from endosomes and lysosomes.

We will consider each type of coated vesicle in the following sections. A summary of the various transport steps along the biosynthetic or secretory pathway that is mediated by each of these coated vesicles is indicated in Figure 8.25*a*.

COPII-Coated Vesicles: Transporting Cargo from the ER to the Golgi Complex

COPII-coated vesicles mediate the first leg of the journey through the biosynthetic pathway—from the ER to the ERGIC and CGN (Figure 8.25*a,b*). The COPII coat contains a number of proteins that were first identified in mutant yeast cells that were unable to carry out transport from the ER to the Golgi complex. Homologues of the yeast proteins were subsequently found in the coats of vesicles budding from the ER in mammalian cells. Antibodies to COPII-coat proteins block budding of vesicles from ER membranes but have no effect on movement of cargo at other stages in the secretory pathway.

COPII coats select and concentrate certain components for transport in vesicles. Certain integral membrane proteins of the ER are selectively captured because they contain "ER export" signals as part of their cytosolic tail. These signals interact specifically with COPII proteins of the vesicle coat (Figure 8.25*b*). Proteins selected by COPII-coated vesicles include (1) enzymes that act at later stages in the biosynthetic pathway, such as the glycosyltransferases of the Golgi complex (indicated as orange membrane proteins in Figure 8.25*b*), (2) membrane proteins involved in the docking and fusion of the vesicle with the target compartment, and (3) membrane proteins that are able to bind soluble cargo (such as the secretory proteins, indicated by the red spheres in Figure 8.25*b*). Mutations in one of these cargo receptors has been linked to an inherited bleeding disorder. Persons with this disorder fail to secrete certain coagulation factors that promote blood clotting.

Among the COPII coat proteins is a small G protein called Sar1, which is recruited specifically to the ER membrane. Like other G proteins, Sar1 plays a regulatory role, in this case by initiating vesicle formation and by regulating the assembly of the vesicle coat. These activities are illustrated in Figure 8.26. In step 1 of Figure 8.26a, Sar1 is recruited to the ER membrane in the GDP-bound form and is induced to exchange its GDP for a molecule of GTP. Upon binding of GTP, Sar1 undergoes a conformational change that causes its N-terminal α helix to insert itself into the cytosolic leaflet of the ER bilayer (step 2). This event has been demonstrated to bend the lipid bilayer, which is an important step in the conversion of a flattened membrane into a spherical vesicle. Membrane bending is probably aided by a change in packing of the lipids that make up the two leaflets of the bilayer.
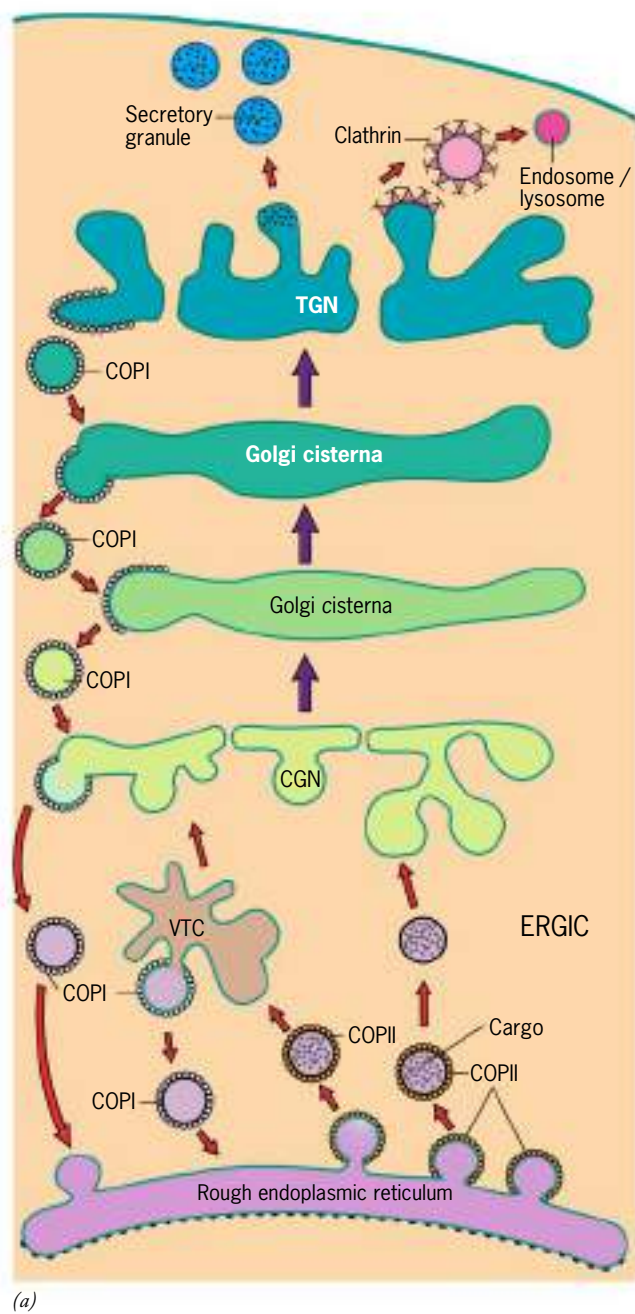
In step 3, Sar1-GTP has recruited two additional polypeptides of the COPII coat, Sec23 and Sec24, which bind as a "banana-shaped" dimer. Because of its curved shape (Figure 8.26b), the Sec23-Sec24 dimer provides additional pressure on the membrane surface to help it further bend into a curved bud. Sec24 also functions as the primary adaptor protein of the COPII coat that interacts specifically with the ER export signals in the cytosolic tails of membrane proteins that are destined to traffic on to the Golgi complex. In Figure 8.26a, step 4, the remaining subunits of the COPII coat, Sec13 and Sec31, bind to the membrane to form the outer structural cage of the protein coat. Figure 8.26b shows a representation of a 40-nm vesicle with the COPII coat bound to its surface. The Sec13-Sec31 cage assembles into a relatively simple lattice in which each vertex is formed by the convergence of four Sec13-Sec31 legs (Figure 8.26b). A certain degree of flexibility is built into the interface between the Sec13-Sec31 subunits that allow them to form cages of varying diameter, thus accommodating vesicles of varying size.

Once the entire COPII coat has assembled, the bud is separated from the ER membrane in the form of a COPII-coated vesicle. Before the coated vesicle can fuse with a target

FIGURE 8.25 **Proposed movement of materials by vesicular transport between membranous compartments of the biosynthetic/secretory pathway.** (a) The three different types of coated vesicles indicated in this schematic drawing are thought to have distinct transport roles. COPII-coated vesicles mediate transport from the ER to the ERGIC and Golgi complex. COPI-coated vesicles return proteins from the ERGIC and Golgi complex to the ER. COPI-coated vesicles also transport Golgi enzymes between cisternae in a retrograde direction. Clathrin-coated vesicles mediate transport from the TGN to endosomes and lysosomes. Transport of materials along the endocytic pathway is not shown in this drawing. (b) Schematic drawing of the assembly of a COPII-coated vesicle. Assembly begins when Sar1 is recruited to the ER membrane and activated by exchange of its bound GDP with a bound GTP. These steps are shown in Figure 8.26. Cargo proteins of the ER lumen (red spheres and diamonds) bind to the luminal ends of transmembrane cargo receptors. These receptors are then concentrated within the coated vesicle through interaction of their cytosolic tails with components of the COPII coat. ER resident proteins (e.g., BiP) are generally excluded from the coated vesicles. Those that do happen to become included in a coated vesicle are returned to the ER as described later in the text. One of the COPII coat proteins, namely Sec24, can exist in at least four different isoforms. It is likely that different isoforms of this protein recognize and bind membrane proteins with different sorting signals, thus broadening the specificity in types of materials that can be transported by COPII vesicles.

(a)
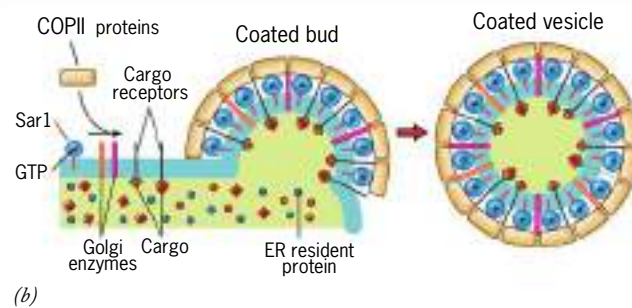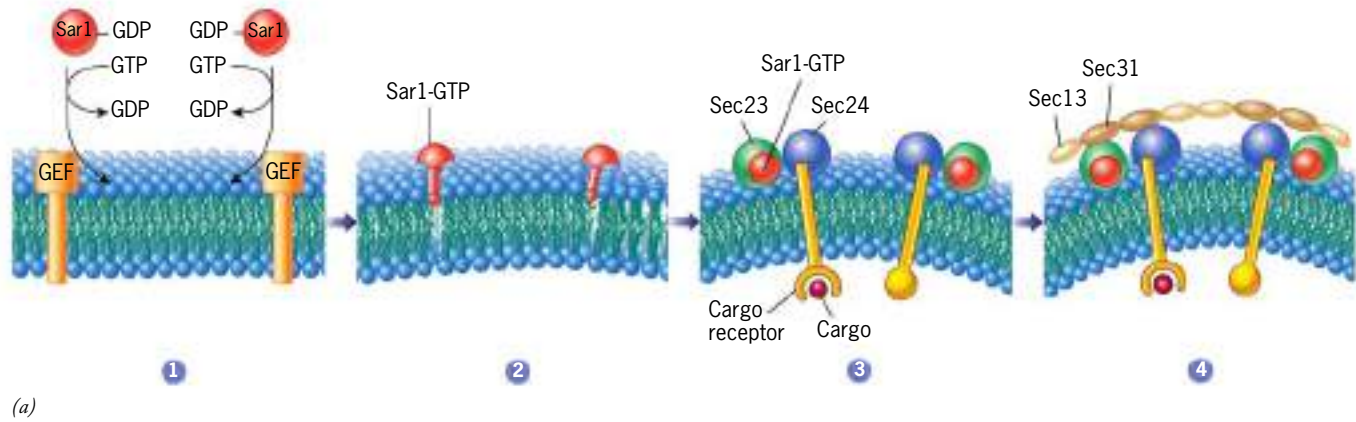
FIGURE 8.26 **Proposed roles of the COPII coat proteins in generating membrane curvature, assembling the protein coat, and capturing cargo.** (*a*) In step 1, Sar1-GDP molecules have been recruited to the ER membrane by a protein called a GEF (guanine-exchange factor) that catalyzes the exchange of the bound GDP with a bound GTP. In step 2, each Sar1-GTP molecule has extended a finger-like α helix along the membrane within the cytosolic leaflet. This event induces the curvature of the lipid bilayer at that site. In step 3, a dimer composed of two COPII polypeptides (Sec23 and Sec24) has been recruited by the bound Sar1-GTP. The Sec23-Sec24 dimer is thought to further induce the curvature of the membrane in the formation of a vesicle. Both Sar1 and Sec23-Sec24 can bring about membrane curvature when incubated with synthetic liposomes in vitro. Transmembrane cargo accumulates within the forming COPII vesicle as their cytosolic tails bind to the Sec24 polypeptide of the COPII coat. In step 4, the remaining COPII polypeptides (Sec13 and Sec31) have joined the complex to form an outer structural scaffold of the coat. (*b*) A molecular model of the outer Sec13-Sec31 cage of the COPII coat as it would assemble around the surface of a 40-nm "vesicle." Each edge or leg of the lattice that makes up the cage consists of a heterotetramer (two Sec31 subunits seen as dark green and light green and two Sec13 subunits seen as orange and red). Four such legs converge to form each vertex of the lattice. Two copies of the Sar1-Sec23-Sec24 complex (shown in red, magenta, and blue, respectively) that would form the inner layer of the COPII coat are also shown in this model. It can be seen how the inner surface of the Sec23-Sec24 complex conforms to the curvature of the vesicle. (B: FROM STEPHAN FATH ET AL., COURTESY OF JONATHAN GOLDBERG, CELL 129:1333, 2007, BY PERMISSION OF CELL PRESS.)

(b)

membrane, the protein coat must be disassembled and its components released into the cytosol. Disassembly is triggered by hydrolysis of the bound GTP to produce a Sar1-GDP subunit, which has decreased affinity for the vesicle membrane. Dissociation of Sar1-GDP from the membrane is followed by the release of the other COPII subunits.

COPI-Coated Vesicles: Transporting Escaped Proteins Back to the ER

COPI-coated vesicles were first identified in experiments in which cells were treated with molecules similar in structure to GTP (GTP analogues) but, unlike GTP, cannot be hydrolyzed. Under these conditions, COPI-coated vesicles accumulated within the cell (Figure 8.27) and could be isolated from homogenized cells by density gradient centrifugation
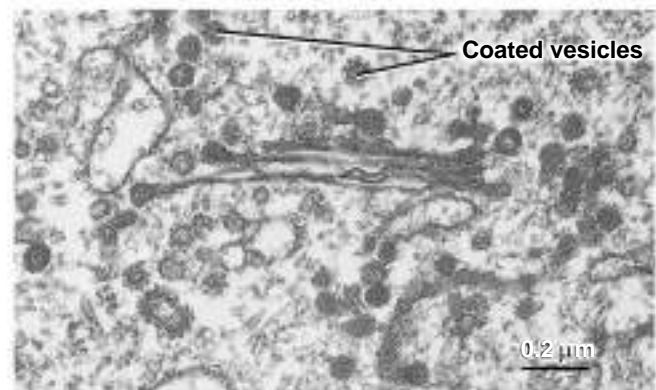
FIGURE 8.27 **Accumulation of COPI-coated vesicles.** This micrograph shows the Golgi complex of a permeabilized cell that had been treated with a nonhydrolyzable GTP analogue (GTPγS). Numerous COPI-coated vesicles and coated buds are visible. The COPI coat contains a preassembled complex of seven distinct proteins in addition to the GTP-binding protein ARF1. (FROM JAMES E. ROTHMAN AND LELIO ORCI, FASEB J. 4:1467, 1990.)

(Section 18.6). COPI-coated vesicles accumulate in the presence of a nonhydrolyzable GTP analogue because, similar to their COPII counterparts, the coat contains a small GTP-binding protein, called ARF1, whose bound GTP must be hydrolyzed before the coat can disassemble. COPI-coated vesicles have been most clearly implicated in the retrograde transport of proteins, including the movement of (1) Golgi resident enzymes in a *trans*-to-*cis* direction (as indicated in Figure 8.23*d*, which shows a gold-labeled mannosidase II molecule in a COPI vesicle) and (2) ER resident enzymes from the ERGIC and the Golgi complex back to the ER (Figure 8.25*a*). To understand the role of COPI-coated vesicles in retrograde transport, we have to consider a more general subject.

Retaining and Retrieving Resident ER Proteins If vesicles continually bud from membrane compartments, how does each compartment retain its unique composition? What determines, for example, whether a particular protein in the membrane of the ER remains in the ER or proceeds on to the Golgi complex? Studies suggest that proteins are maintained in an organelle by a combination of two mechanisms:

1. *Retention* of resident molecules that are excluded from transport vesicles. Retention may be based primarily on the physical properties of the protein. For example, soluble proteins that are part of large complexes or membrane proteins with short transmembrane domains are not likely to enter a transport vesicle.

2. *Retrieval* of "escaped" molecules back to the compartment in which they normally reside.

Proteins that normally reside in the ER, those both in the lumen and in the membrane, contain short amino acid sequences at their C-terminus that serve as *retrieval signals*, ensuring their return to the ER if they should be accidentally carried forward to the ERGIC or Golgi complex. The retrieval of "escaped" ER proteins from these compartments is accomplished by specific receptors that capture the molecules and return them to the ER in COPI-coated vesicles (Figures 8.25*a*, 8.28). Soluble resident proteins of the ER lumen (such as protein disulfide isomerase and the molecular chaperones that facilitate folding) typically possess the retrieval signal "lys-asp-glu-leu" (or KDEL in single-letter nomenclature). As shown in Figure 8.28, these proteins are recognized and returned to the ER by the *KDEL receptor*, an integral membrane protein that shuttles between the *cis* Golgi and the ER compartments. If the KDEL sequence is deleted from an ER protein, the escaped proteins are not returned to the ER but instead are carried forward through the Golgi complex. Conversely, when a cell is genetically engineered to express a lysosomal or secretory protein that contains an added KDEL C-terminus, that protein is returned to the ER rather than being sent on to its proper destination. Membrane proteins that reside in the ER also have a retrieval signal that binds to the COPI coat, facilitating their return to the ER. The most common retrieval sequences for ER membrane proteins involve two closely linked basic residues, most commonly
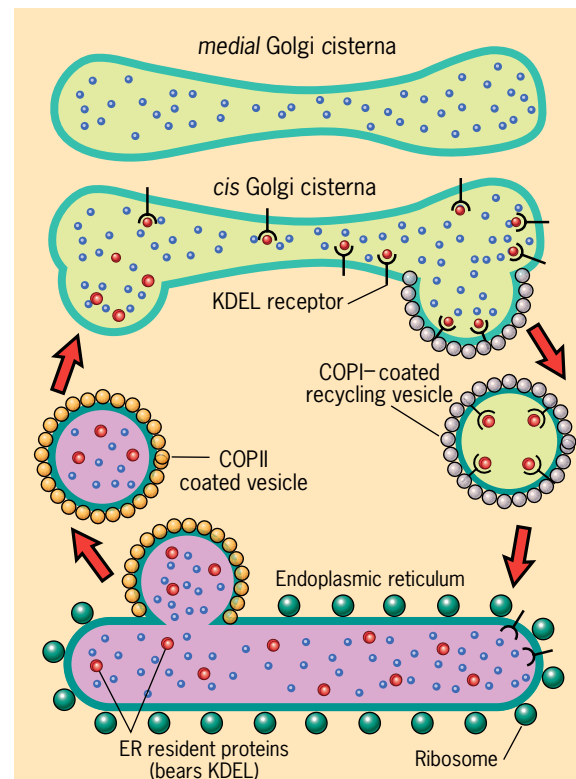
FIGURE 8.28 **Retrieving ER proteins.** Resident proteins of the ER contain amino acid sequences that lead to their retrieval from the Golgi complex if they are accidentally incorporated into a Golgi-bound transport vesicle. Soluble ER proteins bear the retrieval signal KDEL. Retrieval is accomplished as soluble ER proteins bind to KDEL receptors residing in the membranous wall of *cis* Golgi compartments. The KDEL receptors, in turn, bind to proteins of the COPI coat, which allows the entire complex to be recycled back to the ER.

KKXX (where K is lysine and X is any residue). Each membrane compartment in the biosynthetic pathway may have its own retrieval signals, which helps explain how each compartment can maintain its unique complement of proteins despite the constant movement of vesicles in and out of that compartment.

Beyond the Golgi Complex: Sorting Proteins at the TGN

Despite all the discussion of transport vesicles, we have yet to examine how a particular protein that has been synthesized in the ER is targeted toward a particular cellular destination. It is important that a cell be able to distinguish among the various proteins that it manufactures. A pancreatic cell, for example, has to segregate newly synthesized digestive enzymes that will be secreted into a duct, from newly synthesized cell-adhesion molecules that will ultimately reside in the plasma membrane, from lysosomal enzymes that are destined for lysosomes. This is accomplished as the cell sorts proteins destined for different sites into different membrane-bound carriers. The *trans* Golgi network (TGN), which is the last stop in the Golgi complex,
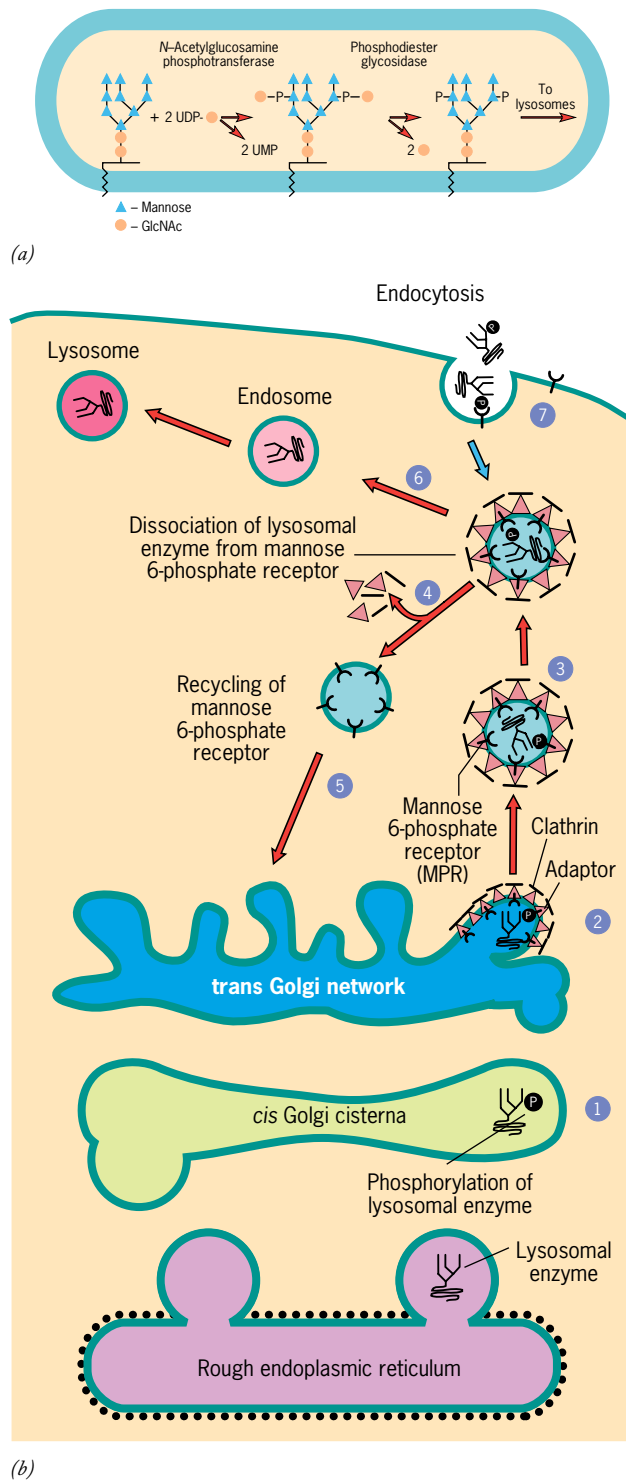
FIGURE 8.29 Targeting lysosomal enzymes to lysosomes. (*a*) Lysosomal enzymes are recognized by an enzyme in the *cis* cisternae that transfers a phosphorylated *N*-acetylglucosamine from a nucleotide sugar donor to one or more mannose residues of *N*-linked oligosaccharides. The glucosamine moiety is then removed in a second step by a second enzyme, leaving mannose 6-phosphate residues as part of the oligosaccharide chain. (*b*) Schematic diagram showing the pathways followed by a lysosomal enzyme (black) from its site of synthesis in the ER to its delivery to a lysosome. The mannose residues of the lysosomal enzyme are phosphorylated in the Golgi cisternae (step 1) and then selectively incorporated into a clathrin-coated vesicle at the TGN (step 2). The mannose 6-phosphate receptors are thought to have a dual role (step 3): they interact specifically with the lysosomal enzymes on the luminal side of the vesicle, and they interact specifically with adaptors on the cytosolic surface of the vesicle (shown in Figure 8.30). The mannose 6-phosphate receptors separate from the enzymes (step 4) and are returned to the Golgi complex (step 5). The lysosomal enzymes are delivered to an endosome (step 6) and eventually to a lysosome. Mannose 6-phosphate receptors are also present in the plasma membrane, where they capture lysosomal enzymes that are secreted into the extracellular space and return the enzymes to a pathway that directs them to a lysosome (step 7).

mal enzymes are specifically recognized by enzymes that catalyze the two-step addition of a phosphate group to certain mannose sugars of the *N*-linked carbohydrate chains (Figure 8.29*a*). Thus, unlike other glycoproteins sorted at the TGN, lysosomal enzymes possess phosphorylated mannose residues, which act as sorting signals. This mechanism of protein sorting was discovered through studies on cells from humans that lacked one of the enzymes involved in phosphate addition (discussed in the Human Perspective on page 299). Lysosomal enzymes carrying a mannose 6-phosphate signal are recognized and captured by *mannose 6-phosphate receptors* (*MPRs*), which are integral membrane proteins that span the TGN membranes (Figure 8.29*b*).

Lysosomal enzymes are transported from the TGN in clathrin-coated vesicles (which are the third and last type of coated vesicles to be discussed). The structure of clathrin-coated vesicles is described in detail on page 303 in connection with endocytosis, a process that is better understood than budding at the TGN. It will suffice at this point to note that the coats of these vesicles contain (1) an outer honeycomb-like lattice composed of the protein clathrin, which forms a structural scaffold, and (2) an inner shell composed of protein adaptors, which covers the surface of the vesicle membrane that faces the cytosol (Figure 8.30). The term *adaptor* describes a molecule that physically links two different components. Lysosomal enzymes are escorted from the TGN by a family of adaptor proteins called **GGAs**.

As indicated in the inset of Figure 8.30, a GGA molecule has several domains, each capable of grasping a different protein involved in vesicle formation. The outer ends of the GGA adaptors bind to clathrin molecules, holding the clathrin scaffolding onto the surface of the vesicle. On their inner surface, the GGA adaptors bind to a sorting signal in the cytosolic tails of the mannose 6-phosphate receptors. The MPRs, in

functions as a major sorting station, directing proteins to various destinations. The best understood of the post-Golgi pathways is one that carries lysosomal enzymes.

Sorting and Transport of Lysosomal Enzymes Lysosomal proteins are synthesized on membrane-bound ribosomes of the ER and carried to the Golgi complex along with other types of proteins. Once in the Golgi cisternae, soluble lysoso-
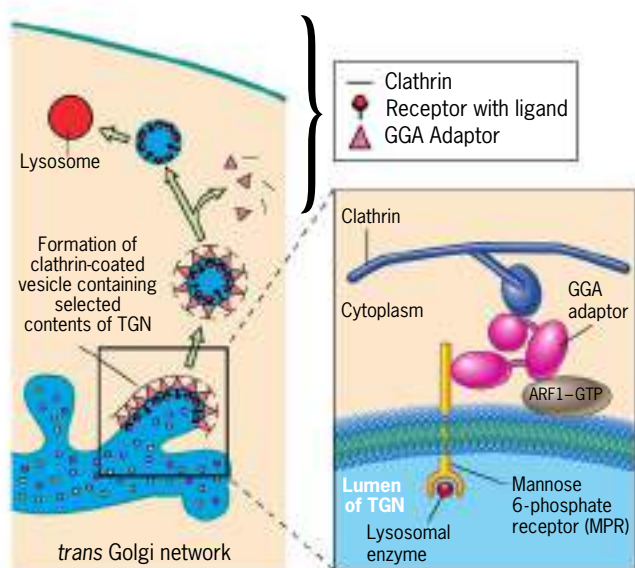
FIGURE 8.30 **The formation of clathrin-coated vesicles at the TGN.** Clathrin-coated vesicles that bud from the TGN contain GGA, an adaptor protein consisting of several distinct domains. One of the GGA domains binds to the cytosolic domains of membrane proteins, including those that will ultimately reside in the boundary membrane of the lysosome and also the MPR that transports lysosomal enzymes. Other GGA domains bind to ARF1 and to the surrounding cytosolic network of clathrin molecules.

turn, bind to soluble lysosomal enzymes within the vesicle lumen (Figure 8.30, inset). As a result of these interactions with GGA adaptors, the MPRs in the TGN membrane and lysosomal enzymes within the TGN lumen become concentrated into clathrin-coated vesicles. As in the formation of COPI and COPII vesicles, production of clathrin-coated vesicles begins with the recruitment to the membrane of a small GTP-binding protein, in this case ARF1, which sets the stage for the binding of the other coat proteins. After the vesicle has budded from the TGN, the clathrin coat is lost and the uncoated vesicle proceeds to its destination, which may be an early endosome, late endosome, or plant vacuole. Before they reach one of these organelles, the MPRs dissociate from the lysosomal enzymes and return to the TGN (Figure 8.29b) for another round of lysosomal enzyme transport.

Sorting and Transport of Nonlysosomal Proteins Lysosomal proteins are not the only materials that are exported from the TGN. As indicated in Figure 8.2, membrane proteins destined for the plasma membrane and secretory materials destined for export from the cell are also transported from the TGN, but the mechanisms are poorly understood. According to one model, membranous carriers are produced as the TGN fragments into vesicles and tubules of various size. This concept fits with the cisternal maturation model, which proposes that the cisternae of the Golgi complex move continually toward the TGN, where they would have to disperse to allow continued maturation of the Golgi stack. Proteins

that are discharged from the cell by a process of regulated secretion, such as digestive enzymes and hormones, are thought to form selective aggregates that eventually become contained in large, densely packed secretory granules. These aggregates are apparently trapped as immature secretory granules bud from the rims of the *trans* Golgi cisternae and TGN. In some cells, long tubules are seen to be pulled out of the TGN by motor proteins that operate along microtubular tracks. These tubules are then split into a number of vesicles or granules by membrane fission. Once they have departed from the TGN, the contents of the secretory granules become more concentrated. Eventually, the mature granules are stored in the cytoplasm until their contents are released following stimulation of the cell by a hormone or nerve impulse.

The targeted delivery of integral proteins to the plasma membrane appears to be based largely on sorting signals in the cytoplasmic domains of the membrane proteins. Considerable research has focused on polarized cells, such as those depicted in Figure 8.11. In these cells, membrane proteins destined to reside in the apical portion of the plasma membrane contain different sorting signals from those destined for the lateral or basal portion. Plasma membrane proteins of nonpolarized cells, such as fibroblasts and white blood cells, may not require special sorting signals. Such proteins may simply be carried from the TGN to the cell surface in vesicles of the constitutive secretory pathway (Figure 8.2b).

Targeting Vesicles to a Particular Compartment

Vesicle fusion requires specific interactions between different membranes. Vesicles from the ER, for example, fuse with the ERGIC or *cis* Golgi network and not with a *trans* cisterna. Selective fusion is one of the factors that ensures a directed flow through the membranous compartments of the cell. Despite a major research effort, we still do not fully understand the mechanisms by which cells target vesicles to particular compartments. It is thought that a vesicle contains specific proteins associated with its membrane that govern the movements and fusion potential of that vesicle. To understand the nature of these proteins, we will consider the steps that occur between the stages of vesicle budding and vesicle fusion.

1. *Movement of the vesicle toward the specific target compartment.* In many cases, membranous vesicles must move considerable distances through the cytoplasm before reaching their eventual target. These types of movement are mediated largely by microtubules, which act like railroad tracks carrying cargo containers along a defined pathway to a predetermined destination. The membranous carriers seen in Figure 8.19, for example, were observed to move from the ERGIC to the Golgi complex over microtubules.

2. *Tethering vesicles to the target compartment.* The initial contacts between a transport vesicle and its target membrane, such as a Golgi cisterna, are thought to be mediated by so-called tethering proteins. Two groups of tethering proteins have been described (Figure 8.31a): rod-shaped, fibrous proteins that are capable of forming a molecular

(a) Tethering

(b) Docking

(c)
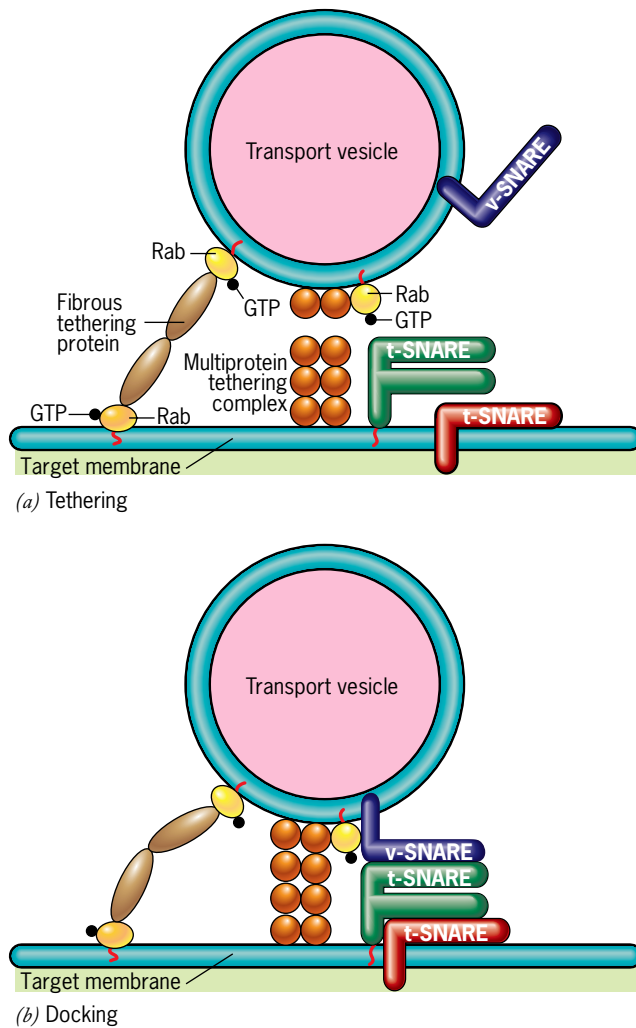
FIGURE 8.31 Proposed steps in the targeting of transport vesicles to target membranes. (*a*) According to this model, Rab proteins on the vesicle and target membrane are involved in recruiting tethering proteins that mediate initial contact between the two membranes. Two types of tethering proteins are depicted: highly elongated fibrous proteins (e.g., p115 and EEA1) and multiprotein complexes (e.g., the exocyst and TRAPPI). (*b*) During the docking stage leading up to membrane fusion, a v-SNARE in the vesicle membrane interacts with the t-SNAREs in the target membrane to form a four-stranded α-helical bundle that brings the two membranes into intimate contact (see next figure). In the cases described in the text, SNAP-25, one of the t-SNAREs, is a peripheral membrane protein that is bound to the lipid bilayer by a lipid anchor rather than a transmembrane domain. SNAP-25 contributes two helices to the four-helix SNARE bundle. (*c*) A model of a synaptic vesicle showing the distribution of only one of its constitutent proteins, the SNARE synaptobrevin. This surface density and structures of the synaptobrevin molecules are based on calculations of the number of these proteins per vesicle and the known structure of the molecule. A complete portrait of the proteins on the surface of a synaptic vesicle is shown in the image on the book cover. (C: FROM SHIGEO TAKAMORI ET AL., CELL 127:841, 2006, COURTESY OF REINHARD JAHN, BY PERMISSION OF CELL PRESS.)
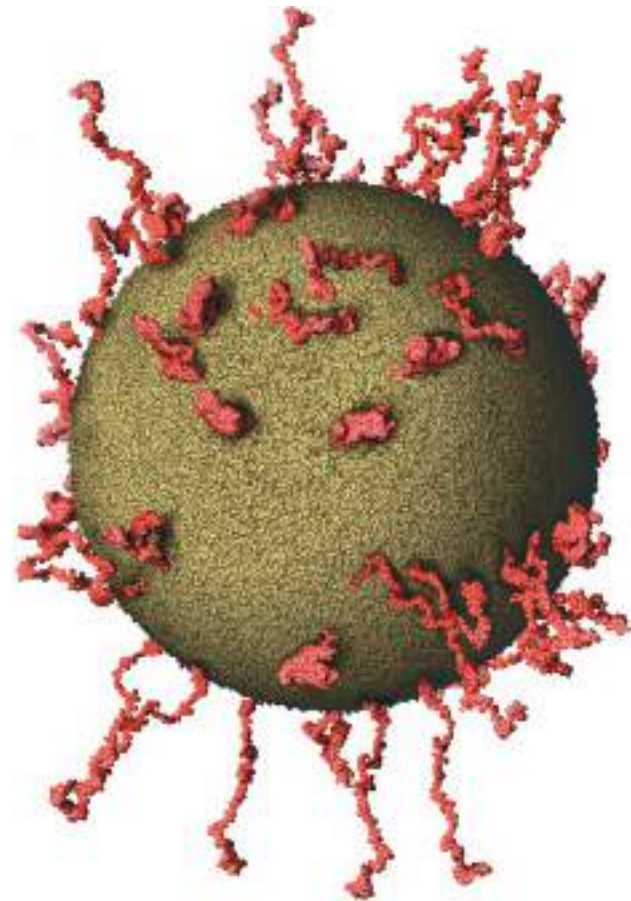
bridge between the two membranes over a considerable distance (50–200 nm) and large multiprotein complexes that appear to hold the two membranes in closer proximity. It is hypothesized that tethering is an early stage in the process of vesicle fusion that requires specificity between the vesicle and target compartment. Much of this specificity may be conferred by a family of small G proteins called **Rabs**, which cycle between an active GTP-bound state and an inactive GDP-bound state. GTP-bound Rabs associate with membranes by a lipid anchor. With over 60 different Rab genes identified in humans, these proteins constitute the most diverse group of proteins involved in

membrane trafficking. More importantly, different Rabs become associated with different membrane compartments. This preferential localization gives each compartment a unique surface identity, which is required to recruit the proteins involved in targeting specificity. In their GTP-bound state, Rabs play a key role in vesicle targeting by recruiting specific cytosolic tethering proteins to specific membrane surfaces (Figure 8.31*a*). Rabs also play a key role in regulating the activities of numerous proteins involved in other aspects of membrane trafficking, including the motor proteins that move membranous vesicles through the cytoplasm (shown in Figure 9.52*b*).

3. ***Docking vesicles to the target compartment.*** At some point during the process leading to vesicle fusion, the membranes of the vesicle and target compartment come into close contact with one another as the result of an interaction between the cytosolic regions of integral proteins of the two membranes. The key proteins that engage in these interactions are called **SNAREs**, and they constitute a family of more than 35 membrane proteins whose members are localized to specific subcellular compartments. Although SNAREs vary considerably in structure and size, all contain a segment in their cytosolic domain called a *SNARE motif* that consists of 60–70 amino acids capable of forming a complex with another SNARE motif. SNAREs can be divided functionally into two categories, **v-SNAREs**, which become incorporated into the membranes of transport vesicles during budding, and **t-SNAREs**, which are located in the membranes of target compartments (Figure 8.31*b*). The best studied SNAREs are those that mediate docking of synaptic vesicles with the presynaptic membrane during the regulated release of neurotransmitters (page 163). In this case, the plasma membrane of the nerve cell contains two t-SNAREs, syntaxin and SNAP-25, whereas the synaptic vesicle membrane contains a single v-SNARE, synaptobrevin.

The distribution of synaptobrevin molecules on a representative synaptic vesicle is shown in Figure 8.31*c*. As the synaptic vesicle and presynaptic membrane approach one another, the SNARE motifs of t- and v-SNARE molecules from apposing membranes interact to form four-stranded bundles as shown in Figure 8.32*a*. Each bundle consists of four α helices, two donated by SNAP-25 and one each donated by syntaxin and synaptobrevin. These parallel α helices zip together to form a tightly interwoven complex that pulls the two apposing lipid bilayers into very close association (Figures 8.31*b* and 8.32*a*). The formation of similar four-stranded helical bundles occurs among other SNAREs at other sites throughout the cell, wherever membranes are destined to fuse. It is interesting to note that the SNAREs of the synaptic vesicle and presynaptic membrane are the targets of two of the most potent bacterial toxins; those responsible for botulism and tetanus. These deadly toxins act as proteases whose only known substrates are SNAREs. Cleavage of the neuronal SNAREs by the toxins blocks the release of neurotransmitters, which causes paralysis.

4. ***Fusion between vesicle and target membranes.*** When artificial lipid vesicles (liposomes) containing purified t-SNAREs are mixed with liposomes containing a puri-

(a)

(b)

(c)

(d)
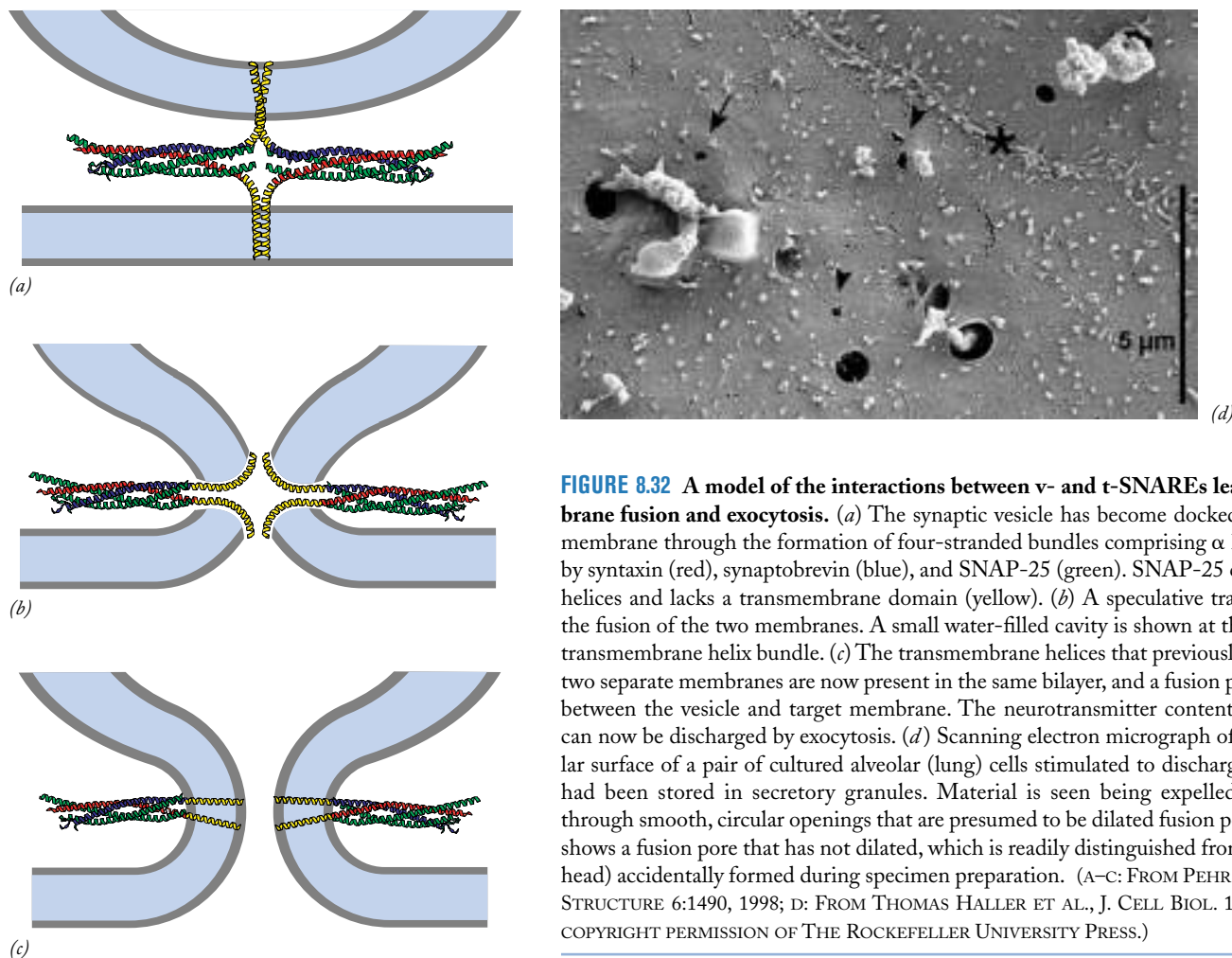
FIGURE 8.32 **A model of the interactions between v- and t-SNAREs leading to membrane fusion and exocytosis.** (*a*) The synaptic vesicle has become docked to the plasma membrane through the formation of four-stranded bundles comprising α helices donated by syntaxin (red), synaptobrevin (blue), and SNAP-25 (green). SNAP-25 contributes two helices and lacks a transmembrane domain (yellow). (*b*) A speculative transition state in the fusion of the two membranes. A small water-filled cavity is shown at the center of the transmembrane helix bundle. (*c*) The transmembrane helices that previously resided in the two separate membranes are now present in the same bilayer, and a fusion pore has opened between the vesicle and target membrane. The neurotransmitter contents of the vesicle can now be discharged by exocytosis. (*d*) Scanning electron micrograph of the extracellular surface of a pair of cultured alveolar (lung) cells stimulated to discharge proteins that had been stored in secretory granules. Material is seen being expelled from the cell through smooth, circular openings that are presumed to be dilated fusion pores. The arrow shows a fusion pore that has not dilated, which is readily distinguished from holes (arrowhead) accidentally formed during specimen preparation. (A–C: FROM PEHR A. B. HARBURY, STRUCTURE 6:1490, 1998; D: FROM THOMAS HALLER ET AL., J. CELL BIOL. 155:286, 2001; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

fied v-SNARE, the two types of vesicles fuse with one another, but not with themselves. This finding indicates that interactions between t- and v-SNAREs are capable of pulling two lipid bilayers together with sufficient force to cause them to fuse (Figure 8.32*b,c*). However, a large body of evidence suggests that, although interaction between v- and t-SNAREs is required for membrane fusion, it is not sufficient by itself to bring about fusion *within a cell*. According to the prevailing view regarding the regulated secretion of neurotransmitter molecules, the four-stranded SNARE bundle remains locked in an inactive conformation by interaction with accessory proteins. Vesicles at this stage remain docked at the membrane and ready to discharge their contents almost instantaneously once they receive an activating signal, in the form of a rise in Ca^{2+} concentration (as discussed below). Regardless of how it is regulated, once the lipid bilayers of the two membranes merge, the SNAREs that previously projected from separate membranes now reside in the same membrane (Figure 8.32*c*). Dissociation of the four-stranded SNARE complex is achieved by a doughnut-shaped, cytosolic protein called NSF that attaches to the SNARE bundle and, using energy from ATP hydrolysis, twists it apart.

Now that we have described the events that occur during the fusion of a vesicle with a target membrane, we can return to the question: How is the specificity of this interaction determined? According to current consensus, the ability of a particular vesicle and target membrane to fuse is determined by the specific combination of interacting proteins, including tethering proteins, Rabs, and SNAREs that can be assembled at that site in the cell. Taken together, these multiple interactions between several types of proteins provide a high level of specificity, ensuring that each membrane compartment can be selectively recognized.

Exocytosis The fusion of a secretory vesicle or secretory granule with the plasma membrane and subsequent discharge of its contents is called **exocytosis**. Exocytosis probably occurs on a rather continual basis in most cells, as proteins and other materials are delivered to both the plasma membrane and extracellular space. However, the best studied examples of exocytosis are those that occur during regulated secretion, most notably the release of neurotransmitters into the synaptic cleft. In these cases, membrane fusion produces an opening through which the contents of the vesicle or granule are released into the extracellular space. It was noted on page 164 that the arrival of a nerve impulse at the terminal knob of a neuron leads to an increase in the influx of Ca^{2+} and the subsequent discharge of neurotransmitter molecules by exocytosis. In this case, fusion is regulated by a calcium-binding protein (synaptotagmin) present in the membrane of the synaptic vesicle. In other types of cells, exocytosis is generally triggered by release of Ca^{2+} from cytoplasmic stores. Contact between the vesicle and plasma membranes is thought to lead to the formation of a small, protein-lined "fusion pore" (Figure 8.32*c*). Some fusion pores may simply reclose, but in

most cases, the pore rapidly dilates to form an opening for discharge of the vesicle contents (Figure 8.32*d*). Regardless of the mechanism, when a cytoplasmic vesicle fuses with the plasma membrane, the luminal surface of the vesicle membrane becomes part of the outer surface of the plasma membrane, whereas the cytosolic surface of the vesicle membrane becomes part of the inner (cytosolic) surface of the plasma membrane (Figure 8.14).

8.6 LYSOSOMES

Lysosomes are an animal cell's digestive organelles. A typical lysosome contains at least 50 different hydrolytic enzymes (Table 8.1) produced in the rough ER and targeted to these organelles. Taken together, lysosomal enzymes can hydrolyze virtually every type of biological macromolecule. The enzymes of a lysosome share an important property: all have their optimal activity at an acid pH and thus are **acid hydrolases**. The pH optimum of these enzymes is suited to the low pH of the lysosomal compartment, which is approximately 4.6. The high internal proton concentration is maintained by a proton pump (an H^+-ATPase) present in the organelle's boundary membrane. Lysosomal membranes contain a variety of highly glycosylated integral proteins whose carbohydrate chains are thought to form a protective lining that shields the membrane from attack by the enclosed enzymes.

Although lysosomes house a predictable collection of enzymes, their appearance in electron micrographs is neither distinctive nor uniform. Figure 8.33 shows a small portion of a Kupffer cell, a phagocytic cell in the liver that engulfs aging red blood cells. The lysosomes of a Kupffer cell exhibit an irregular shape and variable electron density, illustrating how difficult it is to identify these organelles on the basis of morphology alone.

The presence within a cell of what is, in essence, a bag of destructive enzymes suggests a number of possible functions. The best studied role of lysosomes is the breakdown of materials brought into the cell from the extracellular environment. Many single-celled organisms ingest food particles, which are then enzymatically disassembled in a lysosome. The resulting

TABLE 8.1 A Sampling of Lysosomal Enzymes

Enzyme	Substrate
Phosphatases	
Acid phosphatase	Phosphomonoesters
Acid phosphodiesterase	Phosphodiesters
Nucleases	
Acid ribonuclease	RNA
Acid deoxyribonuclease	DNA
Proteases	
Cathepsin	Proteins
Collagenase	Collagen
GAG-hydrolyzing enzymes	
Iduronate sulfatase	Dermatan sulfate
β-Galactosidase	Keratan sulfate
Heparan N-sulfatase	Heparan sulfate
α-N-Acetylglucosaminidase	Heparan sulfate
Polysaccharidases and oligosaccharidases	
α-Glucosidase	Glycogen
Fucosidase	Fucosyloligosaccharides
α-Mannosidase	Mannosyloligosaccharides
Sialidase	Sialyloligosaccharides
Sphingolipid hydrolyzing enzymes	
Ceramidase	Ceramide
Glucocerebrosidase	Glucosylceramide
β-Hexosaminidase	G_{M2} ganglioside
Arylsulfatase A	Galactosylsulfatide
Lipid hydrolyzing enzymes	
Acid lipase	Triacylglycerols
Phospholipase	Phospholipids

0.3 µm

FIGURE 8.33 Lysosomes. Portion of a phagocytic Kupffer cell of the liver showing at least 10 lysosomes of highly variable size. (FROM HANS GLAUMANN ET AL., J. CELL BIOL. 67:887, 1975; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

FIGURE 8.34 Autophagy. Electron micrograph of a mitochondrion and peroxisome enclosed in a double membrane wrapper. This autophagic vacuole (or autophagosome) would have fused with a lysosome and its contents digested. (FROM DON FAWCETT AND DANIEL FRIEND/PHOTO RESEARCHERS.)

nutrients pass through the lysosomal membrane into the cytosol. In mammals, phagocytic cells, such as macrophages and neutrophils, function as scavengers that ingest debris and potentially dangerous microorganisms (page 308). Ingested bacteria are generally inactivated by the low pH of the lysosome and then digested enzymatically. As shown in Figure 17.24, peptides produced by this digestive process are "posted" on the cell surface where they alert the immune system to the presence of a foreign agent.

Lysosomes also play a key role in organelle **turnover**, that is, the regulated destruction of the cell's own organelles and their replacement. During this process, which is called **autophagy**, an organelle, such as the mitochondrion shown in Figure 8.34, is surrounded by a double membrane to produce a structure called an *autophagosome*. The outer membrane then fuses with a lysosome to produce an *autophagolysosome* in which the enclosed organelle is degraded and the breakdown products are made available to the cell. It is calculated that one mitochondrion undergoes autophagy every 10 minutes or so

in a mammalian liver cell. If a cell is deprived of nutrients, a marked increase in autophagy is observed. Under these conditions, the cell acquires energy to maintain its life by cannibal-
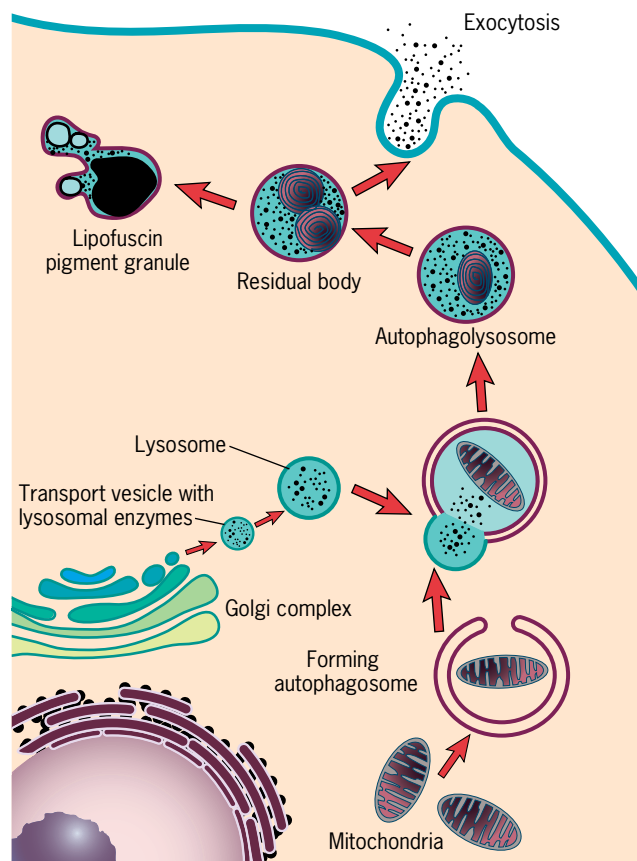
FIGURE 8.35 A summary of the autophagic pathway. The steps are described in the text.

izing its own organelles. In recent years, autophagy has also been shown to help protect an organism against intracellular threats ranging from abnormal protein aggregates to invading bacteria. If autophagy is blocked in a particular portion of the brain of a laboratory animal, that region of the nervous system experiences massive loss of nerve cells. These findings reveal the importance of autophagy in protecting brain cells from the continuous damage to proteins and organelles that is experienced by these long-lived cells.

The role of lysosomes in autophagy is summarized in Figure 8.35. Once the digestive process in the autophagolysosome has been completed, the organelle is termed a *residual body*. Depending on the type of cell, the contents of the residual body may be eliminated from the cell by exocytosis, or they may be retained within the cytoplasm indefinitely as a *lipofuscin granule*. Lipofuscin granules increase in number as an individual becomes older; accumulation is particularly evident in long-lived cells such as neurons, where these granules are considered a major characteristic of the aging process. The role of lysosomes in various diseases is discussed in the accompanying Human Perspective.

REVIEW

1. Describe three distinct roles of lysosomes.
2. Describe the events that occur during the autophagic destruction of a mitochondrion.

THE HUMAN PERSPECTIVE

Disorders Resulting from Defects in Lysosomal Function

Our understanding of the mechanisms by which proteins are targeted to particular organelles began with the discovery that the mannose 6-phosphate residues in lysosomal enzymes act as an "address" for delivery of these proteins to lysosomes. The discovery of the lysosome address was made in studies of patients with a rare and fatal inherited condition known as *I-cell disease*. Many cells in these patients contain lysosomes that are bloated with undegraded materials. Materials accumulate in the lysosomes because of the absence of hydrolytic enzymes. When fibroblasts from these patients were studied in culture, it was found that lysosomal enzymes are synthesized at normal levels but are secreted into the medium and not targeted to lysosomes. Further analysis revealed that the secreted enzymes lacked the mannose phosphate residues that are present on the lysosomal enzymes of cells from normal individuals. The I-cell defect was soon traced to the deficiency of an enzyme (*N*-acetylglucosamine phosphotransferase) required for mannose phosphorylation in the Golgi complex (see Figure 8.29*a*).

In 1965, H. G. Hers of the University of Louvain in Belgium offered an explanation as to how the absence of a seemingly unim-

portant lysosomal enzyme, α-glucosidase, could lead to the development of a fatal inherited condition known as Pompe disease. Hers suggested that, in the absence of α-glucosidase, undigested glycogen accumulated in lysosomes, causing swelling of the organelles and irreversible damage to the cells and tissues. Diseases of this type, characterized by the deficiency of a single lysosomal enzyme and the corresponding accumulation of undegraded substrate (Figure 1), are called **lysosomal storage disorders**. Over 40 such diseases have been described, affecting approximately 1 in 8000 infants. Those diseases resulting from an accumulation of undegraded sphingolipids are listed in Table 1. The symptoms of lysosomal storage diseases can range from very severe to barely detectable, depending primarily on the degree of enzyme dysfunction. Several diseases have also been traced to mutations in lysosomal membrane proteins that impair the transport of substances to the cytosol.

Among the best studied lysosomal storage diseases is Tay-Sachs disease, which results from a deficiency of the enzyme β-*N*-hexosaminidase A, an enzyme that degrades the ganglioside G_{M2} (Figure 4.6). G_{M2} is a major component of the membranes of
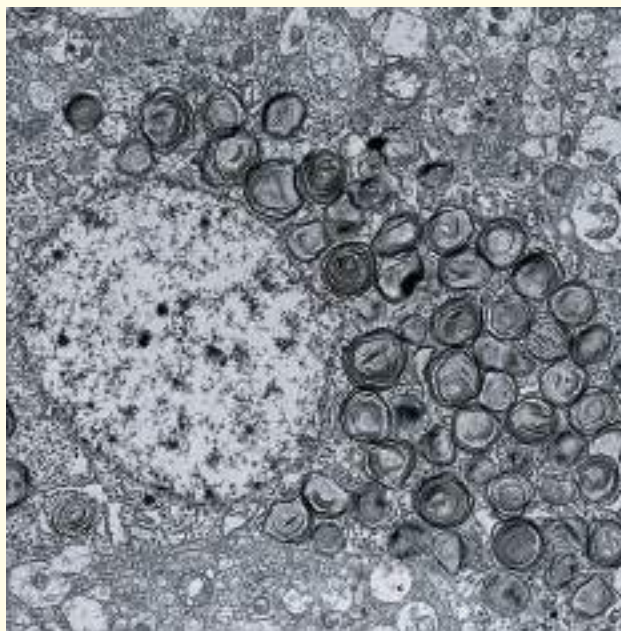
FIGURE 1 Lysosomal storage disorders. Electron micrograph of a section through a portion of a neuron of a person with a lysosomal storage disease characterized by an inability to degrade G_{M2} gangliosides. These cytoplasmic vacuoles stain for both lysosomal enzymes and the ganglioside, indicating they are lysosomes in which undigested glycolipids have accumulated. (COURTESY OF KINUKO SUZUKI.)

brain cells, and in the absence of the hydrolytic enzyme, the ganglioside accumulates in the bloated lysosomes of brain cells (Figure 1), causing dysfunction. In its severe form, which strikes during infancy, the disease is characterized by progressive mental and motor retardation, as well as skeletal, cardiac, and respiratory abnormalities. The disease is very rare in the general population but reached an incidence up to 1 in 3600 newborns among Jews of eastern European ancestry. The incidence of the disease has dropped dramatically in this ethnic population in recent years as the result of identification of carriers, genetic counseling of parents at risk, and prenatal diagnosis by amniocentesis. In fact, all of the known lysosomal storage diseases can be diagnosed prenatally.

In the past few years, the prospects for treatment of lysosomal storage diseases have improved with the demonstration that the symptoms of Gaucher's disease, a deficiency of the lysosomal enzyme glucocerebrosidase, can be alleviated by *enzyme replacement therapy*. Infants with Gaucher's disease accumulate large quantities of glucocerebroside lipids in the lysosomes of their macrophages, causing spleen enlargement and anemia. Initial attempts to correct the disease by infusing a solution of the normal human enzyme into the bloodstream were unsuccessful because the enzyme was taken up by liver cells, which are not seriously affected by the deficiency. To target macrophages, the enzyme was purified from human placental tissue and treated with three different glycosidases to remove terminal sugars on the enzyme's oligosaccharide chains, which exposed underlying mannose residues (see Figure 8.22). Following infusion into the bloodstream, this modified enzyme (marketed under the name Cerezyme) is recognized by mannose receptors on the surface of macrophages and rapidly taken up by receptor-mediated endocytosis (page 302). Because lysosomes are the natural target site of materials brought into the macrophage by endocytosis, the enzymes are efficiently delivered to the precise sites in the cell where the deficiency is manifested. Thousands of victims of this disease have been successfully treated in this way. Enzyme replacement therapy for the treatment of several other lysosomal storage diseases has either been approved or is being investigated in clinical trials. Unfortunately, many of these diseases affect the central nervous system, which is unable to take up circulating enzymes because of the blood–brain barrier (page 256).

An alternate approach that has shown some promise in preclinical trials is referred to as *substrate reduction therapy*, in which small-molecular-weight drugs (e.g., Zavesca) are administered to inhibit the synthesis of the substances that accumulate in the disease. Finally it can be noted that, although it is accompanied by considerable risk to the patient, bone marrow (or cord blood) transplantation has proven relatively successful in treating some of these diseases. It is thought that the foreign transplanted cells, which contain normal copies of the gene in question, secrete a limited amount of the normal lysosomal enzyme. Some of these enzyme molecules are then taken up by the patient's own cells, which lessens the impact of the enzyme deficiency.

TABLE 1 Sphingolipid Storage Diseases

Disease	Enzyme deficiency	Principal storage substance	Consequences
G_{M1} Gangliosidosis	G_{M1} β-Galactosidase	Ganglioside G_{M1}	Mental retardation, liver enlargement, skeletal involvement, death by age 2
Tay-Sachs disease	Hexosaminidase A	Ganglioside G_{M2}	Mental retardation, blindness, death by age 3
Fabry's disease	α-Galactosidase A	Trihexosylceramide	Skin rash, kidney failure, pain in lower extremities
Sandhoff's disease	Hexosaminidases A and B	Ganglioside G_{M2} and globoside	Similar to Tay-Sachs disease but more rapidly progressing
Gaucher's disease	Glucocerebrosidase	Glucocerebroside	Liver and spleen enlargement, erosion of long bones, mental retardation in infantile form only
Niemann-Pick disease	Sphingomyelinase	Sphingomyelin	Liver and spleen enlargement, mental retardation
Farber's lipogranulomatosis	Ceramidase	Ceramide	Painful and progressively deformed joints, skin nodules, death within a few years
Krabbe's disease	Galactocerebrosidase	Galactocerebroside	Loss of myelin, mental retardation, death by age 2
Sulfatide lipidosis	Arylsulfatase A	Sulfatide	Mental retardation, death in first decade

8.7 PLANT CELL VACUOLES

As much as 90 percent of the volume of many plant cells is occupied by a single membrane-bound, fluid-filled central **vacuole** (Figure 8.36). While simple in structure, plant vacuoles carry out a wide spectrum of essential functions. Many of a cell's solutes and macromolecules, including ions, sugars, amino acids, proteins, and polysaccharides, are stored temporarily in the vacuole. Vacuoles may also store a host of toxic compounds. Some of these compounds (such as cyanide-containing glycosides and glucosinolates) are part of an arsenal of chemical weapons that are released when the cell is injured by an herbivore or fungus. Other toxic compounds are simply the by-products of metabolic reactions; because plants lack the type of excretory systems found in animals, they utilize their vacuoles to isolate these by-products from the rest of the cell. Some of these compounds, such as digitalis, have proven to have important clinical value.

The membrane that bounds the vacuole, the **tonoplast**, contains a number of active transport systems that pump ions into the vacuolar compartment to a concentration much higher than that in the cytoplasm or the extracellular fluid. Because of its high ion concentration, water enters the vacuole by osmosis. Hydrostatic (turgor) pressure exerted by the vacuole not only provides mechanical support for the soft tissues of a plant (page 146), it also stretches the cell wall during cell growth.

Plant vacuoles are also sites of intracellular digestion, not unlike lysosomes, which are absent in plants. In fact, plant vacuoles have some of the same acid hydrolases found in lysosomes. The pH of the vacuole is maintained at a low value by a V-type H^+-ATPase (page 156) within the tonoplast that pumps protons into the vacuolar fluid. Like the proteins of the lysosome, many of the proteins of a plant vacuole are synthesized on membrane-bound ribosomes of the RER, transported through the Golgi complex, and sorted at the *trans* face of the Golgi before being targeted to the vacuole.

REVIEW   ?

1. Describe three distinct roles of plant vacuoles.

2. In what ways is a plant vacuole similar to a lysosome? In what ways is it different?

8.8 THE ENDOCYTIC PATHWAY: MOVING MEMBRANE AND MATERIALS INTO THE CELL INTERIOR

We have described at length how a cell transports materials from the RER and Golgi to the plasma membrane and extracellular space. Let us now turn to the movement of materials in the opposite direction. We considered in Chapter 4 how low-molecular-weight solutes pass through the plasma membrane. But how are cells able to take in materials that are
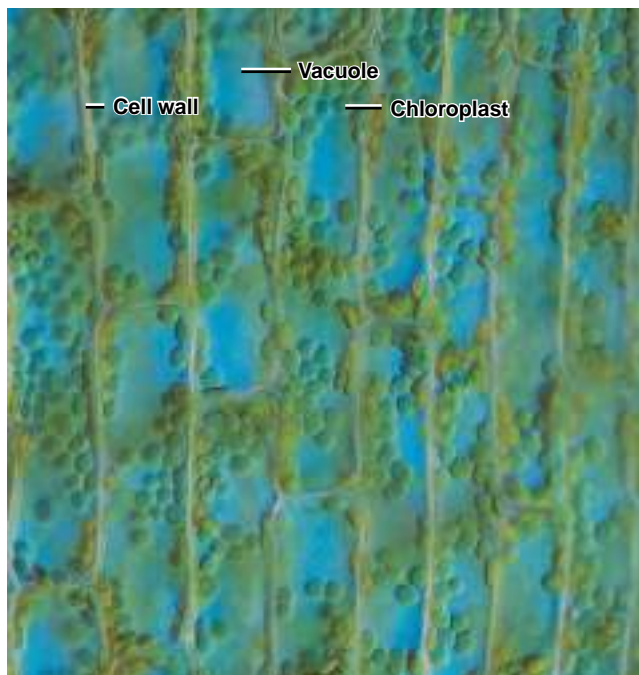
(a)
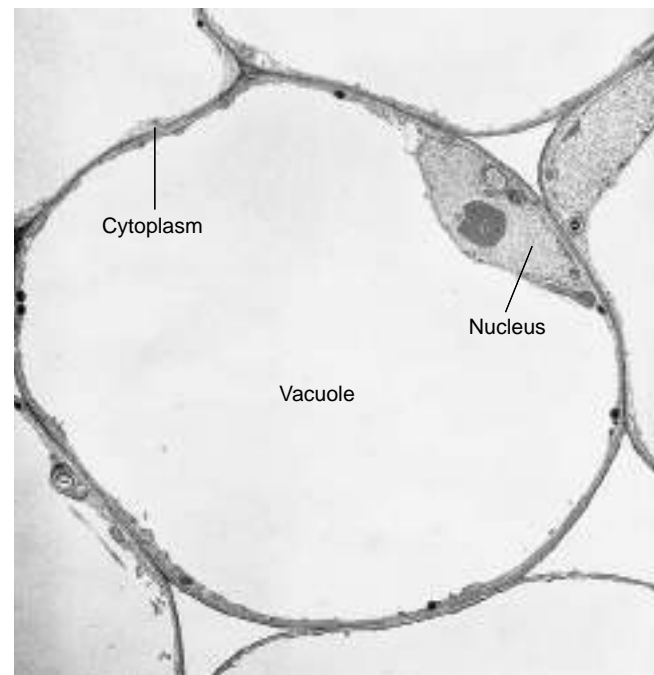
(b)

FIGURE 8.36 Plant cell vacuoles. (*a*) Each of the cylindrical leaf cells of the aquatic plant *Elodea* contains a large central vacuole surrounded by a layer of cytoplasm containing the chloroplasts that are visible in the micrograph. (*b*) Transmission electron micrograph of a soybean cortical cell showing the large central vacuole and thin layer of surrounding cytoplasm. (A: FROM M. I. WALKER/PHOTO RESEARCHERS; B: FROM M. F. BROWN/VISUALS UNLIMITED.)

too large to penetrate this membrane regardless of its permeability properties? And how are proteins that reside in the plasma membrane recycled to the cell's internal compartments? Both of these requirements are met by the endocytic pathway, in which segments of the plasma membrane invaginate to form cytoplasmic vesicles that are transported into the cell interior. We will consider two basic processes in this section of the chapter, endocytosis and phagocytosis, which occur by different mechanisms. **Endocytosis** is primarily a process by which the cell internalizes cell-surface receptors and bound extracellular ligands. **Phagocytosis** describes the uptake of particulate matter.

Endocytosis

Endocytosis can be divided broadly into two categories: bulk-phase endocytosis and receptor-mediated endocytosis. *Bulk-phase endocytosis* (also known as *pinocytosis*) is the nonspecific uptake of extracellular fluids. Any molecules, large or small, that happen to be present in the enclosed fluid also gain entry into the cell. Bulk-phase endocytosis can be visualized by adding a substance to the culture medium, such as the dye lucifer yellow or the enzyme horseradish peroxidase, that is taken up by cells nonspecifically. Bulk-phase endocytosis also removes portions of the plasma membrane and may function primarily in the recycling of membrane between the cell-surface and interior compartments. **Receptor-mediated en-**

docytosis, in contrast, brings about the uptake of specific extracellular macromolecules (ligands) following their binding to receptors on the external surface of the plasma membrane.

Receptor-Mediated Endocytosis and the Role of Coated Pits   Receptor-mediated endocytosis (RME) provides a means for the selective and efficient uptake of macromolecules that may be present at relatively low concentrations in the extracellular fluid. Cells have receptors for the uptake of many different types of ligands, including hormones, growth factors, enzymes, and blood-borne proteins carrying certain nutrients. Substances that enter a cell by means of RME become bound to receptors that collect in specialized domains of the plasma membrane, known as **coated pits**. Receptors are concentrated in coated pits at 10–20 times their level in the remainder of the plasma membrane. Coated pits (Figure 8.37*a*) are recognized in electron micrographs as sites where the surface is indented and the plasma membrane is covered on its cytoplasmic face by a bristly, electron-dense coat containing clathrin, the same protein found in the protein coat of vesicles formed at the TGN (page 293). Coated pits invaginate into the cytoplasm (Figure 8.37*b*) and then pinch free of the plasma membrane to form coated vesicles (Figure 8.37*c,d*). To understand the mechanism of coated vesicle formation, we need to examine the molecular structure of the clathrin coat.

Figure 8.38 shows a coated pit as seen from the extracellular and cytoplasmic surfaces of the plasma membranes of

(a)

(b)

(c)

(d)   0.1 μm
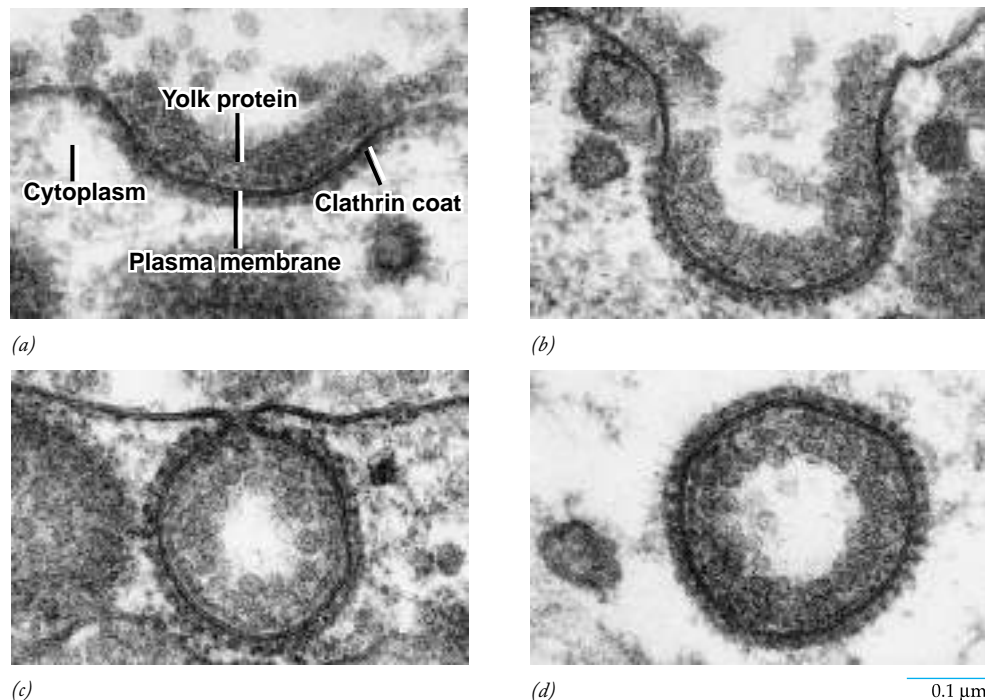
FIGURE 8.37   Receptor-mediated endocytosis. This sequence of micrographs shows the steps in the uptake of yolk lipoproteins by the hen oocyte. (*a*) The proteins to be taken into the cell are concentrated on the extracellular surface of an indented region of the plasma membrane, forming a coated pit. The cytosolic surface of the plasma membrane of the coated pit is covered with a layer of bristly, electron-dense material consisting of the protein clathrin. (*b*) The coated pit has sunk inward to form a coated bud. (*c*) The plasma membrane is about to pinch off as a vesicle containing the yolk protein on its luminal (previously extracellular) surface and clathrin on its cytosolic surface. (*d*) A coated vesicle that is no longer attached to the plasma membrane. The next step in the process is the release of the clathrin coat. (FROM M. M. PERRY AND A. B. GILBERT, J. CELL SCIENCE 39:257, 1979; BY PERMISSION OF THE COMPANY OF BIOLOGISTS, LTD.)

(a)   (b)   (c)   0.1 µm
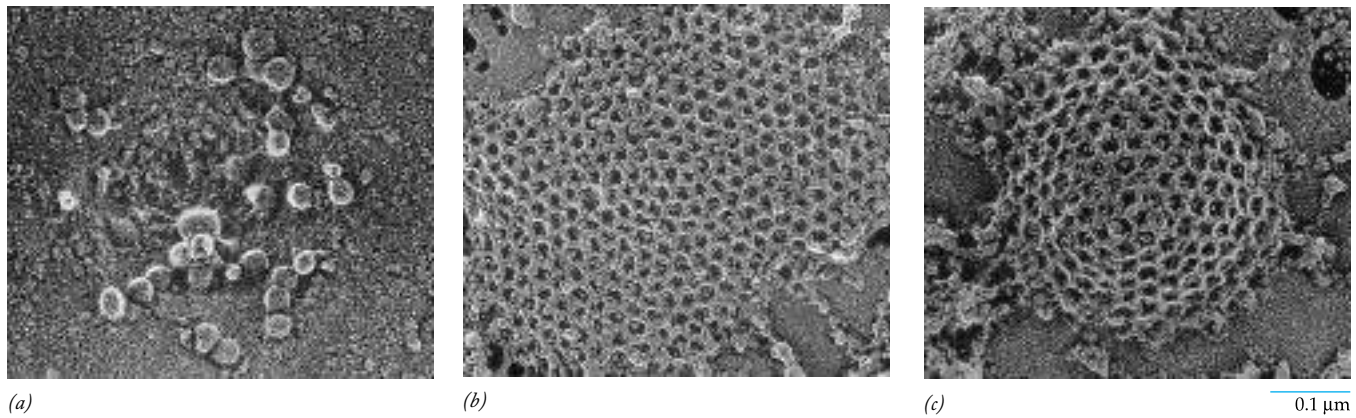
FIGURE 8.38 **Coated pits.** (*a*) Electron micrograph of a replica formed on the extracellular surface of a whole, freeze-dried fibroblast that had been incubated with LDL-cholesterol. Particles of LDL-cholesterol are visible as spherical structures located on the *extracellular surface* of the coated pit. (*b*) Electron micrograph of a replica formed on the *cytosolic surface* of a coated pit of a ruptured fibroblast. The coat is composed of a flattened network of clathrin-containing polygons associated with the inner surface of the plasma membrane. (*c*) Electron micrograph of a replica of the cytosolic surface of a coated bud showing the invaginated plasma membrane surrounded by a clathrin lattice that has assumed a hemispheric shape. (A–C: FROM JOHN HEUSER AND LOUISE EVANS, J. CELL BIOL. 84:568, 1980; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

cells that had been engaged in receptor-mediated endocytosis. When viewed from its cytoplasmic surface (Figure 8.38*b,c*), the bristly coat appears to consist of a network of polygons (hexagons and pentagons) resembling a honeycomb. The geo-

FIGURE 8.39 **Clathrin triskelions.** Electron micrograph of a metal-shadowed preparation of clathrin triskelions. Inset shows the triskelion is composed of three heavy chains. The inner portion of each heavy chain is linked to a smaller light chain. (REPRINTED WITH PERMISSION FROM ERNST UNGEWICKELL AND DANIEL BRANTON, NATURE 289:421, 1981; © COPYRIGHT 1981, MACMILLAN MAGAZINES LTD.)

metric construction of the coat is derived from the structure of its clathrin building blocks. Each clathrin molecule consists of three heavy chains and three light chains, joined together at the center to form a three-legged assembly called a *triskelion* (Figure 8.39). The overlapping arrangement of the triskelions within the clathrin scaffold of a coated vesicle is shown in Figure 8.40. Each leg of a clathrin triskelion extends outward along two edges of a polygon. The clathrin molecules overlap in such a way that each vertex of a polygon contains a center of one of the component triskelions.

Like the clathrin-coated vesicles that bud from the TGN (page 293), the coated vesicles that form during endocytosis also contain a layer of adaptors situated between the clathrin lattice and the surface of the vesicle facing the cytosol. The best studied adaptor operating in connection with clathrin-mediated endocytosis is AP2. Unlike the GGA adaptors utilized at the TGN, which consist of a single subunit with several domains (Figure 8.30), the AP2 adaptors that become incorporated into the vesicles that bud from the plasma membrane contain multiple subunits having different functions (Figure 8.40). The μ subunit of AP2 adaptors engages the cytoplasmic tails of specific plasma membrane receptors, leading to the concentration of these selected receptors—and their bound cargo molecules—into the emerging coated vesicle (discussed further in the Experimental Pathways on page 314). In contrast, the β-adaptin subunit of AP2 adaptors binds and recruits the clathrin molecules of the overlying lattice. A comparison of Figures 8.26 and 8.40 shows both marked differences and similarities between the coats of COPII- and clathrin-coated vesicles. Both coats contain two distinct layers: an outer geometric scaffold and an inner layer of adaptor proteins. The structure of the outer scaffolds, however, are very different; the subunits of the clathrin lattice (three-legged clathrin complexes) overlap extensively, whereas those of the COPII lattice (rod-like Sec13-Sec31 complexes)

(a)

(b)   (c)
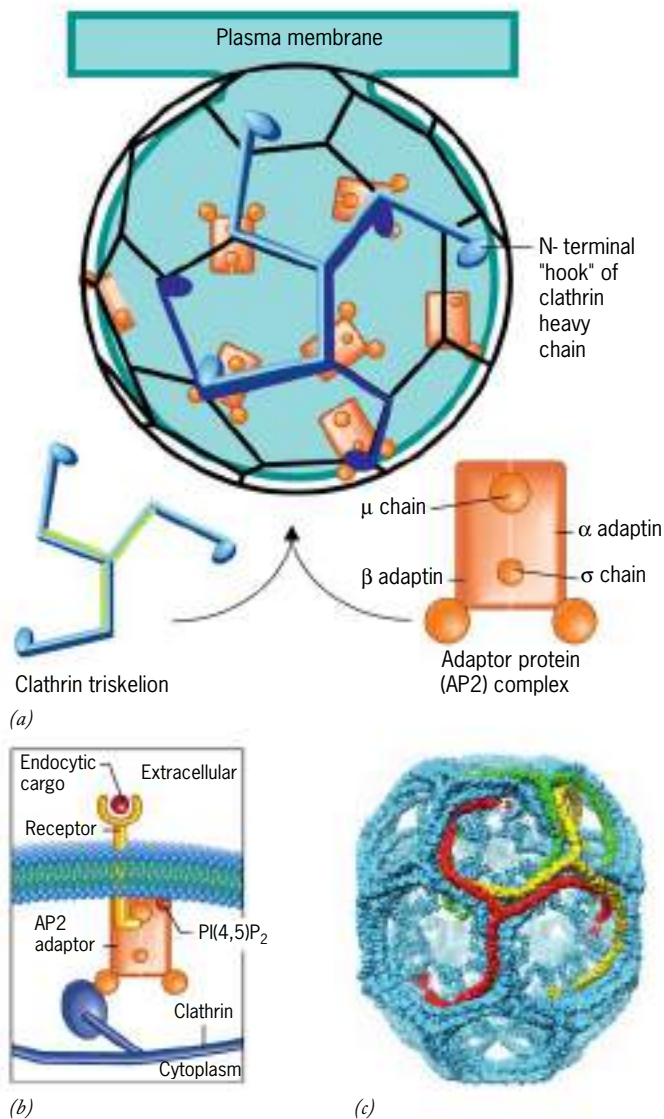
FIGURE 8.40 Molecular organization of a coated vesicle. (*a*) Schematic drawing of the surface of a coated vesicle showing the arrangement of triskelions and adaptors in the outer clathrin coat. The sides of the polygons are formed by parts of the legs of overlapping triskelions. The N-terminus of each clathrin heavy chain forms a "hook" that projects toward the surface of the membrane where it engages an adaptor. Each adaptor, which consists of four different polypeptide subunits, can bind a diverse array of accessory proteins that are not shown in this illustration. Both the hooks and adaptors are situated at the vertices of the polyhedrons. (*Note:* Not all of the triskelions of the lattice are shown in this figure; if they were, every vertex would have a clathrin hub, hook, and associated adaptor.) (*b*) Schematic drawing of a cross section through the surface of a coated vesicle showing the interactions of the AP2 adaptor complexes with both the clathrin coat and membrane receptors. Recruitment of AP2 adaptors to the plasma membrane is facilitated by the presence of PI(4,5)P$_2$ molecules in the inner (cytosolic) leaflet of the membrane. Each receptor is bound to a ligand being internalized. (*c*) Reconstruction of a clathrin cage containing 36 triskelions showing the overlapping arrangement of several of these trimeric molecules (shown in different colors). (A: REPRODUCED, WITH PERMISSION, AFTER S. SCHMID, ANNUAL REVIEW OF BIOCHEMISTRY, VOL. 66; © 1997 BY ANNUAL REVIEWS. C: REPRINTED WITH PERMISSION FROM ALEXANDER FOTIN ET AL., NATURE 432:574, 2004, COURTESY OF STEPHEN C. HARRISON; © COPYRIGHT 2004, MACMILLAN MAGAZINES LTD.)

do not overlap at all. Whether there is a functional basis for these two different types of construction strategies is unclear.

The structure depicted in Figure 8.40 is a greatly simplified version of a "real" coated vesicle, which may contain upwards of two dozen different accessory proteins that form a dynamic network of interacting molecules. These proteins have poorly understood roles in cargo recruitment, coat assembly, membrane curvature and invagination, interaction with cytoskeletal components, vesicle release, and membrane uncoating. The best studied of these accessory proteins is dynamin.

Dynamin is a large GTP-binding protein that is required for the release of a clathrin-coated vesicle from the membrane on which it forms. Dynamin self-assembles into a helical collar around the neck of an invaginated coated pit (Figure 8.41), just before it pinches off from the membrane. In the model depicted in Figure 8.41*a*, steps 3–4, hydrolysis of the bound GTP by the polymerized dynamin molecules induces a twisting motion in the dynamin helix that severs the coated vesicle from the plasma membrane. According to this mechanism, dynamin acts as an enzyme capable of utilizing the chemical energy of GTP to generate mechanical forces.

The Role of Phosphoinositides in the Formation of Coated Vesicles   Although the discussion emphasizes the protein molecules of the coat and vesicle, the phospholipids of the vesicle membrane also play an important role. As discussed in Chapter 15, phosphate groups can be added to different positions of the sugar ring of the phospholipid phosphatidylinositol (PI), converting them into phosphoinositides (see Figure 15.8). Seven distinct phosphoinositides are identified: PI(3)P, PI(4)P, PI(5)P, PI(3,4)P$_2$, PI(4,5)P$_2$, PI(3,5)P$_2$, and PI(3,4,5)P$_3$. The phosphorylated rings of these phosphoinositides reside at the surface of the membrane where they can be recognized and bound by particular proteins. Different phosphoinositides are concentrated in different membrane compartments, which helps give each compartment a unique "surface identity". The inner leaflet of the plasma membrane, for example, tends to contain elevated levels of PI(4,5)P$_2$, which plays an important role in recruiting proteins involved in clathrin-mediated endocytosis, such as dynamin and AP2 (see Figure 8.40*b*). A lipid species such as PI(4,5)P$_2$ can have a dynamic regulatory role because it can be rapidly formed and destroyed by enzymes that are localized at particular places and times within the cell. In the example of endocytosis, PI(4,5)P$_2$ disappears from a site of endocytosis about the time the coated vesicle is pinched away from the plasma membrane. Other PIs that have been associated with the secretory/endocytic pathways include PI(3)P localized at early endosomes and intraluminal vesicles of late endosomes: PI(4)P localized at the TGN, secretory granules, and synaptic vesicles; and PI(3,5)P$_2$ localized at the late endosome boundary membrane.

The Endocytic Pathway   Molecules taken into a cell by endocytosis are routed through a well-defined **endocytic pathway** (Figure 8.42*a*). Before describing events that occur along the endocytic pathway, it is worth considering two different types of receptors that are subjected to endocytosis. One

(a)

(b)   150 nm
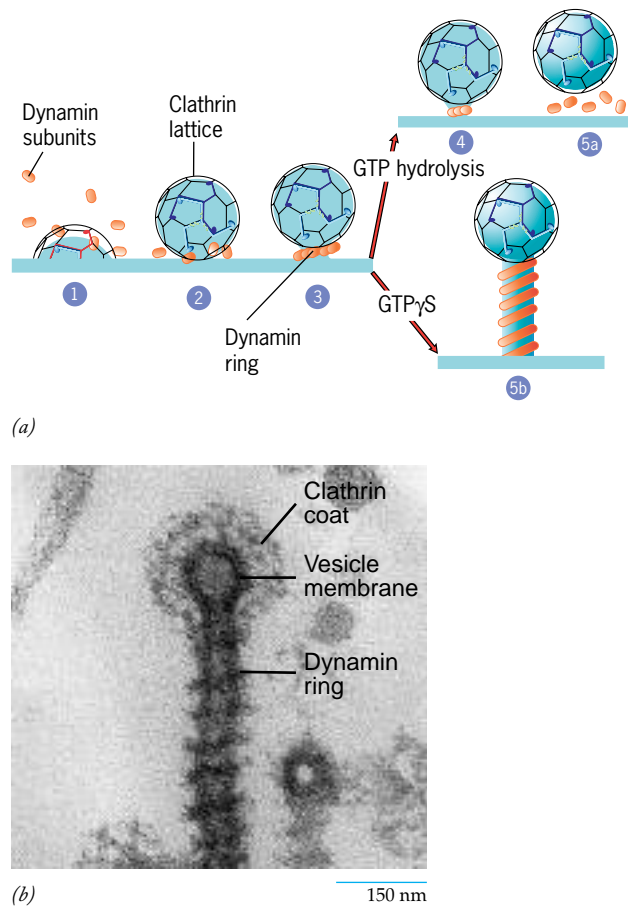
FIGURE 8.41 **The role of dynamin in the formation of clathrin-coated vesicles.** (*a*) The clathrin lattice of the coated pit (step 1) undergoes rearrangement to form an invaginated vesicle connected to the overlying plasma membrane by a stalk (step 2). At this point, the dynamin subunits, which are concentrated in the region, undergo polymerization to form a ring around the stalk (step 3). Changes in the conformation of the ring, which are thought to be induced by GTP hydrolysis (step 4), lead to fission of the coated vesicle from the plasma membrane and disassembly of the dynamin ring (step 5a). If vesicle budding occurs in the presence of GTPγS, a nonhydrolyzable analogue of GTP, dynamin polymerization continues beyond formation of a simple collar, producing a narrow tubule constructed from several turns of the dynamin helix (step 5b). (*b*) Electron micrograph showing a coated vesicle forming in the presence of GTPγS, which corresponds to the stage depicted in step 5b of part *a*. (*Note:* Considerable evidence also points to the involvement of the actomyosin machinery in vesicle formation and scission. See *Curr. Opin. Cell Biol.* 19:453, 2007 and *Nature Revs. Mol. Cell Biol.* 7:404, 2006 for discussion.) (A: FROM P. DE CAMILLI ET AL. CURRENT OPIN. NEUROBIOL. VOL. 5, P. 562, 1995; B: REPRINTED WITH PERMISSION FROM KOHJI TAKEI ET AL., NATURE VOL. 374, COVER 3/9/95; © COPYRIGHT 1995, MACMILLAN MAGAZINES LTD.)

group of receptors, which we will refer to as "housekeeping receptors," is responsible for the uptake of materials that will be used by the cell. The best studied examples are the transferrin and LDL (low-density lipoprotein) receptors, which mediate the delivery to cells of iron and cholesterol, respectively. The LDL receptor is discussed in detail at the end of this section. The red-colored receptor of Figure 8.42*a* represents a housekeeping receptor. The second group of receptors, which we

will refer to as "signaling receptors," is responsible for binding extracellular ligands that carry messages that change the activities of the cell. These ligands, which include hormones such as insulin and growth factors such as EGF, bind to the surface receptor (shown in green in Figure 8.42*a*) and signal a physiologic response inside the cell (discussed at length in Chapter 15). Endocytosis of the first group of receptors leads typically to the delivery of the bound materials, such as iron and cholesterol, to the cell and the return of the receptor to the cell surface for additional rounds of uptake. Endocytosis of the second group of receptors often leads to the destruction of the receptor, a process called *receptor down-regulation*, which has the effect of reducing the sensitivity of the cell to further stimulation by the hormone or growth factor. Receptor down-regulation is a mechanism by which cells regulate their ability to respond to extracellular messengers. "Signaling receptors" are typically marked for endocytosis and subsequent destruction by the covalent attachment of a "tag" to the cytoplasmic tail of the receptor while it resides at the cell surface. The tag is a small protein called ubiquitin, which is added enzymatically. Membrane proteins that are not normally subjected to endocytosis become internalized if they are made to carry an added ubiquitin.

Following internalization, vesicle-bound materials are transported to a dynamic network of tubules and vesicles known collectively as **endosomes**, which represent distribution centers along the endocytic pathway. The fluid in the lumen of endosomes is acidified by a H^+-ATPase in the boundary membrane. Endosomes are divided into two classes: **early endosomes**, which are typically located near the peripheral region of the cell, and **late endosomes**, which are typically located closer to the nucleus. According to the prevailing model, early endosomes progressively mature into late endosomes. This transformation from an early to a late endosome is characterized by a decrease in pH, an exchange of Rab proteins (e.g., from Rab5 to Rab7), and a major change in the internal morphology of the structures. This latter change occurs as the outer boundary membrane of the endosome forms buds on its luminal surface that invaginate inward to create a population of vesicles that crowd the interior of the late endosome. Because of these internal vesicles, which are shown in the electron micrograph of Figure 8.42*b*, late endosomes are also referred to as *multivesicular bodies* (MVBs).

Receptors that are taken up by endocytosis are transported in vesicles to an early endosome, which serves as a sorting station that directs different types of receptors and ligands along different pathways (Figure 8.42*a*). "Housekeeping receptors" typically dissociate from their bound ligands as a result of the high H^+ concentration of the early endosomes. The receptors are then concentrated into specialized tubular compartments of the early endosome, which represent recycling centers. Vesicles that bud from these tubules carry receptors back to the plasma membrane for additional rounds of endocytosis (Figure 8.42*a*). In contrast, released ligands (e.g., LDLs) become concentrated into a sorting compartment before being dispatched to a late endosome and ultimately to a lysosome, where final processing occurs. As noted above, "signaling receptors" with their attached ubiquitin tags

(a)

(b)

(c)

FIGURE 8.42 **The endocytic pathway.** The movement of materials from the extracellular space to early endosomes where sorting occurs. Endocytosis of two types of receptor–ligand complexes is shown. Housekeeping receptors, such as the LDL receptor (shown in red), are typically sent back to the plasma membrane, whereas their ligands (blue spheres) are transferred to late endosomes. Signaling receptors, such as the EGF receptor (shown in green), are typically transported to late endosomes along with their ligands (orange-yellow). Late endosomes also receive newly synthesized lysosomal enzymes (red spheres) from the TGN. These enzymes are carried by mannose 6-phosphate receptors (MPRs), which return to the TGN. The contents of late endosomes are transferred to lysosomes by a number of routes (not shown). The inset on the left shows an enlarged view of a portion of a late endosome with an intraluminal vesicle budding inward from the outer membrane. The membranes of these vesicles contain receptors to be degraded. (*b*) Electron micrograph showing the internal vesicles within the lumen of a late endosome. A number of lysosomes are present in the vicinity. (*c*) The gold particles seen in this electron micrograph are bound to EGF receptors that were internalized by endocytosis and have become localized within the membranes of the internal vesicles of this late endosome. (B: FROM J. PAUL LUZIO, PAUL R. PRYOR, AND NICHOLAS A. BRIGHT, NATURE REVS. MOL. CELL BIOL. 8:625, 2007; © COPYRIGHT 2007, MACMILLAN MAGAZINES LTD., C: COURTESY OF CLARE FUTTER.)

do not recycle back to the membrane. Instead, these ubiquitinated signaling receptors are recognized by a series of protein complexes (called ESCRT complexes) that sort the receptors into the membranes that give rise to the internal vesicles of the late endosomes (Figure 8.42*c*). Ultimately, late endosomes containing these intraluminal vesicles fuse with a lysosome (Figure 8.42*b*), which leads to the degradation of the contents of the endosome by lysosomal enzymes.

LDLs and Cholesterol Metabolism   Among many examples of receptor-mediated endocytosis, the first studied and best understood is one that provides animal cells with exogenous cholesterol. Animal cells use cholesterol as an essential part of their plasma membranes and as a precursor to steroid hormones. Cholesterol is a hydrophobic molecule that is transported in the blood as part of huge lipoprotein complexes, such as the *low-density lipoprotein* (*LDL*) shown in Figure 8.43. Each

LDL particle contains a central core of about 1500 cholesterol molecules esterified to long-chain fatty acids. The core is surrounded by a single layer of phospholipids that contains a single copy of a large protein, called *apolipoprotein B-100*, which binds specifically to LDL receptors on the surfaces of cells.

LDL receptors are transported to the plasma membranes of cells, where they become concentrated in coated pits, even in the absence of LDL ligand. As a result, the receptors are on the cell surface ready to take up the blood-borne lipoproteins if they should become available. Once the LDL is bound to a coated pit, the pit invaginates to form a coated vesicle, the clathrin coat is disassembled, and the LDL receptors pass through the early endosomes and back to the plasma membrane, as depicted in Figure 8.42*a*. Meanwhile, the LDL particles are delivered to late endosomes and lysosomes, where the protein component is degraded and the cholesterol is deesterified and used by the cell in membrane assembly or other metabolic processes (e.g., steroid hormone formation). Persons with a rare inherited disorder called *Niemann–Pick type C disease* lack one of the proteins required to transfer

Apolipoprotein B-100

Phospholipid

Unesterified cholesterol

Esterified cholesterol
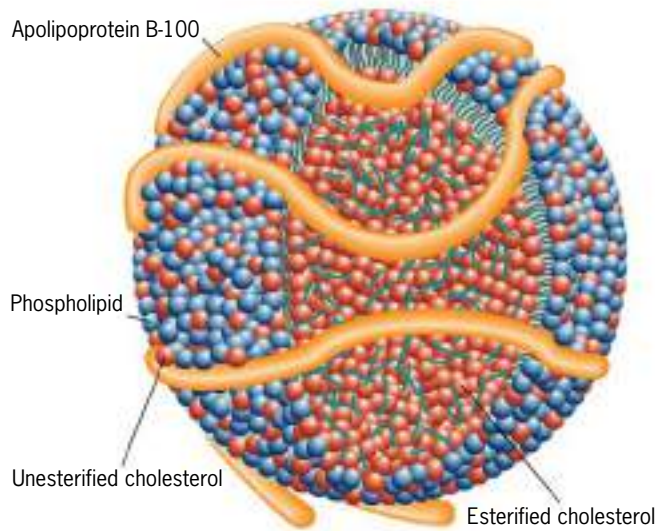
FIGURE 8.43 LDL cholesterol. Each particle consists of esterified cholesterol molecules, surrounded by a mixed monomolecular layer of phospholipids and cholesterol, and a single molecule of the protein apolipoprotein B-100, which interacts specifically with the LDL receptor projecting from the plasma membrane.

cholesterol out of lysosomes. The resulting accumulation of cholesterol in these organelles leads to nerve degeneration and death during early childhood. Studies of a different disease that led to the discovery of receptor-mediated endocytosis and LDL internalization are described in the accompanying Experimental Pathways.

The level of LDL in the blood has been related to the development of atherosclerosis, a condition characterized by formation of plaques on the walls of arteries that reduce the flow of blood through the vessel and act as sites for the formation of blood clots. Blood clots that block the coronary arteries are the leading cause of myocardial infarction (heart attack). Studies suggest that atherosclerosis results from a chronic inflammatory response that is initiated by the deposition of LDL within the inner walls of blood vessels, as indicated in Figure 8.44. Lowering blood LDL levels is most readily accomplished by administration of drugs called *statins* (e.g., lovastatin and Lipitor) that block HMG CoA reductase, a key enzyme in the synthesis of cholesterol (page 312). When blood cholesterol levels are lowered, the risk of heart attack is reduced.

LDLs are not the only cholesterol-transporting agents in the blood. *HDLs* (*high-density lipoproteins*) have a similar construction but contain a different protein (apolipoprotein A-I) and play a different physiologic role in the body. LDL serves primarily to carry cholesterol molecules from the liver, where they are synthesized and packaged, through the blood to the body's cells. HDL carries cholesterol in the opposite direction. Excess cholesterol is transported out of the plasma membrane of the body's cells directly to circulating HDL particles, which carry cholesterol to the liver for excretion. Just as high blood levels of LDL are associated with increased risk of heart disease, high blood levels of HDL are associated with decreased risk, which has led to HDL being called the "good cholesterol." There is little doubt that lowering LDL levels is beneficial, but the consequences of raising HDL levels is less clear-cut. For example, cholesterol molecules can be transferred from HDL to other lipoprotein particles by an enzyme called cholesteryl ester transfer protein (CETP), an activity that tends to lower HDL cholesterol levels. CETP has become a focus of research following the discovery of a population of Japanese families whose members routinely live more than 100 years and carry mutations in the *CETP* gene. At least two small-molecular-weight CETP inhibitors have been tested in clinical trials and found to increase HDL levels in the blood. One of these drug candidates (torcetrapib) was dropped from further study despite the fact that it raised HDL blood levels. For reasons that are unclear, subjects taking the drug and a statin were considerably more likely to die than those in the control group who took only the statin. Another CETP inhibitor (anacetrapib) is still being tested clinically at the time of this writing.

Smooth muscle cell

Macrophage

Fibrous cap

Endothelial cell

Endothelial injury   Leukocyte

Foam cell containing LDL
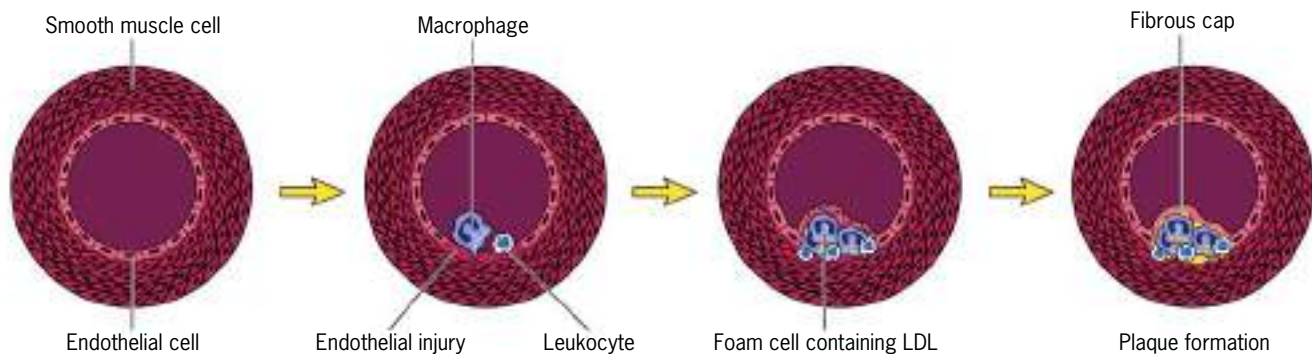
Plaque formation

FIGURE 8.44 A model of atherosclerotic plaque formation. According to this model, plaque formation is initiated by various types of injury to the endothelial cells that line the vessel, including damage inflicted by oxygen free radicals that chemically alter the LDL-cholesterol particles. The injured endothelium acts as an attractant for white blood cells (leukocytes) and macrophages, which migrate beneath the endothelium and begin a process of chronic inflammation. The macrophages ingest the oxidized LDL, which becomes deposited in the cytoplasm as cholesterol-rich fatty droplets. These cells are referred to as macrophage foam cells and they are often already present in the blood vessels of adolescents and young adults. Substances released by the macrophages stimulate the proliferation of smooth muscle cells, which produce a dense, fibrous connective tissue matrix (fibrous cap) that bulges into the arterial lumen. Not only do these bulging lesions restrict blood flow, they are prone to rupture, which can trigger the formation of a blood clot and ensuing heart attack.

(a)
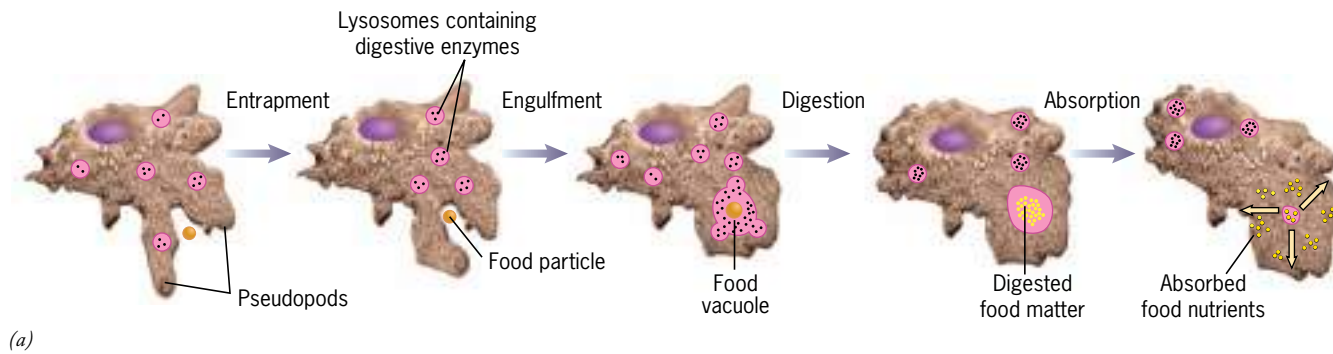
FIGURE 8.45 Phagocytosis. (*a*) Schematic diagram of the steps in the engulfment, digestion, and absorption of materials taken into an amoeba by phagocytosis. (*b*) The process of engulfment as illustrated by a polymorphonuclear leukocyte ingesting a yeast cell (lower left). (B: FROM JANET BOYLES AND DOROTHY F. BAINTON, CELL 24:906, 1981, BY PERMISSION OF CELL PRESS.)

(b)   1.5 μm
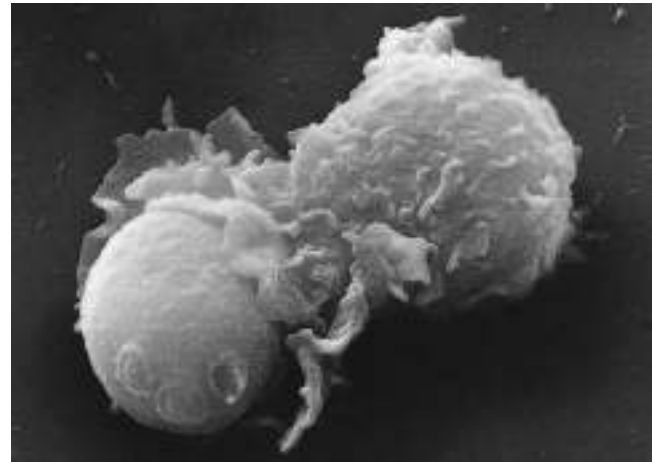
Phagocytosis

Phagocytosis ("cell eating") is carried out extensively by a few types of cells specialized for the uptake of relatively large particles (>0.5 μm diameter) from the environment. Many single-celled protists, such as amoebas and ciliates, make their livelihood by trapping food particles and smaller organisms and enclosing them within folds of the plasma membrane (Figure 8.45a). The folds fuse to produce a vacuole (or *phagosome*) that pinches off inwardly from the plasma membrane. The phagosome fuses with a lysosome, and the material is digested within the resulting *phagolysosome*.

In most animals, phagocytosis is a protective mechanism rather than a mode of feeding. Mammals possess a variety of "professional" phagocytes, including macrophages and neutrophils, that wander through the blood and tissues phagocytizing invading organisms, damaged and dead cells, and debris. These materials are recognized and bound by receptors on the surface of the phagocyte prior to uptake. Once inside the phagocyte, microorganisms may be killed by lysosomal enzymes or by oxygen free radicals generated within the lumen of the phagosome. The process of particle engulfment is pictured in the chapter-opening photograph and in Figure 8.45b. The steps in the digestion of engulfed materials are illustrated in Figure 8.46. The engulfment of particulate material by phagocytosis is driven by contractile activities of the actin-containing microfilaments that underlie the plasma membrane.

Not all bacteria ingested by phagocytic cells are destroyed. In fact, some species hijack the phagocytic machinery to promote their own survival in the body. *Mycobacterium tuberculosis*, the agent responsible for tuberculosis, for example, is taken into the cytoplasm of a macrophage by phagocytosis, but the bacterium is able to inhibit the fusion of its phagosome with a lysosome. Recent studies suggest that, even if the phagosome does become highly acidic, the bacterium is able to maintain its own physiological pH despite the lowered pH of its surrounding medium. The bacterium responsible for Q fever, *Coxiella burnetii*, becomes enclosed in a phagosome that does fuse with a lysosome, but neither the acidic environ-
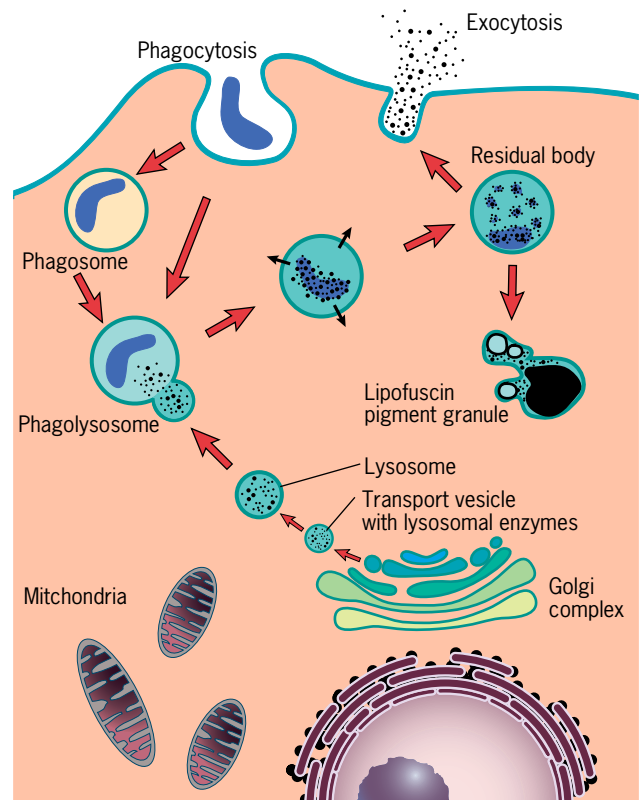
FIGURE 8.46 A summary of the phagocytic pathway.

ment nor the lysosomal enzymes can destroy the pathogen. *Listeria monocytogenes*, a bacterium that causes meningitis, produces proteins that destroy the integrity of the lysosomal membrane, allowing the bacterium to escape into the cell's cytosol (see Figure 9.67).

8.9 POSTTRANSLATIONAL UPTAKE OF PROTEINS BY PEROXISOMES, MITOCHONDRIA, AND CHLOROPLASTS

The division of the contents of a cell into large numbers of compartments presents many organizational challenges to the cell's protein-trafficking machinery. We have seen in this chapter that protein trafficking within a eukaryotic cell is governed by (1) sorting signals, such as the signal peptide of secreted proteins or mannose-phosphate groups of lysosomal enzymes, and (2) receptors that recognize these signals and deliver proteins containing them to the proper compartment. Four of the cell's major organelles—nucleus, mitochondria, chloroplasts, and peroxisomes—import proteins through one or more outer boundary membranes. As in the case of the rough ER, proteins that are imported by these organelles contain amino acid sequences that serve as addresses that are recognized by receptors at the organelle's outer membrane. Unlike the rough ER, which generally imports its proteins cotranslationally, the proteins of these other organelles are imported posttranslationally, that is, following their complete synthesis on free ribosomes in the cytosol.

The import of proteins into the nucleus is a specialized topic and will be discussed separately in Section 12.1. Import of proteins into peroxisomes, mitochondria, and chloroplasts will be discussed in the following sections.

Uptake of Proteins into Peroxisomes

Peroxisomes are very simple organelles having only two subcompartments in which an imported protein can be placed: the boundary membrane and the internal matrix (page 200). Proteins destined for a peroxisome possess a *peroxisomal targeting signal*, either a *PTS* for a peroxisomal matrix protein or an *mPTS* for a peroxisomal membrane protein. Several different PTSs, mPTSs, and PTS receptors have been identified. PTS receptors bind to peroxisome-destined proteins in the cytosol and shuttle them to the surface of the peroxisome. The PTS receptor apparently accompanies the peroxisomal protein through the boundary membrane into the matrix and then recycles back to the cytosol to escort another protein. Unlike mitochondria and chloroplasts, whose imported proteins must assume an unfolded state, peroxisomes are somehow able to import peroxisomal matrix proteins in their native, folded conformation, even those that consist of several subunits. The mechanism by which peroxisomes are able to accomplish this daunting assignment remains a matter of speculation.

Uptake of Proteins into Mitochondria

Mitochondria have four subcompartments into which proteins can be delivered: an outer mitochondrial membrane (OMM), inner mitochondrial membrane (IMM), intermembrane space, and matrix (see Figure 5.3c). Although mitochondria synthesize a few of their own integral membrane polypeptides (13 in mammals), roughly 99 percent of the organelle's proteins are encoded by the nuclear genome, synthesized in the cytosol, and imported posttranslationally. We will restrict the present discussion to proteins of the mitochondrial matrix and inner mitochondrial membrane, which together constitute the vast majority of the proteins targeted to this organelle. As with peroxisomal proteins, and proteins of other compartments, mitochondrial proteins contain signal sequences that target them to their home base. Most mitochondrial matrix proteins contain a removable targeting sequence (called the *presequence*) located at the N-terminus of the molecule (step 1, Figure 8.47a) that includes a number of positively charged residues. In contrast, most proteins destined for the IMM contain internal targeting sequences that remain as part of the molecule.

Before a protein can enter a mitochondrion, several events are thought to take place. First, the protein must be presented to the mitochondrion in a relatively extended, or unfolded, state (steps 1 and A, Figure 8.47a). Several different molecular chaperones (e.g., Hsp70 and Hsp90) have been implicated in preparing polypeptides for uptake into mitochondria, including ones that specifically direct mitochondrial proteins to the cytosolic surface of the OMM (Figure 8.47a). The OMM contains a protein-import complex, the *TOM complex*, which includes (1) receptors that recognize and bind mitochondrial proteins and (2) protein-lined channels through which unfolded polypeptides are translocated across the outer membrane (steps 2 and B).[5] Proteins that are destined for the IMM or matrix must pass through the intermembrane space and engage a second protein-import complex located in the IMM, called a *TIM complex*. The IMM contains two major TIM complexes: TIM22 and TIM23. TIM22 binds integral proteins of the IMM that contain an internal targeting sequence and inserts them into the lipid bilayer (steps C–D, Figure 8.47a). TIM23,

[5]It is interesting to note that, unlike the translocon of the ER or peroxisome, the pore-forming protein of the TOM complex (Tom 40) is a β-barrel protein, like the other integral proteins of the OMM (page 175), reflecting its evolution from the outer membrane of an ancestral bacterium. This has functional consequences, as the β-barrel protein cannot open laterally to allow integral proteins to insert into the OMM. As a result, OMM proteins have to pass into the intermembrane space before entering the OMM bilayer.

(a)

(b)
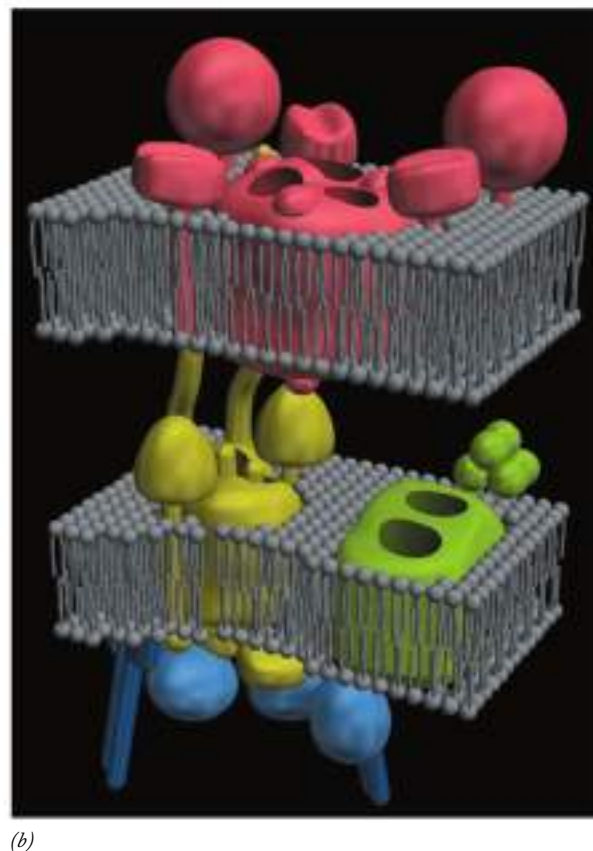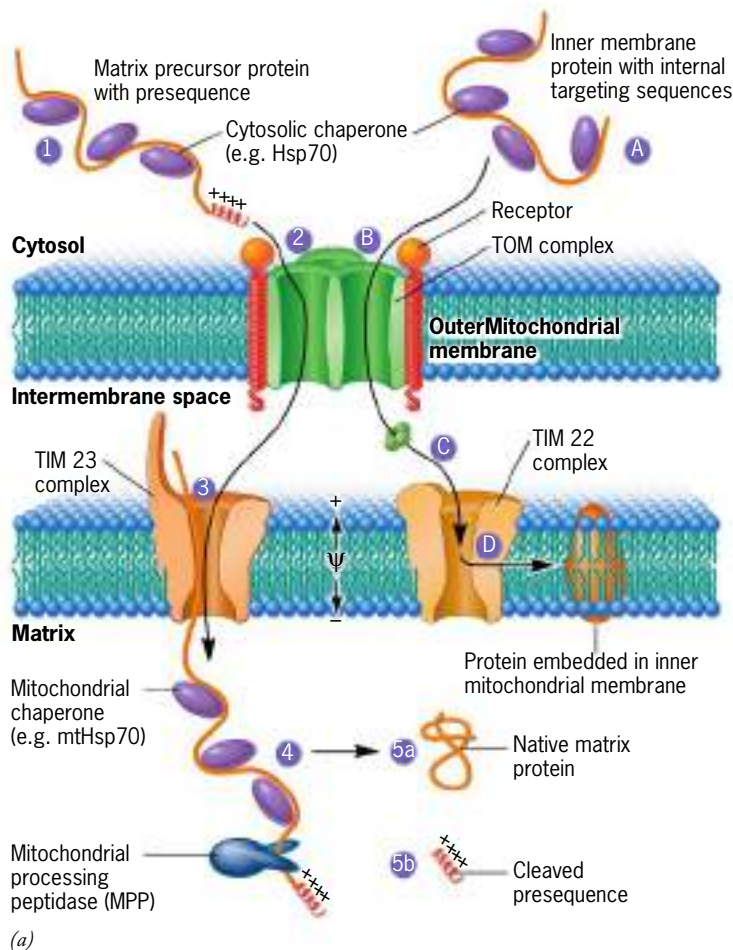
FIGURE 8.47   **Importing proteins into a mitochondrion.** (*a*) Proposed steps taken by proteins imported posttranslationally into either the mitochondrial matrix or inner mitochondrial membrane. The polypeptide is targeted to a mitochondrion by a targeting sequence, which is located at the N-terminus in the matrix protein (step 1) and is located internally in most inner membrane proteins (step A). Cytosolic Hsp70 molecules unfold the polypeptides prior to their entry into the mitochondrion. The proteins are recognized by membrane receptors (red transmembrane proteins) and translocated through the OMM by way of pores in the TOM complex of the OMM (step 2 or B). Most integral proteins of the IMM are directed to the TIM22 complex of the IMM (step C), which steers them into the lipid bilayer of the IMM (step D). Mitochondrial matrix proteins are translocated through the TIM23 complex of the IMM (step 3). Once the protein enters the matrix, it is bound by a mitochondrial chaperone (step 4), which may either pull the polypeptide into the matrix or act like a Brownian ratchet to ensure that it diffuses into the matrix (these alternate chaperone mechanisms are discussed in the text). Once in the matrix, the unfolded protein assumes its native conformation (step 5a) with the help of Hsp60 chaperones (not shown). The presequence is removed enzymatically (step 5b). (*b*) A three-dimensional model of the mitochondrial protein-import machinery, showing the number, relative size, and topology of the various proteins involved in this activity. The TOM complex is a reddish color, the TIM23 complex is yellow-green, the TIM22 complex is green, and the cooperating chaperones are blue. (B: FROM TOSHIYA ENDO, HAYASHI YAMAMOTO, AND MASATOSHI ESAKI, J. CELL SCIENCE, COVER OF VOL. 116, # 16,2003; BY PERMISSION OF THE COMPANY OF BIOLOGISTS, LTD.)

in contrast, binds proteins with an N-terminal presequence, which includes all of the proteins of the matrix (as well as a number of proteins of the IMM that will not be discussed). TIM23 recognizes and translocates the matrix proteins completely through the IMM and into the inner aqueous compartment (step 3, Figure 8.47*a*). Translocation occurs at sites where the outer and inner mitochondrial membranes come into close proximity so that the imported protein can cross both membranes simultaneously. Movement into the matrix is powered by the electric potential across the IMM acting on the positively charged targeting signal; if the potential is dissipated by addition of a drug such as DNP (page 191), translocation ceases.

As it enters the matrix, a polypeptide interacts with mitochondrial chaperones, such as mtHsp70 (step 4, Figure 8.47*a*), that mediate entry into the aqueous compartment. Two mechanisms have been proposed to explain the general action of molecular chaperones involved in the movement of proteins across membranes, which is a widespread phenomenon. According to one view, the chaperones act as force-generating motors that use energy derived from ATP hydrolysis to actively "pull" the unfolded polypeptide through the translocation pore. According to the alternate view, the chaperones aid in the diffusion of the polypeptide across the membrane. Diffusion is a random process in which a molecule can move in any available direction. Con-

sider what would happen if an unfolded polypeptide had entered a translocation pore in the mitochondrial membrane and had "poked its head" into the matrix. Then consider what would happen if a chaperone residing on the inner surface of the membrane were able to bind the protruding polypeptide in such a way that it blocked the diffusion of the polypeptide back through the pore and into the cytosol, but did not block its diffusion further into the matrix. As the polypeptide diffused further into the matrix, it would be bound repeatedly by the chaperone and at each stage prevented from diffusing backward. This mechanism of chaperone action is referred to as *biased diffusion*, and the chaperone is said to be acting as a "Brownian ratchet"; the term *Brownian* implies random diffusion, and a "ratchet" is a device that allows movement in only one direction. Recent studies suggest that both mechanisms of chaperone action are probably utilized and act cooperatively. Regardless of the mechanism of entry, once in the matrix the polypeptide achieves its native conformation (step 5a, Figure 8.47*a*) following enzymatic removal of the presequence (step 5b).

Uptake of Proteins into Chloroplasts

Chloroplasts have six subcompartments into which proteins can be delivered: an inner and outer envelope membrane and intervening intermembrane space, as well as the stroma, thylakoid membrane, and thylakoid lumen (Figure 8.48). Chloroplast and mitochondrial import mechanisms exhibit many similarities, although their translocation machineries have evolved independently. As in the mitochondria,

1. the vast majority of chloroplast proteins are imported from the cytosol,
2. the outer and inner envelope membranes contain distinct translocation complexes (*Toc* and *Tic complexes*, respectively) that work together during import,
3. chaperones aid in the unfolding of the polypeptides in the cytosol and folding of the proteins in the chloroplast, and
4. most proteins destined for the chloroplast are synthesized with a removable N-terminal sequence (termed the *transit peptide*).

The transit peptide does more than simply target a polypeptide to a chloroplast: it provides an "address" that localizes the polypeptide to one of several possible subcompartments within the organelle (Figure 8.48). All proteins translocated through the chloroplast envelope contain a *stroma targeting domain* as part of their transit peptide, which guarantees that the polypeptide will enter the stroma. Once in the stroma, the stroma targeting domain is removed by a processing peptidase located in that compartment. Those polypeptides that belong in a thylakoid membrane or thylakoid lumen bear an additional segment in their transit peptide, called the *thylakoid transfer domain*, that dictates entry into the thylakoids. Several distinct pathways have been identified by which proteins are either inserted into the thylakoid membrane or translocated into the thylakoid lumen. These pathways exhibit striking similarities to transport systems in bacterial cells, the presumed ancestors of chloroplasts. Many of the proteins that
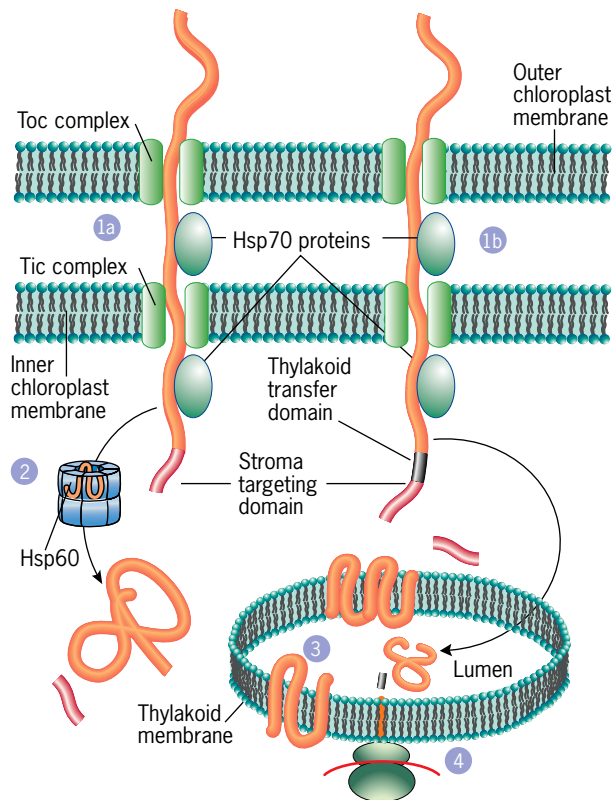
FIGURE 8.48 Importing proteins into a chloroplast. Proteins encoded by nuclear genes are synthesized in the cytosol and imported through protein-lined pores in both membranes of the outer chloroplast envelope (step 1). Proteins destined for the stroma (step 1a) contain a stroma-targeting domain at their N-terminus, whereas proteins destined for the thylakoid (step 1b) contain both a stroma-targeting domain and a thylakoid-transfer domain at their N-terminus. Stromal proteins remain in the stroma (step 2) following translocation through the outer envelope and removal of their single targeting sequence. The presence of the thylakoid transfer domain causes thylakoid proteins to be translocated either into or completely through the thylakoid membrane (step 3). A number of the proteins of the thylakoid membrane are encoded by chloroplast genes and synthesized by chloroplast ribosomes that are bound to the outer surface of the thylakoid membrane (step 4).

reside within the thylakoid membrane are encoded by chloroplast genes and synthesized on membrane-bound ribosomes of the chloroplast, as illustrated in Figure 8.48, step 4.

REVIEW ?

1. How are proteins, such as the enzymes of the TCA cycle, able to arrive in the mitochondrial matrix?
2. What is the role of cytosolic and mitochondrial chaperones in the process of mitochondrial import?
3. Distinguish between two possible import mechanisms: biased diffusion and force-generating motors.
4. Describe the steps by which a polypeptide would move from the cytosol where it is synthesized to the thylakoid lumen.

EXPERIMENTAL PATHWAYS

Receptor-Mediated Endocytosis

Embryonic development begins with the fusion of a minuscule sperm and a much larger egg. Eggs develop from oocytes, which accumulate yolk that has been synthesized elsewhere in the body of the female. How are the high-molecular-weight yolk proteins able to enter the oocyte? In 1964, Thomas Roth and Keith Porter of Harvard University reported on the mechanism by which yolk proteins might be taken into the oocytes of mosquitoes.[1] They noticed that during stages of rapid oocyte growth, there was a dramatic increase in the number of pitlike depressions seen on the surface of the oocyte. The pits, which were formed by invaginations of the plasma membrane, were covered on their inner surface by a bristly coat. In a farsighted proposal, Roth and Porter postulated that yolk proteins were specifically adsorbed onto the outer surface of the membrane of the coated pits, which would then invaginate as coated vesicles. The coated vesicles would lose their bristly coats and fuse with one another to produce the large, membrane-bound yolk bodies characteristic of the mature oocyte.

The first insight into the structure of coated vesicles was obtained in 1969 by Toku Kanaseki and Ken Kadota of the University of Osaka.[2] Electron microscopic examination of a crude vesicle fraction isolated from guinea pig brains showed that coated vesicles were covered by a polygonal basketwork (Figure 1). These investigators suggested that the coatings were an apparatus to control the infolding of the plasma membrane during vesicle formation.

The first studies of the biochemical nature of the vesicle coat were published in 1975 by Barbara Pearse of the Medical Research Council in Cambridge, England.[3] Pearse developed a procedure in which membrane vesicles from pig brains were centrifuged through a succession of sucrose density gradients until a purified fraction of coated vesicles was obtained. Protein from the coated vesicles was solubilized and fractionated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE, Section 18.7). The results indicated that the coat contained one predominant protein species having a molecular mass of approximately 180,000 daltons. Pearse named the protein clathrin. She found the same protein (based on molecular mass and peptide mapping) in preparations of coated vesicles that were isolated from several different types of cells obtained from more than one animal species.[4]

As the research described above was being conducted, an apparently independent line of research was begun in 1973 in the laboratories of Michael Brown and Joseph Goldstein of the University of Texas Medical School in Dallas. Brown and Goldstein had become interested in the inherited condition *familial hypercholesterolemia* (*FH*). People who were homozygous for the defective gene (the *FH* allele) had profoundly elevated levels of serum cholesterol (800 mg/dl vs. 200 mg/dl for a normal person), invariably developed severely blocked (atherosclerotic) arteries, and usually died from heart attack before the age of 20. At the time, very little was known about the fundamental physiologic or biochemical defects in this disorder.

Brown and Goldstein began their studies of FH by examining cholesterol metabolism in cultured fibroblasts derived from the skin of normal and FH-afflicted individuals. They found that the rate-controlling enzyme in cholesterol biosynthesis, HMG CoA reductase, could be inhibited in normal fibroblasts by cholesterol-containing lipoproteins (such as LDL) placed in the medium (Figure 2).[5] Thus, the addition of LDL to the culture medium in which normal fibroblasts were growing led to a decreased level of activity of HMG CoA reductase and a corresponding decrease in the synthesis of cholesterol by the fibroblasts. When HMG CoA reductase levels were measured in FH-derived fibroblasts, they were found to be 40 to 60 times that of normal fibroblasts.[6] In addition, enzyme activity in the FH fibroblasts was totally unaffected by the presence of LDL in the medium (Figure 3).

How could lipoproteins in the medium affect the activity of an enzyme in the cytoplasm of cultured cells? To answer this question, Brown and Goldstein initiated studies on the interaction between the cells and the lipoproteins. They added radioactively labeled LDL to culture dishes containing a single layer of fibroblasts derived from either FH-afflicted or normal human subjects.[7] The normal fibroblasts bound the labeled LDL molecules with high affinity and specificity, but the mutant cells showed virtually no ability to bind these lipoprotein molecules (Figure 4). These results indicated that normal cells have a highly specific receptor for LDL and that this receptor was defective or missing in cells from patients with FH.

To visualize the process of receptor binding and internalization, Brown and Goldstein teamed up with Richard Anderson, who had
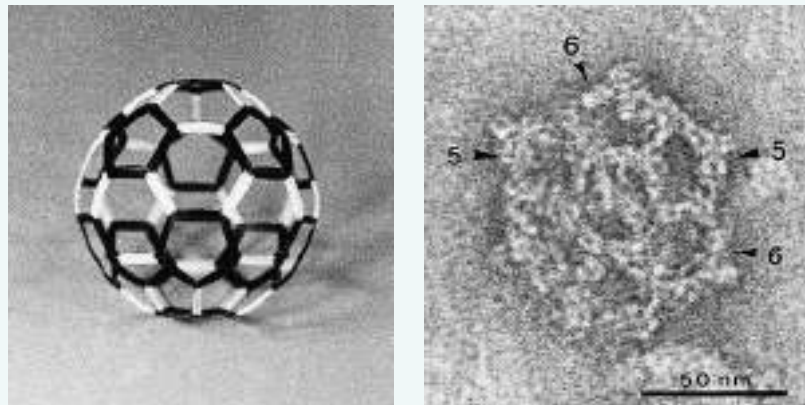
(a)

(b)

FIGURE 1 (*a*) A handmade model of an "empty basket" that would form the surface lattice of a coated vesicle. (*b*) A high-power electron micrograph of an empty proteinaceous basket. The numbers 5 and 6 refer to pentagonal and hexagonal elements of the lattice, respectively. (A–B: FROM TOKU KANASEKI AND KEN KADOTA, J. CELL BIOL. 42:216, 1969; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)
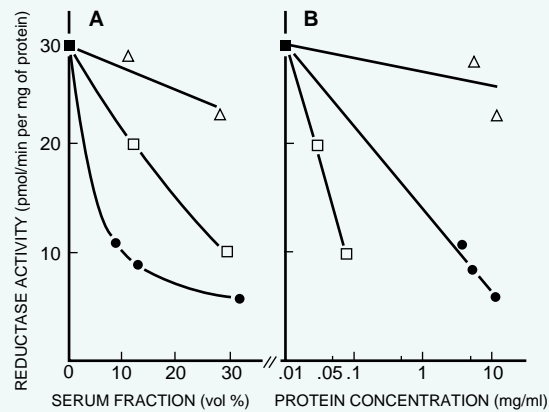
FIGURE 2 HMG CoA reductase activity in normal fibroblasts was measured following addition of the lipoprotein fraction of calf serum (open squares), unfractionated calf serum (closed circles), or nonlipoprotein fraction of calf serum (open triangles). It is evident that the lipoproteins greatly depress the activity of the enzyme, while the nonlipoproteins have little effect. (FROM M. S. BROWN ET AL., PROC. NAT'L. ACAD. SCI. U.S.A. 70:2166, 1973.)

been studying cellular structure with the electron microscope. The group incubated fibroblasts from normal and FH subjects with LDL that had been covalently linked to the iron-containing protein ferritin. Because of the iron atoms, ferritin molecules are able to scatter a beam of electrons and thus can be visualized in the electron microscope. When normal fibroblasts were incubated with LDL-ferritin at 4°C, a temperature at which ligands can bind to the cell surface but cannot be internalized, LDL-ferritin particles were seen to be bound to the cell surface. Close examination revealed that the LDL particles were not randomly scattered over the cell surface but were localized to
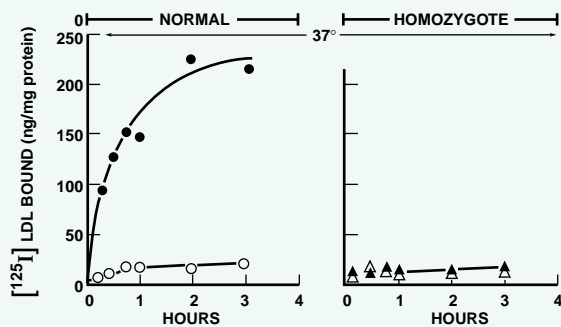
FIGURE 4 Time course of radioactive [^{125}I]-labeled LDL binding to cells from a normal subject (circles) and a homozygote with FH (triangles) at 37°C. The cells were incubated in a buffer containing 5 mg/ml [^{125}I] LDL in the presence (open circles and triangles) and absence (closed circles and triangles) of 250 mg/ml of nonradioactive LDL. It is evident that in the absence of added nonradioactive LDL, the normal cells bind significant amounts of the labeled LDL, while the cells from the FH patients do not. The binding of labeled LDL is greatly reduced in the presence of nonradioactive LDL because the nonlabeled lipoproteins compete with the labeled ones for binding sites. Thus, the binding of the lipoprotein to the cells is specific (i.e., it is not the result of some nonspecific binding event). (FROM M. S. BROWN AND J. L. GOLDSTEIN, PROC. NAT'L. ACAD. SCI. U.S.A. 71:790, 1974.)
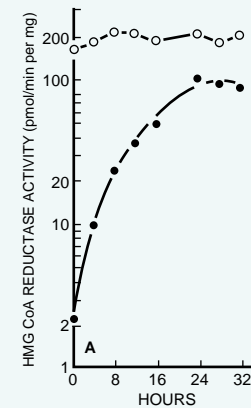
FIGURE 3 Fibroblast cells from either a control subject (closed circles) or a patient with homozygous FH (open circles) were grown in dishes containing fetal calf serum. On day 6 (which corresponds to 0 hours on the graph), the medium was replaced with fresh medium containing lipoprotein-deficient human plasma. At the indicated time, extracts were prepared, and HMG CoA reductase activity was measured. If we look at the control cells, it is apparent that at the beginning of the monitoring period the cells have very little enzyme activity because the medium had contained enough cholesterol-containing lipoproteins that the cells did not need to synthesize their own. Once the medium was changed to the lipoprotein-deficient plasma, the cells were no longer able to use cholesterol from the medium and thus increased the amount of enzyme within the cell. In contrast, the cells from the FH patients showed no response to either the presence or absence of lipoproteins in the medium. (FROM J. L. GOLDSTEIN AND M. S. BROWN, PROC. NAT'L. ACAD. SCI. U.S.A. 70:2805, 1973.)

short segments of the plasma membrane where the membrane was indented and coated by a "fuzzy" material (Figure 5).[8] These segments of the membrane were similar to the coated pits originally described by Roth and Porter and had since been seen in a variety of cell types. Al-
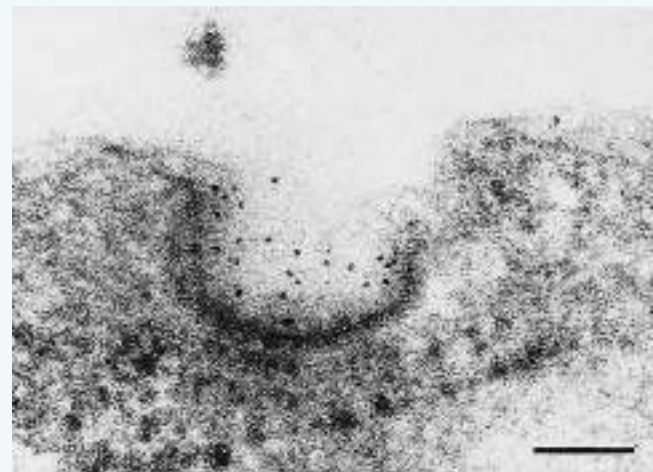
FIGURE 5 Electron micrograph showing the binding of LDL to coated pits of human fibroblasts. The LDL is made visible by conjugating the particles to electron-dense, iron-containing ferritin. (FROM R. G. W. ANDERSON, M. S. BROWN, AND J. L. GOLDSTEIN, CELL 10:356, 1977; BY PERMISSION OF CELL PRESS.)

though cells from patients with FH were found to have a similar number of coated pits on their surface, no LDL-ferritin was bound to these mutant cells. The results supported the proposal that the mutant FH allele encoded a receptor that was unable to bind LDL. Subsequent electron microscopic studies on the internalization of LDL-ferritin revealed the endocytic pathway by which these lipoprotein particles were internalized, as described in the chapter.[9]

Based on these results, the group postulated that the rapid internalization of receptor-bound LDL is strictly dependent on the localization of LDL receptors in coated pits. It follows from this postulate that, if an LDL receptor failed to become localized within a coated pit, it would be unable to deliver its bound ligand to cellular lysosomes and thus would be unable to affect cholesterol biosynthesis within the cell. At about this time, an LDL receptor with a different type of mutation was discovered. LDL receptors bearing this new defect (known as the J. D. mutation after the patient in which it occurred) bound normal amounts of radioactively labeled LDL, yet the receptor-bound lipoprotein failed to be internalized and consequently was not delivered to cytoplasmic lysosomes for processing.[10] Anderson and co-workers postulated that the LDL receptor was a transmembrane protein that normally became localized in coated pits because its cytoplasmic domain was specifically bound by a component of the coated pits, possibly clathrin (but later identified as a likely subunit of an AP adaptor, as discussed below). Because of a defect in its cytoplasmic domain, the J. D. mutant receptor was unable to become localized in a cell's coated pits. People with this mutation exhibit the same phenotype as patients whose receptors are unable to bind LDL.

Subsequent studies determined that the normal LDL receptor is a transmembrane glycoprotein of 839 amino acids, with the 50 amino acids at the C-terminal end of the protein extending inward from the membrane as a cytoplasmic domain. Analysis of the J. D. mutant receptor revealed that the protein contained a single amino acid substitution: a tyrosine residue normally located at position 807 was replaced by a cysteine.[11] This single alteration in amino acid sequence obliterated the ability of the protein to become concentrated in coated pits.

Over the next few years, attention turned to the amino acid sequences of the cytoplasmic tails of other receptors that became localized in coated pits. Studies on a diverse array of membrane receptors turned up several internalization signals that were shared by numerous membrane proteins. The two most common such signals are: (1) a YXXϕ signal (as in the transferrin receptor) where X can be any amino acid and ϕ is an amino acid with a bulky hydrophobic side chain and (2) an acidic dileucine signal containing two adjacent leucine residues (as in the CD4 protein which serves as the receptor for HIV on the surface of T lymphocytes). The YXXϕ sequence of the receptor binds to the μ subunit of the AP2 adaptors, and the dileucine motif binds to the σ subunit of the AP2 adaptors (Figure 8.40).[12] X-ray crystallographic studies have revealed the nature of the interactions between the adaptor and these internalization signals. Included within the μ subunit are two hydrophobic pockets, one that binds the tyrosine residue and the other that binds the bulky hydrophobic side chain of the YXXϕ internalization signal.[13] Similarly, the dileucine signal binds to a hydrophobic pocket in the σ subunit.[14] Meanwhile, the AP2 adaptor complex binds to the clathrin coat by means of its β subunit (Figure 8.40). As a result of these various intermolecular contacts, the adaptor complex and the receptor are trapped in coated pits prior to endocytosis.[Reviewed in 15]

Over the past decade, many different lines of investigation of clathrin-mediated endocytosis have been followed. One of the most rewarding experimental approaches has been to label two or more
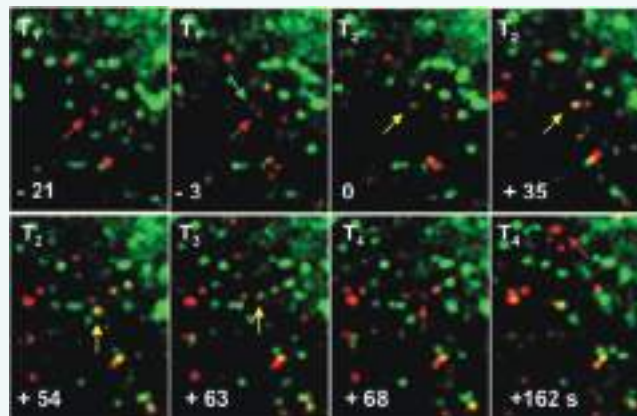
FIGURE 6 A series of fluorescence images showing the capture of a single red-fluorescent LDL particle by a green-fluorescent clathrin-coated pit and its incorporation into a clathrin-coated vesicle, which becomes uncoated and moves into the cytoplasm. The events are described in the text. T_1-T_4 represent the intervals before association (T_1), during association (T_2), joint lateral motion of LDL and clathrin prior to uncoating (T_3), and motion of LDL after uncoating (T_4), respectively. The times are indicated in seconds in each frame, with 0 corresponding to the time when the LDL and clathrin become associated. (FROM MARCELO EHRLICH ET AL., COURTESY OF TOMAS KIRCHHAUSEN, CELL 118:597, 2004; BY PERMISSION OF CELL PRESS.)

components of the endocytic machinery with different fluorescent markers and follow their movements over time within a living cell. Using this approach, researchers have watched dynamin, AP2 adaptors, or various types of cargo receptors bind to clathrin-coated pits, become enveloped within a clathrin-coated vesicle, which then buds off into the cytoplasm and disappears. Figure 6 depicts an example of this type of experiment.[16] The fluorescence micrographs show the surface of a cultured mammalian epithelial cell that is expressing a clathrin light chain that is fused to a variant of the green fluorescent protein. The green patches, for the most part, represent clathrin-coated pits situated at the cell surface. The red-staining spots are individual fluorescently labeled LDL particles that were added to the medium in which the cells were growing. The series of images illustrate a particular clathrin-coated pit (indicated by the green arrows in the early frames marked T_1) that has captured a particular LDL particle (indicated by the red arrows in these same early frames). Once the LDL particle has been bound to an LDL receptor in the coated pit, the overlap of the two fluorescent dyes produces a yellow-orange spot (indicated by the yellow arrows in the frames marked T_2 and T_3). The remaining frames (marked T_4) show the uncoated vesicle containing the red fluorescent LDL particle moving into the adjacent cytoplasm.

References

1. ROTH, T. F. & PORTER, K. R. 1964. Yolk protein uptake in the oocyte of the mosquito *Aedes aegypti. J. Cell Biol.* 20:313–332.

2. KANASEKI, T. & KADOTA, K. 1969. The "vesicle in a basket." *J. Cell Biol.* 42:202–220.

3. PEARSE, B. M. F. 1975. Coated vesicles from pig brain: Purification and biochemical characterization. *J. Mol. Biol.* 97:93–96.

4. PEARSE, B. M. F. 1976. Clathrin: A unique protein associated with the intracellular transfer of membrane by coated vesicles. *Proc. Nat'l. Acad. Sci. U.S.A.* 73:1255–1259.

5. BROWN, M. S., DANA, S. E., & GOLDSTEIN, J. L. 1973. Regulation of HMG CoA reductase activity in human fibroblasts by lipoproteins. *Proc. Nat'l. Acad. Sci. U.S.A.* 70:2162–2166.

6. GOLDSTEIN, J. L. & BROWN, M. S. 1973. Familial hypercholesterolemia: Identification of a defect in the regulation of HMG CoA reductase activity associated with overproduction of cholesterol. *Proc. Nat'l. Acad. Sci. U.S.A.* 70:2804–2808.

7. BROWN, M. S. & GOLDSTEIN, J. L. 1974. Familial hypercholesterolemia: Defective binding of lipoproteins to cultured fibroblasts associated with impaired regulation of HMG CoA reductase activity. *Proc. Nat'l. Acad. Sci. U.S.A.* 71:788–792.

8. ANDERSON, R. G. W., GOLDSTEIN, J. L., & BROWN, M. S. 1976. Localization of low density lipoprotein receptors on plasma membrane of normal human fibroblasts and their absence in cells from a familial hypercholesterolemia homozygote. *Proc. Nat'l. Acad. Sci. U.S.A.* 73:2434–2438.

9. ANDERSON, R. G. W., BROWN, M. S., & GOLDSTEIN, J. L. 1977. Role of the coated endocytic vesicle in the uptake of receptor-bound low density lipoprotein in human fibroblasts. *Cell* 10:351–364.

10. ANDERSON, R. G. W., GOLDSTEIN, J. L., & BROWN, M. S. 1977. A mutation that impairs the ability of lipoprotein receptors to localise in coated pits on the cell surface of human fibroblasts. *Nature* 270:695–699.

11. DAVIS, C. G., ET AL. 1986. The J. D. mutation in familial hypercholesterolemia: Amino acid substitution in cytoplasmic domain impedes internalization of LDL receptors. *Cell* 45:15–24.

12. OHNO, H., ET AL. 1995. Interaction of tyrosine-based sorting signals with clathrin-associated proteins. *Science* 269:1872–1874.

13. OWEN, D. J. & EVANS, P. R. 1998. A structural explanation for the recognition of tyrosine-based endocytic signals. *Science* 282:1327–1332.

14. KELLY, B. T. ET AL. 2008. A structural explanation for the binding of endocytic dileucine motifs by the AP2 complex. *Nature* 456:976–979.

15. UNGEWICKELL, E. J. & HINRICHSEN, L. 2007. Endocytosis: Clathrin-mediated membrane budding. *Curr. Opin. Cell Biol.* 19:417–425.

16. EHRLICH, M. ET AL. 2004. Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell* 118:591–605.

SYNOPSIS

The cytoplasm of eukaryotic cells contains a system of membranous organelles, including the endoplasmic reticulum, Golgi complex, and lysosomes, that are functionally and structurally related to one another and to the plasma membrane. These various membranous organelles are part of a dynamic, integrated endomembrane network in which materials are shuttled back and forth as part of transport vesicles that form by budding from one compartment and fusing with another. A biosynthetic (secretory) pathway moves proteins from their site of synthesis in the ER through the Golgi complex to their final destination (an organelle, the plasma membrane, or the extracellular space), whereas an endocytic pathway moves material in the opposite direction from the plasma membrane or extracellular space to the cell interior. Cargo is directed to its proper destination by specific targeting signals that are part of the proteins themselves. (p. 265)

The endoplasmic reticulum (ER) is a system of tubules, cisternae, and vesicles that divides the cytoplasm into a luminal space within the ER membranes and a cytosolic space outside the membranes. The ER is divided into two broad types: the rough ER (RER), which occurs typically as flattened cisternae and whose membranes have attached ribosomes, and the smooth ER (SER), which occurs typically as tubular compartments and whose membranes lack ribosomes. The functions of the SER vary from cell to cell and include the synthesis of steroid hormones, detoxification of a wide variety of organic compounds, and sequestration of calcium ions. The functions of the RER include the synthesis of secreted proteins, lysosomal proteins, and integral membrane proteins. (p. 273)

Proteins to be synthesized on membrane-bound ribosomes of the RER are recognized by a hydrophobic signal sequence, which is usually situated near the N-terminus of the nascent polypeptide. As the signal sequence emerges from the ribosome, it is bound by a signal recognition particle (SRP) that arrests further synthesis and mediates the binding of the complex to the RER membrane. Following binding, the SRP is released from the membrane, and the nascent polypeptide passes through a protein-lined pore in the ER membrane and into the ER lumen. Lysosomal and secretory proteins are translocated into the lumen in their entirety, whereas membrane proteins become embedded in the lipid bilayer by virtue of one or more hydrophobic transmembrane sequences. Proteins that are not folded correctly are translocated back into the cytosol and destroyed. Once a newly synthesized protein is situated in either the lumen or the membrane of the RER, the protein can be moved from that location to specific destinations along the biosynthetic pathway. If the lumen of the ER becomes "choked" by an overabundance of newly synthesized proteins, a comprehensive response called the UPR stops further synthesis of ER proteins and promotes removal of those already present. (p. 276)

Most of the lipids of a cell's membranes are also synthesized in the ER and moved from that site to various destinations. Phospholipids are synthesized at the cytosolic face of the ER and inserted into the outer leaflet of the ER membrane. Some of these molecules are subsequently flipped into the opposite leaflet. The lipid composition of membranes is subject to modification in a number of ways. For example, phospholipids may be selectively transported from the membrane of one organelle to another, or the head groups of specific lipids may be modified enzymatically. (p. 279)

The addition of sugars (glycosylation) to the asparagine residues of proteins begins in the rough ER and continues in the Golgi complex. The sequences of sugars that make up the oligosaccharide chains of glycoproteins are determined by the types and locations of members of a large family of glycosyltransferases, enzymes that transfer a specific sugar from a nucleotide sugar donor to a specific acceptor. The carbohydrate chains are assembled in the ER one sugar at a time and then transferred as a unit from the dolichol carrier to an asparagine residue of the polypeptide. Almost as soon as the carbohydrate block is transferred, it begins to be modified—first by removal of the terminal glucose residues. Membrane vesicles, with their enclosed cargo, bud from the edges of the ER and are targeted to the Golgi complex. (p. 280)

The Golgi complex functions as a processing plant, modifying the membrane components and cargo synthesized in the ER before it moves on to its target destination. The Golgi complex is also the site of synthesis of the complex polysaccharides that make up the matrix of the plant cell wall. Each Golgi complex consists of a stack of flattened, platelike cisternae with dilated rims and associated vesicles and tubules. Materials enter the stack at the *cis* face and are mod-

ified as they move by vesicle transport toward the opposite, or *trans*, face. As they traverse the stack, sugars are added to the oligosaccharide chains by glycosyltransferases located in particular Golgi cisternae. Once the proteins reach the *trans* Golgi network (TGN) at the end of the stack, they are ready to be sorted and targeted for delivery to their ultimate cellular or extracellular destination. (p. 284)

Most, if not all, of the vesicles that transport material through the endomembrane system are encased initially in a protein coat. Several distinct types of coated vesicles have been identified. COPII-coated vesicles carry materials from the ER to the Golgi complex. COPI-coated vesicles carry materials in the opposite (retrograde) direction, from the Golgi back to the ER. Clathrin-coated vesicles carry materials from the TGN to the endosomes, lysosomes, and plant vacuole. Clathrin-coated vesicles are also part of the endocytic pathway, carrying materials from the plasma membrane to endosomes and lysosomes. (p. 288)

Each compartment along the biosynthetic or endocytic pathway has a characteristic protein composition. Resident proteins tend to be retained in that particular compartment and retrieved if they escape to other compartments. The sorting of proteins in the biosynthetic pathway occurs to a large extent in the last of the Golgi compartments, the TGN. The TGN is the source of vesicles that contain particular membrane proteins that target the vesicle to a particular destination. Lysosomal enzymes produced in the rough ER are sorted in the TGN and targeted for lysosomes in clathrin-coated vesicles. Lysosomal enzymes are enclosed in these budding vesicles because they possess phosphorylated mannose residues on their core oligosaccharides. These modified oligosaccharides are recognized by membrane receptors (MPRs), which in turn are bound to a class of adaptors (GGA proteins) that form a layer between the outer clathrin coat and the vesicle membrane. (p. 289)

Vesicles that bud from a donor compartment possess specific proteins in their membrane that recognize proteins located in the target (acceptor) compartment. The docking of the two membranes is mediated by tethering proteins and regulated by a large family of G proteins called Rabs. Fusion of donor and acceptor membranes is mediated by v-SNAREs and t-SNARES, which interact to form four-stranded bundles. (p. 294)

Lysosomes are diverse-appearing, membrane-bound organelles that contain an array of acid hydrolases capable of digesting virtually every type of biological macromolecule. Among their numerous roles, lysosomes degrade materials, such as bacteria and debris, that are brought into the cell by phagocytosis; they degrade aging cytoplasmic organelles by a process called autophagy; and they digest various macromolecules that are delivered by way of endosomes by receptor-mediated endocytosis. In vertebrates, lysosomes play a key role in immunologic defense. (p. 297)

The vacuoles of plants carry out a diverse array of activities. Vacuoles serve as a temporary storehouse for solutes and macromolecules; they contain toxic compounds used in defense; they provide a container for cellular wastes; they maintain a hypertonic compartment that exerts turgor pressure against the cell wall; and they are sites of intracellular digestion mediated by acid hydrolases. (p. 301)

Endocytosis facilitates the uptake of fluid and suspended macromolecules, the internalization of membrane receptors and their bound ligands, and functions in the recycling of membrane between the cell surface and cytoplasm. Phagocytosis is the uptake of particulate matter. In receptor-mediated endocytosis, specific ligands are bound to plasma membrane receptors. The receptors collect in pits of the membrane that are coated on their cytoplasmic surface by a polygonal scaffold of clathrin molecules. The coated pits give rise to coated vesicles, which lose their coat, and deliver their contents to an endosome and ultimately to a lysosome. Phagocytosis can function as a feeding mechanism or as a cellular system of defense. (p. 302)

Proteins within peroxisomes, mitochondria, or chloroplasts that are encoded by nuclear genes must be imported into the organelle posttranslationally. In all of these cases, the proteins to be imported carry signal sequences that interact with receptors responsible for protein import. All three of these organelles contain protein-lined channels in their membranes that promote the translocation of either folded polypeptides (in peroxisomes) or unfolded polypeptides (in mitochondria and chloroplasts) into the organelle. Mitochondria and chloroplasts contain several different compartments into which newly translocated proteins are directed. (p. 309)

ANALYTIC QUESTIONS

1. Which, if any, of the following polypeptides would be expected to lack a signal peptide: collagen, acid phosphatase, hemoglobin, ribosomal proteins, glycophorin, proteins of the tonoplast?

2. Suppose you suspected that a patient might be suffering from I-cell disease (page 299). How could you decide if this diagnosis was accurate using cells that were cultured from the patient?

3. In what cellular compartment would a glycoprotein be expected to have its greatest mannose content? its greatest *N*-acetylglucosamine content? its greatest sialic acid content?

4. It was noted on page 309 that peroxisomes are able to import folded proteins. Yet, these same organelles are impermeable to relatively small molecules such as NADH and acetyl-CoA. How is it possible to be permeable to one and not the other?

5. Name two proteins you would expect to find as integral components of the RER membrane that would be absent from the SER? Name two proteins of the SER that would not be present in the RER?

6. Suppose you wanted to study the process of regulated secretion in a cell that lacked mature secretory granules, that is, vesicles containing secretory material that was ready to be discharged. How might you obtain a cell that lacked such granules?

7. It was described on page 300 how glucocerebrosidase could be prepared that bore mannose residues at the end of its surface oligosaccharides rather than the usual sugar, sialic acid. A similar version of glucocerebrosidase with exposed mannose residues is being produced by recombinant DNA technology using a special line of cells. What characteristics do you think these cells might exhibit? Clue: Information to answer this question is provided in Figure 8.22.

8. Autoradiography depends on particles emitted from radioactive atoms striking a photographic emulsion that lies on top of the tissue section. When the emulsion is developed, the site where the particle struck the emulsion appears as a silver grain, as in Figure 8.3*a*. How do you think the thickness of the section might affect the resolution of the technique, that is, the ability

to locate the precise site in the cell where radioactivity is incorporated?

9. In which part of a cell would you expect the following compounds to be first incorporated: [^{3}H]leucine, [^{3}H]sialic acid, [^{3}H]mannose, [^{3}H]choline, [^{3}H]glucuronic acid (a precursor of GAGs), [^{3}H]pregnenolone (a precursor of steroid hormones), [^{3}H]rhamnose in a plant cell (rhamnose is a precursor of pectins)?

10. Which of the following cells would you expect to be engaged most heavily in bulk-phase endocytosis: (a) an erythrocyte, (b) a pancreatic acinar cell, (c) a skeletal muscle cell? Why?

11. Would you expect the properties of the cisternal side of Golgi membranes to be more similar to the extracellular or cytosolic side of the plasma membrane? Why?

12. Which compartment(s) of a cell is associated with each of the following: clathrin, calcium ions in a skeletal muscle cell, dolichol phosphate, ribonuclease and lipase, digitalis, LDL receptors, COPI proteins, COPII proteins, unbound SRPs?

13. If a slice of pancreatic tissue had been incubated in [^{3}H]leucine continuously for 2 hours, where in the acinar cells would you expect to find incorporated radioactivity?

14. If you were to add a drug that interfered with the ability of ribosomes to bind to mRNA, what effect would this be expected to have on the structure of the RER?

15. Not all receptors that carry out RME are located in coated pits prior to binding ligand, yet they too become concentrated in coated pits prior to internalization. How do you suppose the binding of a ligand to a receptor would cause it to become concentrated into a coated pit?

16. It was noted in the text that lysosomes lack a distinctive appearance in the electron microscope. How might you determine whether a particular vacuole is in fact a lysosome?

17. Studies of a rare inherited disorder, Dent's disease, revealed that these individuals lacked a chloride ion channel in the endosomes of the cells of their kidney tubules. These patients suffered from a variety of symptoms that suggested that their endosomal compartments were not as acidic as those of normal individuals. How do you suppose that a defect in a chloride channel could account for this condition?

18. Examine the fluorescence micrograph in Figure 8.20c. Although the Golgi complex is brightly stained, there is scattered red fluorescence in other parts of the cell. How can you explain the presence of this dispersed staining?

19. Figure 8.8b shows the localization of GFP-labeled mannosidase II in a cell treated with an siRNA against an mRNA encoding a protein involved in an early step in the secretory pathway. How would you expect the fluorescence pattern to appear in a cell that had been treated with an siRNA against an mRNA encoding the protein GGA? Or the protein clathrin? Or a protein of the signal recognition particle?
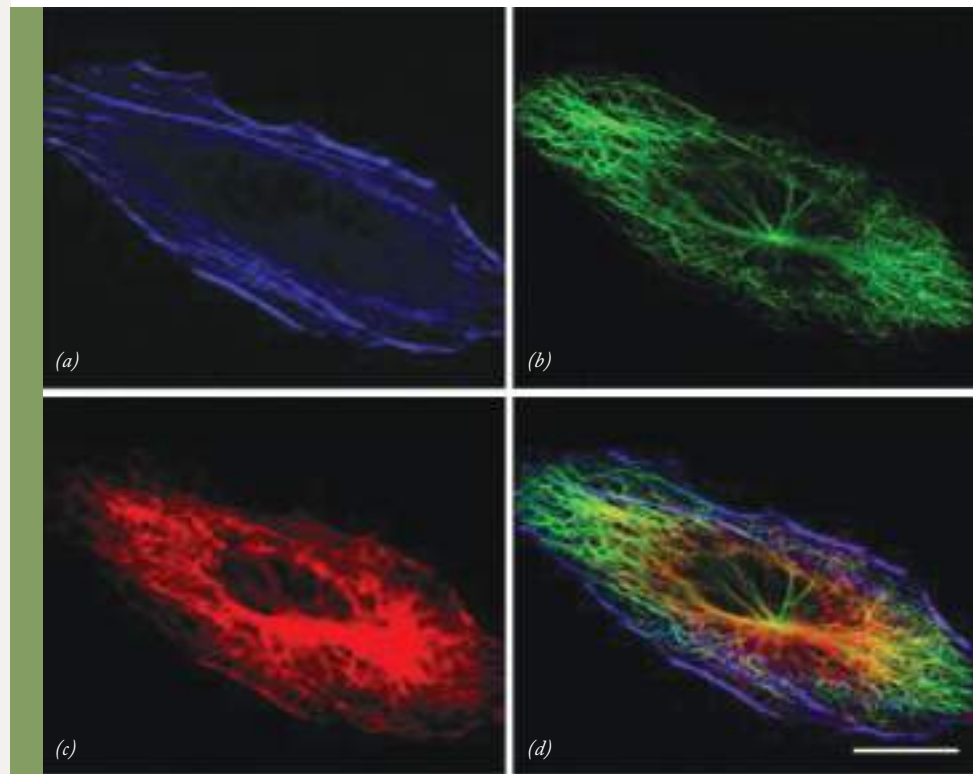
The Cytoskeleton and Cell Motility

The skeleton of a vertebrate is a familiar organ system consisting of hardened elements that support the soft tissues of the body and play a key role in mediating bodily movements. Eukaryotic cells also possess a "skeletal system"—a **cytoskeleton**—that has analogous functions. The cytoskeleton is composed of three well-defined filamentous structures—microtubules, microfilaments, and intermediate filaments—that together form an elaborate interactive network. Each of the three types of cytoskeletal filaments is a polymer of protein subunits held together by weak, noncovalent bonds. This type of construction lends itself to rapid assembly and disassembly, which is dependent on complex cellular regulation. Each cytoskeletal element has distinct properties. **Microtubules** are long, hollow, unbranched tubes composed of subunits of the protein tubulin. **Microfilaments** are solid, thinner structures, often organized into a branching network and composed of the protein actin. **Intermediate filaments** are tough, ropelike fibers composed of a variety of related proteins. The properties of microtubules, intermediate filaments, and actin filaments are summarized in Table 9.1. Although the components of the cytoskeleton appear stationary in micrographs, in reality they are highly dynamic structures capable of dramatic reorganization. Until recently it was widely held that the cytoskeleton was strictly a eukaryotic innovation that was absent from prokaryotic cells. We now know that numerous

The cytoskeleton is constructed from three major types of filaments, which are shown imaged separately in parts a–c. *The distribution of each type of filament within a cultured liver cell (hepatocyte) is revealed by a fluorescent probe that binds specifically to that type of filament. The distribution of microfilaments* (a) *is revealed by fluorescent phalloidin binding; the distribution of microtubules* (b) *by fluorescent anti-tubulin antibodies; and the distribution of intermediate filaments* (c) *by fluorescent anti-keratin antibodies. The image in* d *is a composite of the three previous images and demonstrates that the cytoplasmic organization of each filament type is distinct. Bar equals 10 μm.* (FROM M. BISHR OMARY ET AL., TRENDS BIOCHEM. SCI. 31:385, 2006. COPYRIGHT 2006 WITH PERMISSION FROM ELSEVIER SCIENCE.)

TABLE 9.1 Properties of Microtubules, Intermediate Filaments, and Actin Filaments

	Microtubules	Intermediate filaments	Actin filaments
Subunits incorporated into polymer	GTP-αβ-tubulin heterodimer	~70 different proteins	ATP-actin monomers
Preferential site of incorporation	+ End (β-tubulin)	Internal	+ End (barbed)
Polarity	Yes	No	Yes
Enzymatic activity	GTPase	None	ATPase
Motor proteins	Kinesins, dyneins	None	Myosins
Major group of associated proteins	MAPs	Plakins	Actin-binding proteins
Structure	Stiff, hollow, inextensible tube	Tough, flexible, extensible filament	Flexible, inextensible helical filament
Dimensions	25 nm outer diam.	10-12 nm diam.	8 nm diam.
Distribution	All eukaroytes	Animals	All eukaryotes
Primary functions	Support, intracellular transport, cell organization	Structural support	Motility, contractility
Subcellular distribution	Cytoplasm	Cytoplasm + nucleus	Cytoplasm

prokaryotes contain tubulin- and actin-like proteins that polymerize into cytoplasmic filaments that carry out cytoskeletal-like activities. Proteins distantly related to those of intermediate filaments have also been discovered in certain prokaryotes, so it is apparent that all three types of cytoskeletal elements had their evolutionary roots in prokaryotic structures. We will begin with a brief overview of major cytoskeletal activities. ■

9.1 OVERVIEW OF THE MAJOR FUNCTIONS OF THE CYTOSKELETON

An overview of the major activities of the cytoskeleton in three different nonmuscle cells is shown in Figure 9.1. The cells depicted in this schematic illustration include a polarized epithelial cell, the tip of an elongating nerve cell, and a divid-

Key to Cytoskeletal Functions

(1) Structure and Support   **(2)** Intracellular Transport   **(3)** Contractility and Motility   **(4)** Spatial Organization

(a) **Epithelial cell**   (b) **Nerve cell**   (c) **Dividing cell**
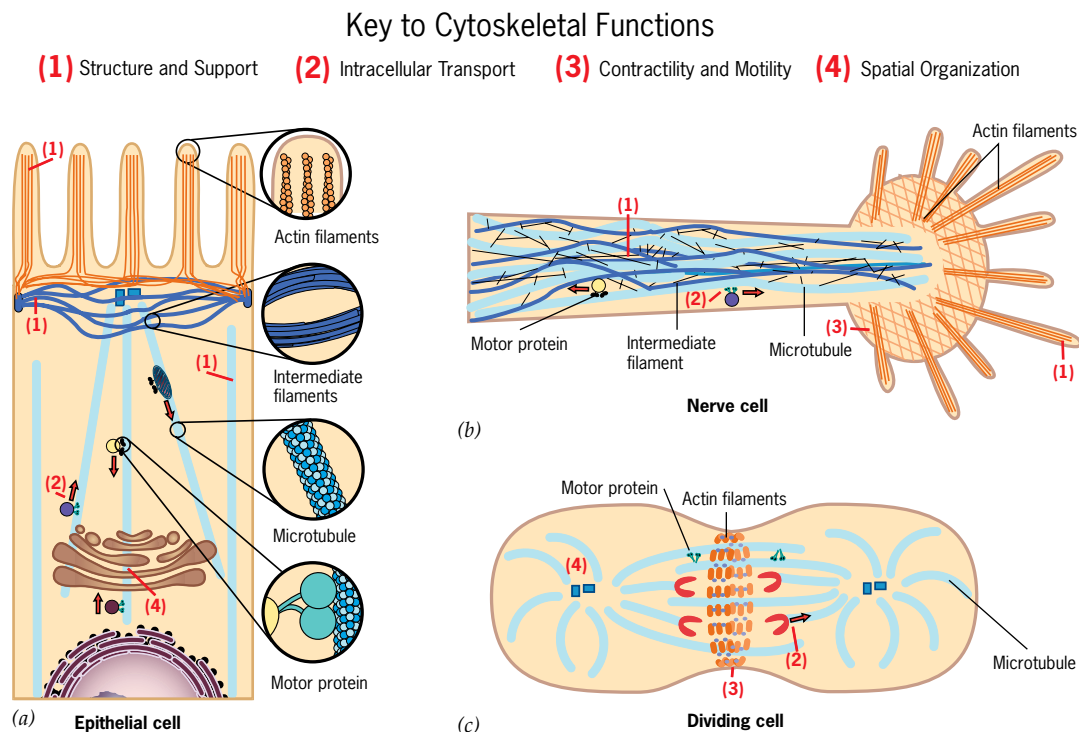
FIGURE 9.1 An overview of the structure and functions of the cytoskeleton. Schematic drawings of (*a*) an epithelial cell, (*b*) a nerve cell, and (*c*) a dividing cell. The microtubules of the epithelial and nerve cells function primarily in support and organelle transport, whereas the microtubules of the dividing cell form the mitotic spindle required for chromosome segregation. Intermediate filaments provide structural support for both the epithelial cell and nerve cell. Microfilaments support the microvilli of the epithelial cell and are an integral part of the motile machinery involved in nerve cell elongation and cell division.

ing cultured cell. As will be discussed in this chapter, the cytoskeleton of these cells functions as

1. A dynamic scaffold providing structural support that can determine the shape of the cell and resist forces that tend to deform it.

2. An internal framework responsible for positioning the various organelles within the interior of the cell. This function is particularly evident in polarized epithelial cells, such as those depicted in Figure 8.11, in which certain organelles are arranged in a defined order from the apical to the basal end of the cell.

3. A network of tracks that direct the movement of materials and organelles within cells. Examples of this function include the delivery of mRNA molecules to specific parts of a cell, the movement of membranous carriers from the endoplasmic reticulum to the Golgi complex, and the transport of vesicles containing neurotransmitters down the length of a nerve cell. Figure 9.2 shows a small portion of a cultured cell, and it is evident that most of the fluorescent green organelles, which are labeled peroxisomes (page 200), are closely associated with the microtubules (red) of the cell's cytoskeleton. Microtubules are the tracks over which the peroxisomes are transported in mammalian cells.

4. The force-generating apparatus that moves cells from one place to another. Single-celled organisms move either by "crawling" over the surface of a solid substratum or by propelling themselves through their aqueous environment with the aid of specialized, microtubule-containing locomotor organelles (cilia and flagella) that protrude from the cell's surface. Multicellular animals have a variety of cells that are capable of independent locomotion,

including sperm, white blood cells, and fibroblasts (Figure 9.3). The tip of a growing axon is also highly motile (Figure 9.1), and its movement resembles a crawling cell.

5. An essential component of the cell's division machinery. Cytoskeletal elements make up the apparatus responsible for separating the chromosomes during mitosis and meiosis and for splitting the parent cell into two daughter cells during cytokinesis. These events are discussed at length in Chapter 14.

9.2 THE STUDY OF THE CYTOSKELETON

The cytoskeleton is currently one of the most actively studied subjects in cell biology, due in large part to the development of techniques that allow investigators to pursue a coordinated morphological, biochemical, and molecular approach. As a result, we know a great deal about the families of proteins that make up the cytoskeleton; how the subunits are organized into their respective fibrous structures; the molecular "motors" that move along these filaments and generate the forces required for motile activities; and the dynamic qualities that control the spatial organization, assembly, and disassembly of the various elements of the cytoskeleton. A few of the most important approaches to the study of the cytoskeleton will be briefly surveyed.

The Use of Live-Cell Fluorescence Imaging

The concept that eukaryotic cells contained a network of cytoskeletal elements was first derived from studies of tissue sections with the electron microscope in the 1950s and 1960s. But the electron microscope produces only static images and failed to provide much insight into the dynamic structure and function of the various components of the cytoskeleton. Our appreciation of the dynamic character of the cytoskeleton has

5 μm

FIGURE 9.2 **An example of the role of microtubules in transporting organelles.** The peroxisomes of this cell (shown in green and indicated by arrows) are closely associated with microtubules of the cytoskeleton (shown in red). Peroxisomes appear green because they contain a peroxisomal protein fused to the green fluorescent protein (page 267). Microtubules appear red because they are stained with a fluorescently labeled antibody. (FROM E. A. C. WIEMER ET AL., J. CELL BIOL. 136:78, 1997, COURTESY OF S. SUBRAMANI; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

FIGURE 9.3 **The capability of the cytoskeleton to change its three-dimensional organization is apparent in this mouse fibroblast that is migrating over the 90° edge of a coverslip.** Bar, 30 μm. (FROM GUENTER ALBRECHT-BUEHLER, INT. REV. CYTOL. 120:211, 1990.)

increased greatly over the past few decades as the result of a revolution in light microscopy. The fluorescence microscope (Section 18.1) has played a leading part in this revolution by allowing researchers to directly observe molecular processes in living cells—an approach known as *live-cell imaging*.

In the most widely used strategy, fluorescently labeled proteins are synthesized within a cell as a fusion protein containing GFP. This technique was discussed in the previous chapter (page 267) and an example is shown in Figure 9.2. In an alternate approach, the protein subunits of cytoskeletal structures (e.g., purified tubulin or keratin) are fluorescently labeled in vitro by covalent linkage to a small fluorescent dye.

The labeled subunits are then microinjected into a living cell where they are incorporated into the polymeric form of the protein, such as a microtubule or an intermediate filament. The dynamic behavior of the fluorescent structure can be followed over time as the cell participates in its normal activities. Figure 9.4 shows the dramatic changes in length and orientation of individual microtubules at the leading edge of a cell that had been injected with fluorescently labeled tubulin.

If particularly small amounts of fluorescently labeled protein are injected, the cytoskeletal filaments are no longer uniformly labeled as in Figure 9.4, but instead contain irregularly spaced fluorescent speckles. These fluorescent speckles serve as fixed markers to follow dynamic changes in length or orientation of the filament. An example of this technique, which is known as *fluorescence speckle microscopy*, is illustrated in Figure 9.27.

Fluorescence microscopy can also be used to reveal the location within a cell of a protein present in very low concentration. This type of experiment is best carried out with fluorescently labeled antibodies that bind with high affinity to the protein being sought. Antibodies are particularly useful because they can distinguish between closely related variants (isoforms) of a protein. The protein can be localized by injecting the labeled antibody into a living cell, which may also reveal the function of the target protein because attachment of the antibody generally destroys the ability of the protein to perform its normal activity. Alternatively, the location of the target protein can be determined by adding the fluorescent antibody to fixed cells or tissue sections, as illustrated in Figure 9.5.

FIGURE 9.4 Dynamic changes in length of microtubules within an epithelial cell. The cell was injected with a small volume of tubulin that had been covalently linked to the fluorescent dye rhodamine. After allowing time for the cell to incorporate the labeled tubulin into microtubules, a small portion of the edge of the living cell was examined under the fluorescence microscope. The elapsed time is shown at the lower right of each image. The microtubules have been false-colored to increase their contrast. A "pioneering" microtubule (arrowhead) can be seen to grow out to the leading edge of the cell, where it becomes bent and then continues its growth in a direction parallel to the cell's edge. The rate of growth of bent parallel microtubules was much greater than those growing perpendicular to the cell's edge. Bar, 10 μm (FROM CLARE M. WATERMAN-STORER AND E. D. SALMON, J. CELL BIOL. 139:423, 1997; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

FIGURE 9.5 The localization of a protein within a cell using fluorescent antibodies. This algal cell was stained with fluorescent antibodies (yellow-green color) directed against a protein called centrin. Centrin is seen to be localized within the cell's flagella and the rootlike structure at the base of the flagella. The red color of the cell results from autofluorescence of chlorophyll molecules of this photosynthetic alga. (FROM MARK A. SANDERS AND JEFFREY L. SALISBURY OF THE MAYO CLINIC, FROM CELL 70:533, 1992; BY PERMISSION OF CELL PRESS.)

The Use of In Vitro and In Vivo Single-Molecule Assays

The digital images captured by modern video cameras possess exceptional contrast and can be enhanced by computer-assisted image processing. These features allow objects that are ordinarily invisible in the light microscope, such as 25-nm microtubules or 40-nm membrane vesicles, to be observed and photographed. The advent of high-resolution video microscopy has led to the development of *in vitro motility assays.* These assays make it possible to detect the activity of an individual protein molecule acting as a molecular motor in "real time."[1] Single-molecule assays have allowed researchers to make measurements that were not possible with standard biochemical techniques that average the results obtained from large numbers of molecules. In some of the earlier assays, microtubules were attached to a glass coverslip. Then, microscopic beads containing attached motor proteins were placed directly onto the microtubules using aimed laser beams. The laser beams are shone through the objective lens of a microscope, producing a weak attractive force near the point of focus. Because it can grasp microscopic objects, this apparatus is referred to as *optical tweezers.* When ATP is present as an energy source, the movements of a bead along a microtubule can be followed by a video camera, revealing the size of individual steps taken by the motor protein. Focused laser beams can also

---

[1]Background information on molecular motors can be found on page 328.

be used to "trap" a single bead and determine the minute forces (measured as a few piconewtons, pN) generated by a single motor protein as it "tries" to move the bead against the force exerted by the optical trap (see Figure 11.5 for an illustration of this type of experiment).

While the optical trap experiments make it possible to detect the motility of motor proteins in vitro, these measurements are indirect because the researchers are monitoring the motility of beads, not protein molecules. By combining fluorescence microscopy with specialized imaging techniques, scientists have been able to directly visualize the movement of individual motor molecules both in vitro and in living cells. Examples of this experimental approach are illustrated in Figure 9.6. For the in vitro assays, the gene for the kinesin motor was fused to GFP (as discussed on page 267), the fusion gene was introduced into cultured cells, and the cells were allowed to produce the labeled kinesin protein, which was then purified. This procedure provided the researchers with green fluorescent motor molecules. Meanwhile, the team prepared microtubules that were labeled with a red fluorescent dye. When these two preparations were combined under the appropriate conditions, the researchers were able to follow the movements of an individual GFP-labeled kinesin molecule down the length of a single microtubele, as seen in the subsequent frames of Figure 9.6a. The clarity of these images is made possible by the use of a specialized type of laser-based fluorescence microscopy called TIRF (total internal reflection microscopy), which allows the observer to focus on a very thin
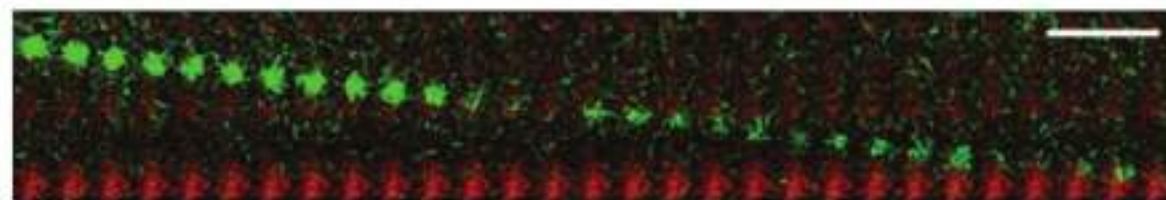
(a)
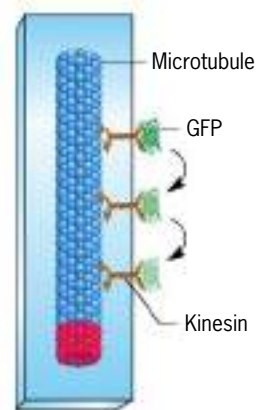
(b)

FIGURE 9.6 **Following the movements of individual, fluorescently labeled kinesin molecules along microtubules both in vitro and in vivo.** (*a*) In this experiment a purified GFP-labeled kinesin molecule (green) is seen to move processively along a microtubule whose plus end is labeled with a red fluorescent dye named Cy5. The successive images, from left to right, represent separate micrographs taken at increasing time periods over the course of the experiment. The kinesin is moving at an average speed of 0.77 μm per second. (The intensity of the green fluorescence diminishes over time as the result of photobleaching that occurs during observation.) Bar, 2 μm. (*b*) Schematic illustration of the experiment shown in part *a*. In the actual experiment, more than one GFP molecule is bound per kinesin molecule. (*c*) The image shows a portion of a living cell that has synthesized Cy5-labeled tubulin (which is assembled into red-labeled microtubules) and GFP-labeled kinesin molecules. Several of the green kinesin molecules can be seen bound to the microtubules. The image seen here represents a single point in time. When observed over a time period, the kinesin molecules are seen to scurry along the microtubules. (Kinesin movement can be seen by watching movie 6 of this paper at www.biophysj.org) (A,C: FROM DAWEN CAI, KRISTEN J. VERHEY, AND EDGAR MEYHÖFER, BIOPHYS. J. 92:4137, 2007; © 2007 BY THE BIOPHYSICAL SOCIETY.)
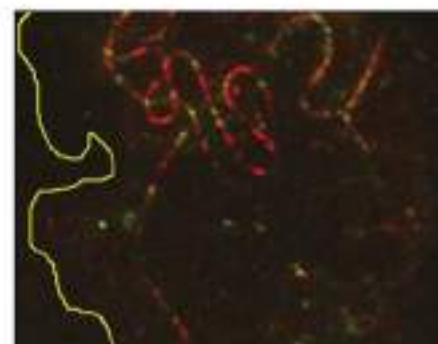
(c)
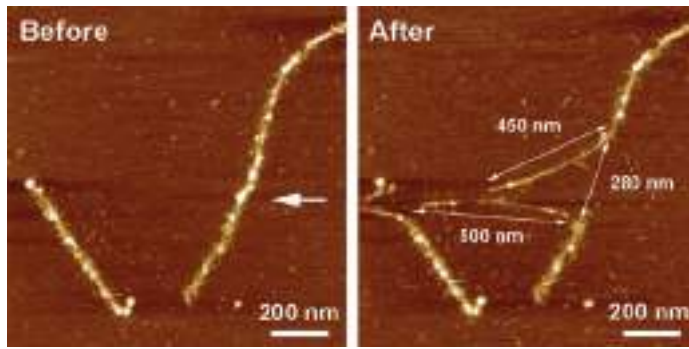
FIGURE 9.7 Determining the mechanical properties of single cytoskeletal filaments. These images show a single intermediate filament before and after it has been subjected to mechanical forces by the tip of an atomic force microscope. The tip is attached at one point along the filament (arrow) and has moved to the left, pulling on the affected segment. This segment has been stretched more than three times its original length (from 280 nm to a stretched length of 500 + 450 nm). It can be noted that the stretched segment of filament is thinner than the unstretched sections, which suggests that the subunits that make up these polymers (see Figure 9.41) may be capable of sliding over one another as the filament is being pulled. (FROM LAURENT KREPLAK, ET AL., J. MOL. BIOL. 354:574, 2005.)

plane that is just above the surface on which the microtubules are lying. As a result, fluorescent light rays from other areas in the field of view do not obscure the information that emanates from the individual protein molecules whose activities are being monitored. Using this type of experimental approach, scientists can compare the properties of different motor proteins, measure the properties of individual motors under different experimental conditions, or ask how a specific mutation of a motor protein affects its motility properties.

Using similar technology, scientists have been able to visualize the movement of individual kinesins inside living cells. An example of this type of experiment is shown in Figure 9.6c. To carry out this experiment, cells are genetically engineered to synthesize both fluorescently labeled microtubule polymers and fluorescently labeled kinesin molecules. Researchers can then watch the interaction of the two types of labeled proteins as they take place within the cell's cytoplasm. As in the in vitro experiments, the movement of individual GFP-kinesin molecules are followed using TIRF microscopy. Together these techniques are giving scientists a detailed look at motor proteins in action.

Techniques have also been developed to measure the mechanical properties of the cytoskeletal elements themselves. The atomic force microscope (AFM, Section 18.3) is an instrument that uses a nanosized tip to probe the surface of a macromolecular specimen. One group of investigators has found that they can embed the tip of an AFM into a single intermediate filament and pull on the end or the middle of the filament to test its extensibility and tensile strength. Figure 9.7 shows the results of such an experiment in which a short section of an immobilized intermediate filament is being pulled out of its linear conformation. It is demonstrated in these experiments that a segment of filament can be mechan-

ically stretched up to 3.5 times its normal length before it breaks into two pieces. These results confirm the suggestion that intermediate filaments are much more extensible than either microfilaments or microtubules.

The development of techniques for working with single molecules has coincided with the creation of a new field of mechanical engineering called **nanotechnology**. A goal of nanotechnologists is the development of tiny "nanomachines" (in the 10–100 nm size range) capable of performing specific activities in a submicroscopic world. There are many perceived uses of nanomachines, including a role in medicine. It is envisioned that, one day, such machines might be introduced into the human body where they could carry out a specific task, such as finding and destroying individual cancer cells. As it happens, nature has already evolved nanosized machines, including the motor proteins that will be discussed in this chapter. Not surprisingly, a number of nanotechnology labs have begun to use these nanosized "trucks" to haul molecular cargoes unlike any found within living organisms.

The Use of Fluorescence Imaging Techniques to Monitor the Dynamics of the Cytoskeleton

Within living cells, the microtubules and actin filaments of the cytoskeleton are both highly dynamic polymer systems. The word "dynamic" is used frequently in this text and it is worthwhile to consider the term's implications. To a cell biologist, "dynamic" implies "ever-changing." Dynamic structures are continually undergoing some type of rebuilding, whether their old parts are being exchanged with new ones or they are being entirely broken down and rebuilt. Thus, even though a microtubule or actin filament appears unchanging from one moment to the next under a microscope, in reality, its molecular components are in a constant state of flux. Just as importantly, the dynamic properties of these structures are closely regulated by the cell and may change from one place in the cytoplasm to another or at different times during the cell cycle. Monitoring these changes requires the use of techniques that can measure polymer dynamics. One powerful technique is called *fluorescence recovery after photobleaching* (or *FRAP*) and was discussed earlier in conjunction with the movement of membrane proteins (page 137). Here we will examine its application to the study of the dynamic properties of microtubules. To carry out a FRAP experiment, a cell is either injected with tubulin coupled to a fluorescent dye or induced to express GFP-tubulin. Using a microscope equipped with a laser that can be focused on a small region of the cell, the fluorescence of the tubulin in that region is bleached by the laser beam. It is then possible to follow the specimen over time and measure the recovery of the fluorescent signal into the bleached zone (Figure 9.8a,c). As with most methods used in cell and molecular biology, there are often several possible interpretations of an experimental result. For example, the recovery of fluorescence can result from either the dynamics of the microtubules turning over in that bleached zone, growth of new microtubules into the bleached zone, or movement of microtubules through the bleached

(a)

(b)

(c)

FIGURE 9.8 **The study of the cytoskeleton using FRAP.** (*a*) A cell expressing GFP-tubulin incorporates the fluorescent tubulin into its microtubule array. When a laser is focused on a box on the array, the fluorescence is bleached (t = 0 sec). Over time the fluorescence is recovered in the bleached zone. (*b*) The fluorescence recovery can come from different properties of microtubules. The dynamic properties of the microtubules are very likely to contribute to fluorescence recovery. Alternatively, if new fluorescent microtubules grow from the centrosome, they can grow into the bleached zone. Finally, it is possible, that a fluorescent translocating microtubule can move through the bleached zone and be seen there at the time of observation. (*c*) Micrographs from a FRAP experiment performed on interphase (non-dividing) cells expressing GFP-tubulin. Two side-by-side cells were bleached in the boxed region with a laser, and the cells were then imaged over time. The fluorescence signal in the bleached region recovers over the time course of the experiment. (COURTESY OF CLAIRE WALCZAK AND RANIA RIZK.)

zone (Figure 9.8*b*). Subsequent experiments are generally required to distinguish among these possible explanations.

REVIEW ?

1. Describe an example in which each of the following methodologies has contributed to our understanding of the nature of the cytoskeleton: fluorescence microscopy, in vitro motility assays, and atomic force microscopy.

2. List some of the major functions of the cytoskeleton.

9.3 MICROTUBULES

Structure and Composition

As the name implies, microtubules are hollow, relatively rigid, tubular structures, and they occur in nearly every eukaryotic cell. Microtubules are components of a diverse array of structures, including the mitotic spindle of dividing cells and the core of cilia and flagella. Microtubules have an

outer diameter of 25 nm and a wall thickness of approximately 4 nm, and may extend across the length or breadth of a cell. The wall of a microtubule is composed of globular proteins arranged in longitudinal rows, termed **protofilaments**, that are aligned parallel to the long axis of the tubule (Figure 9.9*a*). When viewed in cross section, microtubules are seen to consist of 13 protofilaments aligned side by side in a circular pattern within the wall (Figure 9.9*b*). Noncovalent interactions between adjacent protofilaments are thought to play an important role in maintaining microtubule structure.

Each protofilament is assembled from dimeric building blocks consisting of one α-tubulin and one β-tubulin subunit. The two types of globular tubulin subunits have a similar three-dimensional structure and fit tightly together as shown in Figure 9.9*c*. The tubulin dimers are organized in a linear array along the length of each protofilament, as shown in Figure 9.9*d*. Because each assembly unit contains two non-identical components (a heterodimer), the protofilament is asymmetric, with an α-tubulin at one end and a β-tubulin at the other end. All of the protofilaments of a microtubule have the same polarity. Consequently, the entire polymer has polarity. One end of a microtubule is known as the *plus end* and is terminated by a row of β-tubulin subunits (Figure 9.9*d*). The opposite end is the *minus end* and is terminated by a row of α-tubulin subunits. As discussed later in the chapter, the structural polarity of microtubules is an important factor in the

(a)   100 nm   *(b)*   *(c)*

(d)
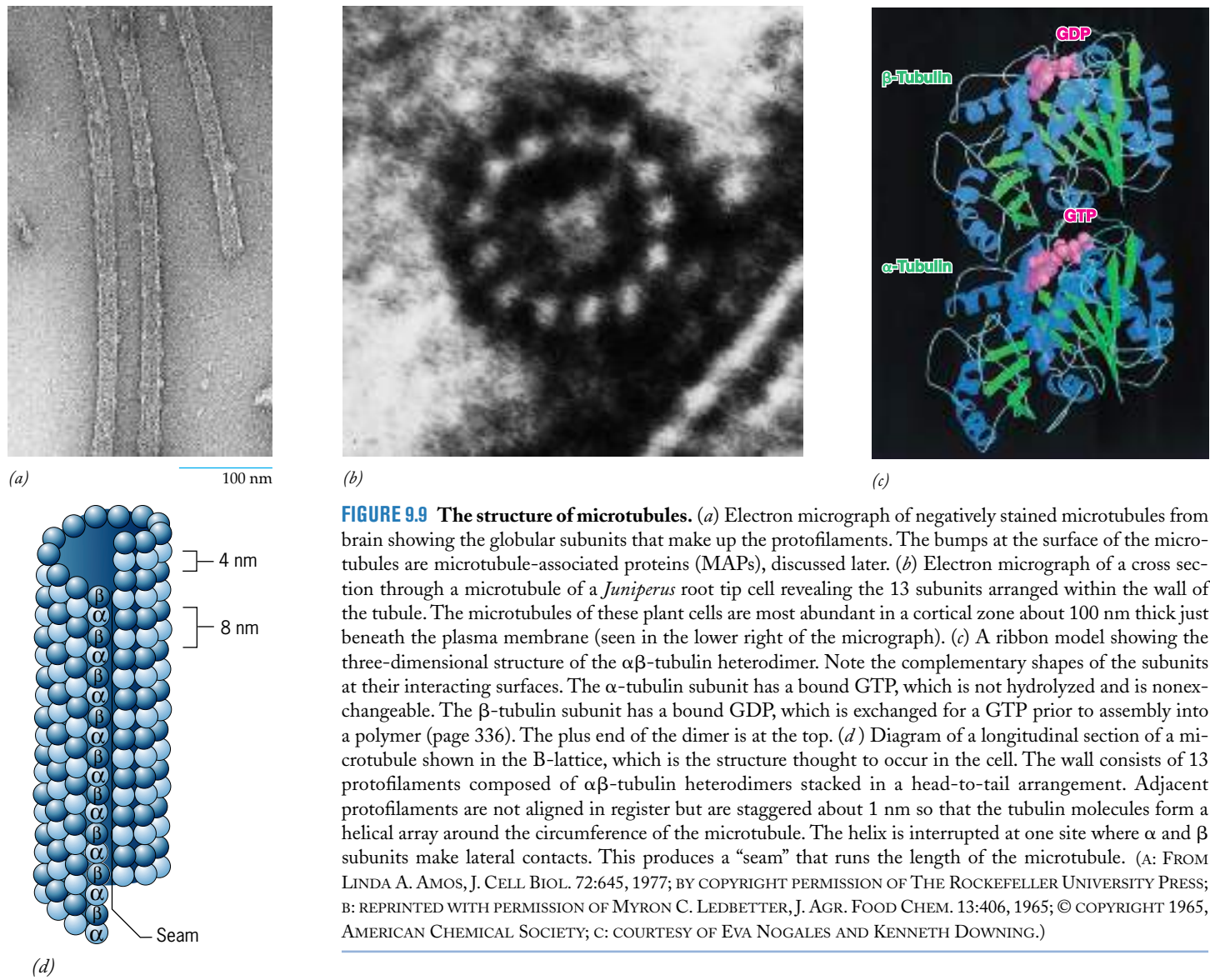
FIGURE 9.9 **The structure of microtubules.** (*a*) Electron micrograph of negatively stained microtubules from brain showing the globular subunits that make up the protofilaments. The bumps at the surface of the microtubules are microtubule-associated proteins (MAPs), discussed later. (*b*) Electron micrograph of a cross section through a microtubule of a *Juniperus* root tip cell revealing the 13 subunits arranged within the wall of the tubule. The microtubules of these plant cells are most abundant in a cortical zone about 100 nm thick just beneath the plasma membrane (seen in the lower right of the micrograph). (*c*) A ribbon model showing the three-dimensional structure of the αβ-tubulin heterodimer. Note the complementary shapes of the subunits at their interacting surfaces. The α-tubulin subunit has a bound GTP, which is not hydrolyzed and is nonexchangeable. The β-tubulin subunit has a bound GDP, which is exchanged for a GTP prior to assembly into a polymer (page 336). The plus end of the dimer is at the top. (*d*) Diagram of a longitudinal section of a microtubule shown in the B-lattice, which is the structure thought to occur in the cell. The wall consists of 13 protofilaments composed of αβ-tubulin heterodimers stacked in a head-to-tail arrangement. Adjacent protofilaments are not aligned in register but are staggered about 1 nm so that the tubulin molecules form a helical array around the circumference of the microtubule. The helix is interrupted at one site where α and β subunits make lateral contacts. This produces a "seam" that runs the length of the microtubule. (A: FROM LINDA A. AMOS, J. CELL BIOL. 72:645, 1977; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS; B: REPRINTED WITH PERMISSION OF MYRON C. LEDBETTER, J. AGR. FOOD CHEM. 13:406, 1965; © COPYRIGHT 1965, AMERICAN CHEMICAL SOCIETY; C: COURTESY OF EVA NOGALES AND KENNETH DOWNING.)

growth of these structures and their ability to participate in directed mechanical activities.

Microtubule-Associated Proteins

Microtubules prepared from living tissue typically contain additional proteins, called **microtubule-associated proteins** (or **MAPs**). MAPs comprise a heterogeneous collection of proteins. The first MAPs to be identified are referred to as "classical MAPs" and typically have one domain that attaches to the side of a microtubule and another domain that projects outward as a filament from the microtubule's surface. The binding of one of these MAPs to the surface of a microtubule is depicted in Figure 9.10. Some MAPs can be seen in electron micrographs as cross-bridges connecting microtubules to each other, thus maintaining their parallel alignment. MAPs generally increase the stability of microtubules and promote their assembly. The microtubule-binding activity of the various MAPs is controlled primarily by the addition and removal
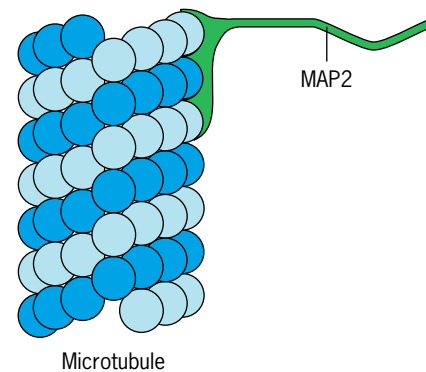
FIGURE 9.10 **Microtubule-associated proteins (MAPs).** Schematic diagram of a brain MAP2 molecule bound to the surface of a microtubule. The MAP2 molecule shown in this figure contains three tubulin-binding sites connected by short segments of the polypeptide chain. (An alternate isoform contains four binding sites). The binding sites are spaced at a sufficient distance to allow the MAP2 molecule to attach to three separate tubulin subunits in the wall of the microtubule. The tails of the MAP molecules project outward, allowing them to interact with other cellular components.

of phosphate groups from particular amino acid residues. An abnormally high level of phosphorylation of one particular MAP, called *tau*, has been implicated in the development of several fatal neurodegenerative disorders, including Alzheimer's disease (page 67). The brain cells of people with these diseases contain strange, tangled filaments (called *neurofibrillary tangles*) consisting of tau molecules that are excessively phosphorylated and unable to bind to microtubules. The neurofibrillary filaments are thought to contribute to the death of nerve cells. Persons with one of these diseases, an inherited dementia called FTDP-17, carry mutations in the *tau* gene, indicating that alteration of the tau protein is the primary cause of this disorder.

Microtubules as Structural Supports and Organizers

Microtubules are stiff enough to resist forces that might compress or bend the fiber. This property enables them to provide mechanical support much as steel girders support a tall office building or poles support the structure of a tent. The distribution of cytoplasmic microtubules in a cell helps determine the shape of that cell. In cultured animal cells, the microtubules extend in a radial array outward from the area around the nucleus, giving these cells their round, flattened shape (Figure 9.11). In contrast, the microtubules of columnar epithelial cells are typically oriented with their long axis parallel to the long axis of the cell (Figure 9.1*a*). This configuration suggests that microtubules help support the cell's elongated shape. The role of microtubules as skeletal elements is particularly evident in certain highly elongated cellular processes, such as cilia and flagella and the axons of nerve cells.

In plant cells, microtubules play an indirect role in maintaining cell shape through their influence on the formation of the cell wall. During interphase, most of a plant cell's microtubules are located just beneath the plasma membrane (as in Figure 9.9*b*), forming a distinct cortical zone. The microtubules of the cortex are thought to influence the movement of the cellulose-synthesizing enzymes located in the plasma membrane (see Figure 7.37). As a result, the cellulose microfibrils of the cell wall are assembled in an orientation that is parallel to the underlying microtubules of the cortex (Figure 9.12). The orientation of these cellulose microfibrils plays
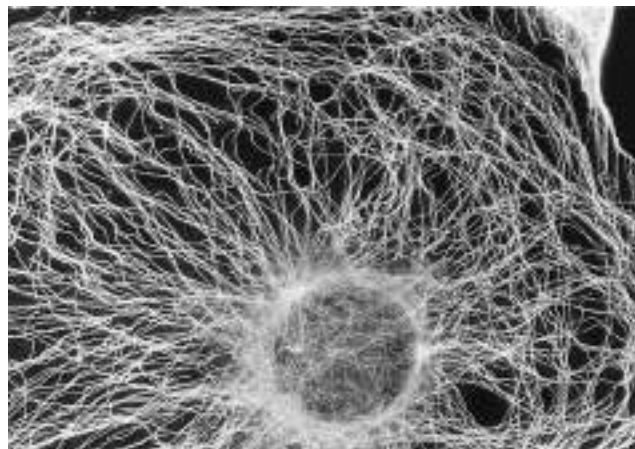
15 μm

FIGURE 9.11 Localization of microtubules of a flattened, cultured mouse cell is revealed by fluorescent anti-tubulin antibodies. Microtubules are seen to extend from the perinuclear region of the cell in a radial array. Individual microtubules can be followed and are seen to curve gradually as they conform to the shape of the cell. (FROM MARY OSBORN AND KLAUS WEBER, CELL 12:563, 1977; BY PERMISSION OF CELL PRESS.)

an important role in determining the growth characteristics of the cell and thus its shape. In most cells, newly synthesized cellulose microfibrils and coaligned microtubules are arranged perpendicular to the long axis of the cell (transversely), like the hoops around a barrel (Figure 9.12). Because the cellulose microfibrils resist lateral expansion, the turgor pressure exerted by the fluid in the cell vacuole is directed to the ends of the cell, causing cell elongation.

Microtubules also play a key role in maintaining the internal organization of cells. Treatment of cells with microtubule-disrupting drugs can seriously affect the location of membranous organelles, including the ER and Golgi complex. The Golgi complex is typically located near the center of an animal cell, just outside the nucleus. Treatment of cells with nocodazole or colchicine, which promote microtubule disassembly, can disperse the Golgi elements into the peripheral regions of the cell. When the drug is removed and microtubules reassemble, the Golgi membranes return to their normal position in the cell interior.

FIGURE 9.12 Spatial relationship between microtubule orientation and cellulose deposition in plant cells. A wheat mesophyll cell that was double stained for microtubules (*a*) and cell wall cellulose (*b*). It is evident that the cortical microtubules and cellulose microfibrils are coaligned. (FROM GEORG JUNG AND WOLFGANG WERNICKE, PROTOPLASMA 53:145, 1990.)

(a)

(b)
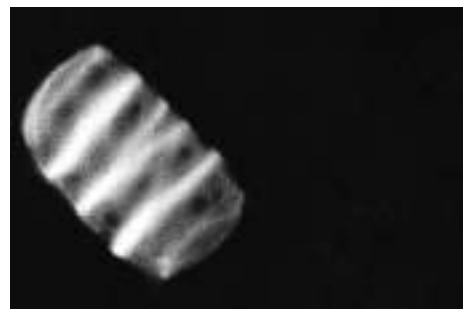
20 μm

Microtubules as Agents of Intracellular Motility

Living cells bristle with activity as macromolecules and organelles move in a directed manner from place to place. While this hustle and bustle can be appreciated by watching the movement of particulate materials in a living cell, the underlying processes are usually difficult to study because most cells lack a highly ordered cytoskeleton. We know, for example, that the transport of materials from one membrane compartment to another depends on the presence of microtubules because specific disruption of these cytoskeletal elements brings the movements to a halt. We will begin our study of intracellular motility by focusing on nerve cells whose intracellular movements rely on a highly organized arrangement of microtubules and other cytoskeletal filaments.

Axonal Transport   The axon of an individual motor neuron may stretch from the spinal cord to the tip of a finger or toe. The manufacturing center of this neuron is its cell body, a rounded portion of the cell that resides within the spinal cord. When labeled amino acids are injected into the cell body, they are incorporated into labeled proteins that move into the axon and gradually travel down its length. Many different materials, including neurotransmitter molecules, are compartmentalized within membranous vesicles in the ER and Golgi complex of the cell body and then transported down the length of the axon (Figure 9.13a). Non-membrane-bound cargo, such as RNAs, ribosomes, and even cytoskeletal elements are also transported along this vast stretch of extended cytoplasm. Different materials move at different rates, with the fastest axonal transport occurring at velocities of 5 μm per second. At this rate, a synaptic vesicle pulled by nanosized motor proteins can travel 0.4 meter in a single day, or about halfway from your spinal cord to the tip of your finger. Structures and materials traveling from the cell body toward the terminals of a neuron are said to move in an *anterograde* direction. Other structures, including endocytic vesicles that form at the axon terminals and carry regulatory factors from target cells, move in the opposite, or *retrograde*, direction—from the synapse toward the cell body. Defects in both anterograde and retrograde transport have been linked to several neurological diseases.

Axons are filled with cytoskeletal structures, including bundles of microfilaments, intermediate filaments, and microtubules interconnected in various ways (Figure 9.13b,c). Using video microscopy, investigators can follow individual vesicles as they move along the microtubules of an axon, either toward or away from the cell body (Figure 9.14). These movements are mediated primarily by microtubules (Figure 9.13c), which serve as tracks for a variety of **motor proteins** that generate

(a)

(b)

Protein connecting microtubules and neurofilaments (e.g., plectin)

MAP

Motor protein

Microtubules (neurotubules)

Intermediate filament (neurofilament)

Organelle

Sidearms interconnecting intermediate filaments

(c)

0.25 μm
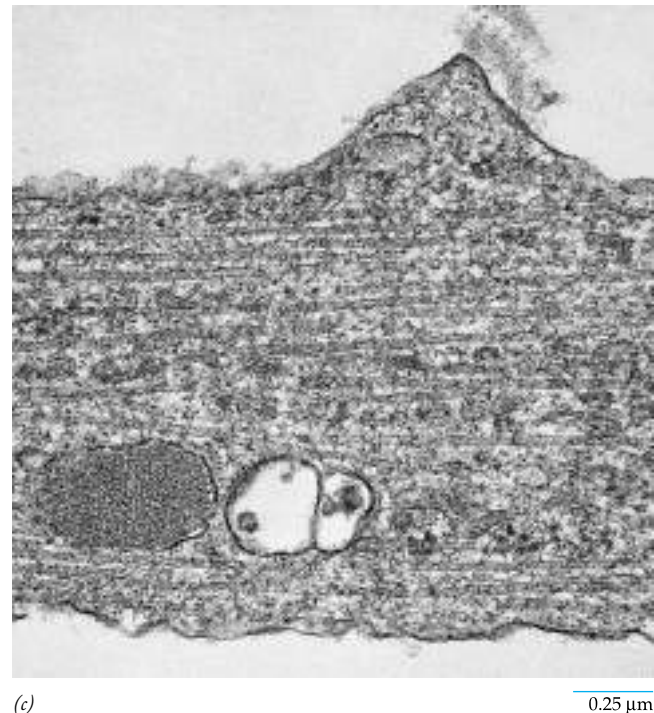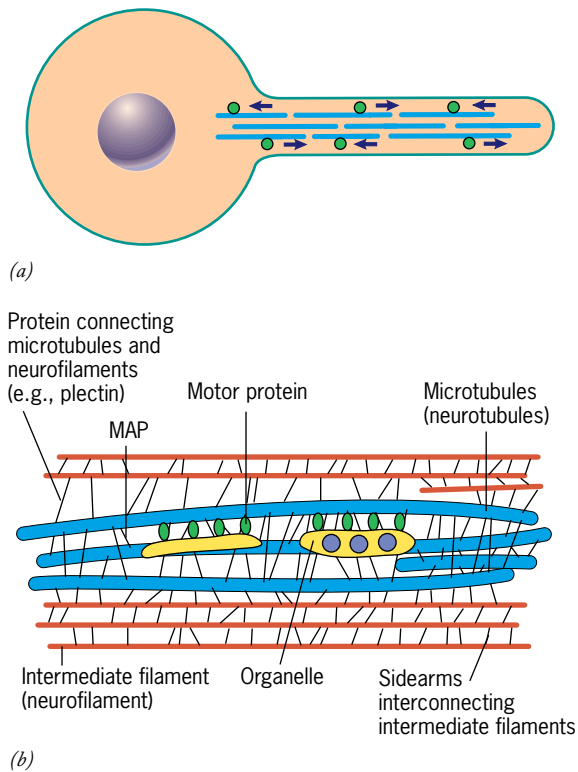
FIGURE 9.13 Axonal transport. (*a*) Schematic drawing of a nerve cell showing the movement of vesicles down the length of an axon along tracks of microtubules. Vesicles move in both directions within the axon. (*b*) Schematic drawing of the organization of the microtubules and intermediate filaments (neurofilaments) within an axon. Vesicles containing transported materials are attached to the microtubules by cross-linking proteins, including motor proteins, such as kinesin and dynein.

(*c*) Electron micrograph of a portion of an axon from a cultured nerve cell, showing the numerous parallel microtubules that function as tracks for axonal transport. The two membrane-bound organelles shown in this micrograph were seen under a light microscope to be moving along the axon at the time the nerve cell was fixed. (C: FROM A. C. BREUER, C. N. CHRISTIAN, M. HENKART, AND P. G. NELSON, J. CELL BIOL. 65:568, 1975; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

(a)   *(b)*   *(c)*
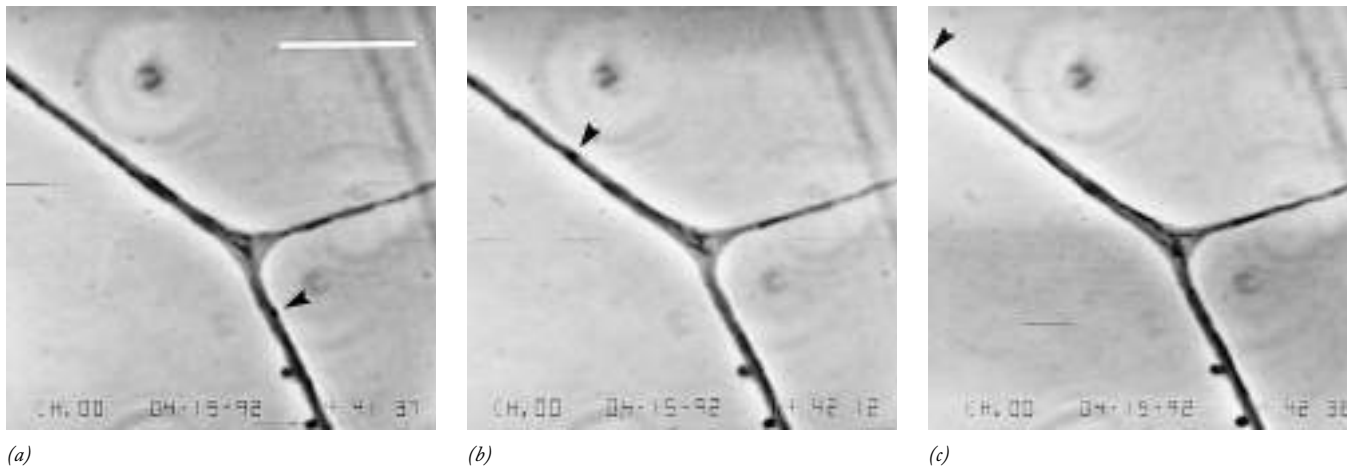
FIGURE 9.14 **Visualizing axonal transport.** (*a–c*) These video micrographs show the progression of a membranous organelle along a branched axon. The cell body is far out of the field to the upper left, while the termini (growth cones, page 373) are out of the field to the bottom right. The position of the organelle is indicated by the arrowheads. The organelle being followed (an autophagic vacuole) moves in the retrograde direction across the branch point and continues to move toward the cell body. Bar, 10 μm. (FROM PETER J. HOLLENBECK, J. CELL BIOL. 121:307, 1993; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

the forces required to move objects within a cell. The study of motor proteins has become a major focus in cell and molecular biology, and a great deal of information has been collected about the nature of these molecules and their mechanism of action, which are the subjects of the following section.

Motor Proteins that Traverse the Microtubular Cytoskeleton

The motor proteins of a cell convert chemical energy (stored in ATP) into mechanical energy, which is used to generate force or to move cellular cargo attached to the motor. Types of cellular cargo transported by these proteins include ribonucleoprotein particles, vesicles, mitochondria, lysosomes, chromosomes, and other cytoskeletal filaments. A single cell may contain a hundred different motor proteins, each presumably specialized for a distinct activity, such as the movement of particular types of cargo in a particular region of the cell. Collectively, motor proteins can be grouped into three broad superfamilies: kinesins, dyneins, and myosins. Kinesins and dyneins move along microtubules, whereas myosins move along microfilaments. No motor protein is known that uses intermediate filament tracks. This is not surprising considering that intermediate filaments are not polarized and thus would not provide directional cues to the motor.

Motor proteins move unidirectionally along their cytoskeletal track in a stepwise manner, from one binding site to the next. As the protein moves along, it undergoes a series of conformational changes that constitute a *mechanical cycle*. The steps of the mechanical cycle are coupled to the steps of a *chemical* (or *catalytic*) *cycle*, which provide the energy necessary to fuel the motor's activity (see Figure 9.61 for an example). The steps in the chemical cycle include the binding of an ATP molecule to the motor, the hydrolysis of the ATP, the release of the products (ADP and P_i) from the motor, and the binding of a new molecule of ATP. The binding and hydrolysis of a single ATP

molecule is used to drive a power stroke that moves the motor a precise number of nanometers along its track. As the motor protein moves to successive sites along the cytoskeletal polymer, the mechanical and chemical cycles are repeated over and over again, pulling the cargo over considerable distances. Keep in mind, we are describing molecular-sized motors, which unlike human-made machines are greatly influenced by their environment. For example, motor proteins have virtually no momentum and are subjected to tremendous frictional resistance from their viscous environment. As a result, a motor protein comes to a stop almost immediately once energy input has ceased.

We will begin by examining the molecular structure and functions of kinesins, which are the smallest and best understood microtubular motors.

Kinesins In 1985, Ronald Vale and colleagues isolated a motor protein from the cytoplasm of squid giant axons that used microtubules as its track. They named the protein **kinesin**, which was subsequently found in virtually all eukaryotic cells. Kinesin is a tetramer constructed from two identical heavy chains and two identical light chains (Figure 9.15*a*). A kinesin molecule has several parts, including a pair of globular heads that bind a microtubule and act as ATP-hydrolyzing, force-generating "engines." Each head (or *motor domain*) is connected to a neck, a rodlike stalk, and a fan-shaped tail that binds cargo to be hauled (Figure 9.15*a*). Surprisingly, the motor domain of kinesin is strikingly similar in structure to that of myosin, despite the fact that kinesin is a much smaller protein and the two types of motors operate over different tracks. Kinesin and myosin almost certainly evolved from a common ancestral protein present in a primitive eukaryotic cell.

When purified kinesin molecules are observed in an in vitro motility assay, the motor proteins move along microtubules toward their plus end (see Figure 9.6). Kinesin is therefore said to be a *plus end–directed microtubular motor*. In an axon, where all of the microtubules are oriented with their minus ends

(a)

(b)

Current Opinion in Cell Biology

(c)
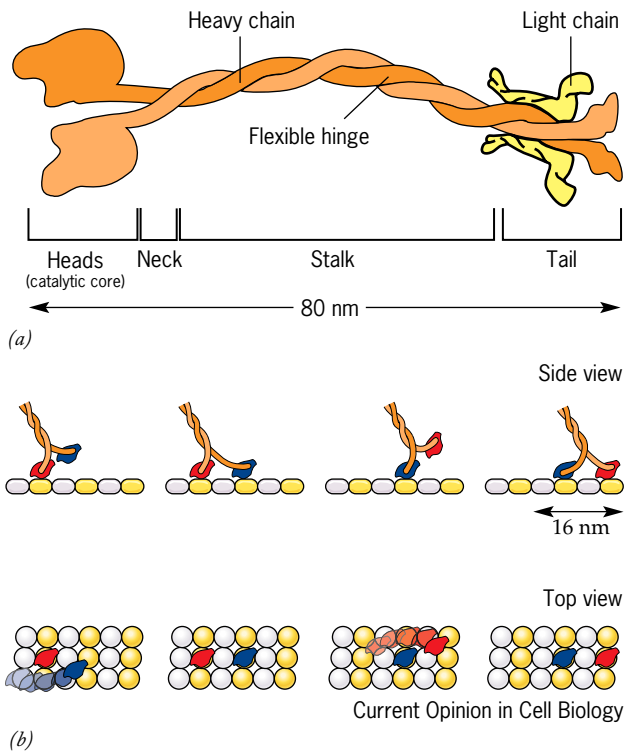
FIGURE 9.15 Kinesin. (*a*) Structure of a conventional kinesin molecule, which consists of two heavy chains that wrap around each other to form a single, common stalk and two light chains associated with the globular ends of the heavy chains. The force-generating heads bind to the microtubule, and the tail binds to the cargo being transported. Having a molecular mass of approximately 380 kDa, kinesin is considerably smaller than the other motor proteins, myosin (muscle myosin, 520 kDa) and dynein (over 1000 kDa). (*b*) Schematic diagram of a kinesin molecule moving along a microtubular track. In the alternate hand-over-hand model shown here, the two heads carry out equivalent but alternating movements, not unlike that of a person walking through a garden on a linear path of stepping stones. As with walking, the trailing head (leg) moves 16 nm forward alternately on the left and right side of the stalk (body). (*c*) The conformational changes that occur in the head (blue) and neck (red) portions of a monomeric kinesin molecule that drive the movement of the protein along a microtubule (yellow contour map). Rather than being connected to a second head, the neck of this truncated kinesin molecule is attached to a GFP molecule (green). The swinging movement of the neck would normally drive the forward motion of the partner head, enabling the dimer to walk toward the plus end of a protofilament. (B: AFTER C. L. ASBURY, CURR. OPIN. CELL BIOL. 17:91, 2005. COPYRIGHT 2005 WITH PERMISSION FROM ELSEVIER SCIENCE; C: FROM RYAN B. CASE ET AL., COURTESY OF RONALD D. VALE, CURR. BIOL., VOL. 10, #3 COVER, 2000.)
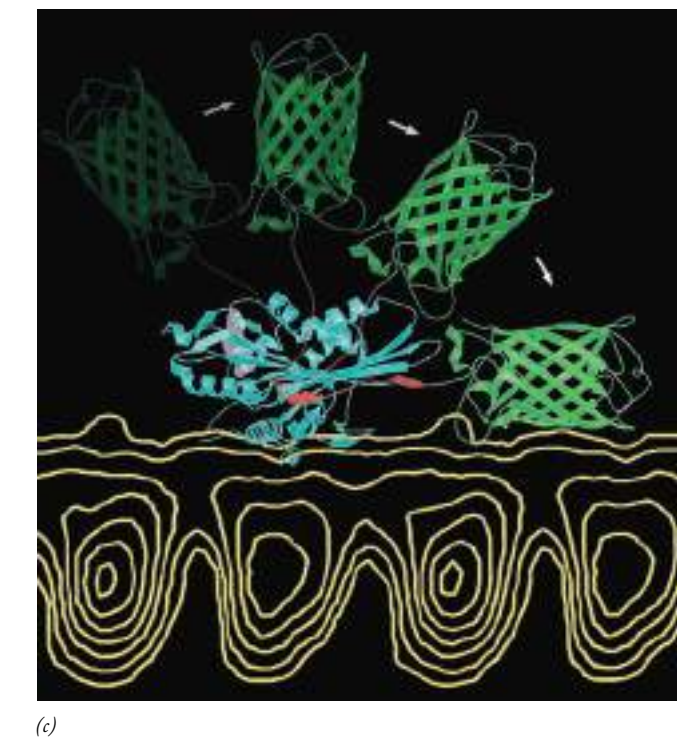
facing the cell body, kinesin transports vesicles and other cargo toward the synaptic terminals. The discovery of kinesin is discussed in the Experimental Pathways, which can be found on the Web at www.wiley.com/college/karp.

A single kinesin molecule moves along a single protofilament of a microtubule at a velocity proportional to the ATP concentration (up to a maximal velocity of about 1 μm per second). At low ATP concentrations, kinesin molecules travel slowly enough for observers to conclude that the protein moves in distinct steps (Figure 9.15*b*). Each step is approximately 8 nm in length, which is also the length of one tubulin dimer in a protofilament, and requires the hydrolysis of a single ATP molecule. It is now generally accepted that kinesin moves by a "hand-over-hand" mechanism depicted in Figure 9.15*b*. According to this model, which is basically similar to a person climbing a rope, the two heads alternate in taking the leading and lagging positions without an accompanying rotation of the stalk and cargo at every step.

The movement of kinesin molecules, both in vitro and in vivo, is highly **processive**, meaning that the motor protein tends to move along an individual microtubule for considerable

distances (over 1 μm) without falling off. A two-headed kinesin molecule can accomplish this feat because at least one of the heads is attached to the microtubule at all times (Figure 9.15*b*). A motor protein with this capability is well adapted for independent, long-distance transport of small parcels of cargo.

The two heads of a kinesin molecule behave in a coordinated manner, so that they are always present at different stages in their chemical and mechanical cycles at any given time. When one head binds to the microtubule, the resulting conformational changes in the adjacent neck region of the motor protein cause the other head to move forward toward the next binding site on the protofilament. The forward-moving head probably finds its precise binding site on the protofilament through a rapid, diffusional (random) search. These actions are illustrated in Figure 9.15*c*, which depicts the head and neck portions of a monomeric kinesin heavy chain associated with a microtubule. Catalytic activity of the head leads to a swinging movement of the neck which, in this illustration, is attached to a molecule of GFP (the green barrel-shaped protein) rather than to a kinesin head. The microtubule is not simply a passive track in these events but plays an active role

in stimulating certain steps in the mechanical and chemical cycles of the kinesin molecule.

The kinesin molecule discovered in 1985, which is referred to as "conventional kinesin" or kinesin-1, is only one member of a superfamily of related proteins, which we will call **KLPs (kinesin-like proteins)**. Based on the analysis of genome sequences, mammals produce approximately 45 different KLPs. The motor portions of KLPs have related amino acid sequences, reflecting their common evolutionary ancestry and their similar role in moving along microtubules. In contrast, the tails of KLPs have diverse sequences, reflecting the variety of cargo these motors haul. A number of different proteins have been identified as potential adaptors that link specific KLPs and their cargoes.

Like kinesin-1, most KLPs move toward the plus end of the microtubule to which they are bound. One small family (called kinesin-14), however, including the widely studied Ncd protein of *Drosophila*, moves in the opposite direction, that is, toward the minus end of the microtubular track. One might expect that the heads of plus end–directed and minus end–directed KLPs would have a different structure because they contain the catalytic motor domains. But the heads of the two proteins are virtually indistinguishable. Instead, differences in directional movement have been traced to differences in the adjacent neck regions of the two proteins. When the head of a minus end–directed Ncd molecule is joined to the neck and stalk portions of a kinesin-1 molecule, the hybrid protein moves toward the plus end of the track. Thus, even though the hybrid has a catalytic domain that normally would move toward the minus end of a microtubule, as long as it is joined to the neck of a plus end motor, it moves in the plus direction.

A third small family (kinesin-13) of kinesin-like proteins are incapable of movement. The KLPs of this group bind to either end of a microtubule and bring about its depolymerization rather than moving along its length. These proteins are often referred to as microtubule *depolymerases*. The important role of these and other KLPs during cell division will be explored in Chapter 14.

Kinesin-Mediated Organelle Transport   We saw in Chapter 8 how vesicles move from one membranous compartment, such as the Golgi complex, to another, such as a lysosome. The routes followed by cytoplasmic vesicles and organelles are largely defined by microtubules (see Figure 9.1), and members of the kinesin superfamily are strongly implicated as force-generating agents that drive the movement of this membrane-bounded cargo. In most cells, as in axons, microtubules are aligned with their plus ends pointed away from the center of the cell. Therefore, members of the kinesin superfamily tend to move vesicles and organelles (e.g., peroxisomes and mitochondria) in an outward direction toward the cell's plasma membrane. This is illustrated in the micrographs in Figure 9.16. The pair of micrographs on the left shows a cell that was

(a)

(b)

(c)

(d)   20 μm
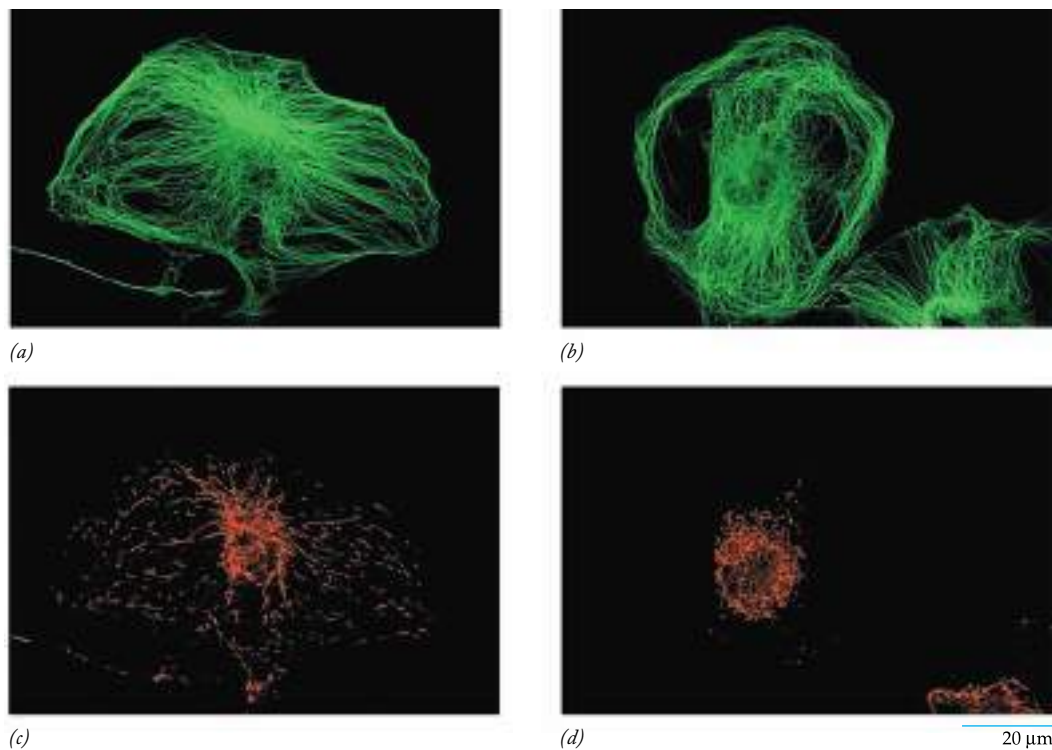
FIGURE 9.16   Alteration in the phenotype of a cell lacking a member of the kinesin superfamily. (*a,c*) Control cell from the extraembryonic tissues of a normal 9.5-day mouse embryo stained in *a* for microtubules (green) and in *c* for mitochondria (red). A significant fraction of the cell's mitochondria are located along microtubules in the peripheral regions of the cell. (*b,d*) A corresponding cell obtained from an embryo that lacked both copies of the gene encoding KIF5B, which is one of three conventional kinesin isoforms in mice and humans. All of the mitochondria are clustered in the central region of the cell, suggesting that KIF5B is responsible for transporting mitochondria in an outward direction. (FROM YOSUKE TANAKA ET AL., COURTESY OF NOBUTAKA HIROKAWA, CELL 93:1150, 1998; BY PERMISSION OF CELL PRESS.)

isolated from a normal 9.5-day mouse embryo and stained to reveal the location of its microtubules (green) and mitochondria (orange). The pair of micrographs on the right shows a cell that was isolated from a 9.5-day mouse embryo lacking both copies of the gene that encodes the KIF5B kinesin heavy chain. The mitochondria of the KIF5B-deficient cell are absent from the peripheral regions of the cell, as would be expected if this plus end-directed kinesin is responsible for the outward movement of membranous organelles (see also Figure 9.17c).

Cytoplasmic Dynein   The first microtubule-associated motor was discovered in 1963 as the protein responsible for the movement of cilia and flagella. The protein was named **dynein**. The existence of cytoplasmic forms was almost immediately suspected, but it took over 20 years before a similar protein was purified and characterized from mammalian brain tissue and called **cytoplasmic dynein**. Cytoplasmic dynein is present throughout the animal kingdom, but there is controversy as to whether or not it is present in plants. Whereas each of us has many different kinesins (and myosins), each adapted

for specific functions, we are able to manage with only two cytoplasmic dyneins, one of which appears responsible for most transport operations.

Cytoplasmic dynein is a huge protein (molecular mass of approximately 1.5 million daltons) composed of two identical heavy chains and a variety of intermediate and light chains (Figure 9.17a). Each dynein heavy chain consists of a large globular head with an elongated projection (stalk). The dynein head, which is an order of magnitude larger than a kinesin head, acts as a force-generating engine. Each stalk contains the all-important microtubule-binding site situated at its tip. The longer projection, known as the stem (or tail), binds the intermediate and light chains, whose functions are not well defined. Structural analyses indicate that the dynein motor domain consists of a number of distinct modules organized in the shape of a wheel (see Figure 9.36), which makes it fundamentally different in both architecture and mode of operation from kinesin and myosin.

In vitro motility assays indicate that cytoplasmic dynein moves processively along a microtubule toward the polymer's
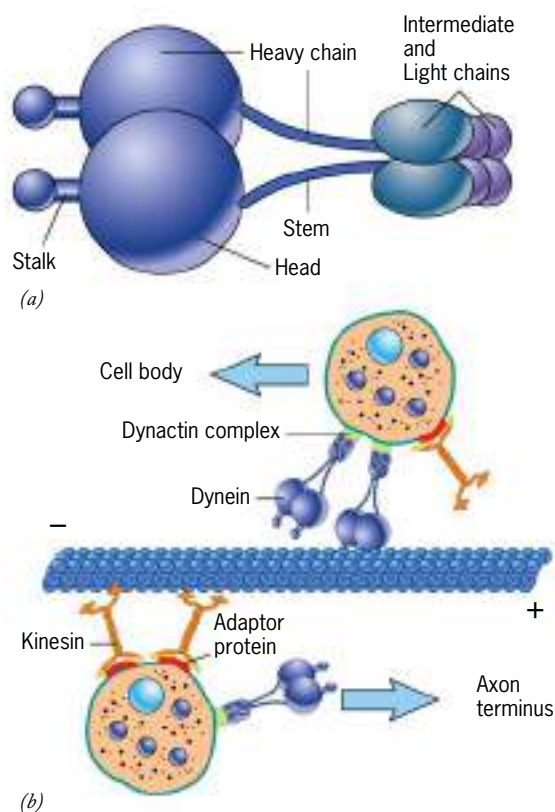
FIGURE 9.17 Cytoplasmic dynein and organelle transport by microtubule-tracking motor proteins. (a) Structure of a cytoplasmic dynein molecule, which contains two dynein heavy chains and a number of smaller intermediate and light chains at the base of the molecule. Each dynein heavy chain contains a large, globular, force-generating head, a protruding stalk containing a binding site for the microtubule, and a stem. (b) Schematic diagram of two vesicles moving in opposite directions along the same microtubule, one powered by kinesin moving toward the plus end of the track, and the other by cytoplasmic dynein moving toward the minus end of the track. In the model shown here, each vesicle contains both types of motor proteins, but the kinesin molecules are inactivated in the upper vesicle and the dynein molecules are inactivated in the lower vesicle. Both motor proteins are attached to the vesicle membrane by an intermediary: kinesin can be attached to vesicles by a variety of integral and peripheral membrane proteins, and dynein by a soluble protein complex called dynactin. (c) Schematic illustration of kinesin-mediated and dynein-mediated transport of vesicles, vesicular-tubular clusters (VTCs), and organelles in a nonpolarized, cultured cell.
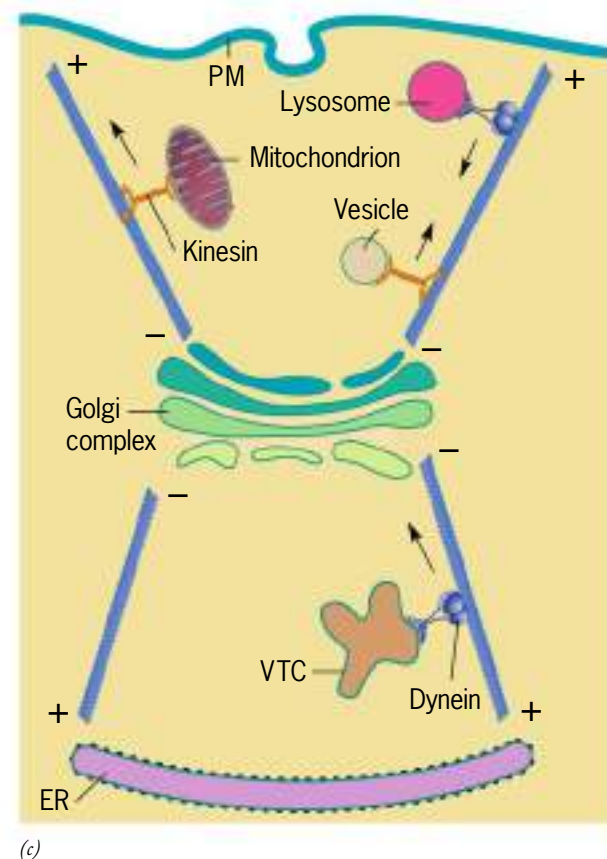
minus end—opposite that of kinesin (Figure 9.17*b*). A body of evidence suggests at least two well-studied roles for cytoplasmic dynein:

1. As a force-generating agent in positioning the spindle and moving chromosomes during mitosis (discussed in Chapter 14).

2. As a minus end–directed microtubular motor with a role in positioning the centrosome and Golgi complex and moving organelles, vesicles, and particles through the cytoplasm.

In nerve cells, cytoplasmic dynein has been implicated in the retrograde movement of membranous organelles and the anterograde movement of microtubules. In fibroblasts and other nonneural cells, cytoplasmic dynein is thought to transport membranous organelles from peripheral locations toward the center of the cell (Figure 9.17*c*). Dynein-driven cargo includes endosomes and lysosomes, ER-derived vesicles heading toward the Golgi complex, RNA molecules, and the HIV virus which is transported to the nucleus of an infected cell. Cytoplasmic dynein does not interact directly with membrane-bounded cargo but requires an intervening multisubunit adaptor, called *dynactin*. Dynactin may also regulate dynein activity and help bind the motor protein to the microtubule, which increases processivity.

According to the simplified model shown in Figure 9.17*c*, kinesin and cytoplasmic dynein move similar materials in opposite directions over the same railway network. As indicated in Figure 9.17*b*, individual organelles may bind kinesin and dynein simultaneously, which endows these organelles with the capability to move in opposite directions depending on conditions. It is not clear how the activity of opposing motors is regulated. As discussed on page 357, myosin may also be present on some of these organelles.
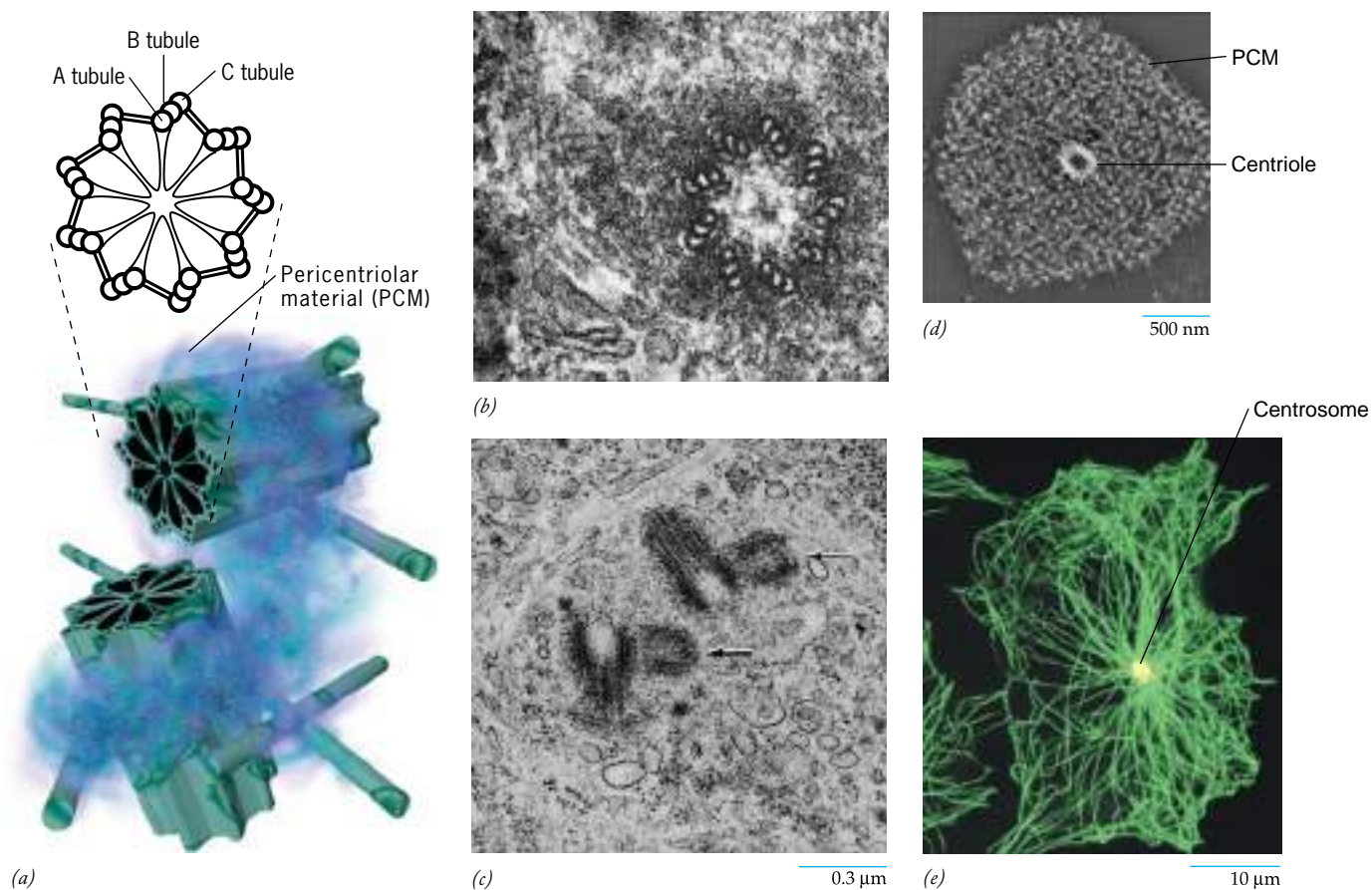
FIGURE 9.18 The centrosome. (*a*) Schematic diagram of a centrosome showing the paired centrioles; the surrounding pericentriolar material (PCM); and microtubules emanating from the PCM, where nucleation occurs. (*b*) Electron micrograph of a cross section of a centriole showing the pinwheel arrangement of the nine peripheral fibrils, each of which consists of one complete microtubule and two incomplete microtubules. (*c*) Electron micrograph showing two pairs of centrioles. Each pair consists of a longer parental centriole and a smaller daughter centriole (arrow), which is undergoing elongation in this phase of the cell cycle (discussed in Section 14.2). (*d*) Electron micrographic reconstruction of a 1.0 M potassium iodide–extracted centrosome, showing the PCM to contain a loosely organized fibrous lattice. (*e*) Fluorescence micrograph of a cultured mammalian cell showing the centrosome (stained yellow by an antibody against a centrosomal protein) at the center of an extensive microtubular network. (A: AFTER S. J. DOXSEY ET AL., CELL 76:643, 1994; BY PERMISSION OF CELL PRESS; B: COURTESY OF B. R. BRINKLEY; C: FROM J. B. RATTNER AND STEPHANIE G. PHILLIPS, J. CELL BIOL. 57:363, 1973; D: FROM BRADLEY J. SCHNACKENBERG ET AL., COURTESY OF ROBERT E. PALAZZO, PROC. NAT'L. ACAD. SCI. U.S.A. 95:9298, 1998; E: FROM TOSHIRO OHTA, ET AL., J. CELL BIOL. 156:88, 2002; COURTESY OF RYOKO KURIYAMA; C AND E: BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

Microtubule-Organizing Centers (MTOCs)

The function of a microtubule within a living cell depends on its location and orientation, which makes it important to understand why a microtubule assembles in one place as opposed to another. When studied in vitro, the assembly of microtubules from αβ-tubulin dimers occurs in two distinct phases: a slow phase of *nucleation*, in which a small portion of the microtubule is initially formed, and a much more rapid phase of *elongation*. Unlike the case in vitro, nucleation of microtubules takes place rapidly inside a cell, where it occurs in association with a variety of specialized structures called **microtubule-organizing centers** (or **MTOCs**). The best studied MTOC is the centrosome.

Centrosomes   In animal cells, the microtubules of the cytoskeleton are typically nucleated by the **centrosome**, a complex structure that contains two barrel-shaped **centrioles** surrounded by amorphous, electron-dense **pericentriolar material** (or **PCM**) (Figure 9.18*a,b*). Centrioles are cylindrical structures about 0.2 μm in diameter and typically about twice as long. Centrioles contain nine evenly spaced fibrils, each of which appears in cross section as a band of three microtubules, designated the A, B, and C tubules. Only the A tubule is a complete microtubule (Figure 9.18*a,b*) and it is connected to the center of the centriole by a radial spoke. The three microtubules of each triplet are arranged in a pattern that gives a centriole a characteristic pinwheel appearance. Centrioles are nearly always found in pairs, with the members situated at right angles to one another (Figure 9.18*a,c*). Extraction of isolated centrosomes with 1.0 M potassium iodide removes about 90 percent of the protein of the PCM, leaving behind a spaghetti-like scaffold of insoluble fibers (Fig. 9.18*d*). As discussed below, the centrosome is the major site of microtubule initiation in animal cells and typically remains at the center of the cell's microtubular network (Figure 9.18*e*).

Figure 9.19*a* shows an early experiment that demonstrates the role of the centrosome in the initiation and organization of the microtubular cytoskeleton. The microtubules of a cultured animal cell were first depolymerized by incubating the cells in colcemid, a drug that binds to tubulin subunits and blocks their use by the cell. Microtubule reassembly was then monitored by removing the chemical, fixing cells at various time intervals, and treating the fixed cells with fluorescent anti-tubulin antibodies. Within a few minutes after removal of the inhibiting conditions, one or two bright fluorescent spots are seen in the cytoplasm of each cell. Within 15 to 30 minutes (Figure 9.19*a*), the number of labeled filaments radiating out of these foci increases dramatically. When these same cells are sectioned and examined in the electron microscope, the newly formed microtubules are found to radiate outward from a centrosome. Close examination shows that the microtubules do not actually penetrate into the centrosome and make contact with the centrioles, but terminate in the dense PCM that resides at the centrosome periphery. It is this material that initiates the formation of microtubules (see Figure 9.20*c*). Although centrioles are not directly involved in microtubule nucleation, they play a role in recruiting the surrounding PCM during assembly of the centrosome and in the general process of centrosome duplication (discussed in Section 14.2).

As illustrated by the experiment just described, centrosomes are sites of microtubule nucleation. The polarity of these microtubules is always the same: the minus end is associated with the centrosome, and the plus (or growing) end is situated at the opposite tip (Figure 9.19*b*). Thus, even though microtubules are nucleated at the MTOC, they are elongated

(a)   10 μm   *(b)*
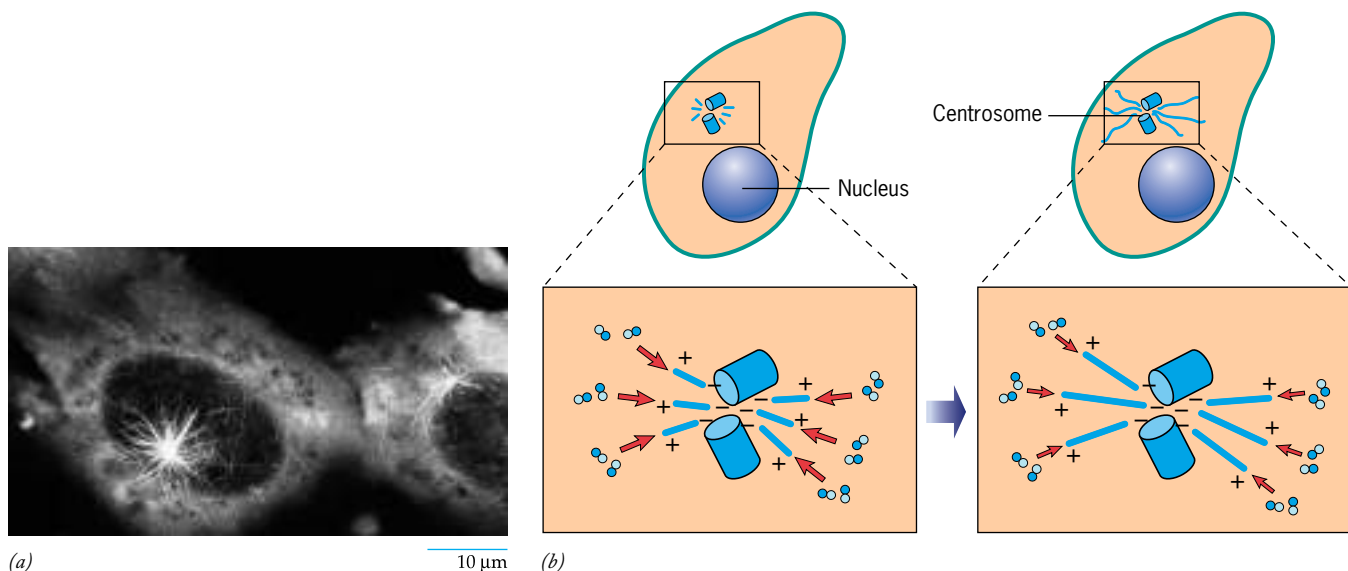
FIGURE 9.19 Microtubule nucleation at the centrosome. (*a*) Fluorescence micrograph of a cultured fibroblast that had been exposed to colcemid to bring about the disassembly of the cell's microtubules and was then allowed to recover for 30 minutes before treatment with fluorescent anti-tubulin antibodies. The bright starlike structure marks the centrosome together with newly assembled microtubules that have begun to grow outward in all directions. (*b*) Schematic diagram of the regrowth of microtubules showing the addition of subunits at the plus end of the polymer away from the centrosome. (A: FROM MARY OSBORN AND KLAUS WEBER, PROC. NAT'L ACAD. SCI. U.S.A., 73:869, 1976.)

at the opposite end of the polymer. The growing end of a microtubule may contain a host of specific proteins that help attach the microtubule to a particular target, such as an endosome or Golgi cisterna in an interphase cell or a condensed chromosome in a mitotic cell.

The fraction of microtubules that remain associated with the centrosome is highly variable from one cell type to another. The centrosome of a nonpolarized cell (e.g., a fibroblast) is typically situated near the center of the cell and tends to remain associated with the minus ends of a large number of microtubules (as in Figure 9.18e). In contrast, many of the microtubules in a polarized epithelial cell are anchored by their minus ends at dispersed sites near the apical end of the cell as their plus ends extend toward the cell's basal surface (Figure 9.1). Similarly, the axon of a nerve cell contains large numbers of microtubules that have no association with the centrosome, which is located in the neuron's cell body. These axonal microtubules are severed from the centrosome where they are initially formed and then transported into the axon by motor proteins. Certain animal cells, including mouse oocytes, lack centrosomes entirely, yet they are still capable of forming complex microtubular structures, such as the meiotic spindle (as discussed in Chapter 14).

Basal Bodies and Other MTOCs   Centrosomes are not the only MTOCs in cells. The outer microtubules in a cilium or flagellum are generated from the microtubules in a structure called a **basal body**, which resides at the base of the cilium or flagellum (see Figure 9.32). Basal bodies are identical in structure to centrioles, and in fact, basal bodies and centrioles can give rise to one another. For example, the basal body that gives rise to the flagellum of a sperm cell is derived from a centriole that had been part of the meiotic spindle of the spermatocyte from which the sperm arose. Conversely, the sperm basal body typically becomes a centriole during the first mitotic division of the fertilized egg. Plant cells lack both centrosomes and centrioles, or any other type of obvious MTOC. Instead, microtubules in a plant cell are nucleated around the surface of the nucleus and widely throughout the cortex (see Figure 9.22).

Microtubule Nucleation   Regardless of their diverse appearance, all MTOCs play similar roles in all cells: they control the number of microtubules, their polarity, the number of protofilaments that make up their walls, and the time and location of their assembly. In addition, all MTOCs share a common protein component—a type of tubulin discovered in the mid-1980s, called **γ-tubulin**. Unlike the α- and β-tubulins, which make up about 2.5 percent of the protein of a nonneural cell, γ-tubulin constitutes only about 0.005 percent of the cell's total protein. Fluorescent antibodies to γ-tubulin stain all types of MTOCs, including the pericentriolar material of centrosomes (Figure 9.20a), suggesting that γ-tubulin is a critical component in microtubule nucle-

(a)   (b)   30 nm   (c)
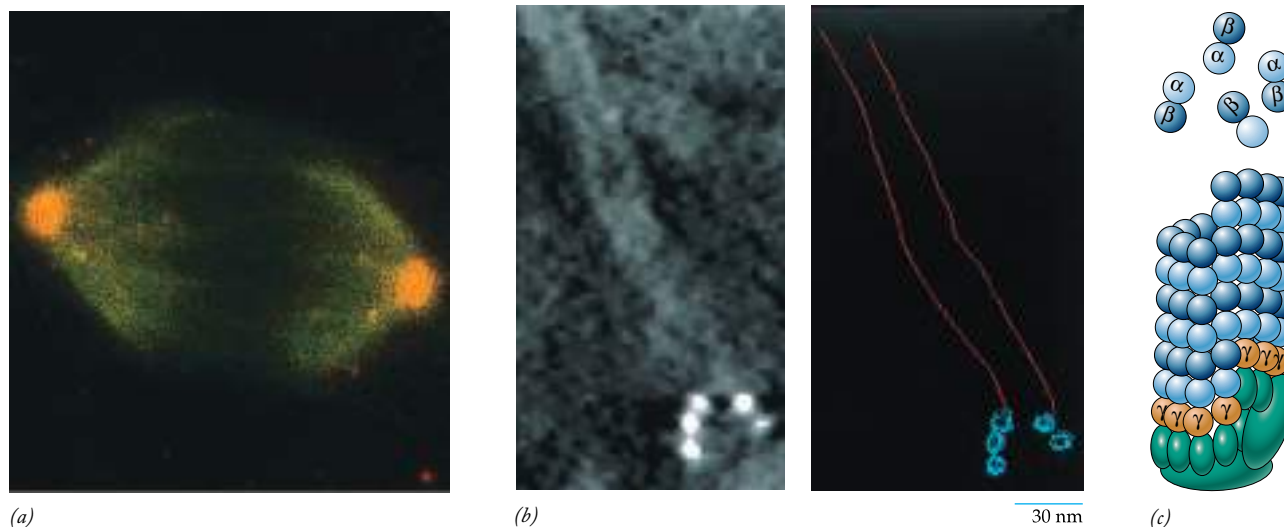
FIGURE 9.20   The role of γ-tubulin in centrosome function. (a) A dividing fibroblast that has been double stained with antibodies against γ-tubulin (red) and β-tubulin (green). Orange staining is due to the coincidence of the two types of tubulin, which occurs in the two centrosomes located at opposite poles of a dividing cell. (b) A reconstruction based on electron micrographs of a portion of a centrosome that had been incubated with purified tubulin in vitro and then labeled with antibodies against γ-tubulin. The antibodies were linked to gold particles to make them visible (as white dots) in the reconstruction. During the incubation with tubulin, the centrosome served as an MTOC to nucleate microtubules whose minus ends are seen to be labeled with clusters of gold, which are often arranged in semicircles or rings. The ac-companying drawing shows the outline of the microtubule seen in the reconstruction. (c) A model for γ-tubulin function during the assembly of microtubules. Nucleation begins as αβ-tubulin dimers bind to an open ring of γ-tubulin molecules (brown), which are held in place by a number of other proteins (green) that make up the large γ-TuRC. Nucleation by the γ-TuRC defines microtubule polarity with a ring of α-tubulin monomers situated at the minus end of the structure. (A: COURTESY OF M. KATHERINE JUNG AND BERL R. OAKLEY; B: REPRODUCED WITH PERMISSION FROM MICHELLE MORITZ ET AL., NATURE 378:639, 1995. PHOTO COURTESY OF DAVID A. AGARD; COPYRIGHT 1995, BY MACMILLAN MAGAZINES LIMITED.)

ation. This conclusion is supported by other studies. For example, microinjection of anti-γ-tubulin antibodies into a living cell blocks the reassembly of microtubules following their depolymerization by drugs or cold temperature.

To understand the mechanism of microtubule nucleation, researchers have focused on the structure and composition of the pericentriolar material (PCM) at the periphery of centrosomes. The insoluble fibers of the PCM (Figure 9.18*d*) are thought to serve as attachment sites for ring-shaped structures that have the same diameter as microtubules (25 nm) and contain γ-tubulin. These ring-shaped structures were discovered when centrosomes were purified and incubated with gold-labeled antibodies that bound to γ-tubulin. The gold particles were seen in the electron microscope to be clustered in semicircles or rings situated at the minus ends of microtubules (Figure 9.20*b*). These are the ends of the microtubules that are embedded in the PCM of the centrosome where nucleation occurs. Similar γ-tubulin ring complexes (or γ-*TuRCs*) have been isolated from cell extracts and shown to nucleate microtubule assembly in vitro. These and other studies have suggested the model shown in Figure 9.20*c* in which a helical array of γ-tubulin (brown) subunits forms an open, ring-shaped template on which the first row of αβ-tubulin dimers assemble. In this model, only the α-tubulin of the heterodimer can bind to a ring of γ-subunits. Thus the γ-TuRC determines the polarity of the entire microtubule and also forms a cap at its minus end.

The Dynamic Properties of Microtubules

Although all microtubules appear quite similar morphologically, there are marked differences in their stability. Microtubules are stabilized by the presence of bound MAPs (page 325) and, likely, by other factors that remain to be determined. Microtubules of the mitotic spindle or the cytoskeleton are extremely labile, that is, sensitive to disassembly. Microtubules of mature neurons are much less labile, and those of centrioles, cilia, and flagella are highly stable. Living cells can be subjected to a variety of treatments that lead to the disassembly of labile cytoskeletal microtubules without disrupting other cellular

structures. Disassembly can be induced by cold temperature; hydrostatic pressure; elevated Ca^{2+} concentration; and a variety of chemicals, including colchicine, vinblastine, vincristine, nocodazole, and podophyllotoxin. The drug taxol stops the dynamic activities of microtubules by a very different mechanism. Taxol binds to the microtubule polymer, inhibiting its disassembly, thereby preventing the cell from assembling new microtubular structures as required. Many of these compounds, including taxol, are used in chemotherapy against cancer because they preferentially kill tumor cells. It had been assumed for years that tumor cells were particularly sensitive to these drugs because of their high rate of cell division. But recent research has revealed there is more to the story. As discussed in Chapter 14, normal cells have a mechanism (or checkpoint) that stops them from dividing in the presence of drugs, such as vinblastine or taxol, that alter the mitotic spindle. As a result, normal cells usually arrest their division activities until the drug has been eliminated from the body. In contrast, many cancer cells lack this mitotic checkpoint and attempt to complete their division even in the absence of a functioning mitotic spindle. This usually results in the death of the tumor cell.

The lability of cytoskeletal microtubules reflects the fact that they are polymers formed by the noncovalent association of protein building blocks. The microtubules of the cytoskeleton are normally subject to depolymerization and repolymerization as the requirements of the cell change from one time to another. The dynamic character of the microtubular cytoskeleton is well illustrated by plant cells. If a typical plant cell is followed from one mitotic division to the next, four distinct arrays of microtubules appear, one after another (Figure 9.21).

1. During most of interphase, the microtubules of a plant cell are distributed widely throughout the cortex, as depicted in Figure 9.21, stage 1. A search for γ-tubulin shows this nucleation factor to be localized along the lengths of the cortical microtubules, suggesting that new microtubules might form directly on the surface of existing microtubules. This idea is supported by studies of tubulin incorporation in living cells (Figure 9.22*a*) and by in vitro assays (Figure 9.22*b*) that show newly formed
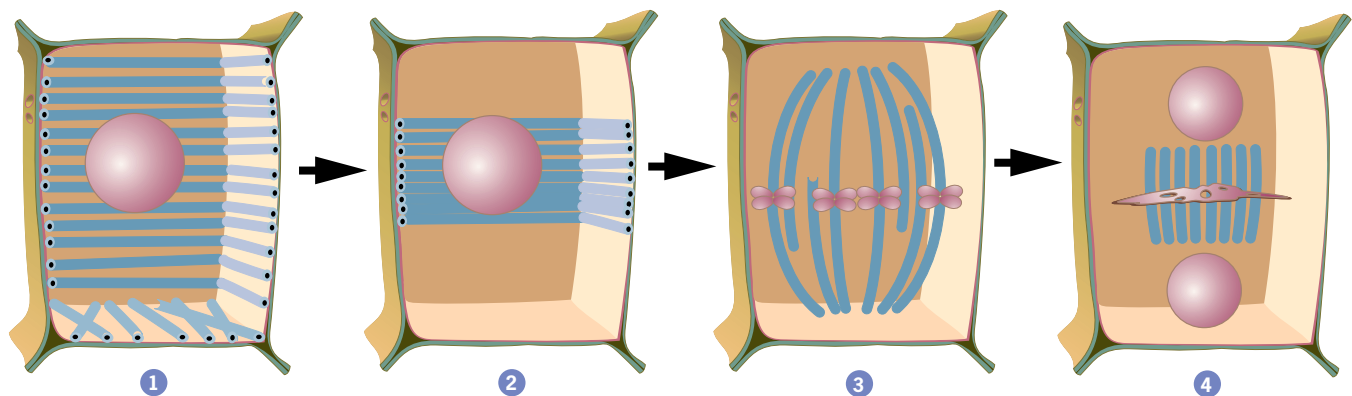
FIGURE 9.21 Four major arrays of microtubules present during the cell cycle of a plant cell. The organization of the microtubules at each stage are described in the text. (AFTER R. H. GODDARD ET AL., PLANT PHYSIOL. 104:2, 1994.)

microtubules branching at an angle off the sides of preexisting microtubules. Once formed, the daughter microtubules are likely severed from the parent microtubule and incorporated into the parallel bundles that encircle the cell (Figures 9.12*a*, 9.21).

2. As the cell approaches mitosis, the microtubules disappear from most of the cortex, leaving only a single transverse band, called the *preprophase band*, that encircles the cell like a belt (Figure 9.21, stage 2). The preprophase band marks the site of the future division plane.

3. As the cell progresses into mitosis, the preprophase band is lost and microtubules reappear in the form of the mitotic spindle (Figure 9.21, stage 3).
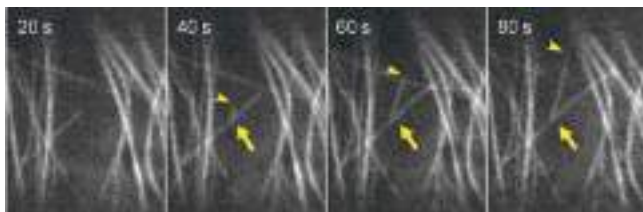
(a)
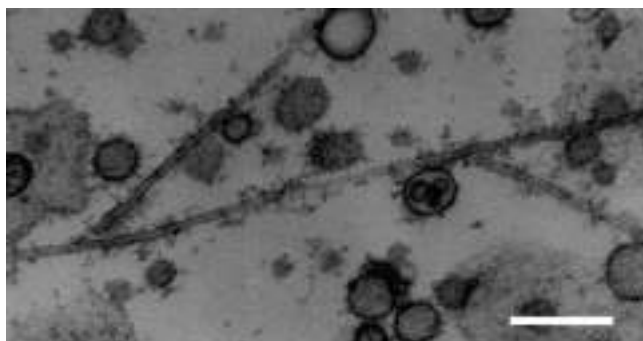
(b)
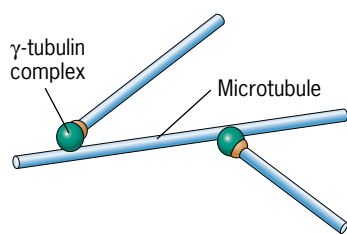
(c)

FIGURE 9.22 Nucleation of plant cortical microtubules. (*a*) The micrographs show a portion of a live cultured tobacco cell that is expressing fluorescently labeled GFP-tubulin. During the period of observation, an existing microtubule of the cortex nucleates the assembly of a daughter microtubule, which grows outward forming a Y-shaped branch. The end of a newly formed microtubule is indicated by the arrowhead, the branchpoint by the arrow. (*b*) Electron micrograph of a microtubule with two daughter microtubules branching from its surface. The branched microtubules were assembled in a cell-free system containing tubulin subunits. Bar, 10 μm. (*c*) A schematic model showing how new microtubules are nucleated at the sites of γ-tubulin present on the surface of an existing microtubule. (A,B: FROM TAKASHI MURATA ET AL., REPRINTED WITH PERMISSION FROM NATURE CELL BIOL. 7:961, 2005. © COPYRIGHT 2005, MACMILLAN MAGAZINES LTD.)

4. After the chromosomes have been separated, the mitotic spindle disappears and is replaced by a bundle of microtubules called the *phragmoplast* (Figure 9.21, stage 4), which plays a role in the formation of the cell wall that separates the two daughter cells (see Figure 14.38).

These dramatic changes in the spatial organization of microtubules are thought to be accomplished by a combination of two separate mechanisms: (1) the rearrangement of existing microtubules and (2) the disassembly of existing microtubules and reassembly of new ones in different regions of the cell. In the latter case, the microtubules that make up the preprophase band are formed from the same subunits that a few minutes earlier were part of the cortical array or, before that, the phragmoplast.

The Underlying Basis of Microtubule Dynamics   Insight into factors that influence the rates of microtubule assembly and disassembly came initially from studies carried out in vitro. The first successful approach to the in vitro assembly of microtubules was taken in 1972 by Richard Weisenberg of Temple University. Reasoning that cell homogenates should possess all the macromolecules required for the assembly process, Weisenberg obtained tubulin polymerization in crude brain homogenates at 37°C by adding Mg^{2+}, GTP, and EGTA (which binds Ca^{2+}, an inhibitor of polymerization). Weisenberg found that microtubules could be disassembled and reassembled over and over simply by lowering and raising the temperature of the incubation mixture. Figure 9.23 shows three microtubules that were assembled in the test tube from purified tubulin. One of the three microtubules in this figure contains only 11 protofilaments (as indicated by its thinner diameter). It is not unexpected that microtubules assembled in vitro might have abnormal protofilament numbers because they lack a proper template (Figure 9.20*c*) normally provided by the γ-tubulin ring complexes in vivo. In vitro assembly of microtubules is aided greatly by the addition of MAPs or by pieces of microtubules or structures that contain microtubules (Figure 9.24), which serve as templates for the addition of free subunits. In these in vitro studies, tubulin subunits are added primarily to the plus end of the existing polymer.

Early in vitro studies established that GTP is required for microtubule assembly. Assembly of tubulin dimers requires that a GTP molecule be bound to the β-tubulin subunit.[2] GTP hydrolysis is not required for the actual incorporation of the dimer onto the end of a microtubule. Rather, the GTP is hydrolyzed to GDP shortly *after* the dimer is incorporated into a microtubule, and the resulting GDP remains bound to the assembled polymer. After a dimer is released from a microtubule during disassembly and enters the soluble pool, the GDP is replaced by a new GTP. This nucleotide exchange "recharges" the dimer, allowing it to serve once again as a building block for polymerization.

Because it includes GTP hydrolysis, assembly of a microtubule is not an inexpensive cellular activity. Why has such a

---

[2]A molecule of GTP is also bound to the α-tubulin subunit, but it is not exchangeable and it is not hydrolyzed following subunit incorporation. The positions of the guanine nucleotides in the αβ-tubulin heterodimer are shown in Figure 9.9*c*.

During times of rapid microtubule growth, tubulin dimers are added more rapidly than their GTP can be hydrolyzed. The presence of a cap of tubulin-GTP dimers at the plus ends of the protofilaments is thought to favor the addition of more subunits and the growth of the microtubule. However, microtubules with open ends, as in step 1, Figure 9.25, are thought to undergo a spontaneous reaction that leads to closure of the tube (steps 2 and 3). In this model, tube closure is accompanied by the hydrolysis of the bound GTP, which generates subunits that contain GDP-bound tubulin. GDP-tubulin subunits have a different conformation from their GTP-tubulin precursors and are less suited to fit into a straight protofilament. The resulting mechanical strain destabilizes the micro-

FIGURE 9.23 **Microtubules assembled in the test tube.** Electron micrograph of frozen, unfixed microtubules that had polymerized in vitro. The individual protofilaments and their globular tubulin subunits are visible. Note that the middle of the three microtubules contains only 11 protofilaments. (COURTESY OF R. H. WADE, INSTITUT DE BIOLOGIE STRUCTURALE, GRENOBLE, FRANCE.)

costly polymerization pathway evolved? To answer this question, it is useful to consider the effect of GTP hydrolysis on microtubule structure. When a microtubule is growing, the plus end is seen under the electron microscope as an open sheet to which GTP-dimers are added (step 1, Figure 9.25).

FIGURE 9.24 **The assembly of tubulin onto an existing microtubular structure.** Electron micrograph showing the in vitro assembly of brain tubulin onto the plus ends of the microtubules of an axoneme from a *Chlamydomonas* flagellum. (COURTESY OF L. I. BINDER AND JOEL L. ROSENBAUM.)

FIGURE 9.25 **The structural cap model of dynamic instability.** According to the model, the growth or shrinkage of a microtubule depends on the state of the tubulin dimers at the plus end of the microtubule. Tubulin-GTP dimers are depicted in red. Tubulin-GDP dimers are depicted in blue. In a growing microtubule (step 1), the tip consists of an open sheet containing tubulin-GTP subunits. In step 2, the tube has begun to close, forcing the hydrolysis of the bound GTP. In step 3, the tube has closed to its end, leaving only tubulin-GDP subunits. GDP-tubulin subunits have a curved conformation compared to their GTP-bound counterparts, which makes them less able to fit into a straight protofilament. The strain resulting from the presence of GDP-tubulin subunits at the plus end of the microtubule is released as the protofilaments curl outward from the tubule and undergo catastrophic shrinkage (step 4). (FROM A. A. HYMAN AND E. KARSENTI, CELL 84:402, 1996; BY PERMISSION OF CELL PRESS.)
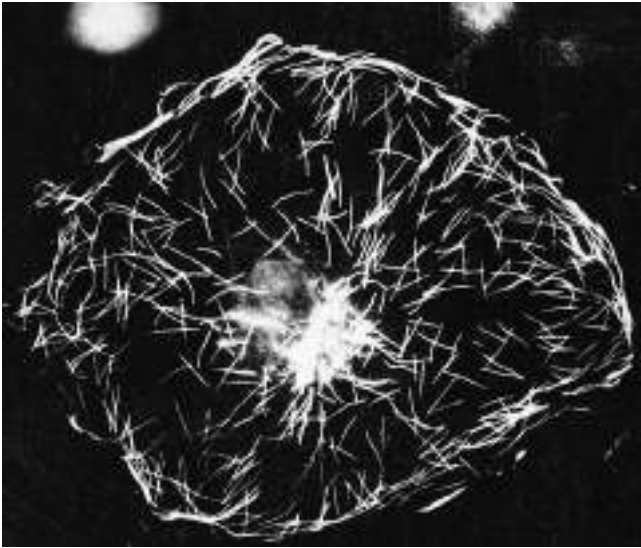
FIGURE 9.26 Microtubule dynamics in living cells. This cultured fibroblast was injected with a small volume of tubulin that had been covalently linked to biotin, a small molecule whose location in the cell is readily determined using fluorescent anti-biotin antibodies. Approximately one minute following injection, the cells were fixed, and the location of biotinylated tubulin that had been incorporated into insoluble microtubules was determined. It is evident from this fluorescence micrograph that, even during periods as short as one minute, tubulin subunits are widely incorporated at the growing ends of cytoskeletal microtubules. (FROM MARC KIRSCHNER, J. CELL BIOL. 102, COVER OF NO. 3, 1986; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

formation. Microtubules can shrink remarkably fast, especially in vivo, which allows a cell to disassemble its microtubular cytoskeleton very rapidly.

The dynamic character of the microtubular cytoskeleton inside a cell can be revealed by microinjecting fluorescently labeled tubulin into a cultured cell. The labeled subunits are rapidly incorporated into the preexisting microtubules of the cytoskeleton, even in the absence of morphological change (Figure 9.26). If individual microtubules are observed under a fluorescence microscope, they appear to grow slowly for a period of time and then shrink rapidly and unexpectedly, as illustrated by the microtubule being followed in Figure 9.27. Because shrinking occurs more rapidly and unexpectedly than elongation, most microtubules disappear from the cell in a matter of minutes and are replaced by new microtubules that grow out from the centrosome.

In 1984, Timothy Mitchison and Marc Kirschner of the University of California, San Francisco, reported on the properties of individual microtubules and proposed that microtubule behavior could be explained by a phenomenon they termed **dynamic instability**. Dynamic instability explains the observation (1) that growing and shrinking microtubules can coexist in the same region of a cell, and (2) that a given microtubule can switch back and forth unpredictably (*stochastically*) between growing and shortening phases, as in Figure 9.27. Dynamic instability is an inherent property of the microtuble itself and, more specifically, of the plus end of the microtubule. As indicated in Figure 9.25, it is the plus end where subunits are added during growth and lost during shrinkage. This does not mean that dynamic instability cannot be influenced by external factors. For example, cells contain a diverse array of proteins (called *+TIPs*) that bind to the dynamic plus ends of microtubules. Some of these +TIPs regulate the rate of the microtubule's growth or shrinkage or the frequency of interconversion between the two phases. Other +TIPs mediate the attachment of the plus end of the microtubule to a specific cellular structure, such as the kinetochore of a mitotic chromosome during cell division or the actin cytoskeleton of the cortex during vesicle transport.

tubule. The strain energy is released as the protofilaments curl outward from the plus end of the tubule and undergo catastrophic depolymerization (step 4). Thus, it would appear that GTP hydrolysis is a fundamental component of the dynamic quality of microtubules. The strain energy stored in a microtubule as a result of GTP hydrolysis makes the microtubule inherently unstable and—in the absence of other stabilizing factors such as MAPs—capable of disassembling soon after its

FIGURE 9.27 Dynamic instability. This series of photographs shows the changes in length of a single microtubule in the growth cone of a neuron. The cell was microinjected with fluorescently labeled tubulin at sufficiently low concentration to produce green fluorescent speckles along the length of microtubules. As discussed on page 321, such speckles provide fixed reference points that can be followed over time. Each of the yellow horizontal lines connects one of these speckles from one time point to another. The blue line indicates the approximate plus end of the microtubule at various time points. From 0 to about 200 seconds, the microtubule undergoes a gradual addition of tubulin subunits at its plus end. Then, from about 200 to 240 seconds the microtubule experiences rapid shrinkage. (FROM ANDREW W. SCHAEFER, NURUL KABIR, AND PAUL FORSCHER, J. CELL BIOL. 158:145, 2002; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

Dynamic instability provides a mechanism by which the plus ends of microtubules can rapidly explore the cytoplasm for appropriate sites of attachment. Attachment temporarily stabilizes microtubules and allows the cell to build the complex cytoskeletal networks discussed in this chapter. Dynamic instability also allows cells to respond rapidly to changing conditions that require remodeling of the microtubular cytoskeleton. One of the most dramatic examples of such remodeling occurs at mitosis when the microtubules of the cytoskeleton are disassembled and remodeled into a bipolar mitotic spindle. This reorganization is associated with a marked change in microtubule stability; microtubules in interphase cells have half-lives that are 5–10 times longer than microtubules in mitotic cells. Unlike the microtubules of the mitotic spindle or cytoskeleton, the microtubules of the organelles to be discussed below lack dynamic activity and, instead, are highly stable.

Cilia and Flagella: Structure and Function

Anyone who has placed a drop of pond water under the lens of a microscope and tried to keep a protozoan from swimming out of the field of view can appreciate the activity of cilia and flagella. **Cilia** and **flagella** are hairlike, motile organelles that project from the surface of a variety of eukaryotic cells. Bacteria also possess structures referred to as *flagella*, but prokaryotic flagella are simple filaments that bear no evolutionary relationship to their eukaryotic counterparts (see Figure 1.14). The discussion that follows pertains only to the eukaryotic organelles.

Cilia and flagella are essentially the same structure. Most biologists use one or the other term based on the type of cell from which the organelle projects and its pattern of movement. According to this distinction, a cilium can be likened to an oar as it moves the cell in a direction perpendicular to the cilium itself. In its power stroke, the cilium is maintained in a rigid state (Figure 9.28a) as it pushes against the surrounding medium. In its recovery stroke, the cilium becomes flexible, offering little resistance to the medium. Cilia tend to occur in large numbers on a cell's surface, and their beating activity is usually coordinated (Figure 9.28b). In multicellular organisms, cilia move fluid and particulate material through various tracts (Figure 9.28c). In humans, for example, the ciliated epithelium lining the respiratory tract propels mucus and trapped debris away from the lungs. Not all cilia are motile: many cells of the body contain a single nonmotile cilium (called a *primary cilium*) that is thought to have a sensory function, monitoring the properties of extracellular fluids. Some of the roles of single motile and nonmotile cilia are discussed in the accompanying Human Perspective.

Direction of locomotion

Power stroke

Recovery stroke

(a)

(b)   10 μm

Cilia

Microvilli

(c)   3.5 μm
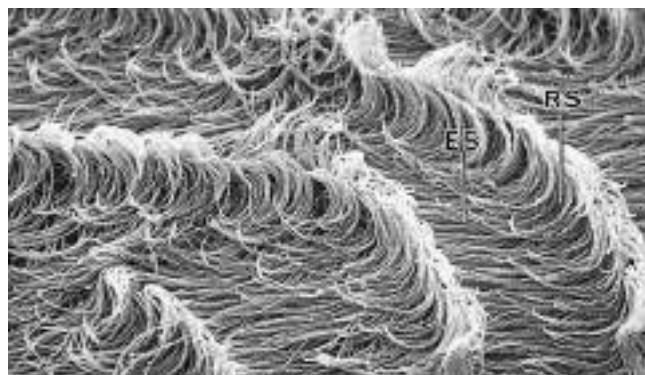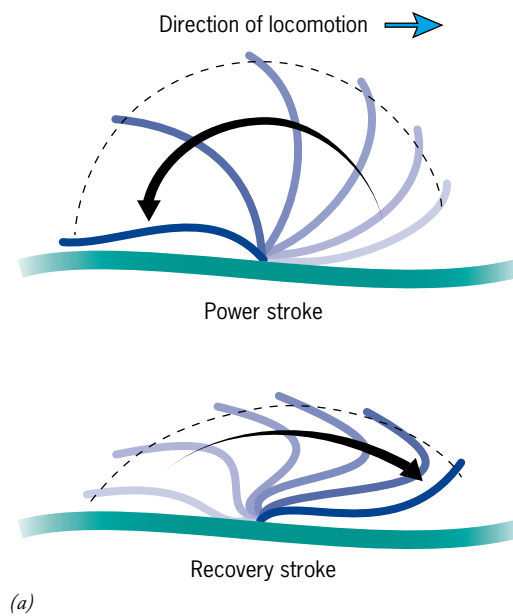
FIGURE 9.28 **Ciliary beat.** (*a*) The various stages in the beat of a cilium. (*b*) The cilia on the surface of a ciliated protozoan beat in metachronal waves in which the cilia in a given row are in the same stage of the beat cycle, but those in adjacent rows are in a different stage. RS, cilia in recovery stroke; ES, cilia in effective power stroke. (*c*) Cilia on the surface of the lip (fimbrium) of a mouse oviduct. (B: REPRINTED WITH PERMISSION FROM G. A. HORRIDGE AND S. L. TAMM, SCIENCE 163:818, 1969; © COPYRIGHT 1969, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE; C: COURTESY OF ELLEN R. DIRKSEN.)

THE HUMAN PERSPECTIVE

The Role of Cilia in Development and Disease

When you look into a mirror, you see a relatively symmetrical organism, one whose left half is essentially a mirror image of the right half. Surgeons, on the other hand, see a strikingly asymmetric organism when they open a person's thoracic or abdominal cavity. The stomach, heart, and spleen, for example, are shifted to the left side of the body, whereas the liver is largely on the right side. On occasion, a physician will see a patient in which the left–right asymmetry of the visceral organs is reversed (a condition called *situs inversus*).

Situs inversus is seen in persons with Kartagener syndrome, which is also characterized by recurrent sinus and respiratory infections and infertility in males. The first clues to the underlying cause of this disorder came in the 1970s when it was discovered that the immotile sperm from these individuals had an abnormal axonemal structure. Depending on the patient, the axonemes may be missing outer or inner dynein arms, central microtubules, or radial spoke structures (see Figure 9.30). Subsequent studies have shown that mutations in a number of genes, including those encoding dynein heavy and intermediate chains, can cause this syndrome. It is understandable that such patients would suffer from respiratory infections, which depend on the clearance of debris and bacteria by respiratory-tract cilia, and from male infertility, but why would roughly half of them exhibit a reversed left–right asymmetry?

The basic body plan of a mammal is laid down during gastrulation in association with a structure called the *embryonic node*. Each cell that makes up the node has a single cilium. These cilia have unusual properties; they are missing the two central microtubules (they have a $9 + 0$ axonemal structure) and exhibit an unusual rotary (twirling) motion. If the motility of these cilia is impaired, as happens in mice harboring mutations in a flagellar dynein gene, roughly half of the animals develop reversed asymmetry, suggesting that left–right asymmetry in these mutants is determined by chance.

The rotation of the nodal cilia moves the surrounding fluid to the left side of the embryonic midline, as was demonstrated by tracking the movement of microscopic fluorescent beads. It was proposed that the extracellular fluid moved by the nodal cilia contains morphogenetic substances (i.e., substances that direct embryonic development) that become concentrated on the left side of the embryo, leading to the eventual formation of different organs on different sides of the midline. This proposal is strongly supported by experimental studies in which mouse embryos were raised in miniaturized chambers in which an artificial flow of fluid across the node could be imposed. When embryos were subjected to a flow of fluid in a direction opposite to that occurring during normal development, the embryos developed with reversed left–right asymmetry.[a]

Many cells of the body have a single, nonmotile primary cilium that is lacking both central microtubules and dynein arms.

These cilia were ignored by researchers for years, but recent studies suggest that they have important functions as "antennae" that sense the chemical and mechanical properties of the fluids into which they project. Consider the primary cilia (Figure 1) on the epithelial cells that line the lumen of microscopic kidney tubules in which urine formation occurs. The importance of these cilia was revealed when it was discovered that a pair of membrane proteins called polycystins are located on the surface of these kidney cilia where they form a Ca^{2+} ion channel. Mutations in *PKD1* and *PKD2*, the genes that encode polycystins, lead to polycystic kidney disease, in which the kidney develops multiple cysts that destroy kidney function. PKD is thought to be a condition that results from a breakdown in the regulation of cell division because the cysts are the result of an abnormally high level of proliferation of the epithelial cells that line parts of the kidney tubules. Mutations in the polycystins are thought to alter the response of the primary cilia to fluid flow, which leads to a disturbance in cal-
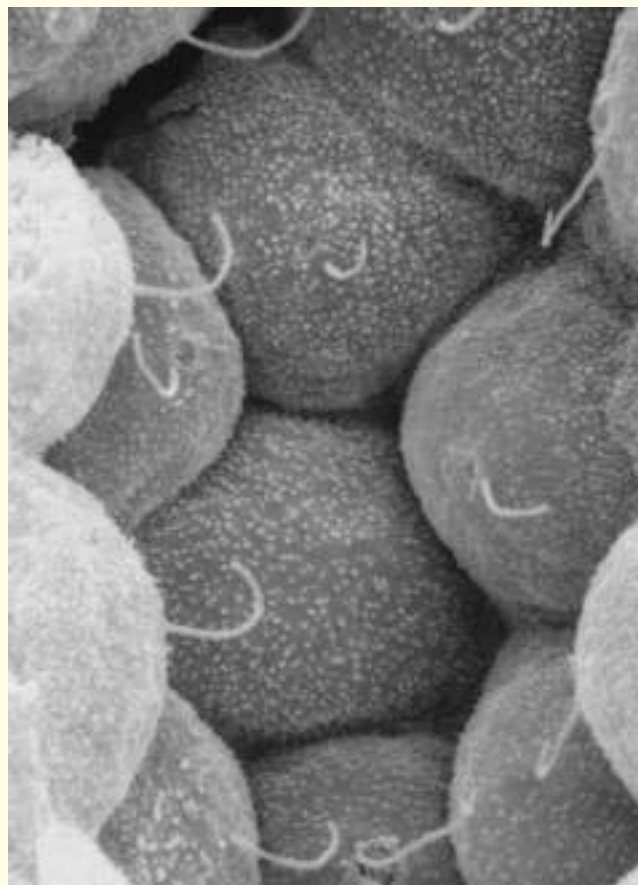
FIGURE 1 Primary cilia. Scanning electron micrograph showing each of these kidney epithelial cells with a single, nonmotile primary cilium. (FROM R. G. KESSEL AND R. H. KARDON, TISSUES AND ORGANS: A TEXTBOOK OF SCANNING ELECTRON MICROSCOPY/VISUALS UNLIMITED.)

---

[a]An alternate but related hypothesis holds that the embryonic node contains two different types of cilia, a population of motile cilia located in the center of the node and nonmotile primary cilia distributed around the node periphery. According to this hypothesis, the motile cilia generate the leftward flow and the nonmotile cilia act as sensory structures that detect the movement and transmit signals that lead to asymmetry.

cium flux across the ciliary membrane, which in turn impairs the transmission of signals to the body of the cell and the nucleus, resulting in abnormal cell proliferation.

The importance of cilia in human development became even more evident with the revelation that Bardet-Biedl syndrome (BBS) is caused by mutations in any one of a number of genes that affect the assembly of basal bodies and cilia. Persons afflicted with BBS exhibit a remarkable range of abnormalities, including polydactyly (extra fingers or toes), situs inversus, obesity, kidney disease, heart defects, mental retardation, genital abnormalities, retinal degeneration, loss of hearing and smell, diabetes, and high blood pressure. The fact that all of these dysfunctions can be traced to abnormalities in basal bodies and cilia illustrates the widespread importance of these structures in organ development and function. Many of the genes responsible for these various cilia-based disorders ("ciliopathies") were first identified in model organisms such as *Chlamydomonas* or *C. elegans*, which provides another example of the importance of basic research on nonvertebrate organisms in furthering our understanding of human disease.

Flagella exhibit a variety of different beating patterns (*waveforms*), depending on the cell type. For example, the single-celled alga pictured in Figure 9.29*a* pulls itself forward by waving its two flagella in an asymmetric manner that resembles the breaststroke of a human swimmer (Figure 9.29*b*). This same algal cell can also push itself through the medium using a symmetric beat resembling that of a sperm (shown in Figure 9.34). The degree of asymmetry in the pattern of the beat in the algal cell is regulated by the internal calcium ion concentration.

An electron micrograph of a cross section of a cilium or flagellum is one of the most familiar images in cell biology (Figure 9.30*a*). The entire ciliary or flagellar projection is covered by a membrane that is continuous with the plasma membrane of the cell. The core of the cilium, called the **axoneme**, contains an array of microtubules that runs longitudinally through the entire organelle. With rare exceptions, the axoneme of a motile cilium or flagellum consists of nine peripheral doublet microtubules surrounding a central pair of single microtubules. This same microtubular structure, which is known as the 9 + 2 array, is seen in axonemes from protists to mammals and serves as another of the many reminders that all living eukaryotes have evolved from a common ancestor. All microtubules of the axoneme have the same polarity: their plus ends are at the tip of the projection, and their minus ends at the base. Each peripheral doublet consists of one complete microtubule, the *A tubule*, and one incomplete microtubule, the *B tubule*, the latter containing 10 or 11 subunits rather than the usual 13.

The basic structure of the axoneme was first described in 1952—in plants by Irene Manton and in animals by Don Fawcett and Keith Porter. As the resolving power of electron microscopes improved, certain of the less obvious components became visible (Figure 9.30*b*). The central tubules were seen to be enclosed by the *central sheath*, which is connected to the A tubules of the peripheral doublets by a set of *radial spokes*. The doublets are connected to one another by an *interdoublet bridge* composed of an elastic protein, nexin. Of particular importance was the observation that a pair of "arms"—an *inner arm* and an *outer arm*—project from the A tubule (Figure 9.30*a*). A longitudinal section, that is, one cut through the

(a)

(b)
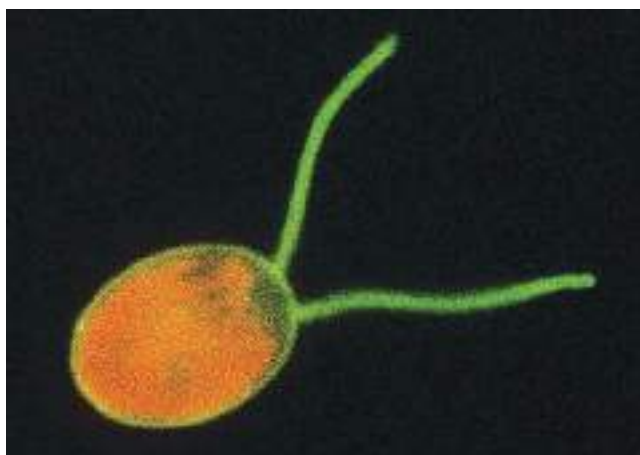
FIGURE 9.29 Eukaryotic flagella. (*a*) The unicellular alga *Chlamydomonas reinhardtii*. The two flagella appear green after binding a fluorescent antibody directed against a major flagellar membrane protein. The red color results from autofluorescence of the cell's chlorophyll. Unlike many flagellated organisms, *Chlamydomonas* does not need its flagella to survive and reproduce, so that mutant strains exhibiting various types of flagellar defects can be cultured. (*b*) The forward movement of *Chlamydomonas* is accomplished by an asymmetric waveform that resembles the breast stroke. A different type of flagellar waveform is shown in Figure 9.34. (A: COURTESY OF ROBERT BLOODGOOD, UNIVERSITY OF VIRGINIA.)
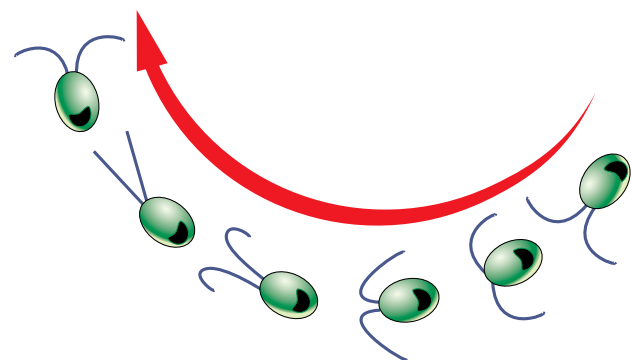
(a)   50 nm

(b)
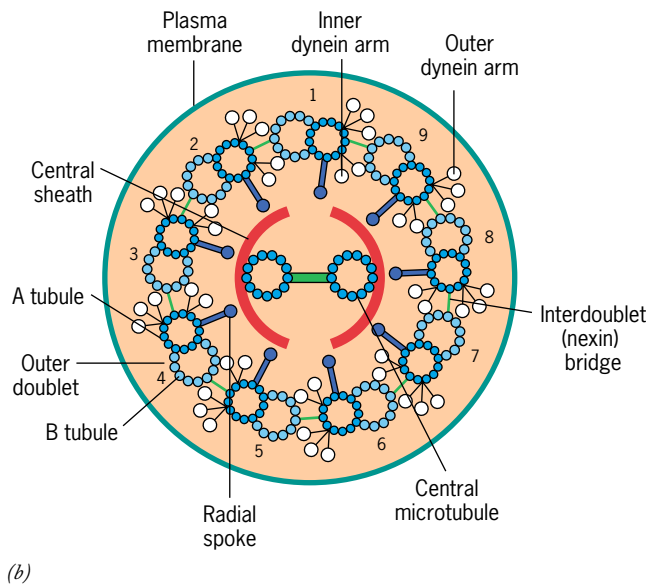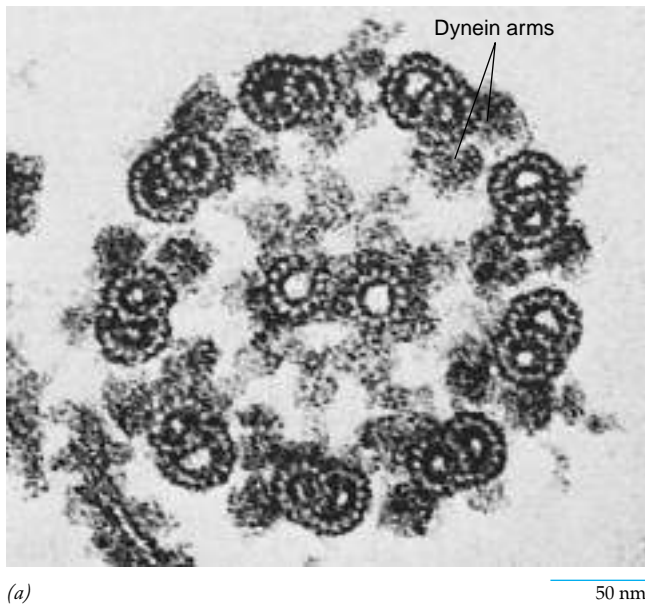
FIGURE 9.30 The structure of a ciliary or flagellar axoneme. (*a*) Cross section through a sperm axoneme. The peripheral doublets are seen to consist of a complete and an incomplete microtubule, whereas the two central microtubules are complete. The dynein arms are seen as "fuzzy" projections from the wall of the complete microtubule. The molecular structure of these arms is discussed in a later section. (*b*) Schematic diagram of an axoneme of a protist showing the structure of the microtubular fibers, the two types of dynein arms (three-headed outer arms and two-headed inner arms), the nexin links between the doublets, the central sheath surrounding the central microtubules, and the radial spokes projecting from the outer doublets toward the central sheath. A more detailed and complex view of the molecular architecture of the axoneme has been obtained with the application of cryoelectron tomography (Section 18.2); see *PNAS* 102:15889, 2005 and *Science* 313:944, 2006. (*Note:* The cilia and flagella of animals typically contain two-headed outer dynein arms.) (A: COURTESY OF LEWIS G. TILNEY AND K. FUJIWARA.)

(a)   60 nm

(b)

FIGURE 9.31 Longitudinal view of an axoneme. (*a*) Electron micrograph of a median longitudinal section of a straight region of a cilium. The radial spokes are seen joining the central sheath with the A microtubule of the doublet. (*b*) Schematic diagram of a longitudinal section of a flagellar doublet. Radial spokes emerge in groups of three, which repeat (at 96 nm in this case) along the length of the microtubule. Outer dynein arms are spaced at intervals of 24 nm. (A: FROM FRED D. WARNER AND PETER SATIR, J. CELL BIOL. 63:41, 1974; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

axoneme parallel to its long axis, reveals the continuous nature of the microtubules and the discontinuous nature of the other elements (Figure 9.31*a*).

As noted earlier, a cilium or flagellum emerges from a *basal body* (Figure 9.32*a*) similar in structure to the centriole of Figure 9.18*a*. The A and B tubules of the basal body elongate to

Cilium

Basal body

(a)

Cilium   Basal body

Central microtubule   Central sheath   Radial spoke   Outer doublet

(b)
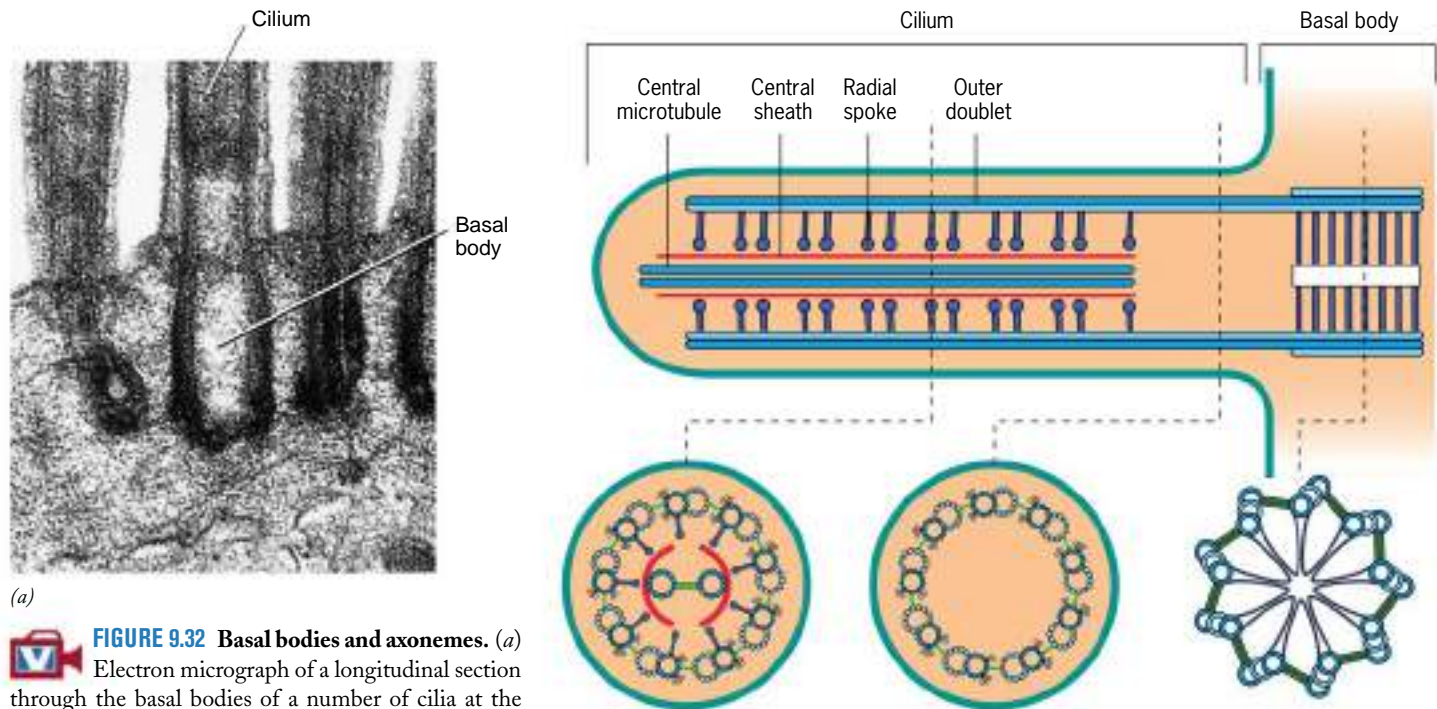
FIGURE 9.32 Basal bodies and axonemes. (*a*) Electron micrograph of a longitudinal section through the basal bodies of a number of cilia at the apical surface of epithelial cells of a rabbit oviduct. These basal bodies arise from centrioles that are generated in the cytoplasm and migrate to sites beneath the plasma membrane. (*b*) Schematic diagram of the structural relationship between the microtubules of the basal body and the axoneme of a cilium or flagellum. (A: COURTESY OF R. G. W. ANDERSON.)

form the doublets of the cilium or flagellum (Figure 9.32*b*). If a cilium or flagellum is sheared from the surface of a living cell, a new organelle is regenerated as an outgrowth of the basal body. As with other microtubular structures, the growth of an axoneme occurs at the plus (outer) ends of its microtubules. How does a cell organize and maintain a construction site at the outer tip of an axoneme situated micrometers from the body of the cell where the synthesis of building materials takes place?

Biologists had been watching the flagella of living cells for over a hundred years, but it wasn't until 1993 that researchers observed the movement of particles in the space between the peripheral doublets and the surrounding plasma membrane. Subsequent studies revealed a process known as **intraflagellar transport (IFT)**, which is responsible for assembling and maintaining these organelles. IFT depends on the activity of both plus end- and minus end-directed microtubular movement (Figure 9.33). Kinesin II moves complex arrays of

FIGURE 9.33 Intraflagellar transport (IFT). Electron micrograph of a longitudinal section of a *Chlamydomonas* flagellum showing two rows of protein particles (bounded by arrowheads) situated between the outer doublet microtubules and the flagellar plasma membrane. As illustrated in the inset, each row of particles and its associated cargo of axonemal proteins are moved along the outer doublet microtubule by a motor protein, either cytoplasmic dynein 1b if they are moving toward the flagellar base or kinesin II if the particles are moving toward the tip of the flagellum. (MICROGRAPH FROM KEITH G. KOZMINSKI ET AL., J. CELL BIOL. 131:1520, 1995, COURTESY OF JOEL L. ROSENBAUM; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

IFT particles

Kinesin II

(−)   (+)

100 nm

(−)   (+)

Cytoplasmic dynein 1b

Flagellar plasma membrane
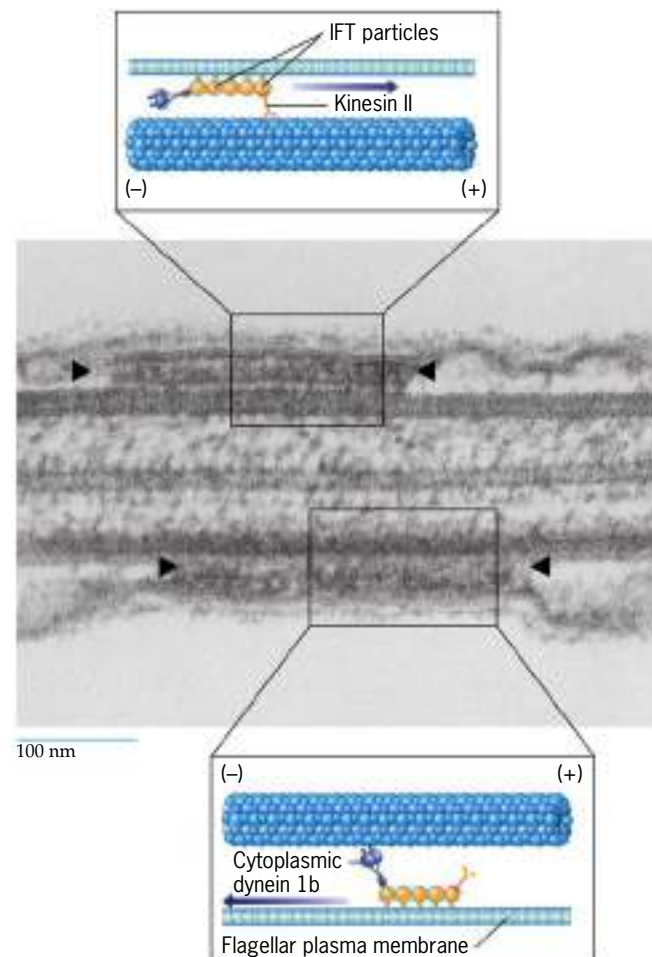
building materials along protofilaments of the peripheral doublets to the assembly site at the tip of the growing axoneme. Kinesin II molecules (and recycled axonemal proteins) are transported back toward the basal body along the same flagellar microtubules by a cytoplasmic dynein-powered mechanism. Recent studies have also implicated IFT in the transport of a number of important signaling molecules. Several of the characteristic features of Bardet-Biedel syndrome (page •••), including obesity and extra digits, can be traced to disturbances in the transport of signaling molecules.

The Dynein Arms The machinery for ciliary and flagellar locomotion resides within the axoneme. This is illustrated in the experiment shown in Figure 9.34 in which a sperm tail axoneme devoid of its overlying membrane is still capable of normal, sustained beating in the presence of added Mg^{2+} and ATP. The greater the ATP concentration, the faster the beat frequency of these "reactivated" organelles.

The protein responsible for conversion of the chemical energy of ATP into the mechanical energy of ciliary locomotion was isolated in the 1960s by Ian Gibbons of Harvard University. Gibbons' experiments provide an elegant example of the relationship between structure and function in biological systems and the means by which the relationship can be revealed by experimental analysis. Using various solutions capable of solubilizing different components, Gibbons carried out a chemical dissection of cilia from the protozoan *Tetrahymena* (Figure 9.35). An intact cilium is shown in cross section in step 1. The dissection began by dissolving the enclosing plasma membrane in the detergent digitonin (step 2). The isolated axonemes of the insoluble fraction were then immersed in a solution containing EDTA, a compound that binds and removes (chelates) divalent ions. When EDTA-treated axonemes were examined in the electron microscope, the central tubules were missing, as were the arms projecting from the A tubules (step 3). Simultaneous with the loss of these structures, the insoluble axonemes lost their ability to hydrolyze ATP, while the supernatant gained it. The ATPase found in the supernatant

was a huge protein (up to 2 million daltons) that Gibbons named *dynein* (from the words *dyne*, meaning "force," and *protein*). This protein is now referred to as **ciliary** (or **axonemal**) **dynein** to distinguish it from cytoplasmic dynein, a related protein involved in organelle and particle transport. When the insoluble parts of the axoneme were mixed with the soluble protein in the presence of Mg^{2+}, much of the ATPase activity once again became associated with the insoluble material in the mixture (step 4). Examination of the insoluble fraction showed that the arms had reappeared on the A tubules of the

FIGURE 9.34 A sea urchin sperm reactivated with 0.2 mM ATP following demembranation with the detergent Triton X-100. This multiple-exposure photomicrograph was obtained with five flashes of light and shows the reactivated flagellum at different stages of its beat. (FROM CHARLES J. BROKAW AND T. F. SIMONICK, J. CELL BIOL. 75:650, 1977; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

FIGURE 9.35 Steps in the chemical dissection of cilia from the protozoan *Tetrahymena*. The numbered steps are described in the text.

axonemes. Gibbons concluded that the arms seen in the micrographs were equivalent to the dynein ATPase molecules recovered in the EDTA-containing solution, and thus it was the arms that released the energy required for locomotion.

In subsequent investigations it was found that treatment of isolated sperm axonemes with high salt (0.6 M NaCl) would selectively remove the outer arms, leaving the inner arms in place (step 5). When ATP was added to axonemes lacking outer arms, they beat at about half the rate of the intact axoneme, though they beat with a normal waveform.

An electron micrograph of an outer dynein molecule (i.e., outer dynein arm) from a *Tetrahymena* axoneme is shown in Figure 9.36*a*. This dynein molecule consists of three heavy chains and a number of intermediate and light chains. As discussed on page 331, each dynein heavy chain is composed of a long stem, a wheel-shaped head, and a stalk. Figure 9.36*b,c* show high-resolution electron micrographs of single dynein heavy chains from *Chlamydomonas* flagella prepared under two different conditions. The difference in conformation between the molecules in these two micrographs (shown merged in Figure 9.36*d*) is proposed to represent the power stroke of a flagellar dynein motor protein. In this model, the rotation of the head depicted in Figure 9.36*b',c'* serves as the basic driving force for ciliary/flagellar motion. To understand the mechanism responsible for this type of motility, we have to reconsider the structure of the axoneme.

The Mechanism of Ciliary and Flagellar Locomotion As will be discussed later in this chapter, the contraction of muscle results from the sliding of actin filaments over adjacent myosin filaments. The sliding force is generated by ratchet-like crossbridges that reside in the head of the myosin molecule. With the muscle system as a model, it was proposed that ciliary movement could be explained by the sliding of adjacent microtubular doublets relative to one another. According to this proposal, the dynein arms pictured in Figure 9.36 act as swinging crossbridges that generate the forces required for ciliary or flagellar movement. The sequence of events is shown in Figure 9.37.

In the intact axoneme, the stem of each dynein molecule (with its associated intermediate and light chains) is tightly anchored to the outer surface of the A tubule, with the globular heads and stalks projecting toward the B tubule of the neighboring doublet. In step 1 of Figure 9.37, the dynein arms anchored along the A tubule of the lower doublet attach to binding sites on the B tubule of the upper doublet. In step 2, the dynein molecules undergo a conformational change, or power stroke, that causes the lower doublet to slide toward the basal end of the upper doublet. This conformational change in a dynein heavy chain is depicted in Figure 9.36*b–d*. In step 3 of Figure 9.37, the dynein arms have detached from the B tubule of the upper doublet. In step 4, the arms have reattached to the upper doublet so that another cycle can begin. (An electron micrograph of an axoneme showing the dynein arms from one doublet attached to the adjacent doublet is shown in Figure 18.18.)

Nexin is an elastic protein that connects adjacent doublets (Figure 9.30*a*). These nexin bridges play an important role in
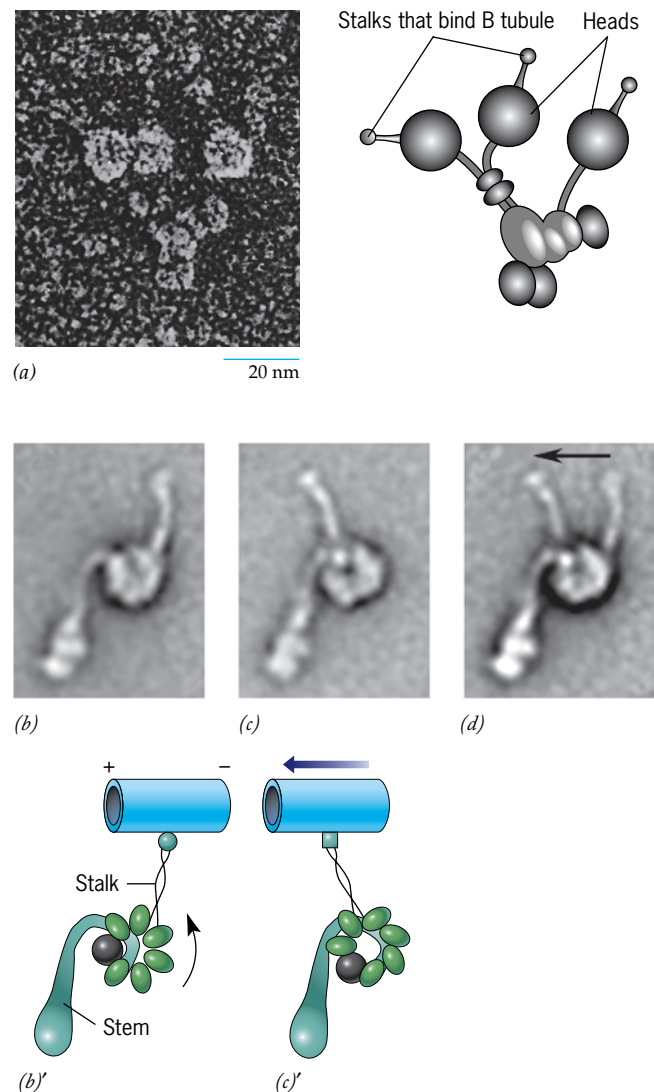
FIGURE 9.36 A model of the structure and function of flagellar/ciliary dynein. (*a*) Platinum replica of a rotary-shadowed, flagellar outer-arm dynein molecule prepared by rapid-freeze, deep-etch electron microscopy. Each of the three heavy chains forms a prominent globular head with an extension (stalk) that functions in linking the dynein arm to the neighboring doublet. An interpretive drawing is shown on the right. (*b–d*) High-resolution micrographs (with *b'*, *c'* interpretive models below) of a flagellar dynein heavy chain before (*b*) and after (*c*) its power stroke. The motor domain, which consists of a number of modules arranged in a wheel, is seen to have rotated, which has caused the stalk to move in a leftward direction. The image shown in (*d*) is a composite of molecules showing the position of the stalk before and after the power stroke. The power stroke causes the microtubule bound to the tip of the stalk to slide 15 nm in a leftward direction relative to the motor domain. In vitro motility assays suggest that dynein may be capable of "shifting gears" so that it can take shorter, more powerful steps as it moves a load of increasing mass. (*Note:* An alternate model of dynein function is discussed in *Science* 322:1647, 2008.) (A: COURTESY OF JOHN E. HEUSER AND URSULA W. GOODENOUGH; B–D: REPRODUCED WITH PERMISSION FROM STAN A. BURGESS ET AL., NATURE 421:717, 2003; COPYRIGHT 2003 BY MACMILLAN MAGAZINES LIMITED.)
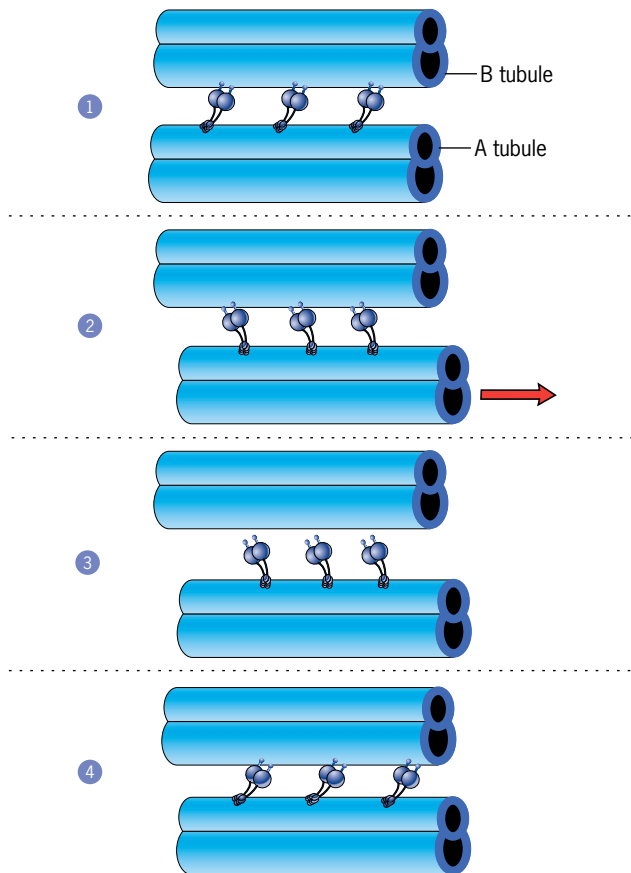
FIGURE 9.37 Schematic representation of the forces that drive ciliary or flagellar motility. The steps are described in the text. A true representation of the power stroke is shown in the previous figure.

FIGURE 9.38 The sliding-microtubule mechanism of ciliary or flagellar motility. Schematic diagram of the sliding of neighboring microtubules relative to one another. When the cilium is straight, all the outer doublets end at the same level (*center*). Cilium bending occurs when the doublets on the inner side of the bend slide beyond those on the outer side (*top and bottom*). The movement of the dynein arms responsible for the sliding of neighboring microtubules was shown in the previous figures. (FROM D. VOET AND J. G. VOET, BIOCHEMISTRY, 2D ED.; COPYRIGHT © 1995 JOHN WILEY AND SONS, INC. REPRINTED BY PERMISSION OF JOHN WILEY & SONS, INC.)

ciliary/flagellar movement by limiting the extent that adjacent doublets can slide over one another. The resistance to sliding provided by the nexin bridges causes the axoneme to bend in response to the sliding force. Sliding on one side of the axoneme alternates with sliding on the other side so that a part of the cilium or flagellum bends first in one direction and then in the opposite direction (Figure 9.38). This requires that, at any given time, the dynein arms on one side of the axoneme are active, while those on the other side are inactive. As a result of this difference in dynein activity, the doublets on the inner side of the bend (at the top and bottom of Figure 9.38) extend beyond those on the opposite side of the axoneme.

Evidence has steadily accumulated in favor of the sliding-microtubule theory and the role of the dynein arms. Microtubule sliding in beating flagellar axonemes has been directly visualized, as illustrated in the photographs of Figure 9.39. In this experiment, isolated axonemes were incubated with tiny gold beads that became attached to the exposed outer surface of the peripheral doublets. The beads served as fixed markers of specific sites on the doublets. The relative position of the beads was then monitored as the axonemes were stimulated to beat by addition of ATP. As the axonemes bent back and forth, the distances between the beads on different doublets increased and decreased in an alternating manner (Fig-

ure 9.39) as would be expected if neighboring doublets slid up and down over one another.

REVIEW ?

1. Describe three functions of microtubules.
2. Contrast the apparent roles of kinesin and cytoplasmic dynein in axonal transport.
3. What is an MTOC? Describe the structure of three different MTOCs and how each of them functions.
4. What is the role of GTP in the assembly of microtubules? What is meant by the term *dynamic instability*? What role does it play in cellular activity?
5. Contrast the movements of a flagellum and a cilium. Describe the power stroke of a dynein heavy chain. Describe the mechanism by which cilia and flagella are able to undergo their bending movements.

FIGURE 9.39 Experimental demonstration of microtubule sliding. Sea urchin sperm were demembranated, reactivated with ATP, and photographed by multiple exposure as in Figure 9.34. In this experiment, gold beads were attached to the exposed outer doublet microtubules, where they could serve as markers for specific sites along different doublets. As the flagellum underwent beating, the relative positions of the beads were monitored. As shown here, the beads moved farther apart and then closer together as the flagellum undulated, indicating that the doublets are sliding back and forth relative to one another. (FROM CHARLES J. BROKAW, J. CELL BIOL. 114, COVER OF NO. 6, 1991; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

9.4 INTERMEDIATE FILAMENTS

The second of the three major cytoskeletal elements to be discussed was seen in the electron microscope as solid, unbranched filaments with a diameter of 10–12 nm. They were named **intermediate filaments** (or **IFs**). To date, intermediate filaments have only been identified in animal cells. Intermediate filaments are strong, flexible ropelike fibers that provide mechanical strength to cells that are subjected to physical stress, including neurons, muscle cells, and the epithelial cells that line the body's cavities. Unlike microfilaments and microtubules, IFs are a chemically heterogeneous group of structures that, in humans, are encoded by approximately 70 different genes. The polypeptide subunits of IFs can be divided into five major classes based on the type of cell in which they are found (Table 9.2) as well as biochemical, genetic, and immunologic criteria. We will restrict the present discussion to classes I–IV, which are found in the construction of cytoplasmic filaments, and consider type V IFs (the lamins), which are present as part of the inner lining of the nucleus, in Section 12.1.

IFs radiate through the cytoplasm of a wide variety of animal cells and are often interconnected to other cytoskeletal filaments by thin, wispy cross-bridges (Figure 9.40). In many cells, these cross-bridges consist of an elongated dimeric protein called *plectin* that can exist in numerous isoforms. Each plectin molecule has a binding site for an intermediate filament at one end and, depending on the isoform, a binding site for another intermediate filament, microfilament, or microtubule at the other end.

Although IF polypeptides have diverse amino acid sequences, all share a similar structural organization that allows them to form similar-looking filaments. Most notably, the polypeptides of IFs all contain a central, rod-shaped, α-helical domain of similar length and homologous amino acid sequence. This long fibrous domain makes the subunits of intermediate filaments very different from the globular tubulin and actin subunits of microtubules and microfilaments. The central fibrous domain is flanked on each side by globular

TABLE 9.2 Properties and Distribution of the Major Mammalian Intermediate Filament Proteins

IF protein	Sequence type	Primary tissue distribution
Keratin (acidic) (28 different polypeptides)	I	Epithelia
Keratin (basic) (26 different polypeptides)	II	Epithelia
Vimentin	III	Mesenchymal cells
Desmin	III	Muscle
Glial fibrillary acidic protein (GFAP)	III	Astrocytes
Peripherin	III	Peripheral neurons
Neurofilament proteins		Neurons of central
NF-L	IV	and peripheral
NF-M	IV	nerves
NF-H	IV	
Nestin	IV	Neuroepithelial
Lamin proteins		All cell types
Lamin A	V	(Nuclear envelopes)
Lamin B	V	
Lamin C	V	

More detailed tables can be found in *Trends Biochem Sci.* 31:384, 2006, *Genes and Development* 21:1582, 2007, and *Trends Cell Biol.* 18:29, 2008.
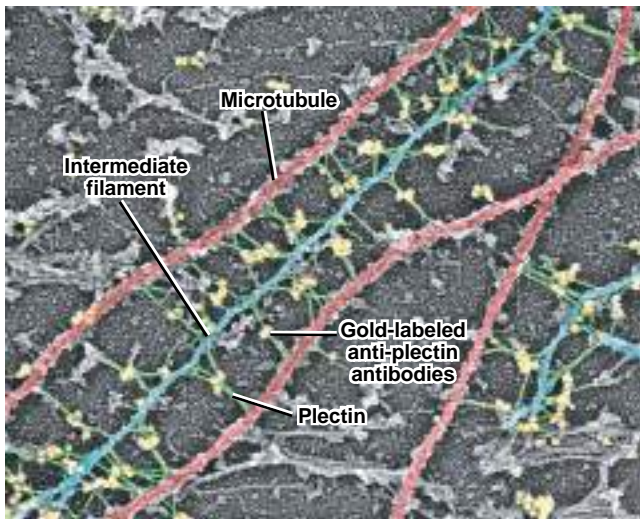
FIGURE 9.40 **Cytoskeletal elements are connected to one another by protein cross-bridges.** Electron micrograph of a replica of a small portion of the cytoskeleton of a fibroblast after selective removal of actin filaments. Individual components have been digitally colorized to assist visualization. Intermediate filaments (blue) are seen to be connected to microtubules (red) by long wispy cross-bridges consisting of the fibrous protein plectin (green). Plectin is localized by antibodies linked to colloidal gold particles (yellow). (Courtesy of Tatyana Svitkina and Gary Borisy.)

FIGURE 9.41 **A model of intermediate filament assembly and architecture.** Each monomer has a pair of globular terminal domains (red) separated by a long α-helical region (step 1). Pairs of monomers associate in parallel orientation with their ends aligned to form dimers (step 2). Depending on the type of intermediate filament, the dimers may be composed of identical monomers (homodimers) or nonidentical monomers (heterodimers). Dimers in turn associate in an antiparallel, staggered fashion to form tetramers (step 3), which are thought to be the basic subunit in the assembly of intermediate filaments. In the model shown here, 8 tetramers associate laterally to form a unit length of the intermediate filament (step 4). Highly elongated intermediate filaments are then formed from the end-to-end association of these unit lengths (step 5).

domains of variable size and sequence (step 1, Figure 9.41). Two such polypeptides spontaneously interact as their α-helical rods wrap around each other to form a ropelike dimer approximately 45 nm in length (step 2). Because the two polypeptides are aligned parallel to one another in the same orientation, the dimer has polarity, with one end defined by the C-termini of the polypeptides and the opposite end by their N-termini.

Intermediate Filament Assembly and Disassembly

The basic building block of IF assembly is thought to be a rodlike tetramer formed by two dimers that become aligned side by side in a staggered fashion with their N- and C-termini pointing in opposite (antiparallel) directions, as shown in Figure 9.41, step 3. Because the dimers point in opposite directions, the tetramer itself lacks polarity. Recent studies of the self-assembly of IFs in vitro suggest that 8 tetramers associate with one another in a side-by-side (lateral) arrangement to form a filament that is one unit in length (about 60 nm) (step 4). Subsequent growth of the polymer is accomplished as these unit lengths of filaments associate with one another in an end-to-end fashion to form the highly elongated intermediate filament (step 5). None of these assembly steps is thought to require the direct involvement of either ATP or GTP. Because the tetrameric building blocks lack polarity, so too does the assembled filament, which is another feature that distinguishes IFs from other cytoskeletal elements.

Intermediate filaments tend to be less sensitive to chemical agents than other types of cytoskeletal elements and more difficult to solubilize. Because of their insolubility, IFs were initially thought to be permanent, unchanging structures, so it came as a surprise to find that they behave dynamically in vivo. When labeled keratin subunits are injected into cultured skin cells, they are rapidly incorporated into existing IFs. Surprisingly, the subunits are not incorporated at the ends of the filament, as might have been expected by analogy with microtubule and microfilament assembly, but rather into the filament's interior (Figure 9.42). The results depicted in Figure 9.42 might reflect the exchange of unit lengths of filament (as shown in Figure 9.41, step 4) directly into an existing IF network. Unlike the other two major cytoskeletal elements, assembly and disassembly of IFs are controlled primarily by phosphorylation and dephosphorylation of the subunits. For
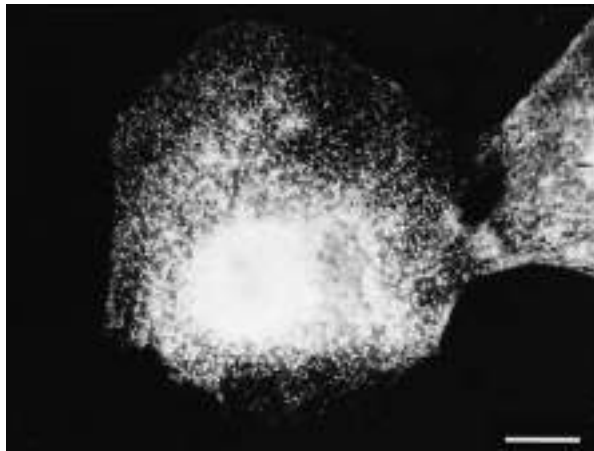
(a)
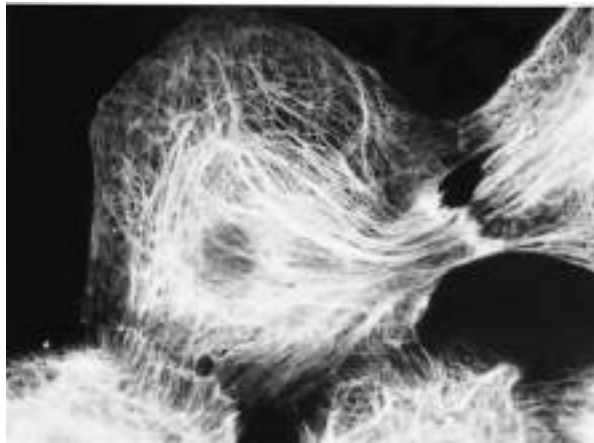
(b)

FIGURE 9.42 **Experimental demonstration of the dynamic character of intermediate filaments.** These photographs show the results of an experiment in which biotin-labeled type I keratin was microinjected into cultured epithelial cells and localized 20 minutes later using immunofluorescence. The photograph in *a* shows the localization of the injected biotinylated keratin (as revealed by anti-biotin antibodies) that had become incorporated into filaments during the 20-minute period following injection. The photograph in *b* shows the distribution of intermediate filaments in the cell as revealed by anti-keratin antibodies. The dotlike pattern of fluorescence in *a* indicates that the injected subunits are incorporated into the existing filaments at sites throughout their length, rather than at their ends. (Compare with a similar experiment with labeled tubulin in Figure 9.26.) Bar, 10 μm. (FROM RITA K. MILLER, KAREN VIKSTROM, AND ROBERT D. GOLDMAN, J. CELL BIOL. 113:848, 1991; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

example, phosphorylation of vimentin filaments by protein kinase A leads to their disassembly.

Types and Functions of Intermediate Filaments

Keratin filaments constitute the primary structural proteins of epithelial cells (including epidermal cells, liver hepatocytes, and pancreatic acinar cells). Figure 9.43*a* shows a schematic

view of the spatial arrangement of the keratin filaments of a generalized epithelial cell, and Figure 9.43*b* shows the actual arrangement within a pair of cultured epidermal cells. Keratin-containing IFs radiate through the cytoplasm, tethered to the nuclear envelope in the center of the cell and anchored at the outer edge of the cell by connections to the cytoplasmic plaques of desmosomes and hemidesmosomes (pages 251 and 244). Figure 9.43*a* also depicts the interconnections between IFs and the cell's microtubules and microfilaments, which transforms these otherwise separate elements into an integrated cytoskeleton. Because of these various physical connections, the IF network is able to serve as a scaffold for organizing and maintaining cellular architecture and for absorbing mechanical stresses applied by the extracellular environment.

The cytoplasm of neurons contains loosely packed bundles of intermediate filaments whose long axes are oriented parallel to that of the nerve cell axon (see Figure 9.13*b*). These IFs, or **neurofilaments**, as they are called, are composed of three distinct proteins: NF-L, NF-H, and NF-M, all of the type IV group of Table 9.2. Unlike the polypeptides of other IFs, NF-H and NF-M have sidearms that project outward from the neurofilament. These sidearms are thought to maintain the proper spacing between the parallel neurofilaments of the axon (see Figure 9.13*b*). In the early stages of differentiation when the axon is growing toward a target cell, it contains very few neurofilaments but large numbers of supporting microtubules. Once the nerve cell has become fully extended, it becomes filled with neurofilaments that provide support as the axon increases dramatically in diameter. Aggregation of NFs is seen in several human neurodegenerative disorders, including ALS and Parkinson's disease. These NF aggregates may block axonal transport, leading to the death of affected neurons.

Efforts to probe IF function have relied largely on genetically engineered mice that fail to produce a particular IF polypeptide (a gene knockout) or produce an altered IF polypeptide. These studies have revealed the importance of intermediate filaments in particular cell types. For example, mice carrying deletions in the gene encoding K14, a type I keratin polypeptide normally synthesized by cells of the basal epidermal layer, have serious health problems. These mice are so sensitive to mechanical pressure that even mild trauma, such as that occurring during passage through the birth canal or during nursing by the newborn, can cause severe blistering of the skin or tongue. This phenotype bears strong resemblance to a rare skin-blistering disease in humans, called *epidermolysis bullosa simplex* (*EBS*).[3] Subsequent analysis of EBS patients has shown that they carry mutations in the gene that encodes the homologous K14 polypeptide (or the K5 polypeptide, which forms dimers with K14). These studies confirm the role of IFs in imparting mechanical strength to cells situated in epithelial layers. Similarly, knockout mice that

---

[3]As mentioned in Chapter 7, similar types of blistering diseases can be caused by defects in proteins of the hemidesmosome, which holds the basal layer of the epidermis to the basement membrane.
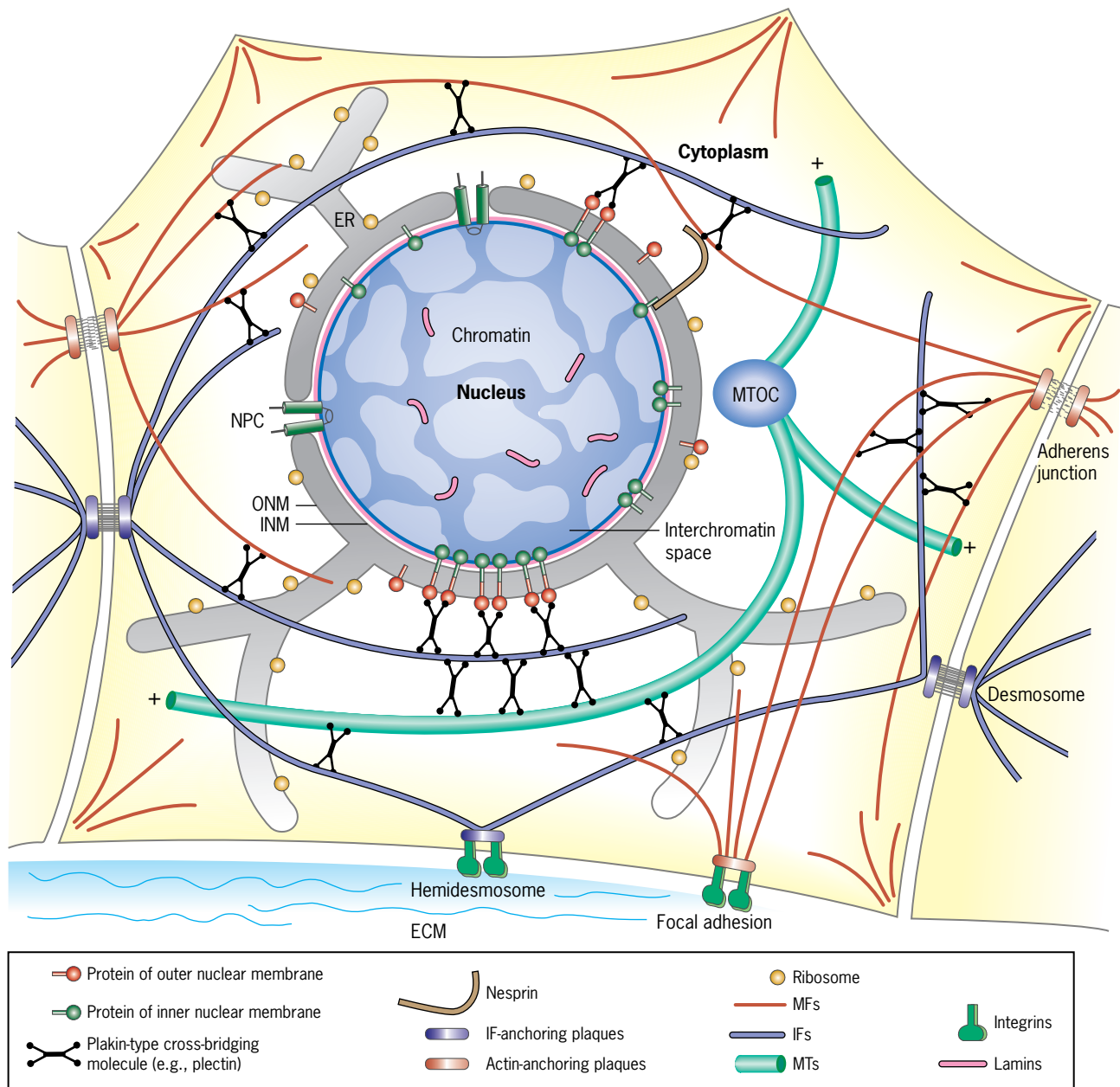
(a)
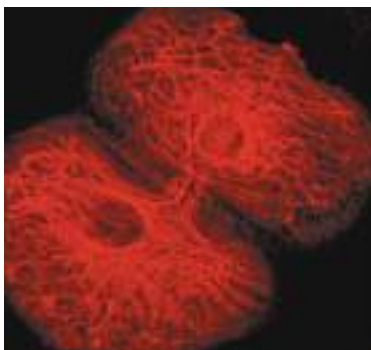
(b)

FIGURE 9.43 The organization of intermediate filaments within an epithelial cell. (*a*) In this schematic drawing, IFs are seen to radiate throughout the cell, being anchored at both the outer surface of the nucleus and the inner surface of the plasma membrane. Connections to the nucleus are made via proteins that span both membranes of the nuclear envelope and to the plasma membrane via specialized sites of adhesion such as desmosomes and hemidesmosomes. IFs are also seen to be interconnected to both of the other types of cytoskeletal fibers. Connections to microtubules and microfilaments are made primarily by members of the plakin family of proteins, such as the dimeric plectin molecule shown in Figure 9.40. (*b*) Distribution of keratin-containing intermediate filaments in cultured skin cells (keratinocytes). The filaments are seen to form a cagelike network around the nucleus and also extend to the cell periphery (A: REPRINTED WITH PERMISSION FROM H. HERRMANN ET AL., NATURE REVS. MOL. CELL BIOL. 8:564, 2007; © COPYRIGHT 2007, MACMILLAN MAGAZINES LTD.; B: FROM PIERRE A. COULOMBE AND M. BISHR OMARY, CURR. OPIN. CELL BIOL. 14:111, 2002.)

fail to produce the desmin polypeptide exhibit serious cardiac and skeletal muscle abnormalities. Desmin plays a key structural role in maintaining the alignment of the myofibrils of a muscle cell, and the absence of these IFs makes the cells extremely fragile. An inherited human disease, named *desmin–related myopathy*, is caused by mutations in the gene that

encodes desmin. Persons with this disorder suffer from skeletal muscle weakness, cardiac arrhythmias, and eventual congestive heart failure. Not all IF polypeptides have such essential functions. For example, mice that lack the vimentin gene, which is expressed in fibroblasts, macrophages, and white blood cells, show relatively minor abnormalities, even though the affected cells lack cytoplasmic IFs. It is evident from these studies that IFs have tissue-specific functions, which are more important in some cells than in others.

REVIEW ?

1. Give some examples that reinforce the suggestion that intermediate filaments are important primarily in tissue-specific functions rather than in basic activities that are common to all cells.
2. Compare and contrast microtubule assembly and intermediate filament assembly.

9.5 MICROFILAMENTS

Cells are capable of remarkable motility. The neural crest cells in a vertebrate embryo leave the developing nervous system and migrate across the entire width of the embryo, forming such diverse products as the pigment cells of the skin, teeth, and the cartilage of the jaws (see Figure 7.11). Legions of white blood cells patrol the tissues of the body searching for debris and microorganisms. Certain parts of cells can also be motile; broad projections of epithelial cells at the edge of a wound act as motile devices that pull the sheet of cells over the damaged area, sealing the wound. Similarly, the leading edge of a growing axon sends out microscopic processes that survey the substratum and guide the cell toward a synaptic target. All of these various examples of motility share at least one component: they all depend on microfilaments, the third major type of cytoskeletal element. Microfilaments are also involved in intracellular motile processes, such as the movement of vesicles, phagocytosis, and cytokinesis. In fact, plant cells rely primarily on microfilaments, rather than microtubules, for the long-distance transport of cytoplasmic vesicles and organelles. This bias toward microfilament-based motility reflects the rather restricted distribution of microtubules in many plant cells (see Figure 9.12).

Microfilaments are approximately 8 nm in diameter and composed of globular subunits of the protein **actin**. In the presence of ATP, actin monomers polymerize to form a flexible, helical filament. As a result of its subunit organization (Figure 9.44*a*), an actin filament is essentially a two-stranded structure with two helical grooves running along its length (Figure 9.44*b*). The terms *actin filament*, *F-actin*, and *microfilament* are basically synonyms for this type of filament. Because each actin subunit has polarity and all the subunits of an actin filament are pointed in the same direction, the entire microfilament has polarity. Consequently, the two ends of an actin filament have different structures and dynamic proper-

(a) *(b)* 100 nm
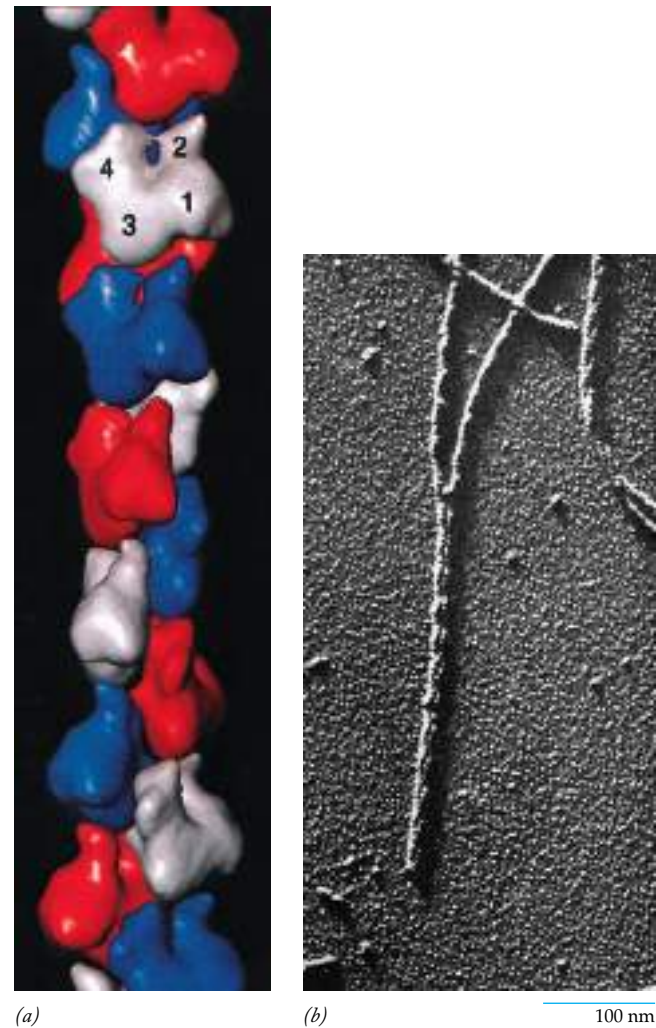
FIGURE 9.44 **Actin filament structure.** (*a*) A model of an actin filament. The actin subunits are represented in three colors to distinguish the consecutive subunits more easily. The subdomains in one of the actin subunits are labeled 1, 2, 3, and 4 and the ATP-binding cleft in each subunit is evident. Actin filaments have polarity, which is denoted as a plus and minus end. The cleft (in the upper red subunit) is present at the minus end of the filament. (*b*) Electron micrograph of a replica of an actin filament showing its double-helical architecture. (A: FROM MICHAEL F. SCHMID ET AL., COURTESY OF WAH CHIU, J. CELL BIOL. 124:346, 1994; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS; B: FROM ROBERT H. DEPUE, JR. AND ROBERT V. RICE, J. MOL. BIOL. 12:302, 1965; COPYRIGHT 1965, BY PERMISSION OF THE PUBLISHER, ACADEMIC PRESS AND ELSEVIER SCIENCE.)

ties. Depending on the type of cell and the activity in which it is engaged, actin filaments can be organized into highly ordered arrays, loose ill-defined networks, or tightly anchored bundles.

The identification of actin filaments in a given cell can be made with certainty using a cytochemical test that takes advantage of the fact that actin filaments, regardless of their source, will interact in a highly specific manner with the protein myosin. To facilitate the interaction, purified myosin (obtained from muscle tissue) is cleaved into fragments by a proteolytic enzyme. One of these fragments, called S1 (see

Figure 9.49), binds to the actin molecules all along the microfilament. In addition to identifying the filaments as actin, the bound S1 fragments reveal the filament's polarity. When S1 fragments are bound, one end of the microfilament appears *pointed* like an arrowhead, while the other end looks *barbed*. An example of this arrowhead "decoration" is shown in the microvilli of intestinal epithelial cells of Figure 9.45. The orientation of the arrowheads formed by the S1–actin complex provides information as to the direction in which the microfilaments are likely to be moved by a myosin motor protein. Actin can also be localized by light microscopy using fluorescently labeled phalloidin (Figure 9.74a), which binds to actin filaments, or by fluorescently labeled anti-actin antibodies.

Actin was identified more than 50 years ago as one of the major contractile proteins of muscle cells. Since then, it has been identified as a major protein in virtually every type of eukaryotic cell that has been examined. Higher plant and animal species possess a number of actin-coding genes whose encoded products are specialized for different types of motile processes. Actins have been remarkably conserved during the evolution of eukaryotes. For example, the amino acid sequences of actin molecules from a yeast cell and from rabbit skeletal muscle are 88 percent identical. In fact, actin molecules from diverse sources can copolymerize to form hybrid filaments.

Microfilament Assembly and Disassembly

Before it is incorporated into a filament, an actin monomer binds a molecule of ATP. Actin is an ATPase, just as tubulin is a GTPase, and the role of ATP in actin assembly is similar to that of GTP in microtubule assembly (page 336). The ATP associated with the actin monomer is hydrolyzed to ADP at some time after it is incorporated into the growing actin filament. As a consequence, the bulk of an actin filament consists of ADP-actin subunits.

Actin polymerization is readily demonstrated in vitro in solutions containing ATP-actin monomers. As in the case of microtubules, the initial stage in filament formation (i.e., *nucleation*) occurs slowly in vitro, whereas the subsequent stage of filament *elongation* occurs much more rapidly. The nucleation stage of filament formation can be bypassed by including preformed actin filaments in the reaction mixture. When preformed actin filaments are incubated with a high concentration of labeled ATP-actin monomers, both ends of the microfilament become labeled, but one end has a higher affinity for monomers and incorporates them at a rate approximately 10 times that of the other end. Decoration with the S1 myosin fragment reveals that the barbed (or plus) end of the microfilament is the fast-growing end, while the pointed (or minus) end is the slow-growing tip (Figure 9.46a).

Figure 9.46b illustrates how the events that occur during actin assembly/disassembly in vitro depend on the concentration of actin monomers. Suppose we were to begin by adding preformed actin filaments (seeds) to a solution of actin in the presence of ATP (step 1). As long as the concentration of ATP-actin monomers remains high, subunits will continue to be added at both ends of the filament (step 2, Figure 9.46b). As the monomers in the reaction mixture are consumed by addition to the ends of the filaments, the concentration of free ATP-actin continues to drop until a point is reached where net addition of monomers continues at the plus end, which has a higher affinity for ATP-actin, but stops at the minus end, which has a lower affinity for ATP-actin (step 3). As filament elongation continues, the free monomer concentration drops further. At this point, monomers continue to be added to the plus ends of the filaments, but a net loss of subunits occurs at their minus end. As the free monomer concentration falls, a point is reached where the two reactions at opposite ends of the filaments are balanced so that both the lengths of the filaments and the concentration of free monomers remain constant (step 4). This type of balance between two opposing activities is an example of *steady state* (page 92) and occurs when the ATP-actin concentration is approximately 0.3 μM.

200 nm

FIGURE 9.45   Determining the location and polarity of actin filaments using the S1 subunit of myosin. Electron micrograph of a replica showing the microfilament bundles in the core of the microvilli of an intestinal epithelial cell. The cell had been fixed, treated with S1 myosin fragments, freeze fractured, and deep etched to expose the filamentous components of the cytoplasm. The intermediate filaments (IFs) at the bottom of the micrograph do not contain actin and therefore do not bind the S1 myosin fragments. These intermediate filaments originate at the desmosomes of the lateral surfaces of the cell. (FROM N. HIROKAWA, L. G. TILNEY, K. FUJIWARA, AND J. E. HEUSER, J. CELL BIOL. 94:430, 1982; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

(a)   0.2 μm

(b)

FIGURE 9.46 Actin assembly in vitro. (*a*) Electron micrograph of a short actin filament that was labeled with S1 myosin and then used to nucleate actin polymerization. The addition of actin subunits occurs much more rapidly at the barbed (plus) end than at the pointed (minus) end of the existing filament. (*b*) Schematic diagram of the kinetics of actin-filament assembly in vitro. All of the orange subunits are part of the original seed; red subunits were present in solution at the beginning of the incubation. The steps are described in the text. Once a steady-state concentration of monomers is reached, subunits are added to the plus end at the same rate they are released from the minus end. As a result, subunits treadmill through the filament in vitro. *Note:* No attempt is made to distinguish between subunits with a bound ATP versus ADP. (A: COURTESY OF M. S. RUNGE AND T. D. POLLARD.)
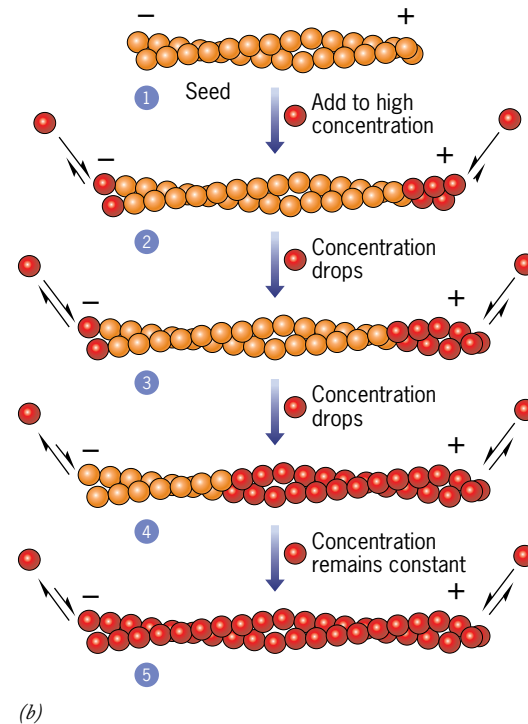
Because subunits are being added to the plus ends and removed from the minus ends of each filament at steady state, the relative position of individual subunits within each filament is continually moving—a process known as "treadmilling" (steps 4–5). Studies on living cells containing fluorescently labeled actin subunits have supported the occurrence of treadmilling in vivo (see Figure 9.54*c*).

Cells maintain a dynamic equilibrium between the monomeric and polymeric forms of actin, just as they do with tubulin and microtubules. As discussed on page 365, the rate of assembly and disassembly of actin filaments in the cell can be influenced by a number of different "accessory" proteins. Changes in the local conditions in a particular part of the cell can push the equilibrium toward either the assembly or the disassembly of microfilaments. By controlling this equilibrium, the cell can reorganize its microfilament cytoskeleton. Such reorganization is required for dynamic processes such as cell locomotion, changes in cell shape, phagocytosis, and cytokinesis.

As noted above, actin filaments play a role in nearly all of a cell's motile processes. The involvement of these filaments is most readily demonstrated by treating the cells with one of the following drugs that disrupt dynamic microfilament-based activities: cytochalasin, obtained from a mold, which blocks the plus ends of actin filaments, allowing depolymerization at the minus end; phalloidin, obtained from a poisonous mushroom, which binds to intact actin filaments and prevents their turnover; and latrunculin, obtained from a sponge, which binds to free monomers and blocks their incorporation into the polymer. Microfilament-mediated processes rapidly grind to a halt when cells contain one of these compounds. The effect of cytochalasin D on the very fine motile processes (*filopodia*) of embryonic sea urchin cells is shown in Figure 9.47.

10 μm
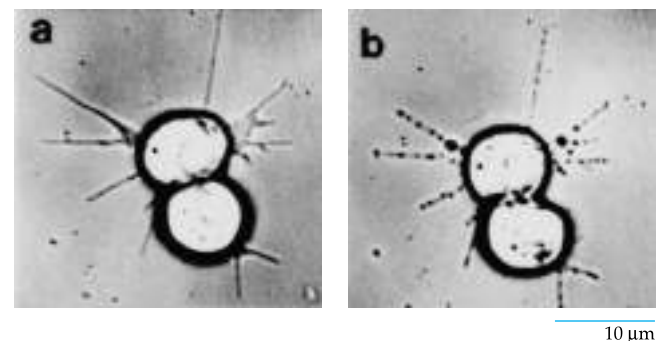
FIGURE 9.47 The effect of cytochalasin on structures containing actin filaments. A pair of sea urchin mesenchyme cells with fine filamentous processes (filopodia) after 30 seconds (*a*) and 5 minutes (*b*) of exposure to cytochalasin D. The drug causes dissolution of the cell processes, which are composed primarily of actin filaments. (FROM GERALD KARP AND MICHAEL SOLURSH, DEV. BIOL. 112:281, 1985.)

Myosin: The Molecular Motor of Actin Filaments

We have previously examined the structure and actions of two molecular motors—kinesin and dynein—that operate in opposite directions over tracks of microtubules. To date, all of the motors known to operate in conjunction with actin filaments are members of the myosin superfamily. Myosins—with the major exception of myosin VI, which is discussed below—move toward the plus end of an actin filament.

Myosin was first isolated from mammalian skeletal muscle tissue and subsequently from a wide variety of eukaryotic cells, including protists, plants, nonmuscle cells of animals, and vertebrate cardiac and smooth muscle tissues. All myosins share a characteristic motor (head) domain. The head contains a site that binds an actin filament and a site that binds and hydrolyzes ATP to drive the myosin motor. Whereas the head domains of various myosins are similar, the tail domains are highly divergent. Myosins also contain a variety of low-molecular-weight (light) chains. Myosins are generally divided into two broad groups: the **conventional** (or **type II**) **myosins**, which were first identified in muscle tissue, and the **unconventional myosins**. The unconventional myosins are subdivided on the basis of amino acid sequence into at least 17 different classes (type I and types III–XVIII). Some of these classes are expressed widely among eukaryotes, whereas others are restricted. Myosin X, for example, is found only in vertebrates, and myosins VIII and XI are present only in plants. Humans contain about 40 different myosins from at least 11 classes, each presumed to have its own specialized function(s). Of the various myosins, type II molecules are best understood.

Conventional (Type II) Myosins   Proteins of the myosin II class are the primary motors for muscle contraction but are also found in a variety of nonmuscle cells. The human genome encodes 16 different myosin II heavy chains, 3 of which function in nonmuscle cells. Among their nonmuscle activities, type II myosins are required for splitting a cell in two during cell division, generating tension at focal adhesions, cell migration, and the turning behavior of growth cones (see Figure 9.76). The effect of inhibiting myosin II activity in an advancing growth cone is demonstrated in Figure 9.48.

An electron micrograph of a pair of myosin II molecules is shown in Figure 9.49a. Each myosin II molecule is composed of six polypeptide chains—one pair of heavy chains and two pairs of light chains—organized in such a way as to produce a highly asymmetric protein (Figure 9.49a). Examination of the molecule in Figure 9.49b shows it to consist of (1) a pair of globular heads that contain the catalytic site of the molecule; (2) a pair of necks, each consisting of a single, uninterrupted α helix and two associated light chains; and (3) a single, long, rod-shaped tail formed by the intertwining of long α-helical sections of the two heavy chains.

Isolated myosin heads (S1 fragments of Figure 9.49b) that have been immobilized on the surface of a glass coverslip are capable of sliding attached actin filaments in an in vitro assay such as that shown in Figure 9.50. Thus, all of the machin-
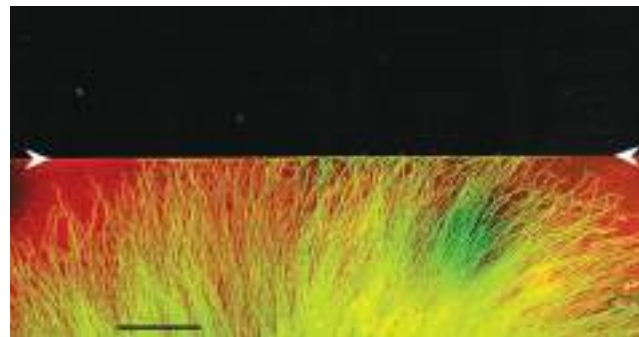
(a)
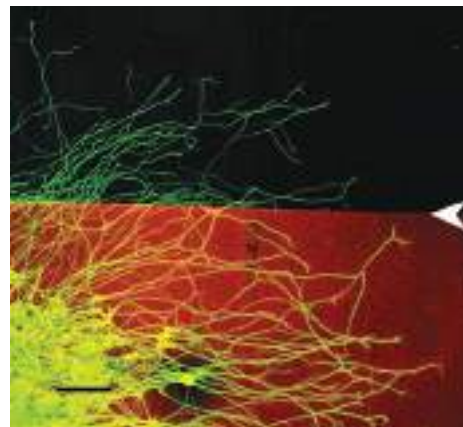
(b)

FIGURE 9.48 Experimental demonstration of a role for myosin II in the directional movement of growth cones. (*a*) Fluorescence micrograph showing fine processes (neurites) growing out from a microscopic fragment of mouse embryonic nervous tissue. The neurites (stained green) are growing outward on a glass coverslip coated with strips of laminin (stained red). Laminin is a common component of the extracellular matrix (page 237). The tip of each nerve process contains a motile growth cone. When the growth cones reach the border of the laminin strip (indicated by the line with arrowheads), they turn sharply and continue to grow over the laminin-coated surface. Bar, 500 μm. (*b*) The tissue in this micrograph was obtained from a mouse embryo lacking myosin IIB. The growth cones no longer turn when they reach the edge of the laminin-coated surface, causing the neurites to grow forward onto a surface (black) lacking laminin. Bar, 80 μm. (REPRINTED WITH PERMISSION FROM STEPHEN G. TURNEY AND PAUL C. BRIDGMAN, NATURE NEUROSCI. 8:717, 2005; © COPYRIGHT 2005, MACMILLAN MAGAZINES LTD.)

ery required for motor activity is contained in a single head. The mechanism of action of the myosin head, and the key role played by the neck, are discussed on page 362. The fibrous tail portion of a myosin II molecule plays a structural role, allowing the protein to form filaments. Myosin II molecules assemble so that the ends of the tails point toward the center of the filament and the globular heads point away from the center (Figures 9.51 and 9.57). As a result, the filament is described as *bipolar*, indicating a reversal of polarity at the filament's center. Because they are bipolar, the myosin heads at the opposite ends of a myosin filament have the ability to pull actin

(a)   20 nm   *(b)*
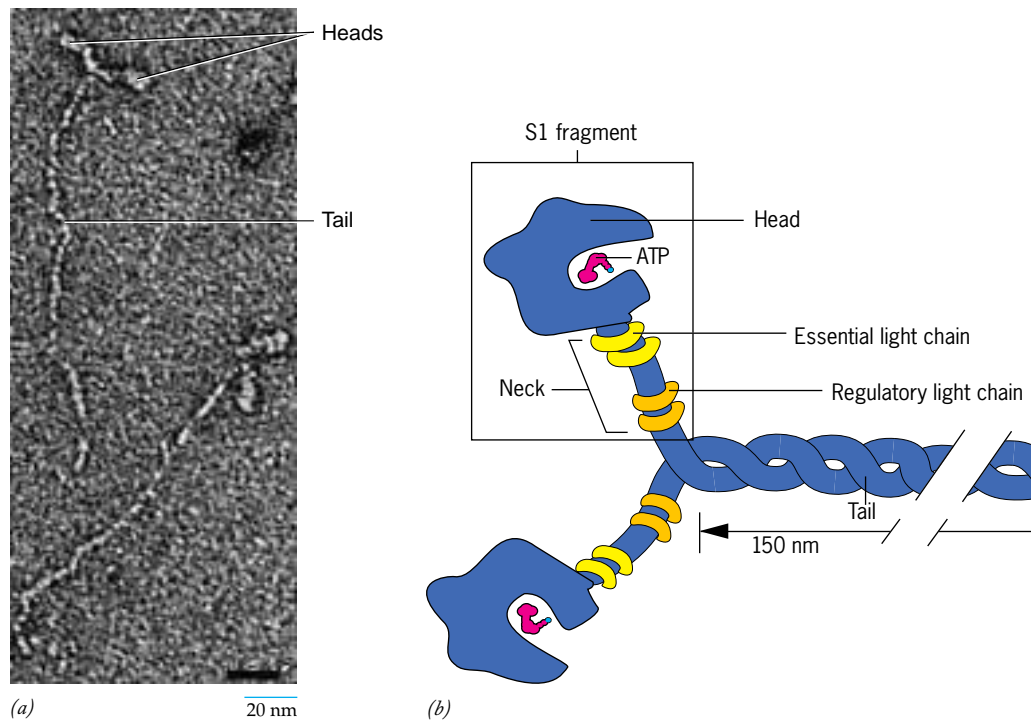
FIGURE 9.49 Structure of a myosin II molecule. (*a*) Electron micrograph of negatively stained myosin molecules. The two heads and tail of each molecule are clearly visible. (*b*) A highly schematic drawing of a myosin II molecule (molecular mass of 520,000 daltons). The molecule consists of one pair of heavy chains (blue) and two pairs of light chains, which are named as indicated. The paired heavy chains consist of a rod-shaped tail in which portions of the two polypeptide chains wrap around one another to form a coiled coil and a pair of globular heads. When treated with a protease under mild conditions, the molecule is cleaved at the junction between the neck and tail, which generates the S1 fragment. (A: FROM S. A. BURGESS, M. L. WALKER, H. D. WHITE, AND J. TRINICK, J. CELL BIOL. 139:676, 1997; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

(a)
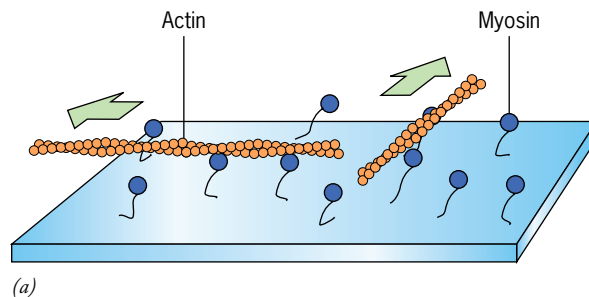
FIGURE 9.50 In vitro motility assay for myosin. (*a*) Schematic drawing in which myosin heads are bound to a silicone-coated coverslip, which is then incubated with a preparation of actin filaments. (*b*) Results of the experiment depicted in *a*. Two video images were taken 1.5 seconds apart and photographed as a double exposure on the same frame of film. The dashed lines with arrowheads show the sliding movement of the actin filaments over the myosin heads during the brief period between exposures. (FROM T. YANAGIDA, ADV. BIOPHYSICS 26:82, 1990.)

filaments toward one another, as occurs in a muscle cell. As described in the following section, the myosin II filaments that assemble in skeletal muscle cells are highly stable components of the contractile apparatus. However, the smaller

(b)
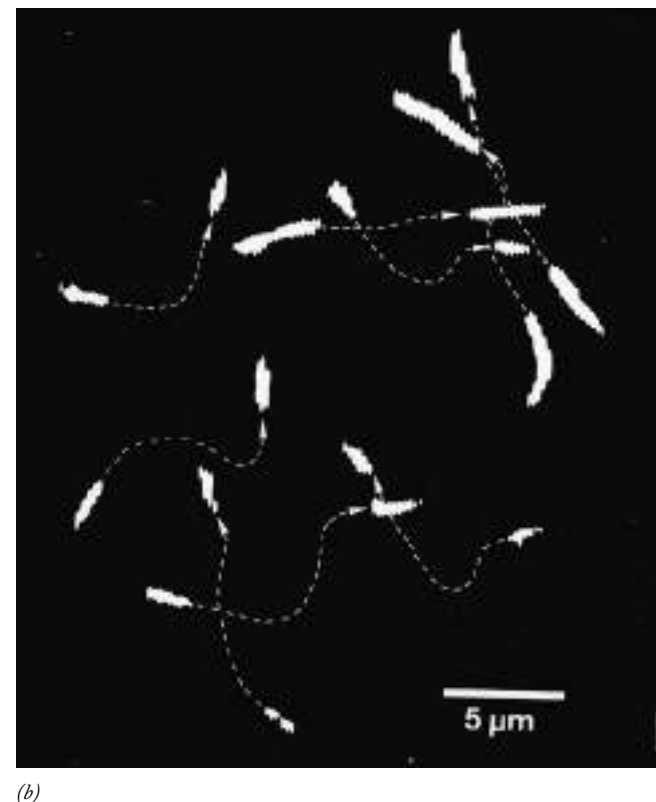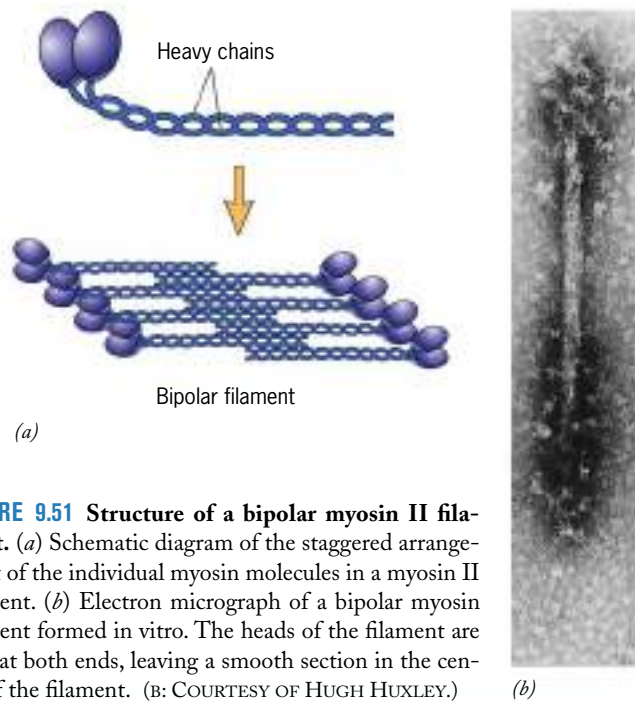
FIGURE 9.51 Structure of a bipolar myosin II filament. (*a*) Schematic diagram of the staggered arrangement of the individual myosin molecules in a myosin II filament. (*b*) Electron micrograph of a bipolar myosin filament formed in vitro. The heads of the filament are seen at both ends, leaving a smooth section in the center of the filament. (B: COURTESY OF HUGH HUXLEY.)

myosin II filaments that form in most nonmuscle cells often display transient construction, assembling when and where they are needed and then disassembling after they have acted.

Unconventional Myosins In 1973, Thomas Pollard and Edward Korn of the National Institutes of Health described a unique myosin-like protein that was extracted from the protist *Acanthamoeba*. Unlike muscle myosin, this smaller unconventional myosin had only a single head and was unable to assemble into filaments in vitro. The protein became known as myosin I, and its localization in microvilli is shown in Figure 9.66. Despite considerable study, the precise role of myosin I in cellular activities remains unclear.

The steps taken by another type of unconventional myosin—myosin V—are shown in a series of electron micrographs that capture the molecule at various stages in its mechanical cycle (Figure 9.52*a*). Myosin V is a dimeric, two-headed myosin that moves processively along actin filaments. Processivity is due to the high affinity of the myosin heads for the actin filament, which ensures that each head remains attached to the filament until the second head reattaches (as shown in the lower photo of Figure 9.52). Myosin V is also noteworthy for the length of its neck, which at 23 nm is about three times that of myosin II. Because of its long neck, myosin V can take very large steps. This is very important for a motor protein that moves processively along an actin filament made up of *helical* strands of subunits. The actin helix repeats itself about every 13 subunits (36 nm), which is just about the step size of a myosin V molecule (Figure 9.52*b*). These and subse-

FIGURE 9.52 Myosin V—a two-headed unconventional myosin involved in organelle transport. (*a*) This series of negatively stained electron micrographs shows individual myosin V molecules that were caught at different stages in their mechanical cycle as each of them was "walking" along an actin filament. The circular arrow shows the sequence as a molecule might move from left to right along the actin filament toward its barbed (plus) end. The micrograph at the bottom center shows a stage in the cycle where both heads are attached to the actin filament

with a spacing of approximately 36 nm, which is equivalent to the helical repeat of the filament. (*b*) Schematic drawing of a dimeric myosin V molecule in the stage equivalent to the bottom center micrograph of part *a*. Rab 27a and melanophilin serve as adaptors that link the globular tail of the motor to the vesicle membrane. (A: FROM MATTHEW L. WALKER ET AL., COURTESY OF PETER J. KNIGHT. REPRINTED WITH PERMISSION FROM NATURE 405:804, 2000; © COPYRIGHT 2000; MACMILLAN MAGAZINES LTD.)
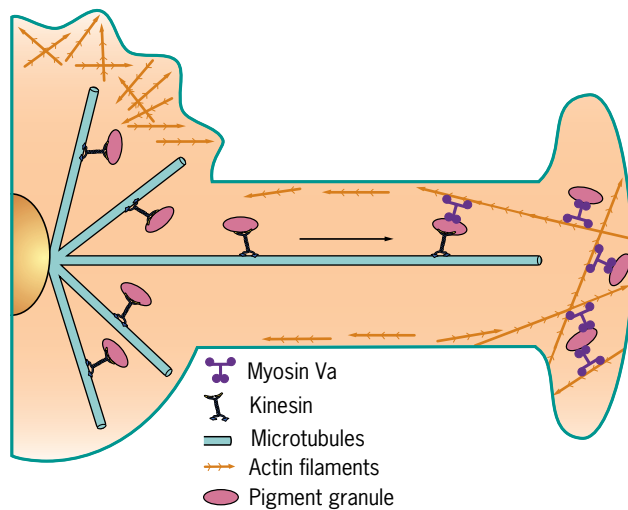
FIGURE 9.53 The contrasting roles of microtubule- and microfilament-based motors in intracellular transport. Most vesicle transport is thought to be mediated by members of the kinesin and dynein families, which carry their cargo over relatively long distances. It is thought that some vesicles also carry myosin motors, such as myosin Va, which transport their cargo over microfilaments, including those present in the peripheral (cortical) regions of the cell. The two types of motors may act in a cooperative manner, as illustrated here in the case of a pigment cell in which pigment granules move back and forth within extended cellular processes. (AFTER X. WU ET AL., J. CELL BIOL. 143:1915, 1998; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

quent studies on the movements of single myosin V molecules suggest that this motor walks along an actin filament by a "hand-over-hand" model, similar in some ways to that depicted in Figure 9.15*b* for kinesin. To accomplish this type of movement, each myosin head must move a distance of 72 nm, twice the distance between two successive binding sites on the actin filament (Figure 9.52*b*). As a result of its giant strides, myosin V can walk along the filament in a straight path even though the underlying "roadway" spirals 360° between its "feet."

Several unconventional myosins (including myosin I, V, and VI) are associated with various types of cytoplasmic vesicles and organelles. Some vesicles have been shown to contain microtubule-based motors (kinesins and/or cytoplasmic dynein) *and* microfilament-based motors (unconventional myosins) and, in fact, the two types of motors may be physically linked to one another. The movement of vesicles and other membranous carriers over long distances within animal cells occurs on microtubules, as previously described. However, once they approach the end of the microtubule, these membranous vesicles are often thought to switch over to microfilament tracks for the local movement through the actin-rich periphery of the cell (Figure 9.53).

Cooperation between microtubules and microfilaments has been best studied in pigment cells (Figure 9.53). In mammals, pigment granules (melanosomes) are transported into fine peripheral processes of the pigment cell by one of the

myosin V isoforms called Va. Melanosomes are then transferred to hair follicles where they become incorporated into a developing hair. Mice lacking myosin Va activity are unable to transfer melanosomes into hair follicles, causing their coat to have a much lighter color. Humans lacking a normal gene encoding myosin Va suffer from a rare disorder called Griscelli syndrome; these individuals exhibit partial albinism (lack of skin coloration) and suffer other symptoms thought to be related to defects in vesicle transport. In 2000 it was discovered that a subset of Griscelli patients had a normal myosin Va gene, but lacked a functional gene encoding a peripheral membrane protein called Rab27a. The Rab family of proteins was discussed on page 295 as molecules that tether vesicles to target membranes. Rabs are also involved in the attachment of myosin (and kinesin) motors to membrane surfaces (Figure 9.52*b*).

Hair cells, whose structure is shown in Figure 9.54*a* have been a particularly good system for studying the functions of unconventional myosins. Hair cells are named for the bundle of stiff, hairlike stereocilia that project from the apical surface of the cell into the fluid-filled cavity of the inner ear. Displacement of the stereocilia by mechanical stimuli leads to the generation of nerve impulses that we perceive as sound. Stereocilia have no relation to the true cilia discussed earlier. Instead of containing microtubules, each stereocilium is supported by a bundle of parallel actin filaments (Figure 9.54*b*) whose barbed ends are located at the outer tip of the projection and pointed ends at the base. Stereocilia have provided some of the most striking images of the dynamic nature of the actin cytoskeleton. While the stereocilia themselves are permanent structures, the actin bundles are in constant flux. Actin monomers continually bind to the barbed end of each filament, treadmill through the body of the filament, and dissociate from the pointed end. This process is captured in the fluorescence micrograph of Figure 9.54*c*, which shows the incorporation of GFP-labeled actin subunits at the barbed end of each filament. Several unconventional myosins (I, V, VI, VII, and XV) are localized at various sites within the hair cells of the inner ear; two of these are shown in Figure 9.54*d* and *e*. The precise roles of these various motor proteins is unclear. Mutations in myosin VIIa are the cause of Usher 1B syndrome, which is characterized by both deafness and blindness. The morphologic effects of mutations in the myosin VIIa gene on the hair cells of the inner ear of mice are shown in Figure 9.54*f,g*. As in humans, mice that are homozygous for the mutant allele encoding this motor protein are deaf. Mysoin VI, a processive organelle transporter in the cytoplasm of many cells, is distinguished by its movement in a "reverse direction," that is, toward the pointed (minus) end of an actin filament. Myosin VI is thought to be involved in the formation of clathrin-coated vesicles at the plasma membrane, the movement of the uncoated vesicles to the early endosomes, and the anchoring of membranes to actin filaments. Mutations in myosin VI are the cause of several inherited diseases.

Now that we have described the basics of actin and myosin structure and function, we can see how these two proteins interact to mediate a complex cellular activity.
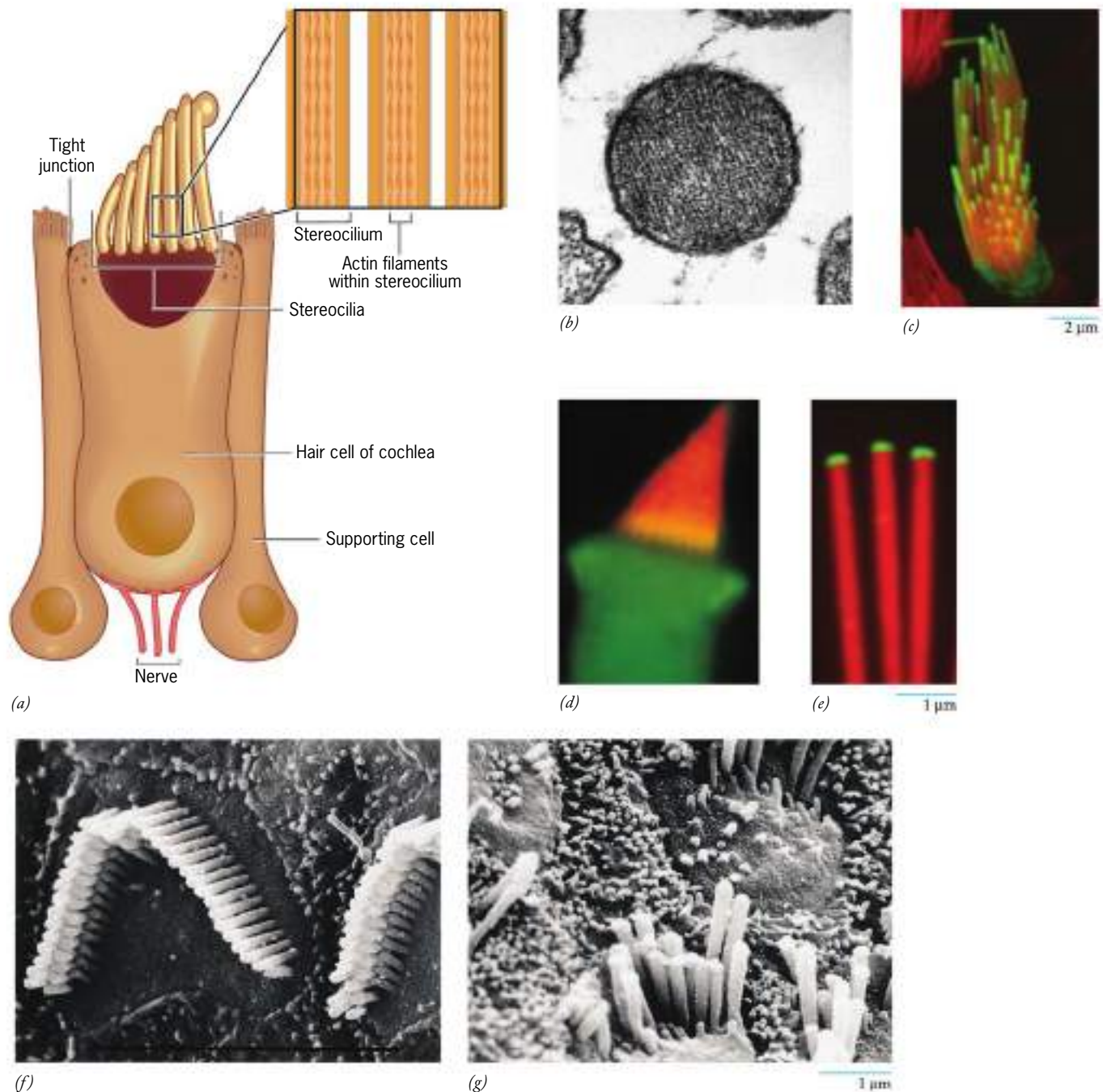
FIGURE 9.54 **Hair cells, actin bundles, and unconventional myosins.** (*a*) Drawing of a hair cell of the cochlea. The inset shows a portion of several of the stereocilia, which are composed of a tightly grouped bundle of actin filaments. (*b*) Transmission electron micrograph of a cross section of a stereocilium showing it is composed of a dense bundle of actin filaments. (*c*) Fluorescence micrograph of a hair cell from the vestibule of a rat inner ear. The tips of the stereocilia are labeled green due to the incorporation of GFP-actin monomers at their barbed ends. Taller stereocilia contain a longer column of GFP-labeled subunits, which reflects a more rapid incorporation of actin monomers. The stereocilia appear red due to labeling by rhodamine-labeled phalloidin, which binds to actin filaments. (*d*) A hair cell from the bullfrog inner ear. The localization of myosin VIIa is indicated in green. The orange bands near the bases of the stereocilia (due to red and green overlap) indicate a con-

centration of myosin VIIa. (*e*) Myosin XVa (green) is localized at the tips of the stereocilia of a rat auditory hair cell. (*f*) Scanning electron micrograph of the hair cells of the cochlea of a control mouse. The stereocilia are arranged in V-shaped rows. (*g*) A corresponding micrograph of the hair cells of a mouse with mutations in the gene encoding myosin VIIa, which causes deafness. The stereocilia of the hair cells exhibit a disorganized arrangement. (A: AFTER T. HASSON CURR. BIOL. 9:R839, 1999; WITH PERMISSION FROM ELSEVIER SCIENCE; B: COURTESY OF A. J. HUDSPETH, R. A. JACOBS, AND P. G. GILLESPIE; C,E: FROM A. K. RZADZINSKA, ET AL., COURTESY OF B. KACHAR, J. CELL BIOL. VOL. 164, 891, 892, 2004; D: FROM PETER GILLESPIE AND TAMA HASSON, J. CELL BIOL. VOL. 137, COVER #6, 1997; C–E, BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS; F,G: FROM T. SELF, ET AL., COURTESY OF K. P. STEEL, DEVELOP. 125:560, 1998; BY PERMISSION OF THE COMPANY OF BIOLOGISTS LTD.)

1. Compare and contrast the characteristics of microtubule assembly versus actin filament assembly.

2. Compare the structure of a fully assembled microtubule, actin filament, and intermediate filament.

3. Describe three functions of actin filaments.

4. Contrast the structure and function of conventional myosin II and the unconventional myosin V.

5. How is it possible for the same vesicle to be transported along both microtubules and microfilaments?

9.6 MUSCLE CONTRACTILITY

Skeletal muscles derive their name from the fact that most of them are anchored to bones that they move. They are under voluntary control and can be consciously commanded to contract. Skeletal muscle cells have a highly unorthodox structure. A single, cylindrically shaped muscle cell is typically 10 to 100 μm thick, over 100 mm long, and contains hundreds of nuclei. Because of these properties, a skeletal muscle cell is more appropriately called a **muscle fiber**. Muscle fibers have multiple nuclei because each fiber is a product of the fusion of large numbers of mononucleated *myoblasts* (premuscle cells) in the embryo.

Skeletal muscle cells may have the most highly ordered internal structure of any cell in the body. A longitudinal section of a muscle fiber (Figure 9.55) reveals a cable made up of hundreds of thinner, cylindrical strands, called **myofibrils**. Each myofibril consists of a repeating linear array of contractile units, called **sarcomeres**. Each sarcomere in turn exhibits a characteristic banding pattern, which gives the muscle fiber a striped or *striated* appearance. Examination of stained muscle fibers in the electron microscope shows the banding pattern to be the result of the partial overlap of two distinct types of filaments, **thin filaments** and **thick filaments** (Figure 9.56*a*). Each sarcomere extends from one *Z line* to the next Z line and contains several dark bands and light zones. A sarcomere has a pair of lightly staining *I bands* located at its outer edges, a more densely staining *A band* located between the outer I bands, and a lightly staining *H zone* located in the center of the A band. A densely staining *M line* lies in the center of the H zone. The I band contains only thin filaments, the H zone only thick filaments, and that part of the A band on

(a)

(b)

(c)
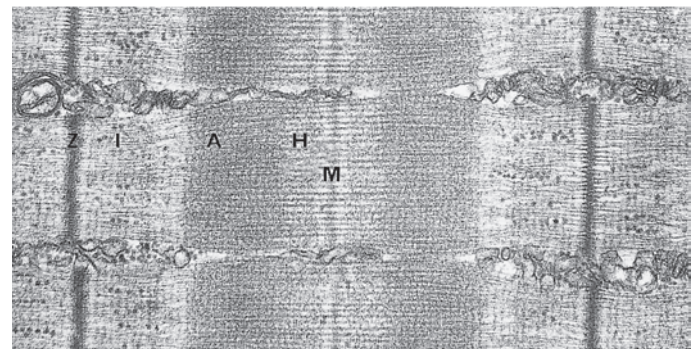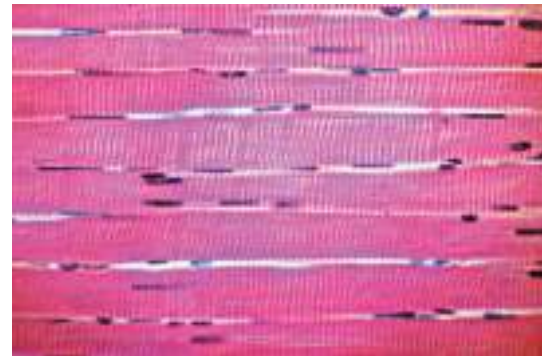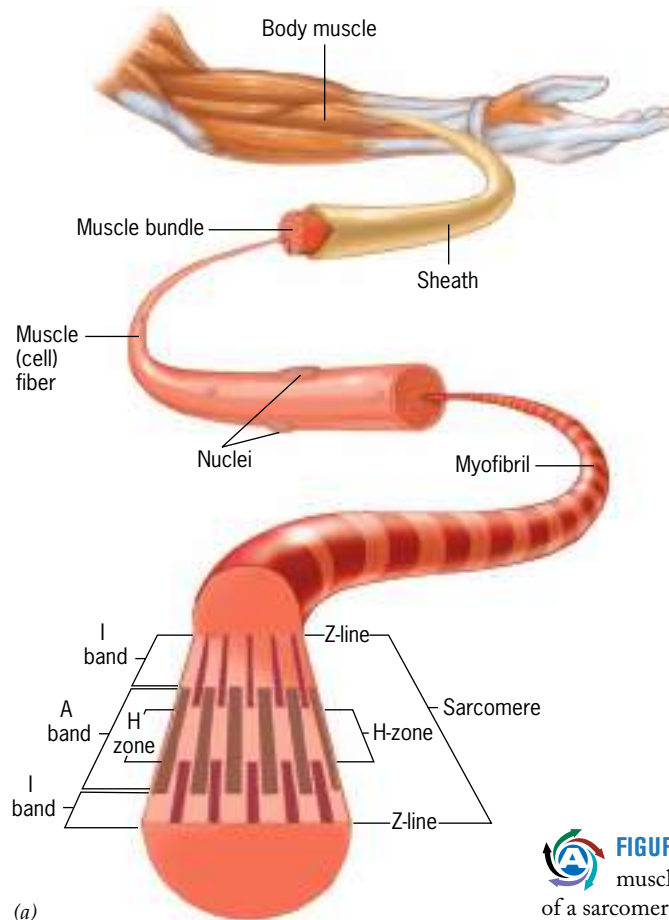
FIGURE 9.55 **The structure of skeletal muscle.** (*a*) Levels of organization of a skeletal muscle. (*b*) Light micrograph of a multinucleated muscle fiber. (*c*) Electron micrograph of a sarcomere with the bands lettered. (B: ERIC GRAVE/PHOTO RESEARCHERS, INC.; C: COURTESY OF GERALDINE F. GAUTHIER.)
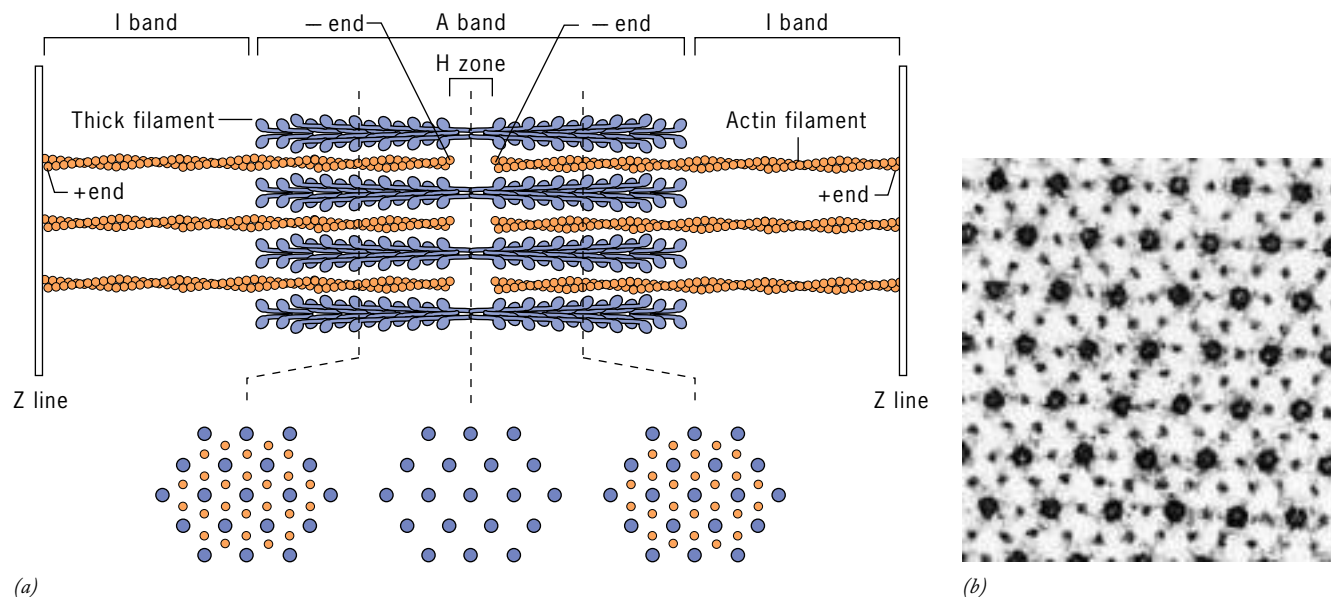
FIGURE 9.56 The contractile machinery of a sarcomere. (*a*) Diagram of a sarcomere showing the overlapping array of thin actin-containing (orange) and thick myosin-containing (purple) filaments. The small transverse projections on the myosin fiber represent the myosin heads (cross-bridges). (*b*) Electron micrograph of a cross section through an insect flight muscle showing the hexagonal arrangement of the thin filaments around each thick filament. (J. AUBER/PHOTO RESEARCHERS.)

either side of the H zone represents the region of overlap and contains both types of filaments.

Cross sections through the region of overlap show that the thin filaments are organized in a hexagonal array around each thick filament, and that each thin filament is situated between two thick filaments (Figure 9.56*b*). Longitudinal sections show the presence of projections from the thick filaments at regularly spaced intervals. The projections represent cross-bridges capable of forming attachments with neighboring thin filaments.

The Sliding Filament Model of Muscle Contraction

All skeletal muscles operate by shortening; there is no other way they can perform work. The units of shortening are the sarcomeres, whose combined decrease in length accounts for the decrease in length of the entire muscle. The most important clue to the mechanism underlying muscle contraction came from observations of the banding pattern of the sarcomeres at different stages in the contractile process. As a muscle fiber shortened, the A band in each sarcomere remained essentially constant in length, while the H band and I bands decreased in width and then disappeared altogether. As shortening progressed, the Z lines on both ends of the sarcomere moved inward until they contacted the outer edges of the A band (Figure 9.57).

Based on these observations, two groups of British investigators, Andrew Huxley and Rolf Niedergerke, and Hugh Huxley and Jean Hanson, proposed a far-reaching model in 1954 to account for muscle contraction. They proposed that the shortening of individual sarcomeres did not result from the shortening of the filaments, but rather from their sliding over one another. The sliding of the thin filaments toward the center of the sarcomere would result in the observed increase in overlap between the filaments and the decreased width of the I and H bands (Figure 9.57). The *sliding-filament model* was rapidly accepted, and evidence in its favor continues to accumulate.

The Composition and Organization of Thick and Thin Filaments   In addition to actin, the thin filaments of a skeletal muscle contain two other proteins, *tropomyosin* and *troponin* (Figure 9.58). Tropomyosin is an elongated molecule (approximately 40 nm long) that fits securely into the grooves within the thin filament. Each rod-shaped tropomyosin molecule is associated with seven actin subunits along the thin filament (Figure 9.58). Troponin is a globular protein complex composed of three subunits, each having an important and distinct role in the overall function of the molecule. Troponin molecules are spaced approximately 40 nm apart along the thin filament and contact both the actin and tropomyosin components of the filament. The actin filaments of each half sarcomere are aligned with their barbed ends linked to the Z line.

Each thick filament is composed of several hundred myosin II molecules together with small amounts of other proteins. Like the filaments that form in vitro (see Figure 9.51), the polarity of the thick filaments of muscle cells is reversed at the center of the sarcomere. The center of the filament is composed of the opposing tail regions of the myosin molecules and is devoid of heads. The myosin heads project from each

(a)

(b)
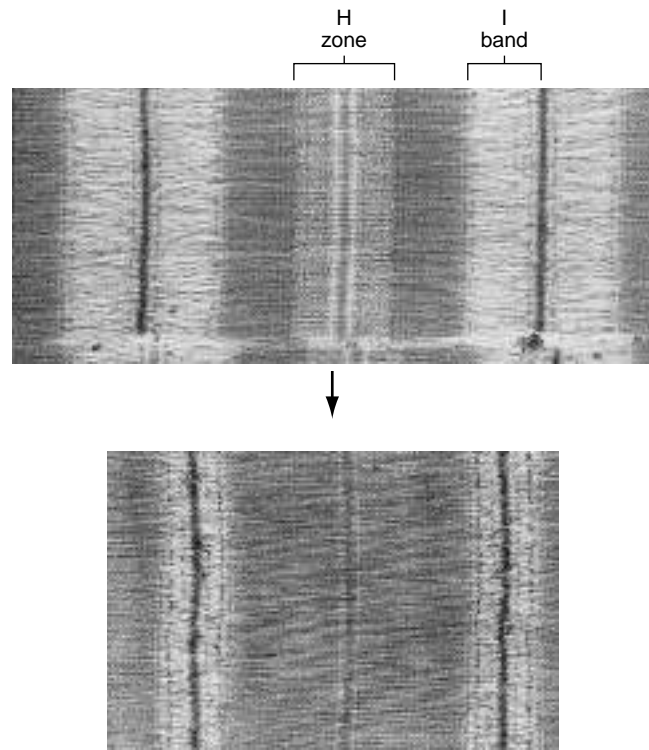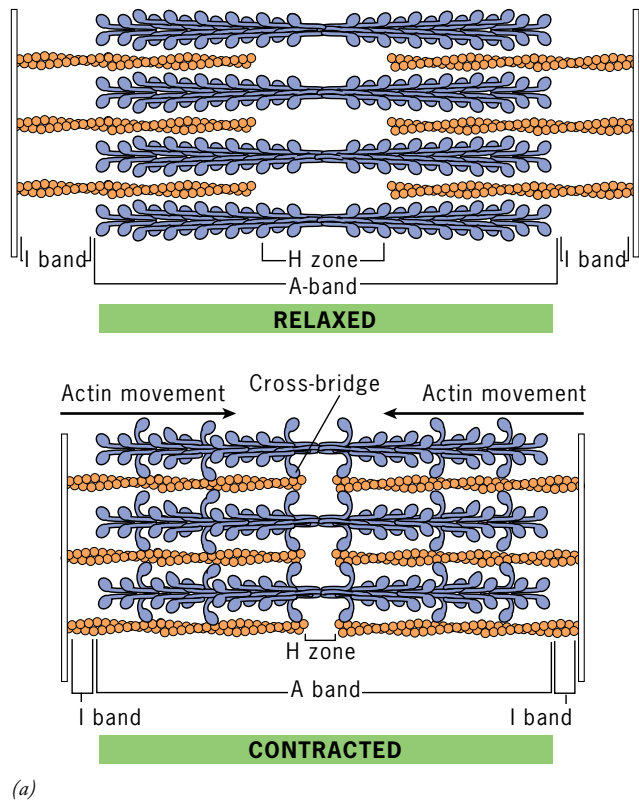
FIGURE 9.57 The shortening of the sarcomere during muscle contraction. (*a*) Schematic diagram showing the difference in the structure of the sarcomere in a relaxed and contracted muscle. During contraction, the myosin cross-bridges make contact with the surrounding thin filaments, and the thin filaments are forced to slide toward the center of the sarcomere. Cross-bridges work asynchronously, so that only a fraction are active at any given instant. (*b*) Electron micrographs of longitudinal sections through a relaxed (*top*) and contracted (*bottom*) sarcomere. The micrographs show the disappearance of the H zone as the result of the sliding of the thin filaments toward the center of the sarcomere. (B: FROM JAMES E. DENNIS/PHOTOTAKE.)

thick filament along the remainder of its length due to the staggered position of the myosin molecules that make up the body of the filament (Figure 9.51).

The third most abundant protein of vertebrate skeletal muscles is *titin*, which is the largest protein yet to be discovered in any organism. The entire titin gene (which can give rise to isoforms of different length) encodes a polypeptide more than 3.5 million daltons in molecular mass and containing more than 38,000 amino acids. Titin molecules originate
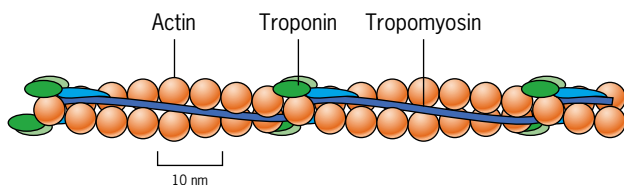
FIGURE 9.58 The molecular organization of the thin filaments. Each thin filament consists of a helical array of actin subunits with rod-shaped tropomyosin molecules situated in the grooves and troponin molecules spaced at defined intervals, as described in the text. The positional changes in these proteins that trigger contraction are shown in Figure 9.63.

at the M line in the center of each sarcomere and extend along the myosin filament, continuing past the A band and terminating at the Z line (Figure 9.59). Titin is a highly elastic protein that stretches like a molecular spring as certain regions within the molecule become uncoiled. Titin is thought to prevent the sarcomere from becoming pulled apart during muscle stretching. Titin also maintains myosin filaments in their proper position within the center of the sarcomere during muscle contraction.

The Molecular Basis of Contraction Following the formulation of the sliding-filament hypothesis, attention turned to the heads of the myosin molecules as the force-generating components of the muscle fiber. During a contraction, each myosin head extends outward and binds tightly to a thin filament, forming the cross-bridges seen between the two types of filaments (Figure 9.57). The heads from a single myosin filament interact with six surrounding actin filaments. While it is bound tightly to the actin filament, the myosin head undergoes a conformational change (described below) that moves the thin actin filament approximately 10 nm toward the center of the sarcomere. Unlike myosin V (depicted in Figure 9.52), muscle myosin (i.e., myosin II) is a *nonprocessive motor*. Muscle
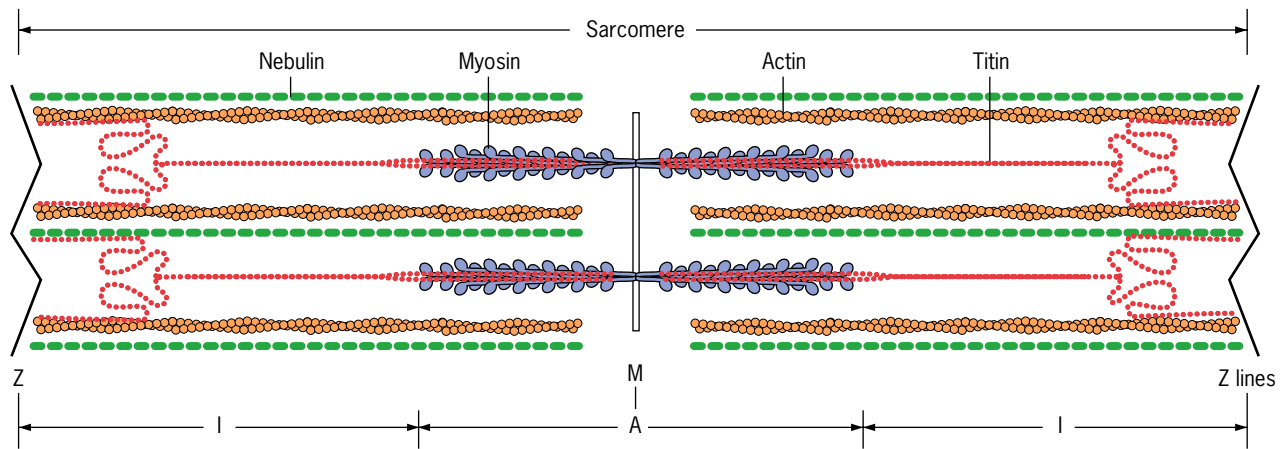
FIGURE 9.59 The arrangement of titin molecules within the sarcomere. These huge elastic molecules stretch from the end of the sarcomere at the Z line to the M band in the sarcomere center. Titin molecules are thought to maintain the thick filaments in the center of the sarcomere during contraction. The I-band portion of the titin molecule contains spring-like domains and is capable of great elasticity. The nebulin molecules (which are not discussed in the text) are thought to act like a "molecular ruler" by regulating the number of actin monomers that are allowed to assemble into a thin filament. (AFTER T. C. S. KELLER, CURR. OPIN. CELL BIOL. 7:33, 1995.)

myosin remains in contact with its track, in this case a thin filament, for only a small fraction (less than 5 percent) of the overall cycle. However, each thin filament is contacted by a team of a hundred or so myosin heads that beat out of synchrony with one another (Figure 9.57a). Consequently, the thin filament undergoes continuous motion during each contractile cycle. It is estimated that a single thin filament in a muscle cell can be moved several hundred nanometers during a period as short as 50 milliseconds.

Muscle physiologists have long sought to understand how a single myosin molecule can move an actin filament approximately 10 nm in a single power stroke. The publication of the first atomic structure of the S1 myosin II fragment in 1993 by Ivan Rayment, Hazel Holden, and their colleagues at the University of Wisconsin focused attention on a proposal to explain its mechanism of action. In this proposal, the energy released by ATP hydrolysis induces a small (0.5 nm) conformational change within the head while the head is tightly bound to the actin filament. The small movement within the head is then amplified approximately 20-fold by the swinging movement of an adjoining α-helical neck (Figure 9.60). According to this hypothesis, the elongated myosin II neck acts as a rigid "lever
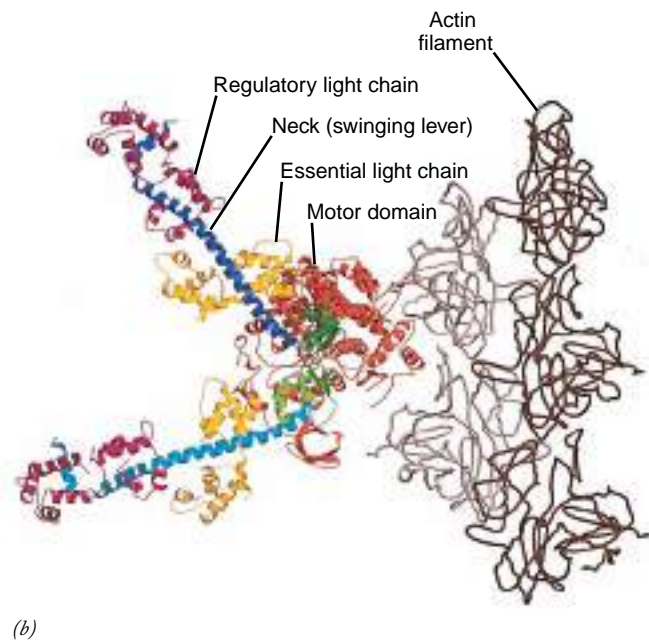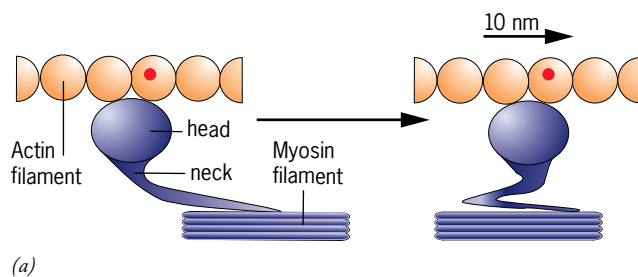
FIGURE 9.60 Model of the swinging lever arm of a myosin II molecule. (a) During the power stroke, the neck of the myosin molecule moves through a rotation of approximately 70°, which would produce a movement of the actin filament of approximately 10 nm. (b) A model of the power stroke of the myosin motor domain consisting of the motor domain (or head) and adjoining neck (or lever arm). An attached actin filament is shown in gray/brown. The long helical neck is shown in two positions, both before and after the power stroke (depicted as the upper dark blue and lower light blue necks, respectively). It is this displacement of the neck region that is thought to power muscle movement. The essential and regulatory light chains, which wrap around the neck, are shown in yellow and magenta, respectively. (B: FROM MALCOLM IRVING ET AL. REPRINTED WITH PERMISSION FROM NATURE STRUCT. BIOL. 7:482, 2000; © COPYRIGHT 2000, MACMILLAN MAGAZINES LTD.)

arm" causing the attached actin filament to slide a much greater distance than would otherwise be possible. The two light chains, which are wrapped around the neck, are thought to provide rigidity for the lever. Support for the "lever-arm" proposal was initially obtained through a series of experiments by James Spudich and colleagues at Stanford University. These researchers constructed genes that encoded altered versions of the myosin II molecule, which contained necks of different length. The genetically engineered myosin molecules were then tested in an in vitro motility assay with actin filaments, similar to that depicted in Figure 9.50. As predicted by the lever-arm hypothesis, the apparent length of the power stroke of the myosin molecules was directly proportional to the length of their necks. Myosin molecules with shorter necks generated smaller displacements, whereas those with longer necks generated greater displacements. Not all studies have supported the correlation between step size and neck length, so the role of the lever-arm remains a matter of controversy.

The Energetics of Filament Sliding   Like the other motor proteins kinesin and dynein, myosin converts the chemical energy of ATP into the mechanical energy of filament sliding. Each cycle of mechanical activity of the myosin cross-bridge takes about 50 msec and is accompanied by a cycle of ATPase activity, as illustrated in the model shown in Figure 9.61. According to this model, the cycle begins as a molecule of ATP binds to the myosin head, an event that induces the dissociation of the cross-bridge from the actin filament (Figure 9.61, step 1). Binding of ATP is followed by its hydrolysis, which occurs before the myosin head makes contact with the actin filament. The products of ATP hydrolysis, namely, ADP and P_i, remain bound to the active site of the enzyme, while the energy released by hydrolysis is absorbed by the protein as a whole (Figure 9.61, step 2). At this point, the cross-bridge is in an energized state, analogous to a stretched spring capable of spontaneous movement. The energized myosin then attaches to the actin molecule (step 3) and releases its bound phosphate, which triggers a large conformational change driven by the stored free energy (step 4). This conformational change shifts the actin filament toward the center of the sarcomere. This movement represents the power stroke of the myosin head as shown in Figure 9.60. The release of the bound ADP (step 5) is followed by the binding of a new ATP molecule so that a new cycle can begin. In the absence of ATP, the myosin head remains tightly bound to the actin filament. The inability of myosin cross-bridges to detach in the absence of ATP is the basis for the condition of *rigor mortis*, the stiffening of muscles that ensues following death.

Excitation–Contraction Coupling   Muscle fibers are organized into groups termed *motor units*. All the fibers of a motor unit are jointly innervated by branches from a single motor neuron and contract simultaneously when stimulated by an impulse transmitted along that neuron. The point of contact of a terminus of an axon with a muscle fiber is called a **neuromuscular junction** (Figure 9.62; see also Figure 4.55 for a closer view of the structure of the synapse). The neuromuscu-
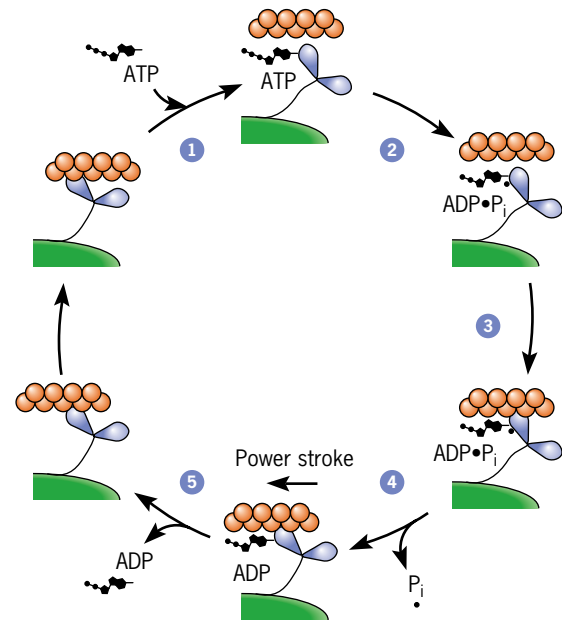
FIGURE 9.61 **A schematic model of the actinomyosin contractile cycle.** The movement of the thin filament by the force-generating myosin head occurs as the result of a linkage between a mechanical cycle involving the attachment, movement, and detachment of the head and a chemical cycle involving the binding, hydrolysis, and release of ATP, ADP, and P_i. In this model, the two cycles begin in step 1 with the binding of ATP to a cleft in the myosin head, causing the detachment of the head from the actin filament. The hydrolysis of the bound ATP (step 2) energizes the head, causing it to bind weakly to the actin filament (step 3). The release of P_i causes a tighter attachment of the myosin head to the thin filament and the power stroke (step 4) that moves the thin filament toward the center of the sarcomere. The release of the ADP (step 5) sets the stage for another cycle. (AFTER M. Y. JIANG AND M. P. SHEETZ, BIOESS. 16:532, 1994.)

lar junction is a site of transmission of the nerve impulse from the axon across a synaptic cleft to the muscle fiber, whose plasma membrane is also excitable and capable of conducting an action potential.

The steps that link the arrival of a nerve impulse at the muscle plasma membrane to the shortening of the sarcomeres deep within the muscle fiber constitute a process referred to as **excitation–contraction coupling**. Unlike a neuron, where an action potential remains at the cell surface, the impulse generated in a skeletal muscle cell is propagated into the interior of the cell along membranous folds called **transverse (T) tubules** (Figure 9.62). The T tubules terminate in very close proximity to a system of cytoplasmic membranes that make up the **sarcoplasmic reticulum (SR),** which forms a membranous sleeve around the myofibril. Approximately 80 percent of the integral protein of the SR membrane consists of Ca^{2+}-ATPase molecules whose function is to transport Ca^{2+} out of the cytosol and into the lumen of the SR, where it is stored until needed.

The importance of calcium in muscle contraction was first shown in 1882 by Sydney Ringer, an English physician.
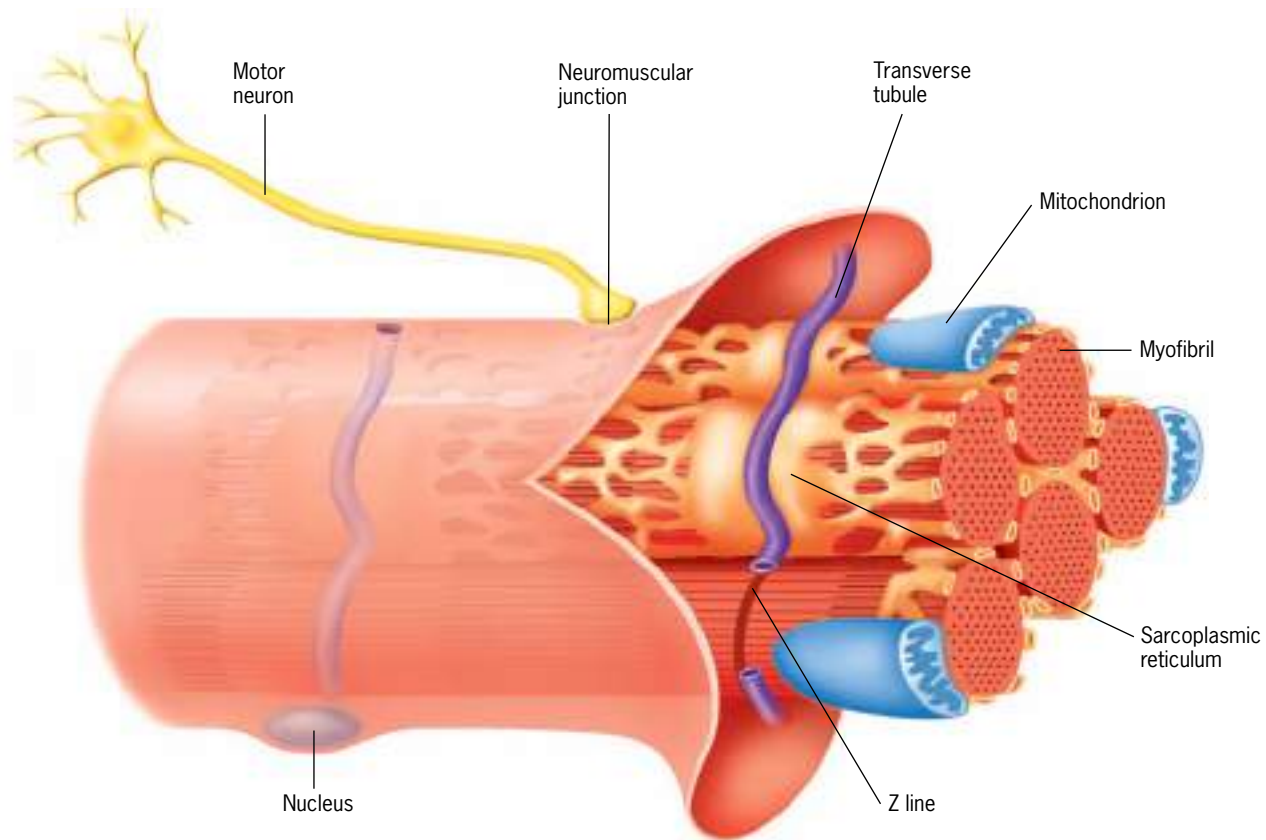
FIGURE 9.62 The functional anatomy of a muscle fiber. Calcium is housed in the elaborate network of internal membranes that make up the sarcoplasmic reticulum (SR). When an impulse arrives by means of a motor neuron, it is carried into the interior of the fiber along the membrane of the transverse tubule to the SR. The calcium gates of the SR open, releasing calcium into the cytosol. The binding of calcium ions to troponin molecules of the thin filaments leads to the events described in the following figure and the contraction of the fiber.

Ringer found that an isolated frog heart would contract in a saline solution made with London tap water, but failed to contract in a solution made with distilled water. Ringer determined that calcium ions present in tap water were an essential factor in muscle contraction. In the relaxed state, the Ca^{2+} levels within the cytoplasm of a muscle fiber are very low (approximately 2×10^{-7} M)—below the threshold concentration required for contraction. With the arrival of an action potential by way of the transverse tubules, calcium channels in the SR membrane are opened, and calcium diffuses out of the SR compartment and over the short distance to the myofibrils. As a result, the intracellular Ca^{2+} levels rise to about 5×10^{-5} M. To understand how elevated calcium levels trigger contraction in a skeletal muscle fiber, it is necessary to reconsider the protein makeup of the thin filaments.

When the sarcomere is relaxed, the tropomyosin molecules of the thin filaments (see Figure 9.58) block the myosin-binding sites on the actin molecules. The position of the tropomyosin within the groove is under the control of the attached troponin molecule. When Ca^{2+} levels rise, these ions bind to one of the subunits of troponin (troponin C), causing a conformational change in another subunit of the troponin molecule. Like the collapse of a row of dominoes, the move-

ment of troponin is transmitted to the adjacent tropomyosin, which moves approximately 1.5 nm closer to the center of the filament's groove (from position b to a in Figure 9.63). This
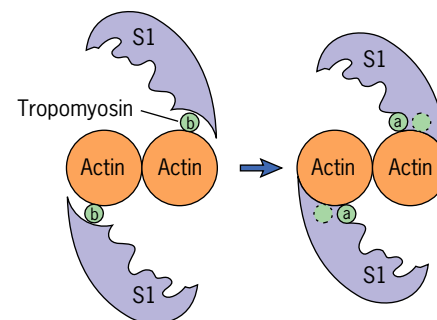
FIGURE 9.63 The role of tropomyosin in muscle contraction. Schematic diagram of the steric hindrance model in which the myosin-binding site on the thin actin filaments is controlled by the position of the tropomyosin molecule. When calcium levels rise, the interaction between calcium and troponin (not shown) leads to a movement of the tropomyosin from position b to position a, which exposes the myosin-binding site on the thin filament to the myosin head.

shift in position of the tropomyosin exposes the myosin-binding sites on the adjacent actin molecules, allowing the myosin heads to attach to the thin filaments. Each troponin molecule controls the position of one tropomyosin molecule, which in turn controls the binding capacity of seven actin subunits in the thin filament.

Once stimulation from the innervating motor nerve fiber ceases, the Ca^{2+} channels in the SR membrane close, and the Ca^{2+}-ATPase molecules in that membrane remove excess calcium from the cytosol. As the Ca^{2+} concentration decreases, these ions dissociate from their binding sites on troponin, which causes the tropomyosin molecules to move back to a position where they block the actin–myosin interaction. The process of relaxation can be thought of as a competition for calcium between the transport protein of the SR membrane and troponin. The transport protein has a greater affinity for the ion, so it preferentially removes it from the cytosol, leaving the troponin molecules without bound calcium.

REVIEW ?

1. Describe the structure of the sarcomere of a skeletal muscle myofibril and the changes that occur during its contraction.

2. Describe the steps that occur between the time that a nerve impulse is transmitted across a neuromuscular junction to the time that the muscle fiber actually begins to shorten. What is the role of calcium ions in this process?

9.7 NONMUSCLE MOTILITY

Skeletal muscle cells are an ideal system for the study of contractility and movement because the interacting contractile proteins are present in high concentration and are part of defined cellular structures. The study of nonmuscle motility is more challenging because the critical components tend to be present in less ordered, more labile, transient arrangements. Moreover, they are typically restricted to a thin *cortex* just beneath the plasma membrane. The cortex is an active region of the cell, responsible for such processes as the ingestion of extracellular materials, the extension of processes during cell movement, and the constriction of a single animal cell into two cells during cell division. All of these processes are dependent on the assembly of microfilaments in the cortex.

In the following pages, we will consider a number of examples of nonmuscle contractility and motility that depend on actin filaments and, in some cases, members of the myosin superfamily. First, however, it is important to survey the factors that govern the rates of assembly, numbers, lengths, and spatial patterns of actin filaments.

Actin-Binding Proteins

Purified actin can polymerize in vitro to form actin filaments, but such filaments cannot interact with one another or per-
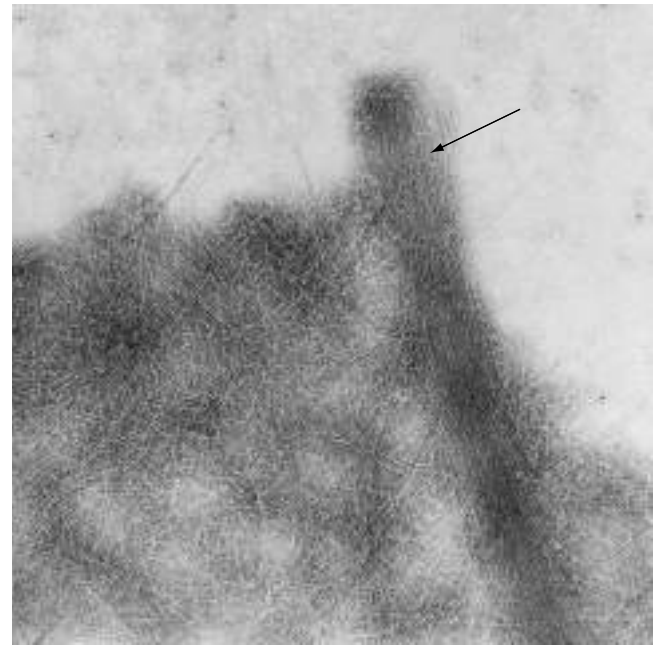
FIGURE 9.64 **Two different arrangements of actin filaments within a cell.** As described later in the chapter, cells move across a substratum by extending various types of processes. This electron micrograph of the leading edge of a motile fibroblast shows the high density of actin filaments. These filaments are seen to be organized into two distinct arrays: as bundles in which the filaments are arranged parallel to one another (arrow) and as a cross-linked meshwork in which the filaments are arranged in various directions. (COURTESY OF J. VICTOR SMALL.)

form useful activities. Under the microscope, they resemble the floor of a barn covered by straw. In contrast, actin filaments in living cells are organized into a variety of patterns, including various types of bundles, thin (two-dimensional) networks, and complex three-dimensional gels (Figure 9.64). The organization and behavior of actin filaments inside cells are determined by a remarkable variety of **actin-binding proteins** that affect the localized assembly or disassembly of the actin filaments, their physical properties, and their interactions with one another and with cellular organelles. More than 100 different actin-binding proteins belonging to numerous families have been isolated from different cell types. Actin-binding proteins can be divided into several categories based on their presumed function in the cell (Figure 9.65).[4]

1. *Nucleating proteins.* The slowest step in the formation of an actin filament is the first step, nucleation, which requires that at least two or three actin monomers come together in the proper orientation to begin formation of the polymer. This is a very unfavorable process for actin molecules left on their own. As noted earlier, the

---

[4]It should be noted that some of these proteins can carry out more than one of the types of activities listed, depending on the concentration of the actin-binding protein and the prevailing conditions (e.g., the concentration of Ca^{2+} and H^+). Most studies of these proteins are conducted in vitro, and it is often difficult to extend the results to activities within the cell.
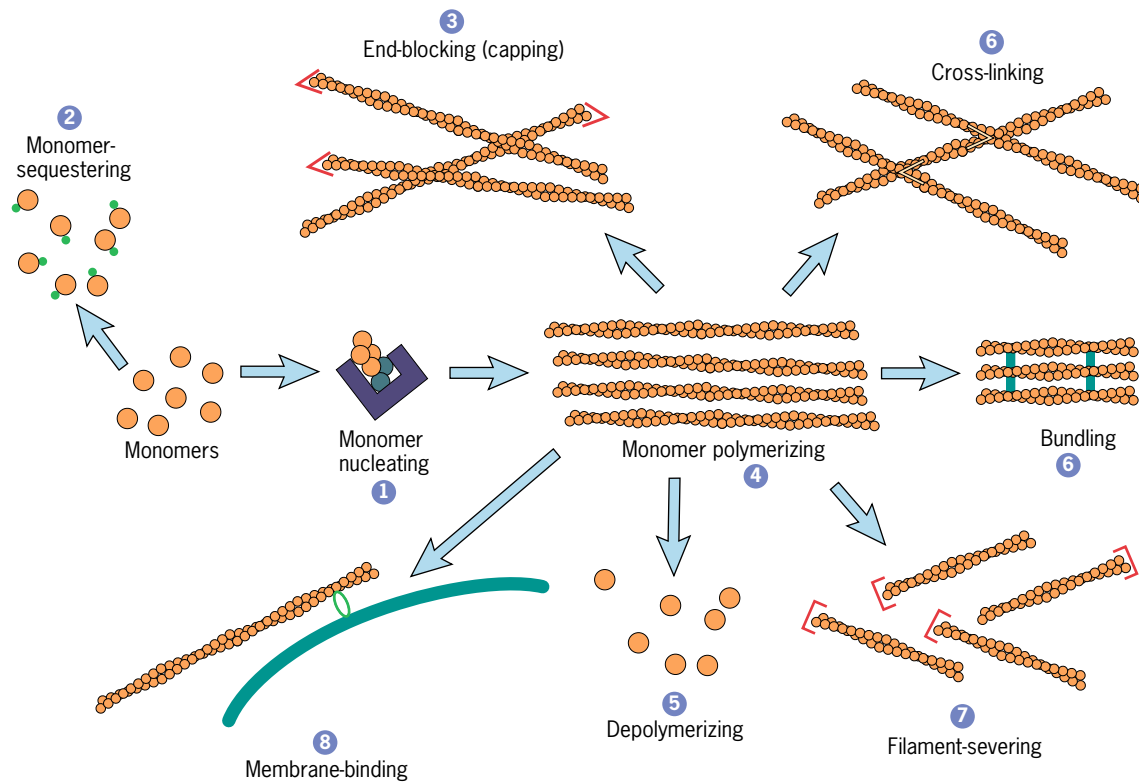
FIGURE 9.65 The roles of actin-binding proteins.

formation of an actin filament is accelerated by the presence of a preexisting seed or nucleus to which monomers can be added (as in Figure 9.46a). Several proteins have been identified that promote nucleation of actin filaments. The best studied is the *Arp2/3 complex*, which contains two "actin-related proteins," that is, proteins that share considerable sequence homology with actins but are not considered "true" actins. Once the complex is activated, the two Arps adopt a conformation that provides a template to which actin monomers can be added, analogous to the way that γ-tubulin is proposed to form a template for microtubule nucleation (Figure 9.20c). As discussed on page 370, the Arp2/3 complex generates networks of short, branched actin filaments. Another nucleating protein, called *formin*, generates unbranched filaments, such as those found at focal adhesions (page 243) and the contractile rings of dividing cells (Section 14.2). Unlike Arp2/3, which remains at the pointed end of the newly formed filament, formin tracks with the barbed end even as new subunits are inserted at that site.

2. ***Monomer-sequestering proteins.*** Thymosins (e.g., thymosin β₄) are proteins that bind to actin-ATP monomers (often called *G-actin*) and prevent them from polymerizing. Proteins with this activity are described as actin monomer-sequestering proteins. These proteins are believed responsible for the relatively high concentration of G-actin in nonmuscle cells (50–200 mM). Without

monomer-sequestering proteins, conditions within the cytoplasm would favor the near complete polymerization of soluble actin monomers into filaments. Because of their ability to bind G-actin and stabilize the monomer pool, changes in the concentration or activity of monomer-sequestering proteins can shift the monomer–polymer equilibrium in a certain region of a cell and determine whether polymerization or depolymerization is favored at the time.

3. ***End-blocking (capping) proteins.*** Proteins of this group regulate the length of actin filaments by binding to one or the other end of the filaments, forming a cap that blocks both loss and gain of subunits. If the fast-growing, barbed end of a filament is capped, depolymerization may proceed at the opposite end, resulting in the disassembly of the filament. If the pointed end is also capped, depolymerization is blocked. The thin filaments of striated muscle are capped at their barbed end at the Z line by a protein called capZ and at their pointed end by the protein tropomodulin. If the tropomodulin cap is disturbed by microinjection of antibodies into a muscle cell, the thin filaments add more actin subunits at their newly exposed pointed end and undergo dramatic elongation into the middle of the sarcomere.

4. ***Monomer-polymerizing proteins.*** Profilin is a small protein that binds to the same site on an actin monomer as does thymosin. However, rather than inhibiting polymer-

ization, profilin promotes the growth of actin filaments. Profilin does this by attaching to an actin monomer and catalyzing the dissociation of its bound ADP, which is rapidly replaced with ATP. The profilin-ATP-actin monomer can then assemble onto the free barbed end of a growing actin filament, which leads to the release of profilin.

5. *Actin filament–depolymerizing proteins.* Members of the cofilin family of proteins (including cofilin, ADF, and depactin) bind to actin-ADP subunits present within the body and at the pointed end of actin filaments (Figure 18.33). Cofilin has two apparent activities: it can fragment actin filaments, and it can promote their depolymerization at the pointed end. These proteins play a role in the rapid turnover of actin filaments at sites of dynamic changes in cytoskeletal structure. They are essential for cell locomotion, phagocytosis, and cytokinesis.

6. *Cross-linking proteins.* Proteins of this group are able to alter the three-dimensional organization of a population of actin filaments. Each of these proteins has two or more actin-binding sites and therefore can cross-link two or more separate actin filaments. Certain of these proteins (e.g., filamin) have the shape of a long, flexible rod and promote the formation of loose networks of filaments interconnected at near right angles to one another (as in Figure 9.64). Regions of the cytoplasm containing such networks have the properties of a three-dimensional elastic gel that resists local mechanical pressures. Other cross-linking proteins (e.g., villin and fimbrin) have a more globular shape and promote the bundling of actin filaments into tightly knit, parallel arrays. Such arrays are found in the microvilli that project from certain epithelial cells (Figure 9.66) and in the hairlike stereocilia (Figure 9.54) that project from receptor cells of the inner ear. Bundling filaments together adds to their rigidity, allowing them to act as a supportive internal skeleton for these cytoplasmic projections.

7. *Filament-severing proteins.* Proteins of this class have the ability to bind to the side of an existing filament and break it in two. Severing proteins (e.g., gelsolin) may also promote the incorporation of actin monomers by creating additional free barbed ends, or they may cap the fragments they generate. As indicated in Figure 9.71, cofilin is also capable of severing filaments.

8. *Membrane-binding proteins.* Much of the contractile machinery of nonmuscle cells lies just beneath the plasma membrane. During numerous activities, the forces generated by the contractile proteins act on the plasma membrane, causing it to protrude outward (as occurs, for example, during cell locomotion) or to invaginate inward (as occurs, for example, during phagocytosis or cytokinesis). These activities are generally facilitated by linking the actin filaments to the plasma membrane indirectly, by means of attachment to a peripheral membrane protein. Two examples were described in previous chapters: the inclusion of short actin polymers into the membrane skele-
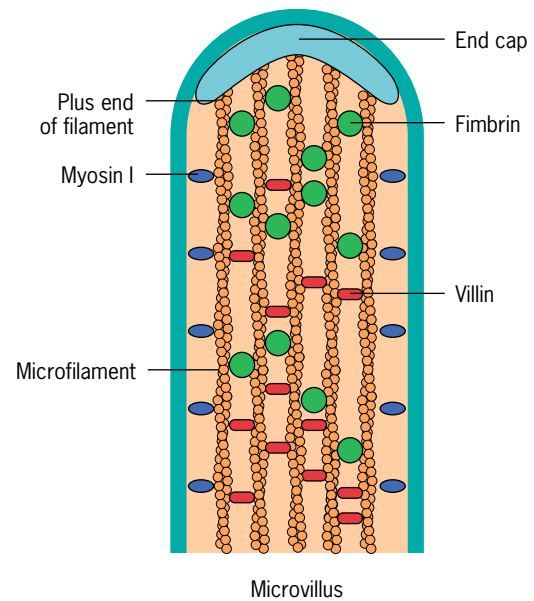
FIGURE 9.66 Actin filaments and actin-binding proteins in a microvillus. Microvilli are present on the apical surface of epithelia that function in absorption of solutes, such as the lining of the intestine and wall of the kidney tubule. Actin filaments are maintained in a highly ordered arrangement by the bundling proteins villin and fimbrin. The role of myosin I, which is present between the plasma membrane of the microvillus and the peripheral actin filaments, remains unclear.

ton of erythrocytes (see Figure 4.32*d*), and the attachment of actin filaments to the membrane at focal adhesions and adherens junctions (see Figures 7.17 and 7.26). Proteins that link membranes to actin include vinculin, members of the ERM family (ezrin, radixin, and moesin), and members of the spectrin family (including dystrophin, the protein responsible for muscular dystrophy).

Examples of Nonmuscle Motility and Contractility

Actin filaments, often working in conjunction with myosin motors, are responsible for a variety of dynamic activities in nonmuscle cells, including cytokinesis, phagocytosis, cytoplasmic streaming (the directed bulk flow of cytoplasm that occurs in certain large plant cells), vesicle trafficking, blood platelet activation, lateral movements of integral proteins within membranes, cell–substratum interactions, cell locomotion, axonal outgrowth, and changes in cell shape. Nonmuscle motility and contractility are illustrated by the following examples.

Actin Polymerization as a Force-Generating Mechanism
Some types of cell motility occur solely as the result of actin polymerization and do not involve the activity of myosin. Consider the example of *Listeria monocytogenes*, a bacterium that infects macrophages and can cause encephalitis or food poisoning. *Listeria* is propelled like a rocket through the cyto-
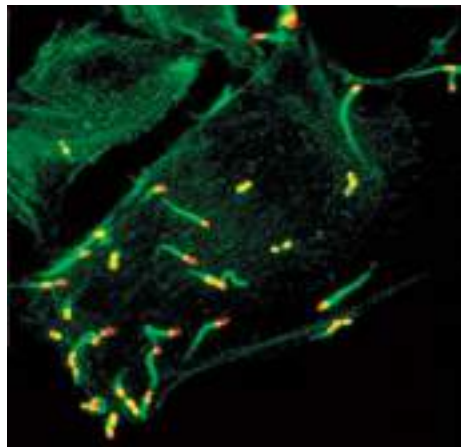
(a)   5 μm

FIGURE 9.67 Cell motility can be driven by actin polymerization.
(*a*) Fluorescence micrograph of a portion of a cell infected with the bacterium *L. monocytogenes*. The bacteria appear as red-stained objects just in front of the green-stained filamentous actin tails. (*b*) Electron micrograph of a cell infected with the same bacterium as in *a*, showing the actin filaments that form behind the bacterial cell and push it through the cytoplasm. The actin filaments have a bristly appearance because they have been decorated with myosin heads. Bar at the upper left, 0.1 μm. (A: COURTESY OF PASCALE COSSART; B: FROM LEWIS G. TILNEY ET AL., J. CELL BIOL. 118:77, 1992; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)
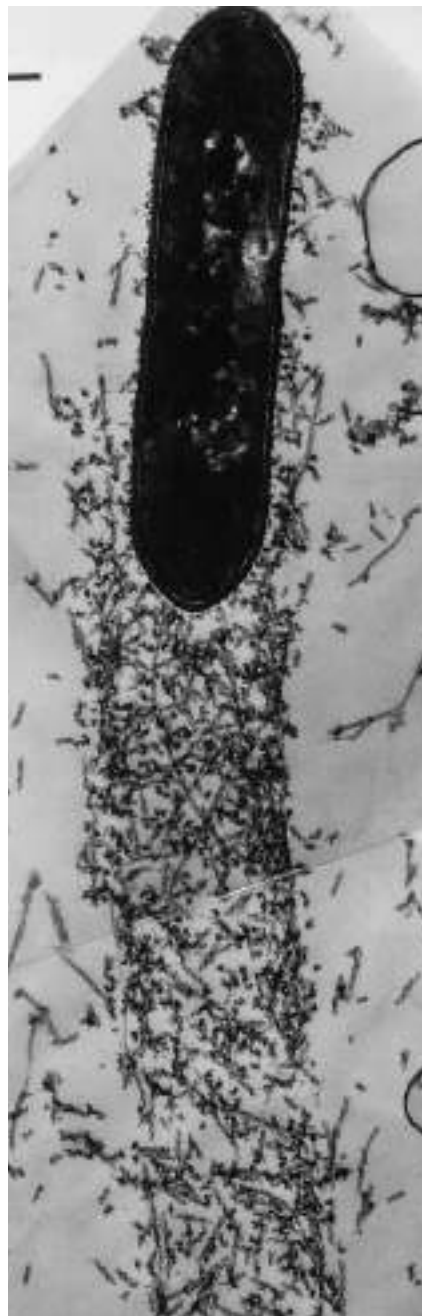
(b)

plasm of an infected cell by the polymerization of actin monomers just behind the bacterium (Figure 9.67). How is the bacterial cell able to induce the formation of actin filaments at a particular site on its surface? Questions about location are important in the study of any type of motile process because motility depends on a cell being able to assemble the necessary machinery at a particular site and a particular time. *Listeria* can accomplish this feat because it contains a surface protein called ActA that is present only at one end of the bacterium. When ActA is exposed within the host cytoplasm, it recruits and activates a number of host proteins (including the Arp2/3 complex, discussed below) that work together to direct the process of actin polymerization. The process of *Listeria* propulsion has been reconstituted in vitro, which has allowed researchers to demonstrate conclusively that actin polymerization by itself, that is, without the participation of myosin motors, is capable of providing the force required for motility.

The same events that occur during *Listeria* propulsion are utilized for normal cellular activities, ranging from the propulsion of cytoplasmic vesicles and organelles to the movement of cells themselves, which is the subject of the following section.

Cell Locomotion   Cell locomotion is required for many activities in higher vertebrates, including tissue and organ development, formation of blood vessels, development of axons, wound healing, and protection against infection. Cell locomotion also contributes to the spread of cancerous tumors. In the following discussion, we will concentrate on studies of cul-

tured cells moving over a flat (i.e., two-dimensional) substrate because these are the experimental conditions that have dominated this field of research. Keep in mind that cells in the body don't move over bare, flat substrates, and there is increasing evidence that some of the findings from these studies may not apply to cells traversing more complicated terrain. Researchers have recently begun to develop more complex substrata, including various types of three-dimensional extracellular matrices (see Figure 18.22), which may lead to revision of some aspects of the mechanism of cell locomotion discussed below.

Figure 9.68 shows a single fibroblast that was in the process of moving toward the lower right corner of the field
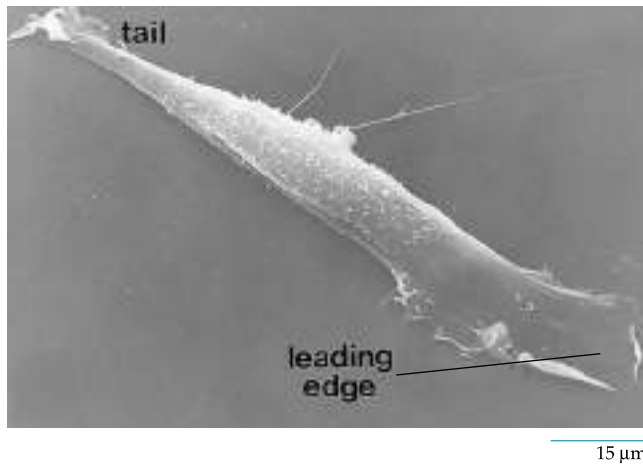
FIGURE 9.68 **Scanning electron micrograph of a mouse fibroblast crawling over the surface of a culture dish.** The leading edge of the cell is spread into a flattened lamellipodium whose structure and function are discussed later in the chapter. (FROM GUENTER ALBRECHT-BUEHLER, INT. REV. CYTOL. 120:194, 1990.)

when it was prepared for microscopy. Cell locomotion, as exhibited by the fibroblast in Figure 9.68, shares properties with other types of locomotion, for example, walking. As you walk, your body performs a series of repetitive activities: first, a leg is extended in the direction of locomotion; second, the bottom of your foot makes contact with the ground, which
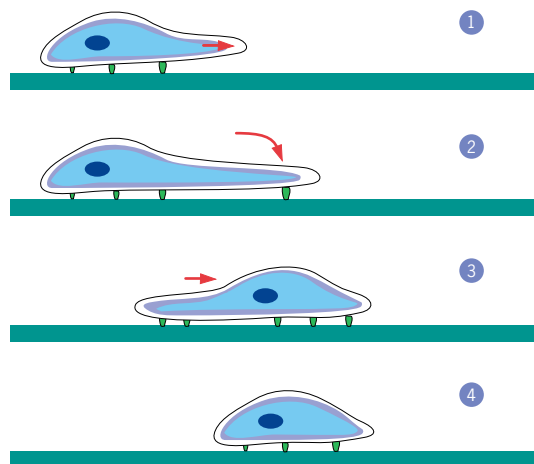
FIGURE 9.69 **The repetitive sequence of activities that occurs as a cell crawls over the substratum.** Step 1 shows the protrusion of the leading edge of the cell in the form of a lamellipodium. Step 2 shows the adhesion of the lower surface of the lamellipodium to the substratum, an attachment that is mediated by integrins residing in the plasma membrane. The cell uses this attachment to grip the substratum. Step 3 shows the movement of the bulk of the cell forward over the site of attachment, which remains relatively stationary. This movement is accomplished by a contractile (traction) force exerted against the substratum. Step 4 shows the cell after the attachments with the substratum have been severed and the rear of the cell has been pulled forward.

acts as a point of temporary adhesion; third, force is generated by the muscles of your legs that moves your entire body forward past the stationary foot, all the while generating traction against the point of adhesion; fourth, your foot—which is now behind your body rather than in front of it—is lifted from the ground in anticipation of your next step. Even though motile cells may assume very different shapes as they crawl over a substratum, they display a similar sequence of activities (Figure 9.69). (1) Movement is initiated by the protrusion of a part of the cell surface in the direction in which the cell is to move. (2) A portion of the lower surface of the protrusion attaches to the substratum, forming temporary sites of anchorage. (3) The bulk of the cell is pulled forward over the adhesive contacts, which eventually become part of the rear of the cell. (4) The cell breaks its rear contacts with the substratum, causing retraction of the trailing edge, or "tail."

Cells that Crawl over the Substratum   When a small piece of living tissue, such as skin or liver, is placed in a culture dish in an appropriate culture medium, individual cells migrate out of the specimen and onto the surface of the dish. Examination of these cells under the microscope typically shows them to be fibroblasts, which are the predominant cells present in connective tissue (see Figure 7.1). As it moves, a fibroblast flattens itself close to the substratum and becomes fan-shaped, with a broadened frontal end and a narrow "tail" (as in Figure 9.68). Its movement is erratic and jerky, sometimes advancing and other times withdrawing. On a good day, a fibroblast may move a distance of about 1 mm. The key to the fibroblast's locomotion is seen by examining its leading edge, which is extended out from the cell as a broad, flattened, veil-like protrusion, called a **lamellipodium** (Figure 9.70a). Lamellipodia are typically devoid of cytoplasmic vesicles and other particulate structures, and the outer edge often exhibits an undulating motion, giving it a ruffled appearance (Figure 9.70b). As a lamellipodium is extended from the cell, it adheres to the underlying substratum at specific points, providing temporary sites of anchorage for the cell to pull itself forward.

We saw on page 368 how the polymerization of actin monomers can provide the force that propels a *Listeria* bacterium through the cytoplasm. This type of intracellular movement is accomplished without the involvement of molecular motors. A similar type of actin-polymerization mechanism is thought to provide the motile force required for the protrusion of the leading edge of a lamellipodium. This type of nonmuscle motility also demonstrates the importance of actin-binding proteins (depicted in Figure 9.65) in orchestrating the assembly and disassembly of actin-filament networks at a particular site within the cell at a particular time.

Suppose we begin with a rounded white blood cell that receives a chemical signal coming from one particular direction where the body has been wounded. Once the stimulus is received at the plasma membrane, it triggers the localized polymerization of actin, which leads to the polarization of the cell and its movement toward the source of the stimulus (Fig-

(a)
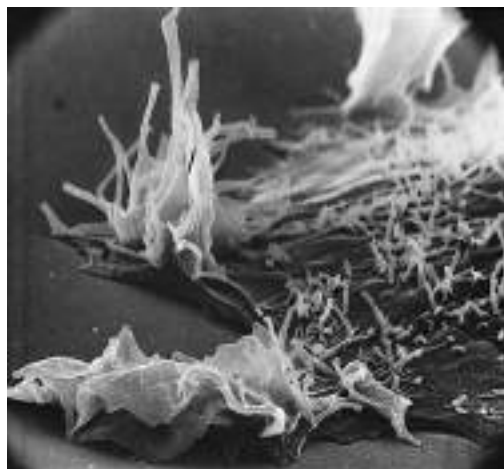
(b)

FIGURE 9.70 The leading edge of a motile cell. (*a*) The leading edge of this motile fibroblast is flattened against the substratum and spread out into a veil-like lamellipodium. (*b*) Scanning electron micrograph of the leading edge of a cultured cell, showing the ruffled membranes of the lamellipodium. (A: COURTESY OF J. VICTOR SMALL; B: FROM JEAN PAUL REVEL, SYMP. SOC. EXP. BIOL. 28:447, 1974.)

ure 9.71*a*).[5] Just as *Listeria* has a protein (ActA) that activates polymerization at the bacterial cell surface, mammalian cells have a family of proteins (the WASP/WAVE family) that activates the Arp2/3 complex at the site of stimulation near the

[5]This type of response can be seen in a remarkable film on the Web showing a neutrophil chasing a bacterium. It can be found using the search words: "neutrophil crawling."

plasma membrane. WASP, the founding member of the family, was discovered as the product of a gene responsible for Wiskott-Aldrich syndrome. Patients with this disorder have a crippled immune system because their white blood cells lack a functional WASP protein and consequently fail to respond to chemotactic signals.

Figure 9.71*b* depicts a model for the major steps in formation of a lamellipodium that would guide a cell in a particular direction. A stimulus is received at one end of the cell (step 1, Figure 9.71*b*), which leads to the activation of Arp2/3 protein complexes by activated WASP proteins (step 2). In its activated state, the Arp2/3 complex adopts a conformation that resembles the free surface at the barbed end of an actin filament. As a result, free ATP-actin monomers bind to the Arp2/3 template, leading to the nucleation of actin filaments (step 3). Polymerization of ATP-bound actin monomers onto the free barbed ends of the growing filaments is promoted by profilin molecules (page 366). Once new actin filaments have formed, Arp2/3 complexes bind to the sides of these filaments (step 4) and nucleate the formation of additional actin filaments that form as branches (step 5). The Arp2/3 complexes remain at the pointed ends, which are situated at the branchpoints. Meanwhile, growth of the barbed ends of older filaments is blocked by the addition of capping protein (step 5). In contrast, addition of actin subunits to the barbed ends of more recently formed filaments of the network pushes the membrane of the lamellipodium outward in the direction of the attractive stimulus (steps 5 and 6). As newer filaments are growing by addition of subunits to their barbed ends, the older capped filaments undergo disassembly from their pointed ends (step 6). Disassembly is promoted by cofilin, which binds to actin-ADP subunits along the filaments (step 6). Actin-ADP subunits released from the disassembling filaments are recharged by conversion into profilin-ATP-actin monomers, which can be reutilized in the assembly of actin filaments at the leading edge.

Figure 9.72 illustrates some of the major structural features of cell locomotion. The electron micrograph of Figure 9.72 shows the branched, cross-linked nature of the filamentous actin network that resides just beneath the plasma membrane of an advancing lamellipodium. The circular insets in Figure 9.72 show a succession of short actin-filament branches, with Arp2/3 complexes highlighted by immunogold labeling. The Arp2/3 complexes are seen to reside at the Y-shaped junctions where the newly polymerized filaments have branched off of preexisting filaments.

Lamellipodial movement is a dynamic process. As actin filament polymerization and branching continue at the very front edge of the lamellipodium, actin filaments are depolymerizing toward the rear of the lamellipodium (step 6, Figure 9.71). Thus taken as a whole, the entire actin-filament array undergoes a type of treadmilling (page 353) in which actin subunits are added to barbed ends of the array at its front end and lost from pointed ends of the array toward the rear.

According to the sequence of events depicted in Figure 9.69, protrusion of the leading edge is followed by movement of the bulk of the cell. The major forces involved in cell

(a)

(b)
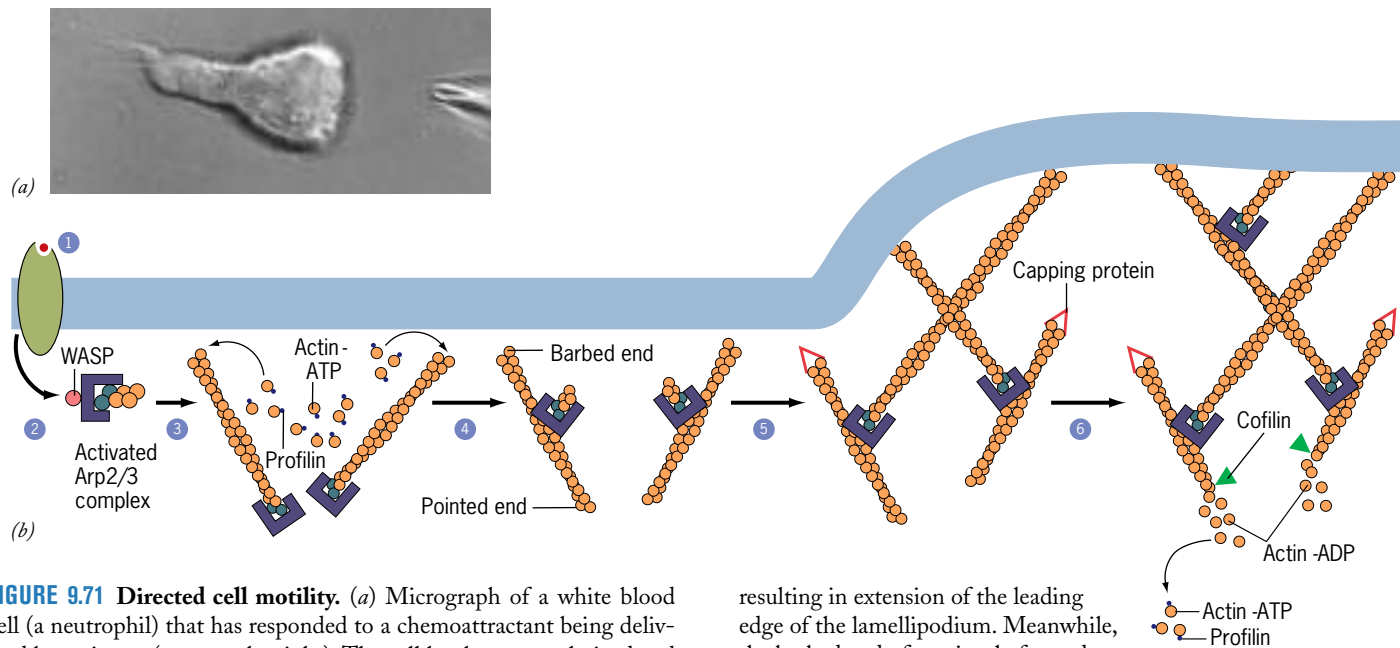
FIGURE 9.71 **Directed cell motility.** (*a*) Micrograph of a white blood cell (a neutrophil) that has responded to a chemoattractant being delivered by a pipette (seen on the right). The cell has become polarized and is moving toward the source of the stimulus. (*b*) A proposed mechanism for the movement of a cell in a directed manner. A stimulus is received at the cell surface (step 1) that leads to the activation of Arp2/3 complex by a member of the WASP/WAVE family (step 2). Activated Arp2/3 complexes serve as nucleating sites for the formation of new actin filaments (step 3). Once filaments have formed, Arp2/3 complexes attach to the sides of the filaments (step 4), which stimulates their nucleating activity. As a result, the bound Arp2/3 complexes initiate side branches that extend outward (step 5) at an angle of about 70° relative to the existing filaments to which they are anchored. As these filaments polymerize, they are thought to push the plasma membrane outward, resulting in extension of the leading edge of the lamellipodium. Meanwhile, the barbed end of previously formed filaments are bound by a capping protein, which prevents further growth of these filaments, keeping them short and rigid. Eventually, the pointed ends of the older, preexisting actin filaments undergo depolymerization, releasing ADP-actin subunits (step 6). Depolymerization is promoted by cofilin, which binds to ADP-actin subunits within the filament and stimulates dissociation of the subunits from the pointed end of the filament. Released subunits bind profilin and become recharged by ATP/ADP exchange, which makes them ready to engage in actin polymerization (as in step 3). (A: FROM CAROLE A. PARENT, CURR. OPIN. CELL BIOL. 16:5, 2004; COPYRIGHT 2004, WITH PERMISSION FROM ELSEVIER SCIENCE.)

locomotion are those generated at sites of adhesion that are required to pull or "tow" the main body of the cell forward (step 3 of Figure 9.69). They are often described as "traction forces" because they occur at sites where the cell grips the substrate. When cells are allowed to migrate over a thin sheet of elastic material, movements of the cell are accompanied by deformation of the underlying substratum (see Figure 7.18). The magnitude of the traction forces exerted at various locations within a live migrating cell can be calculated from the dynamic patterns of substrate deformation and portrayed as shown in Figure 9.73*a*. As seen by examination of this computerized image of a migrating fibroblast, the greatest traction forces are exerted just behind the cell's leading edge where the cell adheres strongly to the underlying substratum. The presence of these sites of attachment is best revealed by following the localization of fluorescently labeled vinculin within a liv-
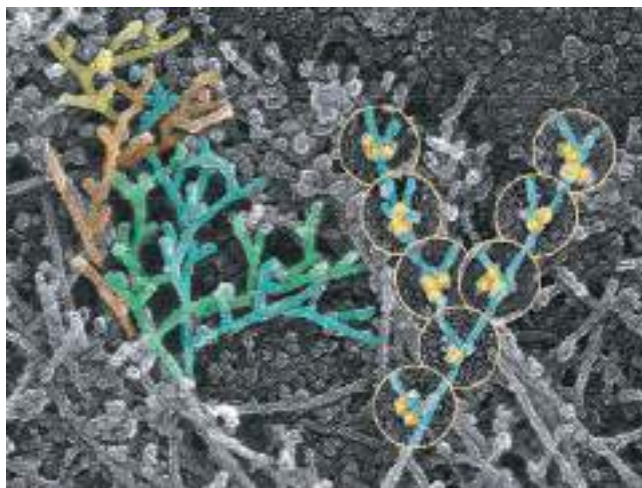
FIGURE 9.72 **The structural basis of lamellipodial extension.** Electron micrograph of a replica of the cytoskeleton at the leading edge of a motile mouse fibroblast. The actin filaments are seen to be arranged in a branched network, which has been colorized to indicate individual "trees." The circular insets show a succession of Y-shaped junctions between branched actin filaments. Arp2/3 complexes are localized at the base of each branch by antibodies linked to colloidal gold particles (yellow). (FROM TATYANA M. SVITKINA AND GARY G. BORISY, FROM J. CELL BIOL., VOL. 145, #5, 1999; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

(a)   20 μm

(b)
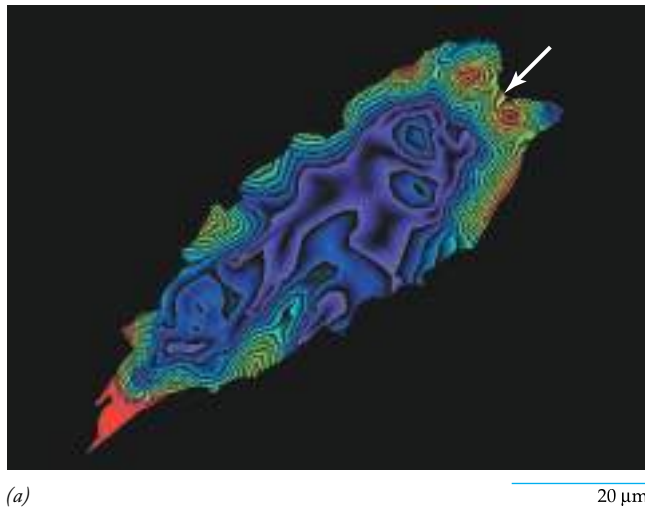
FIGURE 9.73 Distribution of traction forces within a migrating fibroblast. (*a*) As a cell migrates it generates traction (pulling) forces against its substrate. The present image shows the traction forces generated per unit area by the surface of a migrating fibroblast. Traction forces were calculated at different sites on the surface based on the degree of substrate deformation (see Figure 7.18). The magnitude of the traction forces are expressed by varying colors with red representing the strongest forces. The largest forces are generated at sites of small focal complexes that form transiently behind the leading edge of the cell where the lamellipodium is being extended (arrow). Deformation at the rear of the cell (shown in red) occurs as the front end actively pulls against the tail, which is passively anchored. (*b*) A living, migrating
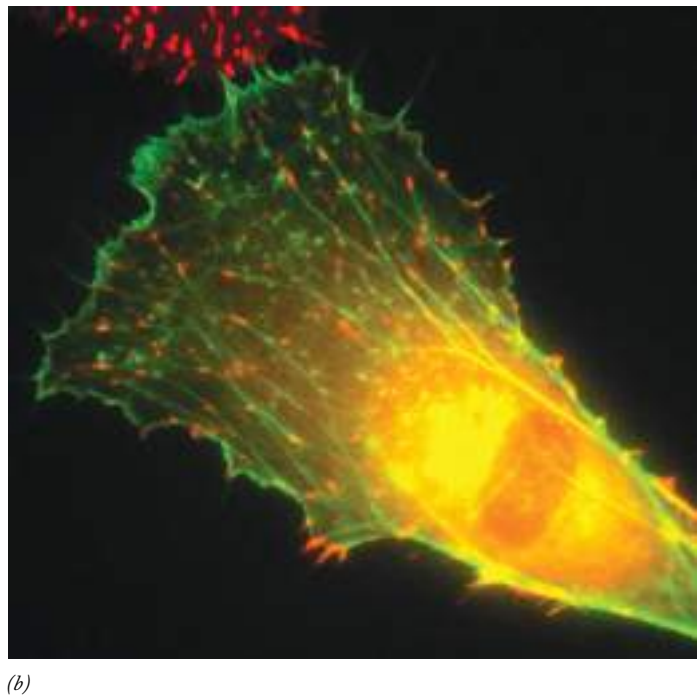
fibroblast exhibiting a well-developed lamellipodium that is adhering to the underlying substratum at numerous sites (red). This cell is expressing GFP-actin (green) and had been injected with rhodamine-tagged vinculin (red). The fluorescently labeled vinculin is incorporated into dot-like focal complexes near the leading edge of the cell. Some of these focal complexes disassemble, whereas others mature into focal adhesions, which are situated farther from the advancing edge. (A: FROM KAREN A. BENINGO ET AL., J. CELL BIOL. 153:885, 2001, COURTESY OF YU-LI WANG; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS; B: REPRINTED WITH PERMISSION FROM J. VICTOR SMALL, ET AL., IMAGE COURTESY OF OLGA KRYLYSHKINA, NATURE REVS. MOL. CELL BIOL. 3:957, 2002; © COPYRIGHT 2002, MACMILIAN MAGAZINES LTD.)

ing cell. This allows investigators to specifically visualize structures where the cell makes contact with the underlying substratum. Figure 9.73*b* is a fluorescence micrograph showing the presence of red fluorescent vinculin molecules concentrated just behind the leading edge of a migrating cell. Vinculin is a major component of focal adhesions, the actin-containing sites illustrated in Figure 7.17. The vinculin-containing sites where the leading edge of a migrating cell adheres to the substratum tend to be smaller and simpler than the mature focal adhesions seen in highly spread, stationary cultured cells and are often referred to as *focal complexes*. The focal complexes that form near the leading edge of a motile cell exert traction force through their associated actin filaments and then typically disassemble as the cell moves forward.

A large body of evidence indicates that actin polymerization is responsible for pushing the leading edge of a cell outward (step 1, Figure 9.69), whereas myosin (in conjunction with actin filaments) is responsible for pulling the remainder of the cell forward (step 3, Figure 9.69). These contrasting roles of actin and myosin are best illustrated in studies on fish keratocytes, which are cells derived from the epidermis that

covers the fish's scales. Keratocytes have been a favored system for studying locomotion because their rapid gliding movement depends on the formation of a very broad, thin lamellipodium. Figure 9.74 shows a moving keratocyte that has been fixed and stained for actin (Figure 9.74*a*) and myosin II (Figure 9.74*b*). As expected from the previous discussion, the advancing edge of the lamellipodium is filled with actin. Myosin, on the other hand, is concentrated in a band where the rear of the lamellipodium joins the remainder of the cell. Electron micrographs of this region show the presence of clusters of small, bipolar myosin II filaments bound to the actin network (Figure 9.74*c*). Contractile forces generated by these myosin molecules are presumed to pull the bulk of the cell along behind the advancing lamellipodium. Myosin I and other unconventional myosins are also thought to generate forces for cell locomotion in some organisms.

Axonal Outgrowth   In 1907, Ross Harrison of Yale University performed one of the classic experiments in biology. Harrison removed a small piece of tissue from the developing nervous system of a frog embryo and placed the fragment into

(a)   (b)   10 μm   (c)
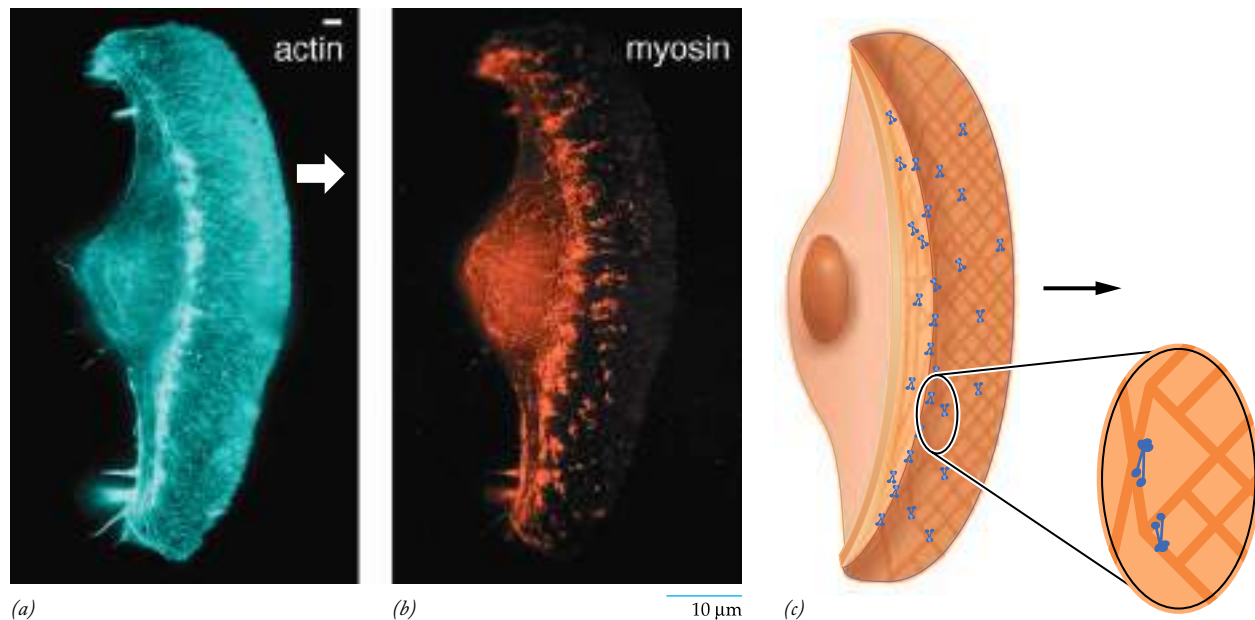
FIGURE 9.74 The roles of actin and myosin in the lamellipodial-based movement of fish keratocytes. (*a,b*) Fluorescence micrographs of a fish keratocyte moving over a culture dish by means of a broad, flattened lamellipodium. The arrow shows the direction of movement, which can occur at rates of 10 μm/min. The distribution of filamentous actin is revealed in part *a*, which shows the localization of fluorescently labeled phalloidin, which binds only to actin filaments. The distribution of myosin in the same cell is revealed in part *b*, which shows the localization of fluorescent antimyosin antibodies. It is evident that the body of the lamellipodium contains actin filaments but is virtually devoid of myosin. Myosin is concentrated, instead, in a band that lies just behind the lamellipodium, where it merges with the body of the cell. (*c*) A schematic drawing depicting the filamentous actin network of the lamellipodium and the actin–myosin interactions toward the rear of the lamellipodium. The actin network is shown in red, myosin molecules in blue. (BY ALEXANDER B. VERKHOVSKY, FROM TATYANA M. SVITKINA ET AL., J. CELL BIOL. 139:397, 1997; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

a tiny drop of lymphatic fluid. Harrison watched the tissue under a microscope over the next few days and found that the nerve cells not only remained healthy, but many of them sprouted processes that grew out into the surrounding medium. Not only was this the first time that cells had been kept alive in tissue culture, the experiment provided strong evidence that axons develop by a process of active outgrowth and elongation.

The tip of an elongating axon is very different from the remainder of the cell (see Figure 9.1*b*). Although the bulk of the axon shows little outward evidence of motile activity, the tip, or **growth cone**, resembles a highly motile, crawling fibroblast. Close examination of a living growth cone reveals several types of locomotor protrusions: a broad, flattened lamellipodium that creeps outward over the substratum; short, stiff *microspikes* (Figure 9.75*a*) that point outward to the edge of the lamellipodium; and highly elongated *filopodia* that extend and retract in a continuous display of motile activity. Fluorescence microscopy shows all of these structures in the peripheral domain of the growth cone to be filled with actin filaments (shown in green, Figure 9.75*b*). These actin filaments are presumed to be responsible for the motile activities of the growth cone. Microtubules, on the other hand, fill the axon and the central domain of the growth cone, providing support for the thin, elongating axon. A number of individual microtubules are seen to penetrate into the actin-rich periph-

ery (shown in orange, Figure 9.75*b*). These penetrating microtubules are highly dynamic and may play an important role in determining the direction of axonal growth.

The growth cone is a highly motile region of the cell that explores its environment and elongates the axon. Within the embryo, the axons of developing neurons grow along defined paths, following certain topographical features of the substratum or responding to the presence of certain chemicals that diffuse into their path. The lamellipodia and filopodia from the growth cone respond to the presence of these physical and chemical stimuli, causing the pathfinding axons to turn toward attractive factors and away from repulsive factors. Figure 9.76*a* shows a cultured neuron whose advancing tip has made a direct turn toward a diffusible protein called netrin, which acts as an attractant for axons growing within the early embryo. Ultimately, the correct wiring of the entire nervous system depends on the uncanny ability of embryonic growth cones to make the proper steering "decisions" that lead them to the target organ they must innervate.

Changes in Cell Shape during Embryonic Development
Each part of the body has a characteristic shape and internal architecture that arises during embryonic development: the spinal cord is basically a hollow tube, the kidney consists of microscopic tubules, each lung is composed of microscopic air spaces, and so forth. Numerous cellular activities are necessary
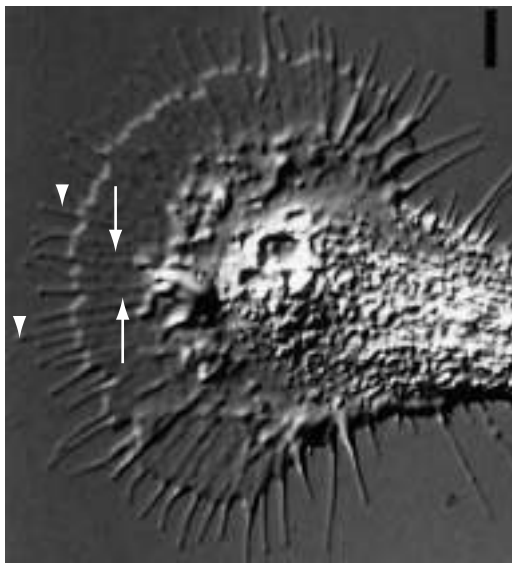
(a)
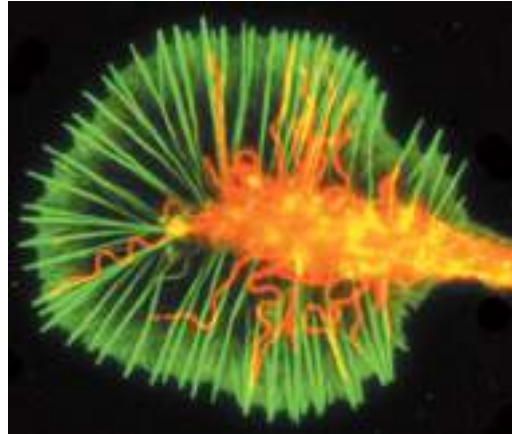
(b)

FIGURE 9.75 The structure of a growth cone: the motile tip of a growing axon. (*a*) A video image of a live growth cone. The terminus is spread into a flattened lamellipodium that creeps forward over the substratum. Rodlike microspikes (arrows) can be seen within the transparent veil of the lamellipodium, and fine processes called filopodia (arrowheads) can be seen projecting ahead of the leading edge of the lamellipodium. Bar, 5 μm. (*b*) Fluorescence micrograph of the growth cone of a neuron showing the actin filaments (green) concentrated in the peripheral domain and the microtubules (orange) concentrated in the central domain. A number of microtubules can be seen to invade the peripheral domain, where they interact with actin-filament bundles. (A: FROM PAUL FORSCHER AND STEPHEN J. SMITH, J. CELL BIOL. 107:1508, 1988; B: FROM FENG-QUAN ZHOU, CLARE M. WATERMAN-STORER, AND CHRISTOPHER S. COHAN, J. CELL BIOL. VOL. 157, #5, COVER, 2002. BOTH BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

for the development of the characteristic morphology of an organ, including programmed changes in cell shape. Changes in cell shape are brought about largely by changes in the orientation of cytoskeletal elements within the cells. One of the best examples of this phenomenon is seen in the early stages of the development of the nervous system.

Toward the end of gastrulation in vertebrates, the outer (ectodermal) cells situated along the embryo's dorsal surface elongate and form a tall epithelial layer called the *neural plate* (Figure 9.77*a,b*). The cells of the neural plate elongate as microtubules become oriented with their long axes parallel to that of the cell (inset, Figure 9.77*b*). Following elongation, the cells of the neural epithelium become constricted at one end, causing them to become wedge shaped and the entire layer of cells to curve inward (Figure 9.77*c*). This latter change in cell shape is brought about by the contraction of a band of microfilaments that assemble in the cortical region of the cells just beneath the apical cell membrane (inset, Figure 9.77*c*). Eventually, the curvature of the neural tube causes the outer edges to contact one another, forming a cylindrical, hollow tube (Figure 9.77*d,e*) that will give rise to the animal's entire nervous system.
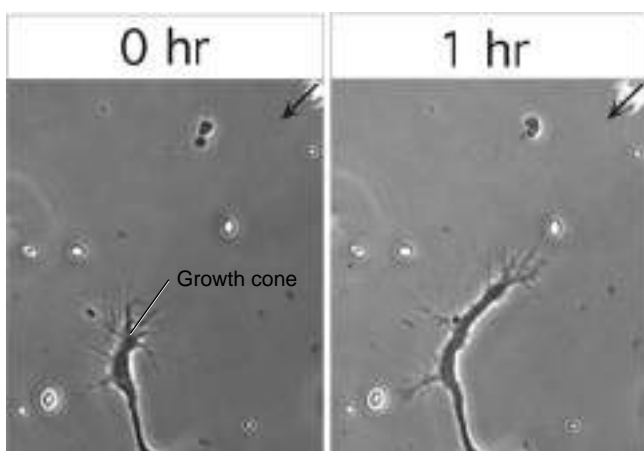
(a)   *(b)*

FIGURE 9.76 The directed movements of a growth cone. A video image of a live growth cone of a *Xenopus* neuron that has turned toward a diffusible protein (netrin-1) released from a pipette whose position is indicated by the arrow. (REPRINTED WITH PERMISSION FROM ELKE STEIN AND MARC TESSIER-LAVIGNE, SCIENCE 291:1929, 2001; © COPYRIGHT 2001, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

REVIEW ?

1. List the various types of actin-binding proteins and one function of each type.

2. Describe the steps taken by a mammalian cell crawling over a substratum.

3. Describe the role of actin filaments in the activities of the growth cone of a neuron.

SYNOPSIS

**The cytoskeleton is composed of three distinct types of fibrous
structures: microtubules, intermediate filaments, and microfila-
ments (actin filaments), which participate in a number of cellular
activities.** Collectively, the elements of the cytoskeleton function as
a structural support that helps maintain the shape of the cell; as an
internal framework responsible for positioning the various organelles
within the cell interior; as part of the machinery required for the
movement of materials and organelles within cells; and as force-
generating elements responsible for the movement of cells from one
place to another. (p. 318)

**Microtubules are hollow, tubular structures 25 nm in diameter
that are assembled from the protein tubulin and, in addition to
the cytoskeleton, form part of the mitotic spindle, centrioles, and
the core of cilia and flagella.** Microtubules are polymers assembled
from $\alpha\beta$-tubulin heterodimers that are arranged in rows, or
protofilaments. Many of the properties of microtubules, including
their, stability and capabilities for interaction, are influenced by
members of a group of microtubule-associated proteins (MAPs).
Because of their stiffness, microtubules often act in a supportive
capacity not unlike the way that steel girders provide support for a

tall building. The structural role of microtubules is most evident on examination of highly elongated processes, such as axons, which are filled with microtubules oriented parallel to the long axis of the process. Microtubules also function in such diverse activities as influencing the deposition of cellulose in a plant cell wall; maintaining the position of the membranous organelles of the biosynthetic pathway including the ER and Golgi complex; and the movement of vesicles and other materials between the cell body and axon terminals of a nerve cell. (p. 324)

Three families of motor proteins have been identified and characterized: kinesins and dyneins, which move along microtubules, and myosins, which move along microfilaments. These motor proteins are able to convert chemical energy stored in ATP into mechanical energy that is used to move cellular cargo attached to the motor. Forces are generated as conformational changes in the motor protein are coupled to a chemical cycle involving the binding and hydrolysis of nucleotides and the release of the bound products. (p. 327)

In most cases, kinesins and cytoplasmic dynein move materials along microtubules in opposite directions. Both kinesin and cytoplasmic dyneins are motor proteins with globular heads that interact with microtubules and act as force-generating engines and an opposite end that binds to specific types of cargo to be transported. Kinesin moves materials toward the plus end of a microtubule, cytoplasmic dynein toward the minus end. Kinesin has been implicated in the movement of ER-derived vesicles, endosomes, lysosomes, and secretory granules and is the primary motor protein mediating anterograde transport in an axon (from cell body to axon terminus). (p. 330)

The nucleation of microtubules in vivo occurs in association with a variety of MTOCs (microtubule organizing centers). In animal cells, the microtubules of the cytoskeleton typically form in association with the centrosome, a complex structure that contains two barrel-shaped centrioles surrounded by amorphous, electron-dense pericentriolar material. Centrioles contain nine evenly spaced fibrils, each of which is composed of three microtubules, and they usually occur in pairs whose members are oriented at right angles to one another. Microtubules typically radiate outward from the pericentriolar material, which contains the components required for microtubule nucleation. The microtubules that form the fibers of a cilium or flagellum originate from a basal body, which has essentially the same structure as a centriole. Centrosomes, basal bodies, and other MTOCs, share a common protein component called γ-tubulin, which plays a key role in nucleation of microtubules. (p. 333)

The microtubules of the cytoskeleton are dynamic polymers that are subject to shortening, lengthening, disassembly, and reassembly. Disassembly of the microtubular cytoskeleton can be induced by a number of agents including colchicine, cold temperature, and elevated Ca^{2+} concentration. The microtubules of the cytoskeleton are normally disassembled prior to cell division, and the tubulin subunits are reassembled as part of the mitotic spindle. This process of disassembly and reassembly is reversed following division. At any given moment, some microtubules of the cytoskeleton are growing in length while others are shrinking. When individual microtubules are followed over time, they are seen to switch back and forth between growing and shrinking phases, a phenomenon known as dynamic instability. Both growth and shrinkage occur predominantly, if not exclusively, at the plus end of the polymer—the end opposite the MTOC. Tubulin dimers that are polymerized into a microtubule contain a GTP molecule that is hydrolyzed soon after its incorpora-

tion into the polymer. During periods of rapid polymerization, the hydrolysis of GTP bound to incorporated tubulin dimers lags behind the incorporation of new dimers, so that the end of the microtubule contains a cap of tubulin-GTP dimers that favors the addition of subunits and growth of the microtubule. Assembly and disassembly may also be governed by Ca^{2+} concentration and the presence of specific MAPs. (p. 335)

Cilia and flagella contain a core structure, the axoneme, that is composed of an array of microtubules that support the organelle as it projects from the cell surface and provide the machinery for generating the forces required for locomotion. In cross section, an axoneme is seen to consist of nine outer doublet microtubules (a complete A microtubule and an incomplete B microtubule) surrounding a pair of single microtubules. A pair of arms projects from the A microtubule of each doublet. The arms are composed of ciliary dynein, a motor protein that uses the energy released by ATP hydrolysis to generate the forces required for ciliary or flagellar bending. This is accomplished as the dynein arms of one doublet attach to the B microtubule of the neighboring doublet and then undergo a conformational change that causes the A microtubule to slide a perceptible distance. Sliding on one side of the axoneme alternates with sliding on the other side so that a part of the cilium or flagellum first bends in one direction and then in the opposite direction. Microtubule sliding has been demonstrated directly by attaching beads to the doublets of demembranated axonemes and following their relative movements during reactivation. (p. 339)

Intermediate filaments (IFs) are ropelike cytoskeletal structures approximately 10 nm in diameter that, depending on the cell type, may be composed of a variety of different protein subunits capable of assembling into similar types of filaments. Unlike microtubules, intermediate filaments are composed of symmetric building blocks (tetrameric subunits), which assemble into filaments that lack polarity. IFs are resistant to tensile forces and relatively insoluble; yet, like the other two types of cytoskeletal elements, they are dynamic structures that rapidly incorporate fluorescently labeled subunits that are injected into a cell. Assembly and disassembly are thought to be controlled primarily by phosphorylation and dephosphorylation. IFs provide mechanical stability to cells and are required for specialized, tissue-specific functions. (p. 347)

Actin filaments (or microfilaments) are 8 nm in diameter, composed of a double-helical polymer of the protein actin, and play a key role in virtually all types of contractility and motility within cells. Depending on the type of cell and the activity, actin filaments can be organized into highly ordered arrays, loose ill-defined networks, or tightly held bundles. Actin filaments are often identified by their ability to bind to the S1 myosin fragment, which also reveals the polarity of the filament. Although both ends of an actin filament can gain or lose subunits, the barbed (or plus end) of the filament is the preferred site of subunit addition, and the pointed (or minus end) is the preferred site of subunit loss. To be incorporated onto the growing end of a filament, an actin subunit must be bound to an ATP, which is hydrolyzed soon after its incorporation. Cells maintain a dynamic equilibrium between the monomeric and polymeric forms of actin, which can be altered by changes in a variety of local conditions. The role of actin filaments in a particular process is most readily tested by treating the cells with cytochalasin, which promotes the depolymerization of the filament, or with phalloidin, which prevents it from disassembling and participating in dynamic activities. (p. 351)

The forces responsible for microfilament-dependent processes may be generated by the assembly of the actin filament or, more often, as the result of interaction with the motor protein myosin.

The propulsion of certain bacteria through the cytoplasm of an infected phagocyte is a process driven by actin polymerization. Myosins are generally divided into two classes: conventional (type II) myosins and unconventional (types I and III–XVIII) myosins. Myosin II is the molecular motor that generates force in the various types of muscle tissues and also a variety of nonmuscle activities, including cytokinesis. Myosin II molecules contain a long, rod-shaped tail attached at one end to two globular heads. The heads bind the actin filament, hydrolyze ATP, and undergo the conformational changes required for force generation. The neck is thought to act as a lever arm that amplifies conformational changes of the head. The fibrous tail mediates the assembly of myosin into bipolar filaments. Most unconventional myosins have a single head and variable tail domains; they have been implicated in cell motility and organelle transport. (p. 354)

The contraction of a skeletal muscle fiber results from the sliding of actin-containing thin filaments toward the center of the individual sarcomeres of a myofibril, and it is driven by forces generated in the myosin cross-bridges that extend outward from the thick filaments. Changes that occur during the shortening of a muscle fiber are reflected in changes in the banding pattern of the sarcomeres as the Z lines at the ends of the sarcomere are moved toward the outer edges of the A bands. Contraction is triggered as an impulse penetrates into the interior of the muscle fiber along the membranous transverse tubules, stimulating the release of Ca^{2+} from storage sites in the sarcoplasmic reticulum (SR). The binding of calcium ions to troponin molecules of the thin filaments causes a change in conformation that moves the tropomyosin molecules to a position that exposes the myosin-binding sites on the actin subunits of the thin filament. The subsequent interaction between myosin and actin triggers filament sliding. (p. 359)

Nonmuscle motility and contractility depend on some of the same proteins found in muscle cells, but they are arranged in less ordered, more labile, and transient configurations. Nonmuscle motility depends on actin, usually in conjunction with myosin. The organization and behavior of actin filaments are determined by a variety of actin-binding proteins that affect the assembly of the actin filaments, their physical properties, and their interactions with one another and with other cellular organelles. Included on the list are proteins that sequester actin monomers and prevent their polymerization; proteins that form a cap at one end of an actin filament, which may block filament growth or lead to filament disassembly; proteins that cross-link actin filaments into bundles, loose meshworks, or three-dimensional gels; proteins that sever actin filaments; and proteins that link actin filaments to the inner surface of the plasma membrane. (p. 365)

Examples of nonmuscle motility and contractility include the crawling of cells over a substratum and axonal outgrowth. Cell crawling is typically accomplished by a flattened, veil-like projection, or lamellipodium, that forms at the cell's leading edge. As a lamellipodium extends from the cell, it adheres to the underlying substratum at specific points, providing temporary sites of anchorage for the cell to crawl over. The protrusion of a lamellipodium is associated with the nucleation and polymerization of actin filaments and their association with various types of actin-binding proteins. The forces required for lamellipodial protrusion result from actin polymerization. The tip of an elongating axon consists of a growth cone, which resembles a highly motile, crawling fibroblast and contains several types of locomotor protrusions, including a lamellipodium, microspikes, and filopodia. The growth cone acts both to explore the environment and to elongate the axon along an appropriate pathway. (p. 368)

ANALYTIC QUESTIONS

1. If a myofibril were pulled so that the sarcomeres increased in length by approximately 50 percent, what effect would you expect this to have on the contractile ability of the myofibril? Why? What effects would this have on the H, A, and I bands?

2. What are three different radioactively or fluorescently labeled substances that you might inject into a cell that would label the microtubules of a cell without labeling the other cytoskeletal elements?

3. What are two types of nonmuscle motility that might be unaffected by antibodies against both myosin I and myosin II? Why?

4. A centriole contains _____ complete microtubules, and a cilium contains _____ complete microtubules.

5. Microtubules can be formed in vitro from tubulin that is bound to GTP analogues that (unlike GTP) cannot be hydrolyzed. What properties would you expect these microtubules to possess?

6. List two things you could do that would shift the dynamic equilibrium of an in vitro preparation of tubulin and microtubules toward the formation of microtubules. List two treatments that would shift the equilibrium in the opposite direction.

7. It was noted that a demembranated ciliary or flagellar axoneme is able to beat with a normal frequency and pattern. Can you conclude that the plasma membrane is not important in ciliary or flagellar function?

8. Because cytoplasmic vesicles are seen to move in both directions within an axon, can you conclude that some microtubules are oriented with their plus end facing the axon terminus and others oriented with the opposite polarity? Why or why not?

9. Would you agree with the statement that the centrosome plays a key role in determining the rates of lengthening and shortening of the microtubules of an animal cell? Why or why not?

10. If you were comparing the molecular structure of kinesin and myosin, which are thought to have been derived from a common ancestral protein, which part (heads or tails) would you expect to be most similar between them? Why?

11. Figure 9.30*a* shows the appearance of a cross section of a ciliary axoneme cut at an interior portion of the cilium. How might the image of a cross section differ if it had been cut very near the tip of a cilium at the beginning of the recovery stroke?

12. If an individual kinesin molecule can move at a rate of 800 nm/sec in an in vitro motility assay, what is the maximal turnover rate (ATP molecules hydrolyzed per second) by one of the molecule's motor domains?

13. Why do you suppose more can be learned about microtubule dynamics by injecting fluorescent tubulin into a cell than radioactively labeled tubulin? Can you think of a question that would be better answered by radioactively labeled tubulin?

14. Suppose you discovered that a mouse lacking copies of the conventional kinesin gene appeared to show no ill effects and lived to a ripe old age. What could you conclude about the role of kinesin in intracellular locomotion?

15. Which type of vertebrate tissue would you expect to be an excellent source of tubulin? of actin? of keratin? Which protein would you expect to be the least soluble and most difficult to extract? What types of protein would you expect as contaminants in a preparation of tubulin? Which in a preparation of actin?

16. Actin is one of the most evolutionarily conserved proteins. What does this tell you about the structure and function of this protein in eukaryotic cells?

17. According to data from genome sequences, cytoplasmic dynein is absent in some plants (e.g., *Arabidopsis*) and present in others (e.g., rice). Does this finding surprise you? What else might you do to confirm or refute such a statement? How is it possible that higher plant cells could operate without cytoplasmic dynein?

18. Myosin II action (Figure 9.61) differs from that of kinesin (Figure 9.15) in that one of the kinesin heads is always in contact with a microtubule, whereas both myosin heads become completely detached from the actin filament. How are these differences correlated with the two types of motor activities in which these proteins engage?

19. The microtubules of an axon are thought to be nucleated at the centrosome, then severed from their nucleation site and moved into the axon. It was noted in the text that cytoplasmic dynein is responsible for the retrograde movement of organelles in axons, yet this same motor is thought to mediate the anterograde movement of microtubules in these same cellular processes. How is it possible that the same minus end–directed motor can be involved in both anterograde and retrograde movements?

20. Motor proteins are often described as mechanoenzymes. Why? Could this same term be applied to virtually all enzymes? Why or why not?

10
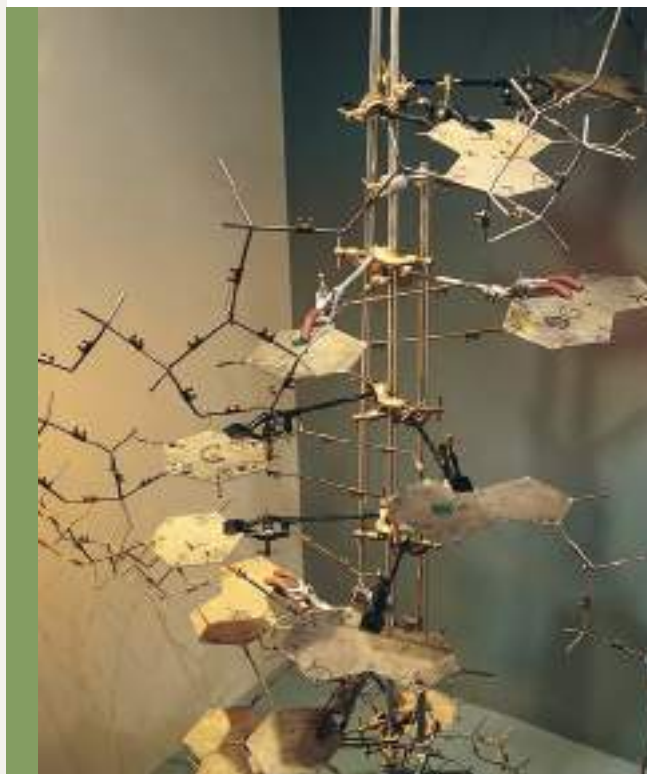
The Nature of the Gene and the Genome

Our concept of the gene has undergone a remarkable evolution as biologists have learned more and more about the nature of inheritance. The earliest studies revealed genes to be discrete factors that were retained throughout the life of an organism and then passed on to each of its progeny. Over the following century, these hereditary factors were shown to reside on chromosomes and to consist of DNA, a macromolecule with extraordinary properties. Figure 10.1 provides an overview of some of the early milestones along this remarkable journey of discovery, capped by the description of the double helical structure of DNA in 1953. In the decades that followed this turning point, a major branch of molecular biology began to focus on the **genome**, which is the collective body of genetic information that is present in a species. A genome contains all of the genes required to "build" a particular organism. During the past decade or so, collaborations by laboratories around the world have uncovered the complete nucleotide sequences of many different genomes, including that of our own species and the chimpanzee, our closest living relative. For the first time in human history, we have the means to reconstruct the genetic path of human evolution by comparing corresponding regions of the genome in related organisms. We can learn which regions of our genome have been duplicated and which have been lost since our split

Model of DNA built by James Watson and Francis Crick at Cambridge University, 1953. Inset shows a photograph taken by Rosalind Franklin of the X-ray diffraction pattern of a DNA fiber that suggested the helical nature of DNA. (COURTESY SCIENCE & SOCIETY PICTURE LIBRARY, SCIENCE MUSEUM, LONDON; INSET: REPRINTED WITH PERMISSION FROM R. E. FRANKLIN AND R. G. GOSLING, NATURE 171:740, 1953 © COPYRIGHT 1953; BY MACMILLAN MAGAZINES LIMITED.)
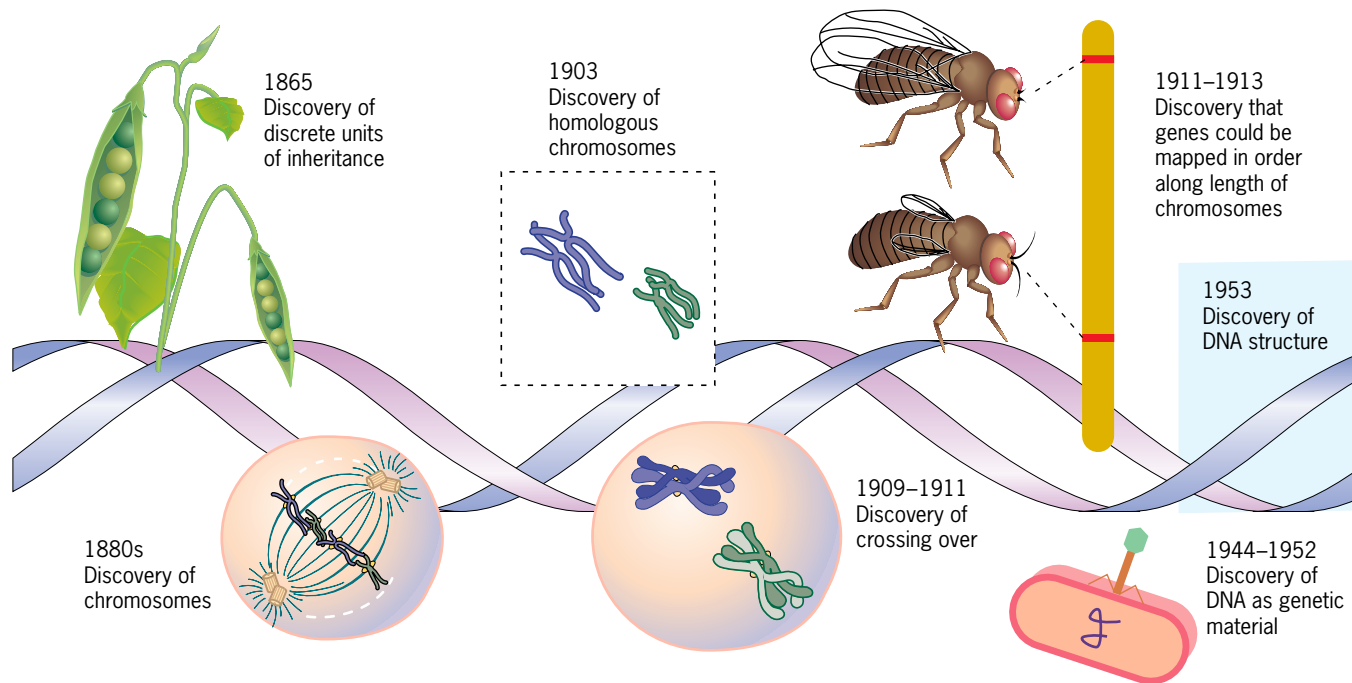
FIGURE 10.1 **An overview depicting several of the most important early discoveries on the nature of the gene.** Each of these discoveries is discussed in the present chapter.

from a common ancestor; we can observe which nucleotides in a particular gene or regulatory region have undergone change and which have remained constant; most important, we can infer which parts of our genome have been subject to natural selection and which have been free to drift randomly as time has passed. We have also begun to use this information to learn more about our history as a species: when and where we arose, how we are related to one another, and how we came to occupy the regions of the Earth that we have. ■

10.1 THE CONCEPT OF A GENE AS A UNIT OF INHERITANCE

The science of genetics began in the 1860s with the work of Gregor Mendel, a friar at the abbey of St. Thomas located in today's Czech Republic. Mendel's laboratory was a small garden plot on the grounds of the abbey. We don't know exactly what motivated Mendel to begin his studies, but he evidently had a clear experimental plan in mind: his goal was to mate, or *cross*, pea plants having different inheritable characteristics and to determine the pattern by which these characteristics were transmitted to the offspring. Mendel chose the garden pea for a number of practical reasons, not the least of which was that he could purchase a variety of seeds that would give rise to plants with distinct characteristics. Mendel chose to focus on seven clearly definable traits, including plant height and flower color, each of which occurred in two alternate and clearly identifiable forms (Table 10.1). Mendel crossbred plants through several generations and counted the number of

individuals having various characteristics. After several years of research, Mendel drew the following conclusions, which are expressed below in modern genetic terminology.

1. The characteristics of the plants were governed by distinct factors (or units) of inheritance, which were later termed **genes**. An individual plant possessed two copies of a gene that controlled the development of each trait, one derived from each parent. The two copies could be either identical to one another or nonidentical. Alternate forms of a gene are called **alleles**. For each of the seven traits studied, one of the two alleles was dominant over the other. When both were present together in the same plant, the existence of the recessive allele was masked by the dominant one.

2. Each reproductive cell (or *gamete*) produced by a plant contained only one copy of a gene for each trait. A particular gamete could have either the recessive or the domi-

TABLE 10.1 **Seven Traits of Mendel's Pea Plants**

Trait	Dominant allele	Recessive allele
Height	Tall	Dwarf
Seed color	Yellow	Green
Seed shape	Round	Angular (wrinkled)
Flower color	Purple	White
Flower position	Along stem	At stem tips
Pod color	Green	Yellow
Pod shape	Inflated	Constricted

(see www.mendelweb.org for a discussion of Mendel's work.)

nant allele for a given trait, but not both. Each plant arose by the union of a male and female gamete. Consequently, one of the alleles that governed each trait in a plant was inherited from the female parent, and the other allele was inherited from the male parent.

3. Even though the pair of alleles that governed a trait remained together throughout the life of an individual plant, they became separated (or *segregated*) from one another during the formation of the gametes. This finding formed the basis of Mendel's law of segregation.

4. The segregation of the pair of alleles for one trait had no effect on the segregation of alleles for another trait. A particular gamete, for example, could receive a paternal gene governing seed color and a maternal gene governing seed shape. This finding formed the basis of Mendel's law of independent assortment.

Mendel presented the results of his work to the members of the Natural History Society of Brünn; the minutes of the meeting record no discussion of his presentation. Mendel's experiments were published in the Brünn society's journal in 1866, where they generated no interest whatsoever until 1900, 16 years after his death. In that year, three European botanists *independently* reached the same conclusions, and all three rediscovered Mendel's paper, which had been sitting on the shelves of numerous libraries throughout Europe for 35 years.

10.2 CHROMOSOMES: THE PHYSICAL CARRIERS OF THE GENES

Although Mendel provided convincing evidence that inherited traits were governed by discrete factors, or genes, his studies were totally unconcerned with the physical nature of these elements or their location within the organism. During the time between Mendel's work and its rediscovery, a number of biologists became concerned with this other aspect of heredity—its physical basis within the cell.

The Discovery of Chromosomes

By the 1880s, a number of European biologists were closely following the activities of cells and using rapidly improving light microscopes to observe newly discernible cell structures. None of these scientists was aware of Mendel's work, yet they understood that whatever it was that governed inherited characteristics would have to be passed from cell to cell and from generation to generation. This, by itself, was a crucial realization; all the genetic information needed to build and maintain a complex plant or animal had to fit within the boundaries of a single cell. Observations of dividing cells by the German biologist Walther Flemming in the early 1880s revealed that the cytoplasmic contents were shuttled into one daughter cell or the other as a matter of chance, depending on the plane through which the furrow happened to divide the cell. In contrast, the cell appeared to go to great lengths to divide the

nuclear contents equally between the daughters. During cell division, the material of the nucleus became organized into visible "threads," which were named **chromosomes**, meaning "colored bodies."

At about this time, the process of fertilization was observed, and the roles of the two gametes—the sperm and the egg—were described (Figure 10.2). Even though the sperm is a tiny cell, it was known to be as important genetically as the much larger egg. What was it that these two very different cells had in common? The most apparent feature was the nucleus and its chromosomes. The importance of the chromosomes contributed by the male became evident in a study by the German biologist Theodore Boveri on sea urchin eggs that had been fertilized by two sperm rather than just one, as is normally the case. This condition, referred to as *polyspermy*, is characterized by disruptive cell divisions and the early death of the embryo. Why should the presence of one extra tiny sperm nucleus within the very large egg have such drastic consequences? The second sperm donates an extra set of chromosomes and an extra centriole (page 333). These additional components lead to abnormal cell divisions in the embryo during which daughter cells receive variable numbers of chromosomes. Boveri concluded that the orderly process of normal development is "dependent upon a particular combination of chromosomes; and this can only mean that the individual chromosomes must possess different qualities." This was the first evidence of a *qualitative* difference among chromosomes.

Events occurring after fertilization were followed most closely in the roundworm *Ascaris*, whose few chromosomes are large and were as readily observed in the nineteenth century as in introductory biology laboratories today. In 1883, the

Second polar body
Maternal pronucleus
Paternal pronucleus
(a) (b) (c)
(d) (e) (f)
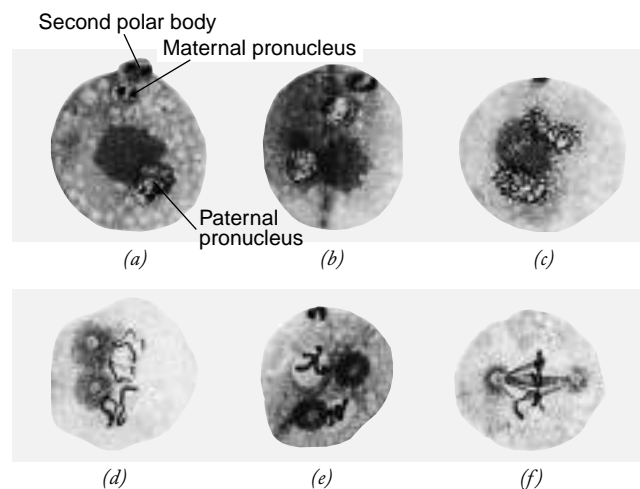
FIGURE 10.2 Events occurring in the roundworm *Ascaris* following fertilization, as reported in a classic nineteenth-century investigation. Both the male and female gamete are seen to contain two chromosomes. Fusion of the sperm and egg nuclei (called pronuclei) in the egg cytoplasm (between *e* and *f*) produces a zygote containing four chromosomes. The second polar body shown in *a* is a product of the previous meiosis as described in Section 14.3. (FROM T. BOVERI, JENAISCHE ZEIT 22:685, 1888.)

Belgian biologist Edouard van Beneden noted that the cells of the worm's body had four large chromosomes, but that the male and female nuclei present in the egg just after fertilization (before the two nuclei fuse with one another) had but two chromosomes apiece (Figure 10.2). About this same time, the process of meiosis was described, and in 1887 the German biologist August Weismann proposed that meiosis included a "reduction division" during which the chromosome number was reduced by half prior to formation of the gametes. If a reduction division did not occur, and each gamete contained the same number of chromosomes as an adult cell, then the union of two gametes would double the number of chromosomes in the cells of the progeny. The number of chromosomes would double with each and every generation, which obviously does not and cannot occur.[1]

Chromosomes as the Carriers of Genetic Information

The rediscovery of Mendel's work and its confirmation had an immediate influence on research in cell biology. Whatever their physical nature, the carriers of the hereditary units would have to behave in a manner consistent with Mendelian principles. In 1903, Walter Sutton, a graduate student at Columbia University, published a paper that pointed directly to the chromosomes as the physical carriers of Mendel's genetic factors. Sutton followed the development of sperm cells in the grasshopper, which like *Ascaris*, has large, easily observable chromosomes. The cells that give rise to sperm are called spermatogonia, and they can undergo two very different types of cell division. A spermatogonium can divide by mitosis to produce more spermatogonia, or it can divide by meiosis to produce cells that differentiate into sperm (see Figure 14.41). On examination of the mitotic stages of grasshopper spermatogonia, Sutton counted 23 chromosomes. Careful examination of the shapes and sizes of the 23 chromosomes suggested that they were present in "look-alike" pairs. Eleven pairs of chromosomes could be distinguished together with one additional chromosome, termed an *accessory chromosome* (and later shown to be a sex-determining X chromosome), that was not part of a pair. Sutton realized that the presence in each cell of pairs of chromosomes, or **homologous chromosomes**, as they were soon called, correlated perfectly with the pairs of inheritable factors uncovered by Mendel.

When Sutton examined the chromosomes in cells just beginning meiosis, he found that the members of each pair of chromosomes were associated with one another, forming a complex called a *bivalent*. Eleven bivalents were visible, each showing a cleft along its length where the two homologous chromosomes were associated (Figure 10.3). The ensuing first meiotic division separated the two homologous chromosomes into separate cells. This was the reduction division proposed 15 years earlier by Weismann on theoretical grounds. Here

[1]Readers who have forgotten the events that occur during meiosis might want to read through Section 14.3, which discusses this pivotal event in the life cycle of eukaryotic organisms.
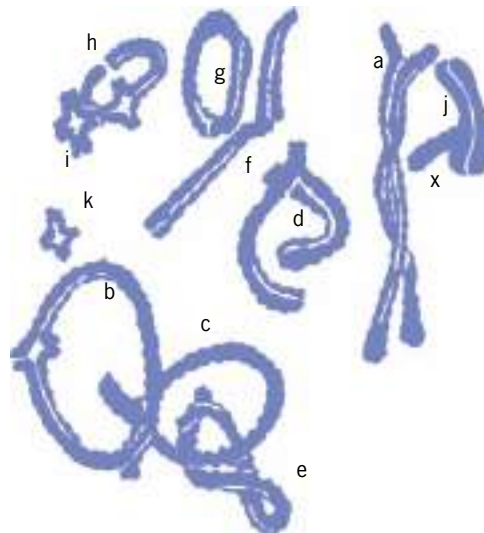
FIGURE 10.3 Homologous chromosomes. Sutton's drawing of the homologous chromosomes of the male grasshopper that have associated during meiotic prophase to form bivalents. Eleven pairs of homologous chromosomes (*a–k*) and an unpaired X chromosome were observed. (FROM W. S. SUTTON, BIOL. BULL. 4:24, 1902.)

also was the physical basis for Mendel's proposal that hereditary factors exist in pairs that remain together throughout the life of an individual and then separate from one another upon formation of the gametes. The reduction division observed by Sutton explained several of Mendel's other findings: gametes could contain only one version (allele) of each gene; the number of gametes containing one allele was equal to the number containing the other allele; and the union of two gametes at fertilization would produce an individual with two alleles for each trait. But many questions remained unanswered. For instance, how were the genes organized within the chromosomes, and could the location of specific genes be determined?

The Chromosome as a Linkage Group As clearly as Sutton saw the relationship between chromosome behavior and Mendel's alleles, he saw one glaring problem. Mendel had examined the inheritance of seven traits and found that each was inherited independently of the others. This observation formed the basis of Mendel's law of independent assortment. But if genes are packaged together on chromosomes, like beads on a string, then genes should be passed from parent to offspring in packages, just as intact chromosomes are passed to the next generation. Genes on the same chromosome should act as if they are *linked* to one another; that is, they should be part of the same **linkage group**.

How is it that all seven of Mendel's traits assorted independently? Were they all on different linkage groups, that is, different chromosomes? As it happens, the garden pea has seven pairs of homologous chromosomes. The genes that govern each of Mendel's traits either occur on different chromosomes or are so far apart from one another on the same chromosome that they act independently (page 385). Sutton's

prophecy of linkage groups was soon confirmed. Within a couple of years, genes for two traits in sweet peas (flower color and pollen shape) were shown to be linked; other evidence for chromosomal linkage rapidly accumulated.

Genetic Analysis in *Drosophila*

Research in genetics soon came to focus on one particular organism, the fruit fly, *Drosophila* (Figure 10.4). Fruit flies have a generation time (from egg to sexually mature adult) of about 10 days and can produce up to 1000 eggs in a lifetime. In addition, fruit flies are very small, so a large number could be kept on hand; they were easy to house and breed; and they were very inexpensive to maintain. In 1909, fruit flies seemed like the ideal organism to Thomas Hunt Morgan of Columbia University, and he began what was to be the start of a new era in genetic research. There was one major disadvantage in beginning work with this insect—there was only one "strain" of fly available, the **wild type**. Whereas Mendel had simply purchased varieties of pea seeds, Morgan had to generate his own varieties of fruit flies. Morgan expected that variants from the wild type would appear if he bred sufficient numbers of flies. Within a year—after observing thousands of flies—he found his first **mutant**, that is, an individual having an inheritable characteristic that distinguished it from the wild type. The mutant had white eyes rather than the normal red-colored ones (Figure 10.4).

By 1915, Morgan and his students had found 85 different mutants with a wide variety of affected structures. It was apparent that on rare occasions a spontaneous change, or **mutation**, occurs within a gene, altering it permanently, so that the altered trait is passed from generation to generation. The demonstration of a spontaneous, inheritable alteration in a gene had consequences far beyond the study of *Drosophila*

genetics. It suggested a mechanism for the origin of variation that exists within populations—evidence for a vital link in the theory of evolution. If variants of genes could arise spontaneously, then isolated populations could become genetically different from one another and ultimately give rise to new species.

Whereas mutations are a necessary occurrence for evolution, they are a tool for geneticists, because they provide contrast to the wild-type condition. As *Drosophila* mutants were isolated, they were bred and crossbred and kept as stocks within the lab. As expected, the 85 mutations did not all assort independently; instead, Morgan found that they belonged to four different linkage groups, one of which contained very few mutant genes (only two in 1915). This finding correlated perfectly with the observation that *Drosophila* cells have four pairs of homologous chromosomes, one of which is very small (Figure 10.5). Little doubt remained that genes reside on chromosomes.

Crossing Over and Recombination

Even though the association of genes into linkage groups was confirmed, linkage between alleles on the same chromosome was found to be *incomplete*. In other words, alleles of two different genes, such as short wings and black body (as in Figure 10.7), that were originally present together on a given chromosome did not always remain together during the production of gametes. Thus maternal and paternal characteristics that were inherited by an individual on separate homologous chromosomes could be reshuffled so that they ended up on the same chromosome of a gamete. Conversely, two characteristics that were inherited together on the same chromosome could become separated from one another and end up in separate gametes.

In 1911, Morgan offered an explanation for the "breakdown" in linkage. Two years earlier, F. A. Janssens had observed that the homologous chromosomes of bivalents became

FIGURE 10.4 **The fruit fly *Drosophila melanogaster*.** Photograph of a wild-type female fruit fly and a mutant male that carries a mutation leading to white eyes. (COURTESY OF STANLEY J. P. IYADURAI.)
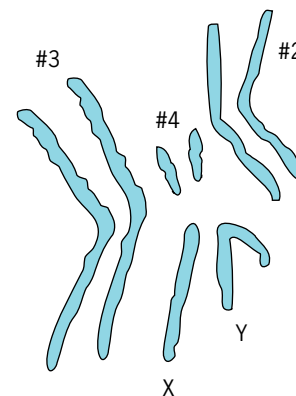
FIGURE 10.5 **Fruit flies have four pairs of homologous chromosomes, one of which is very small.** The two dissimilar homologues are the chromosomes that determine sex. As in humans, male fruit flies are XY and females are XX.
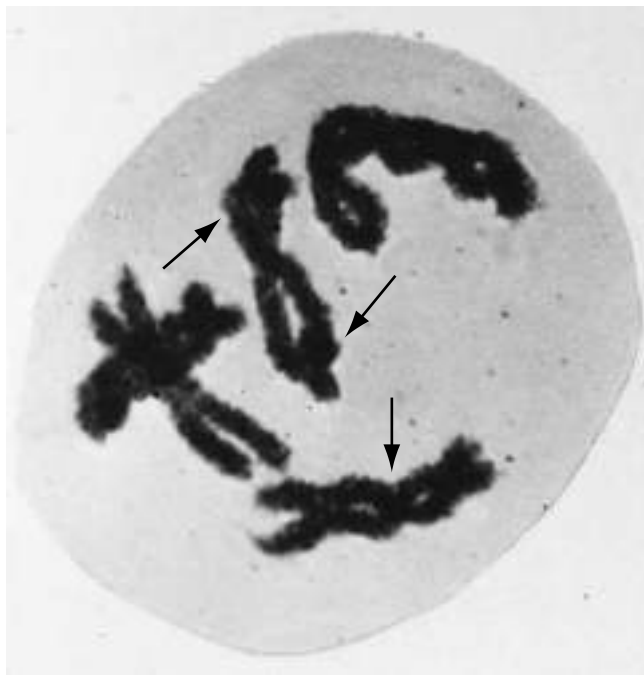
FIGURE 10.6 Visualizing sites of crossing over. Homologous chromosomes wrap around each other during meiosis, as seen in this micrograph of a meiotic cell of a lily. The points at which the homologues are crossed are termed *chiasmata* (arrows) and (as discussed in Chapter 14) are sites at which crossing over had occurred at an earlier stage. (COURTESY OF A. H. SPARROW.)

(a)

Homologous chromosome
pair in tetrad formation

(b)

FIGURE 10.7 Crossing over provides the mechanism for reshuffling alleles between maternal and paternal chromosomes. (*a*) Simplified representation of a single crossover in a *Drosophila* heterozygote (*BbWw*) at chromosome number 2 and the resulting gametes. If either of the crossover gametes participates in fertilization, the offspring will have a chromosome with a combination of alleles that was not present on a single chromosome in the cells of either parent. (*b*) Bivalent (tetrad) formation during meiosis, showing three crossover intersections (chiasmata, indicated by red arrows).

wrapped around each other during meiosis (Figure 10.6). Janssens had proposed that this interaction between maternal and paternal chromosomes resulted in the breakage and exchange of pieces. Capitalizing on this proposal, Morgan suggested that this phenomenon, which he termed **crossing over** (or **genetic recombination**), could account for the appearance of offspring (*recombinants*) having unexpected combinations of genetic traits. An example of crossing over is shown in Figure 10.7.

Analyses of offspring from a large number of crosses between adults carrying a variety of alleles on the same chromosome indicated that (1) the percentage of recombination between a given pair of genes on a chromosome, such as eye color and wing length, was essentially constant from one experiment to another, and (2) the percentages of recombination between different pairs of genes, such as between eye color and wing length versus eye color and body color, could be very different.

The fact that a given pair of genes gave the same approximate frequency of recombination in every cross strongly suggested that the positions of the genes along the chromosome (their **loci**) were fixed and did not vary from one fly to the next. If the locus of each gene is fixed, then the frequency of recombination between two genes provides a measure of the distance that separates those two genes. The more room between two sites on a chromosome for breakage to occur, the more likely it is that a break would occur between those two sites and, therefore, the greater the recom-

bination frequency. In 1911, Alfred Sturtevant, an undergraduate at Columbia University who was working in Morgan's laboratory, conceived the idea that recombination frequencies could be used to map the relative positions of individual genes along a given chromosome. An example of

the underlying principles of this mapping procedure is illustrated in Figure 10.7. In this example, the genes for wing length and body color in *Drosophila* are situated at a considerable distance from one another on the chromosome and, therefore, are likely to become unlinked by an intervening breakage and crossover. In contrast, the genes for eye color and body color are situated very close to one another on the chromosome and consequently are less likely to become unlinked. Using recombination frequencies, Sturtevant—who became one of the century's most prominent geneticists—began to construct detailed maps of the serial order of genes along each of the four chromosomes of the fruit fly. Recombination frequencies have since been used to prepare chromosomal maps from viruses and bacteria as well as a large variety of eukaryotic species.

Mutagenesis and Giant Chromosomes

During the early period of genetics, the search for mutants was a slow, tedious procedure dependent on the *spontaneous* appearance of altered genes. In 1927, using a special strain of fruit flies designed to reveal the presence of recessive alleles, H. J. Muller found that flies subjected to a sublethal dose of X-rays displayed more than 100 times the spontaneous mutation rate exhibited by nonirradiated controls. This finding had important consequences. On the practical side, the use of mutagenic agents, such as X-rays and ultraviolet radiation, greatly increased the number of mutants available for genetic research. The finding also pointed out the hazard of the increasing use of radiation in the industrial and medical fields. Today, mutations in *Drosophila* are most often generated by adding a chemical mutagen (ethyl methane sulfonate) to the animals' feed.

The rediscovery in 1933 of giant chromosomes in certain insect cells by Theophilus Painter of the University of Texas illustrates a basic principle in biology. There is such remarkable variability among organisms, not only at the obvious macroscopic levels, but also at the cellular and subcellular levels, that often one particular type of cell may be much better suited for a particular type of research than others. Cells from the larval salivary gland of *Drosophila* contain chromosomes that are about 100 times thicker than the chromosomes found in most other cells of the organism (Figure 10.8*a*). During larval development, these cells cease dividing, but they keep growing. DNA replication continues, providing the additional genetic material needed to support the high level of secretory activity of these enormous cells. The duplicated DNA strands remain attached to each other in perfect side-by-side alignment (inset of Figure 10.8*a*), producing giant chromosomes with as many as 1024 times the number of DNA strands of normal chromosomes.

These unusual **polytene chromosomes**, as they are called, are rich in visual detail, displaying about 5000 bands when stained and examined microscopically. The banding pattern is essentially constant from one individual to the next, but considerable differences are observed between the chromosomes from flies of different species of the *Drosophila* genus. Painter soon found that individual bands could be correlated with
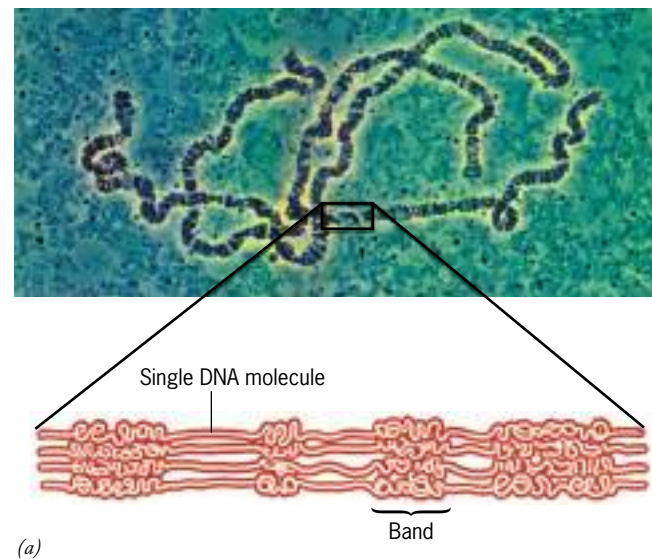
(a)
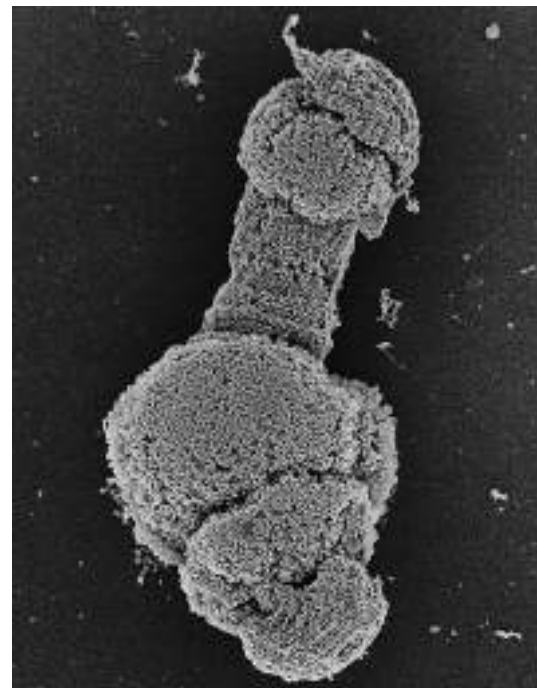
(b)

FIGURE 10.8 Giant polytene chromosomes of larval insects. (*a*) These giant polytene chromosomes from the salivary gland of a larval fruit fly show several thousand distinct, darkly staining bands. The bands have been identified as the loci of particular genes. The inset shows how polytene chromosomes consist of a number of individual DNA molecules. The stained bands on the chromosomes correspond to sites where the DNA is more tightly compacted. (*b*) Scanning electron micrograph of a giant polytene chromosome from a *Chironomus* larva showing how specific sites are expanded to form a "puff." Chromosome puffs are sites where DNA is being very actively transcribed. (A: FROM BIOLOGICAL PHOTO SERVICE; B: COURTESY OF TERRY D. ALLEN AND CLAUS PELLING, J. CELL SCIENCE, COVER OF VOL. 93, PART 4, 1989.)

specific genes. The relative positions of these genes on the giant chromosomes agreed with those predicted on the basis of genetic maps prepared from recombination frequencies, thus

providing visual confirmation of the validity of the entire mapping procedure.

Giant insect chromosomes have been useful in other ways. Comparisons of banding patterns among polytene chromosomes of different species have provided an unparalleled opportunity for the investigation of evolutionary changes at the chromosome level. In addition, these chromosomes are not inert cellular objects but dynamic structures in which certain regions become "puffed out" at particular stages of development (Figure 10.8*b*). These chromosome puffs are sites where DNA is being transcribed at a very high level, providing one of the best systems available for the direct visualization of gene expression (see Figure 18.21*a*).

REVIEW ?

1. What is a linkage group? What is its relationship to a chromosome? How can one determine the number of linkage groups in a species?
2. How do genetic mutations help in mapping the location of genes on a chromosome?
3. What is meant by incomplete linkage? What does this have to do with pairing of homologous chromosomes during meiosis?
4. How do polytene chromosomes of an insect differ from normal chromosomes?

10.3 THE CHEMICAL NATURE OF THE GENE

Classical geneticists uncovered the rules governing the transmission of genetic characteristics and the relationship between genes and chromosomes. In his acceptance speech for the Nobel Prize in 1934, T. H. Morgan stated, "At the level at which the genetic experiments lie, it does not make the slightest difference whether the gene is a hypothetical unit or whether the gene is a material particle." By the 1940s, however, a new set of questions was being considered, foremost of which was, "What is the chemical nature of the gene?" The experiments answering this question are outlined in the Experimental Pathways at the end of this chapter. Once it became evident that genes were made of DNA, biologists were faced with a host of new questions. These questions will occupy us for the remainder of the chapter.

The Structure of DNA

To understand the workings of a complex macromolecule—whether it is a protein, polysaccharide, lipid, or nucleic acid—it is essential to know how that molecule is constructed. The mystery of DNA structure was investigated by a number of laboratories in both the United States and England in the early 1950s and was solved by James Watson and Francis Crick at Cambridge University in 1953. Before describing their proposed DNA structure, let us consider the facts available at the time.

Base Composition   The basic building block of DNA was known to be a **nucleotide** (Figure 10.9*a,b*) consisting of the five-carbon sugar *deoxyribose* to which one phosphate is esterified at the 5′ position of the sugar ring and one nitrogenous base is attached at the 1′ site.[2] Two types of nitrogenous bases are present in a nucleic acid: **pyrimidines**, which contain a single ring, and **purines**, which contain two rings (Figure 10.9*c*). DNA contains two different pyrimidines, *thymine* (*T*) and *cytosine* (*C*), and two different purines, *guanine* (*G*) and *adenine* (*A*). The nucleotides are covalently linked to one another to form a linear polymer, or *strand*, with a backbone composed of alternating sugar and phosphate groups joined by *3′–5′-phosphodiester bonds* (Figure 10.9*c*). The bases attached to each sugar were thought to project from the backbone like a column of stacked shelves.

A nucleotide has a polarized structure: one end, where the phosphate is located, is called the *5′ end* (pronounced "five-prime end"), while the other end is the *3′ end* (Figure 10.9*b*). Because each of the stacked nucleotides in a strand faces the same direction, the entire strand has a direction. One end of the strand is the 3′ end, the other is the 5′ end (Figure 10.9*c*). X-ray diffraction analysis indicated that the distance between the nucleotides of the stack was 3.4 Å (0.34 nm) and suggested the presence of a large structural repeat every 3.4 nm.

As discussed on page 413, DNA was thought for many years to consist of a simple repeating tetranucleotide (e.g., —ATGCATGCATGC—), which made it unlikely to serve as an informational macromolecule. In 1950, Erwin Chargaff of Columbia University reported an important finding that delivered the final blow to the tetranucleotide theory and provided vital information about DNA structure. Chargaff, believing that the sequence of nucleotides of the DNA molecule held the key to its importance, determined the relative amounts of each base in various samples of DNA, that is, the **base composition** of the samples. Base composition analysis of a DNA sample was performed by hydrolyzing the bases from their attached sugars, separating the bases in the hydrolysate by paper chromatography, and determining the amount of material in each of the four spots to which the bases migrated.

If the tetranucleotide theory were correct, each of the four bases in a DNA sample would be present as 25 percent of the total number. Chargaff found that the ratios of the four component bases were quite variable from one type of organism to another, often being very different from the 1:1:1:1 ratio

---

[2]It is useful to introduce a bit of terminology at this point. A molecule consisting simply of one of the four nitrogenous bases of Figure 10.9 linked to a pentose sugar moiety is known as a nucleoside. If the sugar is deoxyribose, the nucleoside is a deoxyribonucleoside. There are four major deoxyribonucleosides distinguished by the attached base: deoxyadenosine, deoxyguanosine, deoxythymidine, and deoxycytidine. If the nucleoside has one or more attached phosphate groups (generally at the 5′ position, but alternatively at the 3′ position), the molecule is a nucleotide. There are nucleoside 5′-monophosphates, nucleoside 5′-diphosphates, and nucleoside 5′-triphosphates, depending on the number of phosphates in the molecule. Examples of each are deoxyadenosine 5′-monophosphate (dAMP), deoxyguanosine 5′-diphosphate (dGDP), and deoxycytidine 5′-triphosphate (dCTP). A similar set of nucleosides and nucleotides involved in RNA metabolism contain the sugar ribose rather than deoxyribose. The nucleotides employed in energy metabolism, such as adenosine triphosphate (ATP), are ribose-containing molecules.

(a)

(b)
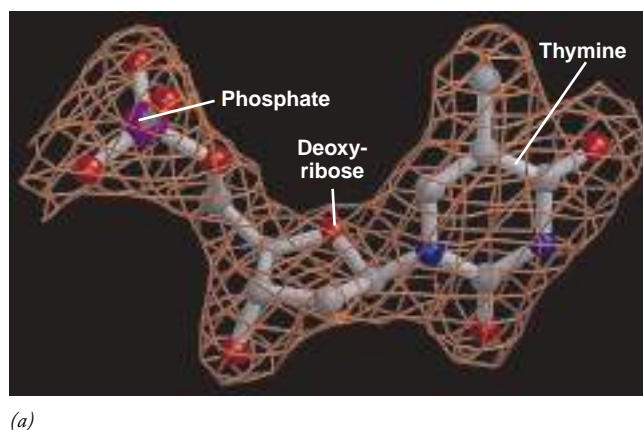
FIGURE 10.9 The chemical structure of DNA. (*a*) Model of a DNA nucleotide containing the base thymine; the molecule is deoxythymidine 5′-monophosphate (dTMP). The netlike cage represents the electron density of the atoms that make up the molecule. (*b*) Chemical structure of a DNA nucleotide containing the base adenosine; the molecule is deoxyadenosine 5′-monophosphate (dAMP). A nucleotide is composed of a nucleoside linked to a phosphate; the nucleoside portion of the molecule (i.e., deoxyadenosine) is enclosed by the dashed line. (*c*) The chemical structure of a small segment of a single DNA strand showing all four nucleotides. (A: REPRINTED WITH PERMISSION FROM ARNON LAVIE ET AL., NATURE STR. BIOL. 4:604, 1997; © COPYRIGHT 1997 BY MACMILLAN MAGAZINES LIMITED.)

(c)

predicted by the tetranucleotide theory. For example, the A:G ratio of the DNA of a tubercle bacillus was 0.4, whereas the A:G ratio of human DNA was 1.56. It made no difference which plant or animal tissue was used as the source of the DNA; the base composition remained constant for that species. Amid this great variability in base composition of different species' DNA, an important numerical relationship was discovered. The number of purines always equaled the number of pyrimidines in a given sample of DNA. More specifically, the number of adenines always equaled the number of thymines, and the number of guanines always equaled the number of cytosines. In other words, Chargaff discovered the following rules of DNA base composition:

$$[A] = [T], [G] = [C], [A] + [T] \neq [G] + [C]$$

Chargaff's findings put the DNA molecule in a new light, giving it specificity and individuality from one organism to another. The significance of the base equivalencies, however, remained obscure.

The Watson-Crick Proposal

When protein structure was discussed in Chapter 2, the importance of secondary and tertiary structure as determinants of the protein's activity was stressed. Similarly, information about the three-dimensional organization of DNA was needed if its biological activity was to be understood. Using X-ray diffraction data (obtained by Rosalind Franklin and Maurice Wilkins at King's College London) and models constructed from cutouts of the four types of nucleotides, Watson
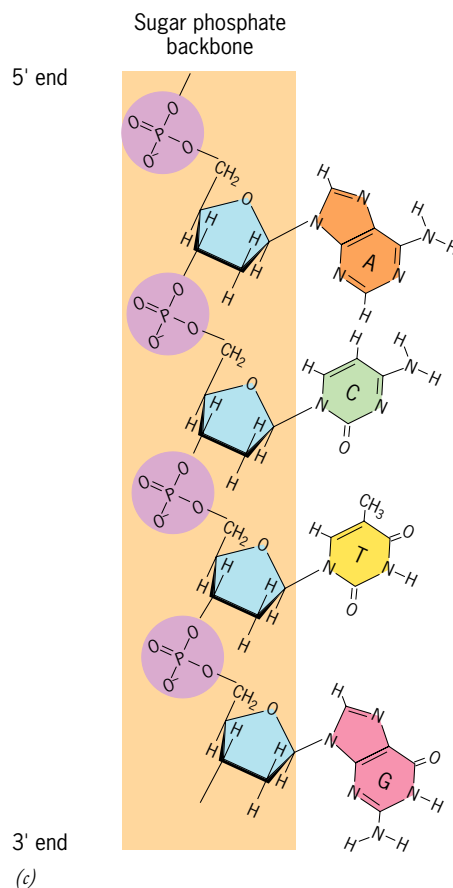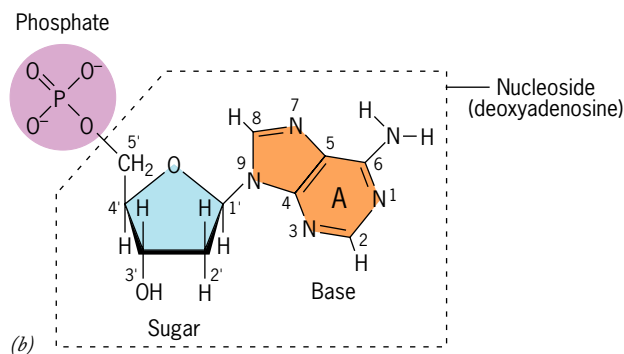
and Crick proposed a structure of DNA that included the following elements (Figure 10.10):

1. The molecule is composed of two chains of nucleotides. This conclusion followed on the heels of an erroneous proposal by Linus Pauling, who had suggested that DNA was composed of three nucleotide strands.

2. The two chains spiral around each other to form a pair of right-handed helices. In a right-handed helix, an observer looking down the central axis of the molecule would see that each strand follows a clockwise path, as it moves away from the observer. The helical nature of DNA was revealed in the pattern of spots produced by Franklin's X-ray diffraction image (seen on page 379), which was shown to Watson during a visit to King's College.
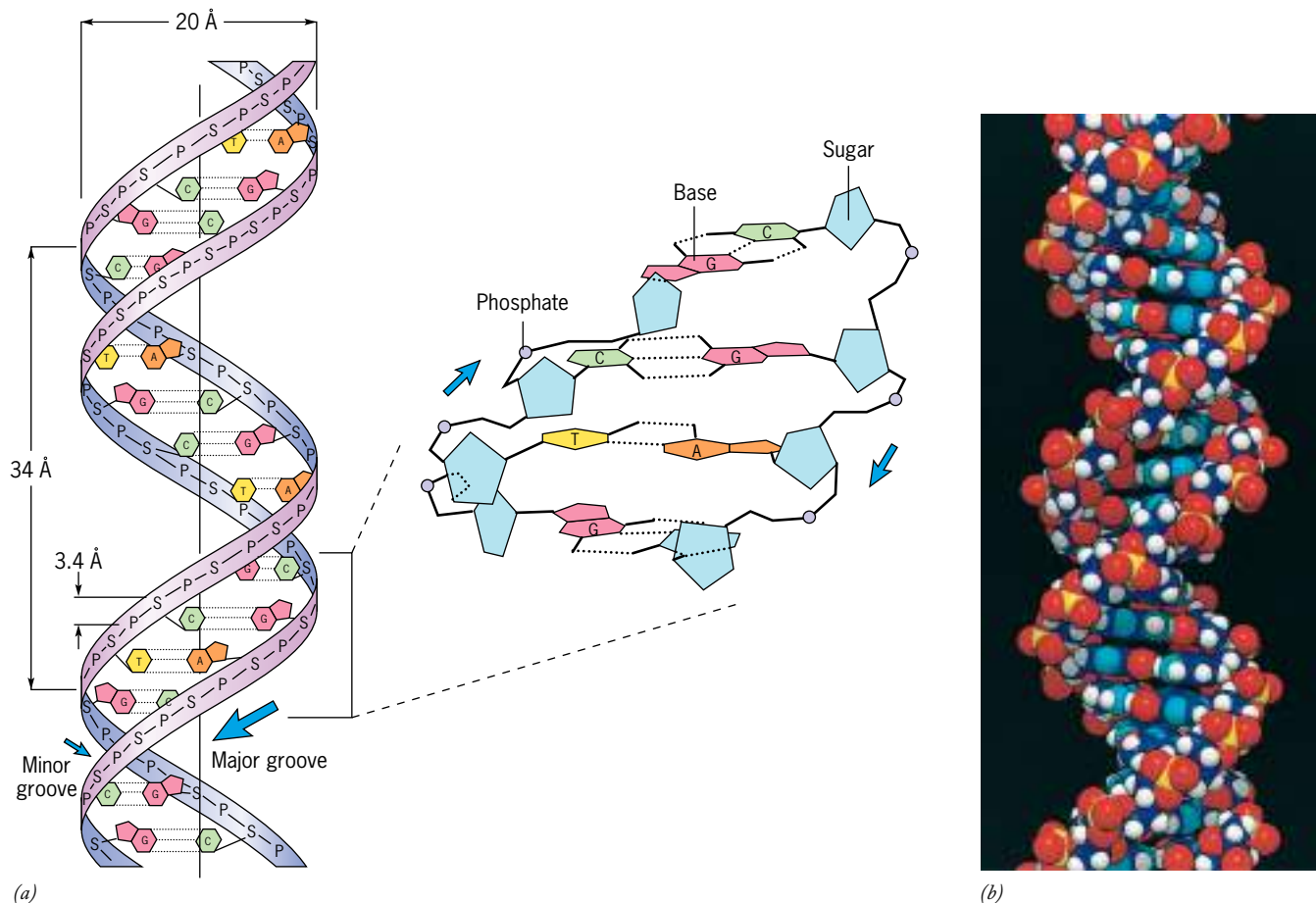
FIGURE 10.10 The double helix. (*a*) Schematic representation of the DNA double helix. (*b*) Space-filling model of the B form of DNA.

continued

3. The two chains comprising one double helix run in opposite directions; that is, they are *antiparallel*. Thus, if one chain is aligned in the $5' \rightarrow 3'$ direction, its partner must be aligned in the $3' \rightarrow 5'$ direction.

4. The –sugar–phosphate–sugar–phosphate– backbone of each strand is located on the outside of the molecule with the two sets of bases projecting toward the center. The phosphate groups give the molecule a large negative charge.

5. The bases occupy planes that are approximately perpendicular to the long axis of the molecule and are, therefore, stacked one on top of another like a pile of plates. Hydrophobic interactions and van der Waals forces (page 36) between the stacked, planar bases provide stability for the entire DNA molecule. Together, the helical turns and planar base pairs cause the molecule to resemble a spiral staircase. This manner of construction is evident in the chapter-opening photograph, which shows the original Watson-Crick model.

6. The two strands are held together by hydrogen bonds between each base of one strand and an associated base on the other strand. Because individual hydrogen bonds are weak and easily broken, the DNA strands can become separated during various activities. But the strengths of hydrogen bonds are additive, and the large numbers of hydrogen bonds holding the strands together make the double helix a stable structure.

7. The distance from the phosphorus atom of the backbone to the center of the axis is 1 nm (thus the width of the double helix is 2 nm).

8. A pyrimidine in one chain is always paired with a purine in the other chain. This arrangement produces a molecule that is 2 nm wide along its entire length.

9. The nitrogen atoms linked to carbon 4 of cytosine and carbon 6 of adenine are predominantly in the amino (NH_2) configuration (Figure 10.9c) rather than the imino (NH) form. Similarly, the oxygen atoms linked to carbon 6 of guanine and carbon 4 of thymine are predominantly in a keto (C=O) configuration rather than the enol (C—OH) form. These structural restrictions on the configurations of the bases suggested that adenine was the only purine structurally capable of bonding to thymine and that guanine was the only purine capable of bonding to cytosine. Therefore, the only possible pairs

(c)

(d)
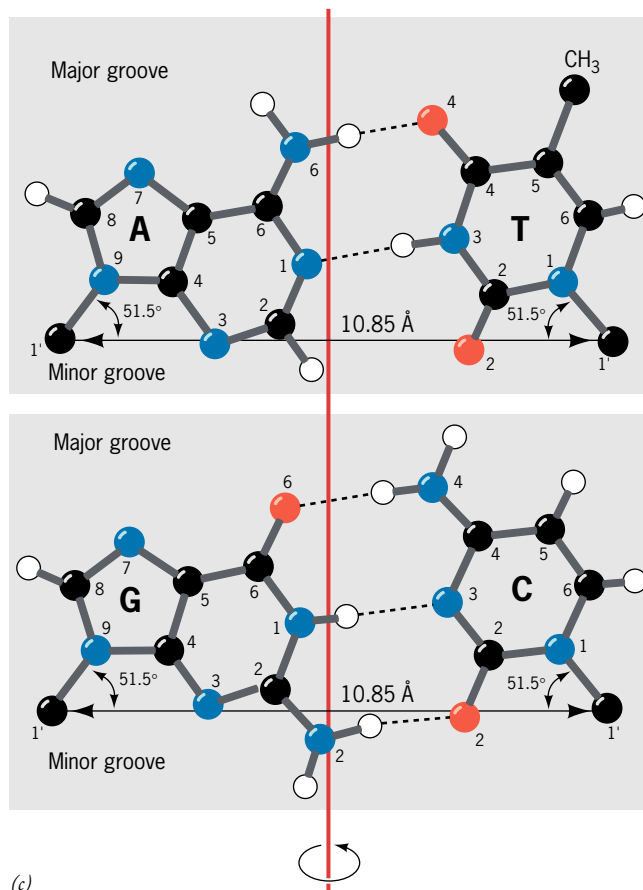
FIGURE 10.10 **The double helix.** (*continued*) (*c*) The Watson-Crick base pairs. The original model showed both A-T and G-C pairs with two hydrogen bonds; the third hydrogen bond in the G-C pair was subsequently identified by Linus Pauling. (*d*) Electron micrograph of DNA being released from the head of a T2 bacteriophage. This linear DNA molecule (note the two free ends) measures 68 μm in length, approximately 60 times longer than the phage head in which it is contained. (c: FROM D. VOET AND J. G. VOET, BIOCHEMISTRY, 2D ED.; COPYRIGHT © 1995, JOHN WILEY & SONS, INC. REPRINTED BY PERMISSION. D: COURTESY OF A. K. KLEINSCHMIDT ET AL., BIOCHIM. BIOPHYS. ACTA 61:861, 1962.)
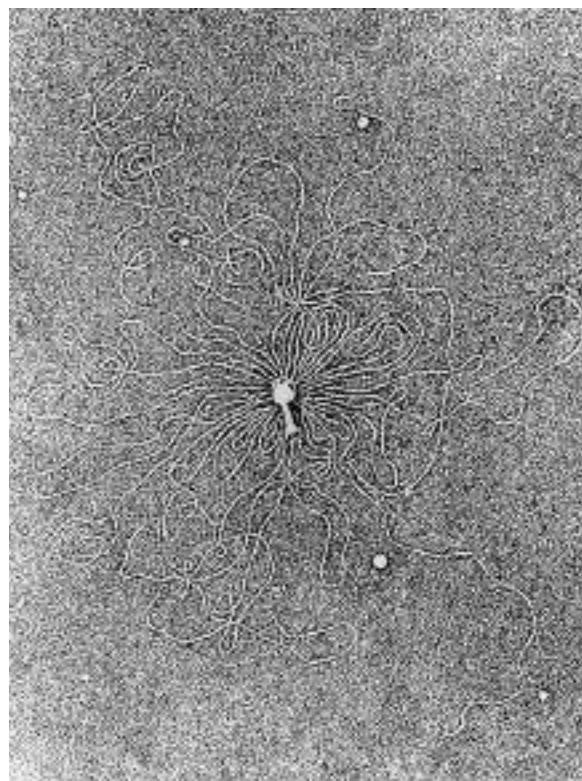
were A-T and G-C (Figure 10.10c). This fit perfectly with the base composition analysis carried out by Chargaff whose results were communicated to Watson and Crick during a meeting of the three scientists in Cambridge in 1952. Because an A-T and G-C base pair had the same geometry (Figure 10.10c), there were no restrictions on the sequence of bases; a DNA molecule could have any one of an unlimited variety of nucleotide sequences.

10. The spaces between adjacent turns of the helix form two grooves of different width—a wider *major groove* and a more narrow *minor groove*—that spiral around the outer surface of the double helix. Proteins that bind to DNA often contain domains that fit into these grooves. In many cases, a protein bound in a groove is able to read the sequence of nucleotides along the DNA without having to separate the strands.

11. The double helix makes one complete turn every 10 residues (3.4 nm), or 150 turns per million daltons of molecular mass.

12. Because an A on one strand is always bonded to a T on the other strand, and a G is always bonded to a C, the nucleotide sequences of the two strands are always fixed relative to one another. Because of this relationship, the two chains of the double helix are said to be **complementary** to one another. For example, A is complementary to T, 5′-AGC-3′ is complementary to 3′-TCG-5′, and one entire chain is complementary to the other. As we shall see, complementarity is of overriding importance in nearly all the activities and mechanisms in which nucleic acids are involved.

The Importance of the Watson-Crick Proposal   From the time biologists first considered DNA as the genetic material, it was expected to fulfill three primary functions (Figure 10.11):

1. *Storage of genetic information.* As genetic material, DNA must contain a stored record of instructions that determine all the inheritable characteristics that an organism exhibits. In molecular terms, DNA must contain the in-
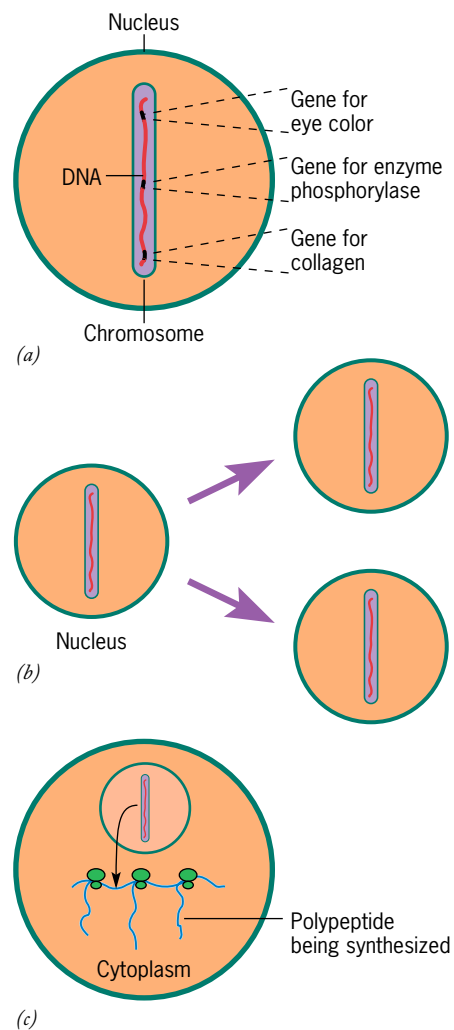
FIGURE 10.11 **Three functions required of the genetic material**. (*a*) DNA must contain the information that encodes inheritable traits. (*b*) DNA must contain the information that directs its own duplication. (*c*) DNA must contain the information that directs the assembly of specific proteins.

formation for the specific order of amino acids in all the proteins that are synthesized by an organism.

2. *Replication and inheritance.* DNA must contain the information for synthesis of new DNA strands (replication). DNA replication allows genetic instructions to be transmitted from one cell to its daughter cells and from one individual to its offspring.

3. *Expression of the genetic message.* DNA is more than a storage center; it is also a director of cellular activity. Consequently, the information encoded in DNA has to be expressed in some form that can take part in events that are taking place within the cell. More specifically, the information in DNA must be used to direct the order by which specific amino acids are incorporated into a polypeptide chain.

The Watson-Crick model of DNA structure suggested a way in which the first two of these three genetic functions

might be accomplished. The model strongly supported the suspicion that the information content of DNA resided in the linear sequence of its bases. A given segment of DNA would correspond to each gene. The specific sequence of nucleotides in that segment would dictate the sequence of amino acids in a corresponding polypeptide. A change in the linear sequence of nucleotides within that segment would correspond to an inheritable mutation in that gene. Differences in nucleotide sequence were presumed to form the basis of genetic variation, whether between individuals of a species or between species themselves.

As for the second function, Watson and Crick's initial publication of DNA structure included a proposal as to how such a molecule might replicate. This may have been the first time that a study of molecular structure led directly to a hypothesis of a basic molecular mechanism. Watson and Crick proposed that during replication, the hydrogen bonds holding the two strands of the DNA helix were sequentially broken, causing the gradual separation of the strands much like the separation of two halves of a zipper. Each of the separated strands, with its nitrogenous bases exposed, would then serve as a *template* directing the order in which complementary nucleotides become assembled to form the complementary strand. When complete, the process would generate two double-stranded DNA molecules that were (1) identical to one another and (2) identical to the original DNA molecule. According to the Watson-Crick proposal, each DNA double helix would contain one strand from the original DNA molecule and one newly synthesized strand (see Figure 13.1). As we will see in Chapter 13, the Watson-Crick proposal fared quite well in predicting the mechanism of DNA replication. Of the three primary functions listed above, only the mechanism by which DNA governs the assembly of a specific protein remained a total mystery.

Not only was the elucidation of DNA structure important in its own right, it provided the stimulus for investigating all the activities in which the genetic material must take part. Once the model for its structure was accepted, any theory of a genetic code, DNA synthesis, or information transfer had to be consistent with that structure.

DNA Supercoiling

In 1963, Jerome Vinograd and his colleagues at the California Institute of Technology discovered that two closed, circular DNA molecules of identical molecular mass could exhibit very different rates of sedimentation during centrifugation (Section 18.10). Further analysis indicated that the DNA molecule sedimenting more rapidly had a more compact shape because the molecule was twisted upon itself (Figure 10.12*a,b*), much like a rubber band in which the two ends are twisted in opposite directions or a tangled telephone cord after extended use. DNA in this state is said to be **supercoiled**. Because supercoiled DNA is more compact than its relaxed counterpart, it occupies a smaller volume and moves more rapidly in response to a centrifugal force or an electric field (Figure 10.12*c*).

(a)   *(b)*   *(c)*
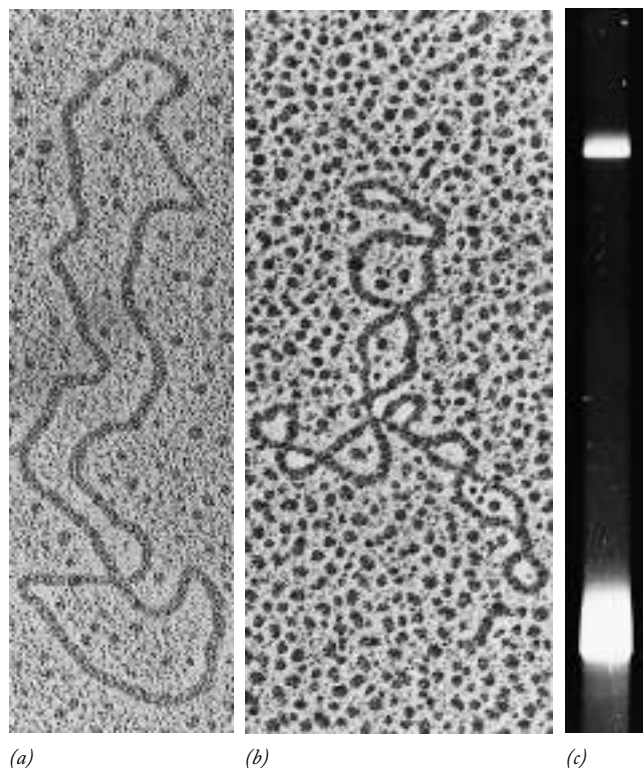
FIGURE 10.12 **Supercoiled DNA.** (*a,b*) Electron micrographs showing the differences in conformation between a relaxed, circular molecule of phage DNA (*a*) and the same type of molecule in a supercoiled state (*b*). (*c*) When a mixture of relaxed and supercoiled SV40 DNA molecules are subjected to gel electrophoresis, the highly compact, supercoiled form (seen at the bottom of the gel) moves much more rapidly than the relaxed form. The DNA molecules are visualized by staining the gel with ethidium bromide, a fluorescent molecule that inserts itself into the double helix. (A,B: COURTESY OF JAMES C. WANG; C: FROM WALTER KELLER, PROC. NAT'L. ACAD. SCI. U.S.A. 72:2553, 1975.)

FIGURE 10.13 **Underwound DNA.** The DNA molecule at the left is underwound; that is, it has more than an average of 10 base pairs per turn of a helix. An underwound molecule spontaneously assumes a negatively supercoiled conformation, as shown on the right.

Supercoiling is best understood if one pictures a length of double-helical DNA lying free on a flat surface. A molecule in this condition has the standard number of 10 base pairs per turn of the helix and is said to be *relaxed*. The DNA would still be relaxed if both ends of the two strands were simply fused to form a circle. Consider, however, what would happen if the molecule were twisted *before* the ends were sealed. If the length of DNA were twisted in a direction opposite to that in which the duplex was wound, the molecule would tend to unwind. An *underwound* DNA molecule has a greater number of base pairs per turn of the helix (Figure 10.13). Because the molecule is most stable with 10 residues per turn, it tends to resist the strain of becoming underwound by becoming twisted upon itself into a supercoiled conformation (Figure 10.13).

DNA is referred to as *negatively supercoiled* when it is underwound and *positively supercoiled* when it is overwound. Circular DNAs found in nature (e.g., mitochondrial, viral, bacterial) are invariably negatively supercoiled. Supercoiling is not restricted to small, circular DNAs but also occurs in linear,

eukaryotic DNA. For example, negative supercoiling plays a key role in allowing the DNA of the chromosomes to be compacted so as to fit inside the confines of a microscopic cell nucleus (Section 12.1). Because negatively supercoiled DNA is underwound, it exerts a force that helps separate the two strands of the helix, which is required during both replication (DNA synthesis) and transcription (RNA synthesis).

Cells rely on enzymes to change the supercoiled state of a DNA duplex. These enzymes are called **topoisomerases** because they change the *topology* of the DNA. Cells contain a variety of topoisomerases, which can be divided into two classes. *Type I topoisomerases* change the supercoiled state of a DNA molecule by creating a transient break in one strand of the duplex. A model for the mechanism of action of human topoisomerase I is shown in Figure 10.14*a*. The enzyme cleaves one strand of the DNA and then allows the intact, complementary strand to undergo a controlled rotation, which relaxes the supercoiled molecule. Topoisomerase I is essential for processes such as DNA replication and transcription. It functions in these activities by preventing excessive supercoiling from building up as the complementary strands of a DNA duplex separate and unwind (as in Figure 13.6). *Type II topoisomerases* make a transient break in both strands of a DNA duplex. Another segment of the DNA molecule (or a separate molecule entirely) is then transported through the break, and the severed strands are resealed. As expected, this complex reaction mechanism is accompanied by a series of dramatic conformational changes that are depicted in the model in Figure 10.14*b*. These enzymes are capable of some remarkable "tricks." In addition to being able to supercoil and relax DNA (Figure 10.14*c*, panel 1), type II topoisomerases can tie a DNA molecule into knots or untie a DNA knot (Figure 10.14*c*, panel 2). They can also cause a population of independent DNA circles to become interlinked (*catenated*), or separate interlinked circles into individual components

Supercoil removal
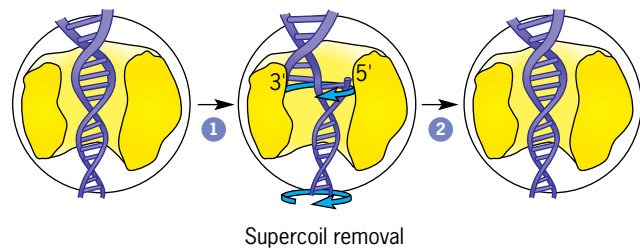
(a)
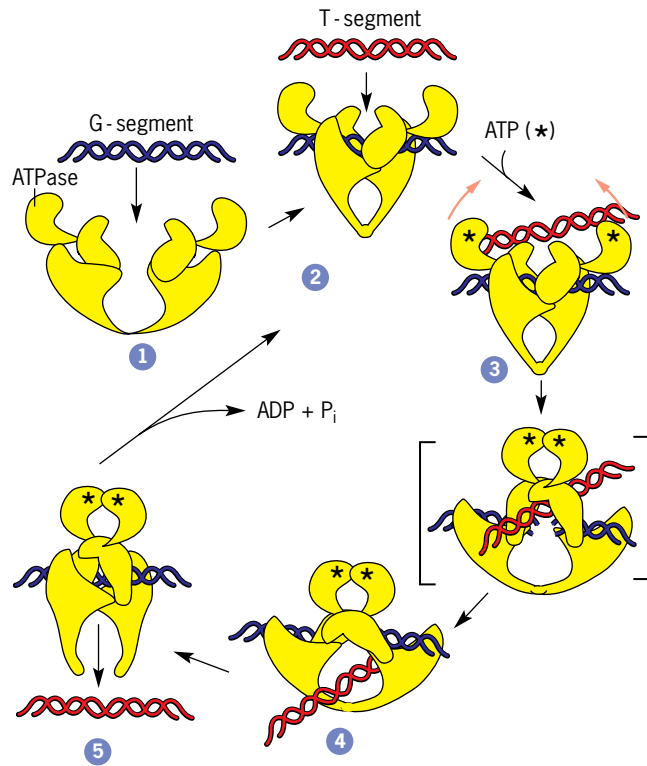
(b)
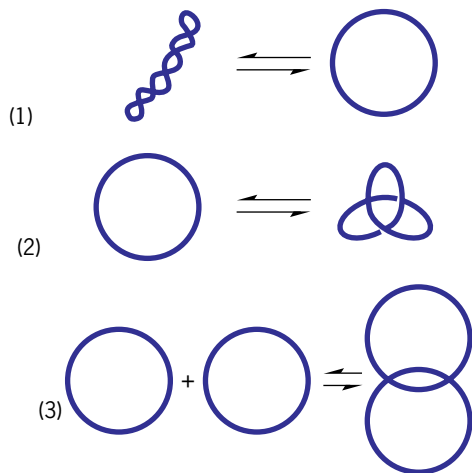
(c)
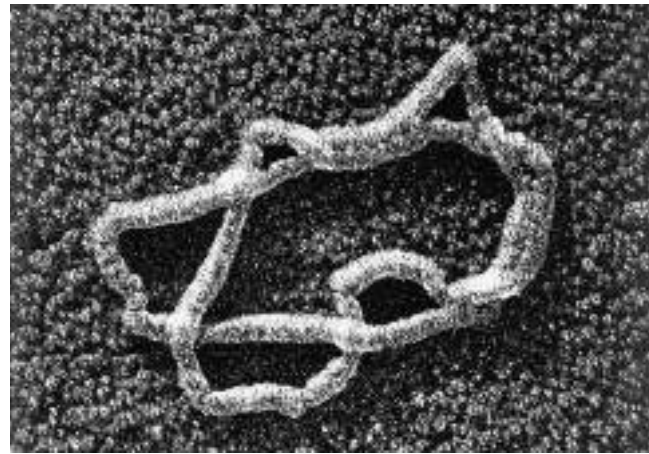
(d)

FIGURE 10.14 **DNA topoisomerases.** (*a*) A model depicting the action of human topoisomerase I. The enzyme (yellow) cuts one of the strands of the DNA (step 1), which rotates around a phosphodiester bond in the intact strand. The cut strand is then resealed (step 2). (*Note:* The drawing depicts a type IB topoisomerase; type IA enzymes found in bacteria act by a different mechanism.) (*b*) A molecular model based on X-ray crystallography depicting the action of topoisomerase II, a dimeric enzyme consisting of two identical halves. In step 1, the enzyme is in an "open" conformation ready to bind the G-DNA segment, so named because it will form the gate through which the T-DNA (or transported DNA) segment will pass. In step 2, the enzyme has undergone a conformational change as it binds the G-segment. In steps 3 and 4, the enzyme binds a molecule of ATP, the G-segment is cleaved, and the T-segment is passed through the open "gate." The bracketed stage represents a hypothetical intermediate carrying out the step in which the T-segment is transported through the G-segment. At this stage, both cut ends of the G-segment are covalently bound to the enzyme. In step 5, the two ends of the G-segment are rejoined, and the T-segment is released. ATP hydrolysis and release of ADP and P_i are proposed to occur as the starting state is regenerated. (*c*) Types of reactions that are catalyzed by topoisomerases. Part 1 shows supercoiling–relaxation reactions; part 2 shows knotting–unknotting reactions; part 3 shows catenation–decatenation reactions. (*d*) Electron micrograph of a pair of interconnected (catenated) circular DNA molecules. Molecules of this type collect in bacteria lacking a specific topoisomerase. (A: REPRINTED WITH PERMISSION FROM D. A. KOSTER ET AL., NATURE 434:671, 2005; © COPYRIGHT 2005, MACMILLAN MAGAZINES LTD; B: REPRINTED WITH PERMISSION FROM J. M. BERGER ET AL., NATURE 379:231, 1997; © COPYRIGHT 1997, MACMILLAN MAGAZINES LTD; D: FROM NICHOLAS COZZARELLI, CELL VOL. 71, COVER #2, 1992; BY PERMISSION OF CELL PRESS.)

(Figure 10.14*c*(3),*d*). Topoisomerase II is required to disentangle DNA molecules before duplicated chromosomes can be separated during mitosis. Human topoisomerase II is a target for a number of drugs (e.g., etoposide and doxorubicin) that bind to the enzyme and prevent the cleaved DNA strands from being resealed. In doing so, these drugs preferentially kill rapidly dividing cells and are therefore used in the treatment of cancer.

REVIEW ?

1. What is meant by saying that a DNA strand has polarity? That the two strands are antiparallel? That the molecule has a major groove and a minor groove? That the strands are complementary to one another?

2. What is the relationship between Chargaff's analysis of DNA base composition and the structure of the double helix?

3. How do underwound and overwound DNAs differ from one another? How do the two classes of topoisomerases differ?

1O.4 THE STRUCTURE OF THE GENOME

DNA is a macromolecule—an assemblage of a large number of atoms bonded together in a defined arrangement whose three-dimensional structure can be described by techniques such as X-ray crystallography. But DNA is also an information repository, which is a property more difficult to describe in simple molecular terms. If, as noted above, a gene corresponds to a particular segment of DNA, then the sum of the genetic information that an individual organism inherits from its parents is equivalent to the sum of all of the DNA segments present in the fertilized egg at the start of life. All the individuals that make up a species population share the same set of genes, even though different individuals invariably possess slightly different versions (alleles) of many of these genes. Thus each species of organism has a unique content of genetic information, which is known as its **genome**. For humans, the genome is essentially equivalent to all of the genetic information that is present in a single (**haploid**) set of human chromosomes, which includes 22 different autosomes and both the X and Y sex chromosomes.

The Complexity of the Genome

To understand how the complexity of the genome is determined, it is first necessary to consider one of the most important properties of the DNA double helix: its ability to separate into its two component strands, a property termed **denaturation**.

DNA Denaturation   As first suggested by Watson and Crick, the two strands of a DNA molecule are held together by weak, noncovalent bonds. When DNA is dissolved in saline solution and the solution is slowly warmed, a temperature is reached when strand separation begins. Within a few degrees, the process is generally complete and the solution contains single-stranded molecules that are completely separated from their original partners. The progress of thermal denaturation (or *DNA melting*) is usually monitored by following the increase in absorbance of ultraviolet light by the dissolved DNA. Once the two DNA strands have separated, the hydrophobic interactions that result from base stacking are greatly decreased, which changes the electronic nature of the bases and increases their absorbance of ultraviolet radiation. The rise in absorbance that accompanies DNA denaturation is shown in Figure 10.15. The temperature at which the shift in absorbance is half completed is termed the *melting temperature* (T_m). The higher the GC content (%G + %C) of the DNA, the higher the T_m. This increased stability of
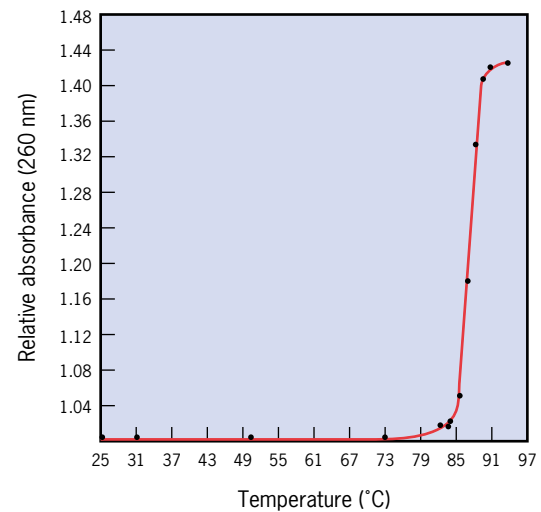
FIGURE 10.15   Thermal denaturation of DNA. A thermal denaturation curve for native bacteriophage T6 DNA in 0.3 M sodium citrate. The "melting" of the DNA (strand separation) occurs over a narrow range of temperature, particularly for the simpler DNAs of small viruses. The temperature corresponding to half the increase in absorbance is termed the T_m (FROM J. MARMUR AND P. DOTY, J. MOL. BIOL. 3:593, 1961; COPYRIGHT 1961, BY PERMISSION OF THE PUBLISHER ACADEMIC PRESS.)

GC-containing DNA reflects the presence of the extra hydrogen bond between the bases as compared with AT pairs (Figure 10.10c).

DNA Renaturation   The separation of the two strands of the DNA duplex by heat is not an unexpected finding, but the *reassociation* of single strands into stable double-stranded molecules with correct base pairs seems almost inconceivable. However, in 1960, Julius Marmur and co-workers at Harvard University found that if they slowly cooled a solution of bacterial DNA that had been thermally denatured, the DNA regained the properties of the double helix; it absorbed less ultraviolet light and once again behaved like genetic material in being able to transform bacterial cells (discussed on page 414). It became apparent from these studies that complementary single-stranded DNA molecules were capable of reassociating, an event termed **renaturation**, or **reannealing**. This finding has proved to be one of the most valuable observations ever made in molecular biology. On one hand, reannealing has served as the basis for an investigation into the complexity of the genome, a subject discussed in the following section. On the other hand, reannealing has led to the development of a methodology called *nucleic acid hybridization*, in which complementary strands of nucleic acids from different sources can be mixed to form double-stranded (hybrid) molecules. Examples of the types of questions that have been answered by allowing single-stranded nucleic acids to hybridize are discussed later in the chapter (page 397) and illustrated in Figure 10.19. Nucleic acid hybridization plays a key role in our most important modern biotechnologies, including DNA sequencing, DNA cloning, and DNA amplification.

The Complexity of Viral and Bacterial Genomes Several factors determine the rate of renaturation of a given preparation of DNA. They are (1) the ionic strength of the solution, (2) the incubation temperature, (3) the concentration of DNA, (4) the period of incubation, and (5) the size of the interacting molecules. With these factors in mind, consider what might happen if one were to compare the rate of renaturation of three different DNAs, each constituting the entire genome of (1) a small virus, such as MS-2 (4×10^3 nucleotide pairs); (2) a larger virus, such as T4 (1.8×10^5 nucleotide pairs); and (3) a bacterial cell, such as *E. coli* (4.5×10^6 nucleotide pairs). The primary difference between these DNAs is their length. To compare their renaturation, it is important that the reacting molecules are of equivalent length, typically about 1000 base pairs. DNA molecules can be fragmented into pieces of this length in various ways, one of which is to force them through a tiny orifice at high pressure.

When these three types of DNA—all present at the same lengths and DNA concentration (e.g., mg/ml)—are allowed to reanneal, they do so at distinctly different rates (Figure 10.16). The smaller the genome, the faster the renaturation. The reason becomes apparent when one considers the concentration of complementary sequences in the three preparations. Because all three preparations have the same amount of DNA in a given volume of solution, it follows that

the smaller the genome size, the greater the number of genomes present in a given weight of DNA and the greater the chance of collision between complementary fragments.

The Complexity of Eukaryotic Genomes The renaturation of viral and bacterial DNAs occurs along single, symmetrical curves (Figure 10.16). Renaturation occurs along such curves because all of the sequences are present (with the exception of a very few sequences in bacterial DNA) at the same concentration. Consequently, every nucleotide sequence in the population is as likely to find a partner in a given time as any other sequence. This is the result that would have been predicted from the various gene mapping studies that suggested the DNA of a chromosome contained one gene after another in a linear array. These results on viral and bacterial genomes contrast sharply with those obtained on mammalian DNA by Roy Britten and David Kohne of the California Institute of Technology. Unlike those of the simpler genomes, the DNA fragments from a mammalian genome reanneal at markedly different rates (Figure 10.17). These differences reflect the fact that the various nucleotide sequences in a preparation of eukaryotic DNA fragments are present at markedly different concentrations. This was the first insight that eukaryotic DNA is not just a simple progression of one gene after another, as in a bacterium or virus, but has a much more complex organization.

When DNA fragments from plants and animals are allowed to reanneal, the curve typically shows three more-or-less distinct steps (Figure 10.17), which correspond to the reannealing of three broad classes of DNA sequences. The three classes reanneal at different rates because they differ as to the number of times their nucleotide sequence is repeated within the population of fragments. The three classes are termed the **highly repeated fraction**, the **moderately repeated fraction**, and the **nonrepeated fraction**.

FIGURE 10.16 **The kinetics of renaturation of viral and bacterial DNAs.** The curves show the renaturation of sheared strands of DNA from two viruses (MS-2 and T4) and a bacterium (*E. coli*). (The formation of double-stranded DNA is plotted against C_0t, which is a term that combines two variables: initial DNA concentration (C_0) and time of incubation (t). A solution containing a high concentration of DNA incubated for a short time will have the same C_0t as one of low concentration incubated for a correspondingly longer time; both will have the same percentage of reannealed DNA.) The genome size, that is, the number of nucleotide base pairs in one copy of total genetic information of the organism, is indicated by the arrows near the upper numerical scale. The shape of each of the renaturation curves is very simple and occurs with a single slope. However, the time over which renaturation occurs is very different and depends on the concentration of complementary fragments, which in turn depends on the size of the genome. The larger the genome, the lower the concentration of complementary fragments in solution, and the greater the time required for renaturation to be complete.

FIGURE 10.17 **An idealized plot showing the kinetics of renaturation of eukaryotic DNA.** When single-stranded DNA is allowed to reanneal, three classes of fragments can usually be distinguished by their frequency of repetition within the genome: a highly repeated DNA fraction, a moderately repeated DNA fraction, and a nonrepeated (single-copy) DNA fraction. (*Note:* This is an idealized plot: the three classes of sequences are not as clearly separated in an actual renaturation curve.)

Highly Repeated DNA Sequences The highly repeated fraction, which consists of sequences present in at least 10^5 copies per genome, constitutes anywhere from about 1 to 10 percent of the total DNA. Highly repeated sequences are typically short (a few hundred nucleotides at their longest) and present in clusters in which the given sequence repeats itself over and over again without interruption. A sequence arranged in this end-to-end manner is said to be present *in tandem*. Highly repeated sequences fall into several overlapping categories, including satellite DNAs, minisatellite DNAs, and microsatellite DNAs.

■ Satellite DNAs. Satellite DNAs consist of short sequences (about five to a few hundred base pairs in length) that form very large clusters, each containing up to several million base pairs of DNA. In many species, the base composition of these DNA segments is sufficiently different from the bulk of the DNA that fragments containing the sequence can be separated into a distinct "satellite" band during density gradient centrifugation (hence the name *satellite* DNA). Satellite DNAs tend to evolve very rapidly, causing the sequences of these genomic elements to vary even between closely related species. The localization of satellite DNA within the centromeres of chromosomes is described on page 398 and in further detail in Section 12.1.

■ **Minisatellite DNAs.** Minisatellite sequences range from about 10 to 100 base pairs in length and are found in sizeable clusters containing as many as 3000 repeats. Thus, minisatellite sequences occupy considerably shorter stretches of the genome than do satellite sequences. Minisatellites tend to be unstable, and the number of copies of a particular sequence often increases or decreases from one generation to the next as the result of unequal crossing over (see Figure 10.22). Consequently, the length of a particular minisatellite locus is highly variable in the population, even among members of the same family. Because they are so variable (or *polymorphic*) in length, minisatellite sequences form the basis for the technique of *DNA fingerprinting*, which is used to identify individuals in criminal or paternity cases (Figure 10.18).

■ **Microsatellite DNAs.** Microsatellites are the shortest sequences (1 to 5 base pairs long) and are typically present in small clusters of about 10 to 40 base pairs in length, which are scattered quite evenly through the genome. DNA replicating enzymes have trouble copying regions of the genome that contain these small, repetitive sequences, which causes these stretches of DNA to change in length through the generations. Because of their variable lengths within the population, microsatellite DNAs have been used to analyze the relationships between different human populations, as illustrated in the following example. Many anthropologists argue that modern humans arose in Africa. If this is true, then members of different African populations should exhibit greater DNA sequence variation than human populations living on other continents because genomes of African populations have had longer to diverge. The argument for African genesis has received support from studies on human DNA sequences. In one
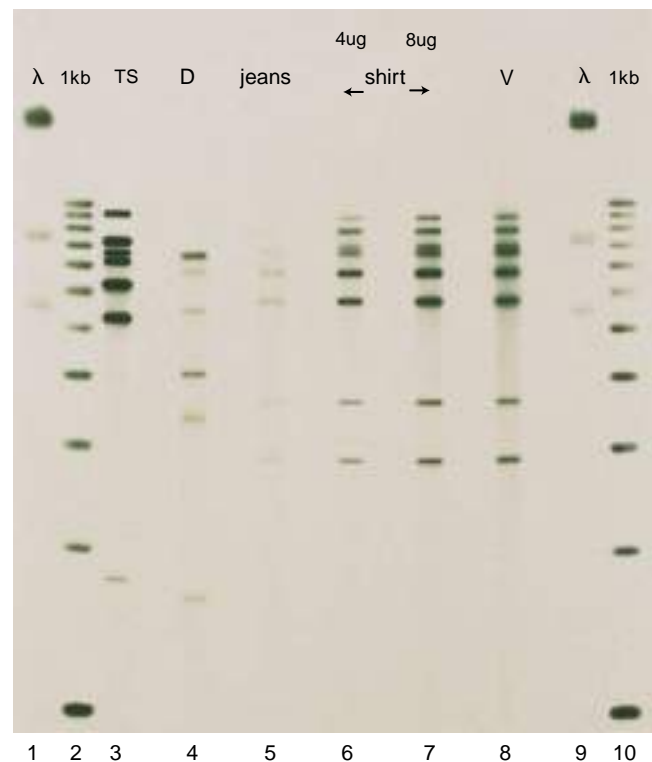
FIGURE 10.18 DNA fingerprinting. In this technique, which is used widely to identify an individual from a sample of DNA, the DNA is digested by treatment with specific nucleases (called restriction endonucleases, described in Section 18.13), and the DNA fragments are separated on the basis of length by gel electrophoresis. The location in the gel of DNA fragments containing specific DNA sequences is determined using labeled probes with sequences complementary to those being sought. The DNA fragments that bind these probes have variable lengths from one person to the next because of the presence of variable numbers of tandem repeats (VNTRs) in the genome. Forensic labs typically analyze about 13 VNTR loci that are known to be highly polymorphic. The chance that two individuals will have identical VNTRs at these loci is astronomically small. The fingerprint shown in this figure was used in a criminal case in which the defendant was charged with the fatal stabbing of a young woman. The bloodstains on the pants and shirt of the defendant were compared to the known blood standards from the victim and defendant. DNA from the bloodstains on the defendant's clothing did not match his own blood standard, but they did match that of the victim. The lanes contain DNA from the following sources: 1, 2, 3, 9, and 10 are control DNA samples that serve as quality control checks; 4, the defendant's blood; 5, bloodstains from the defendant's pants; 6 and 7, bloodstains from the defendant's shirt; and 8, the victim's blood. With the advent of DNA amplification techniques (i.e., PCR), minuscule samples of DNA can be used for these analyses. (COURTESY OF ORCHID CELLMARK, PRINCETON, NJ.)

study that analyzed 60 different microsatellite loci, it was found that members of African populations had a significantly greater genetic divergence than Asian or European populations. The vast majority of microsatellite loci occur outside of genes, and changes in their length generally go unnoticed. This is not the case for the microsatellite sequences discussed in the accompanying Human Perspective.

Diseases that Result from Expansion
of Trinucleotide Repeats

For decades, biologists believed that genes were always transmitted from generation to generation as stable entities. On rare occasion, a change occurs in the nucleotide sequence of a gene in the germ line, creating a mutation that is subsequently inherited. This is one of the basic tenets of Mendelian genetics. Then in 1991, several laboratories reported a new type of "dynamic mutation" in which the nucleotide sequence of particular genes changed dramatically between parents and offspring. In each case, these mutations affected genes that contained a repeating trinucleotide unit (e.g., CCG or CAG) as part of their sequence. In most members of the population, these particular genes contain a relatively small (but variable) number of repeated trinucleotides, and they are transmitted from one generation to the next without a change in the number of repeats. In contrast, a small fraction of the population possesses a mutant version of the gene that contains a larger number of these repeating units. Unlike the normal version, the mutant alleles are highly unstable, and the number of repeating units tends to increase as the gene passes from parents to offspring. When the number of repeats increases beyond a critical number, the individual inheriting the mutant allele develops a serious disease.

At the present time, more than a dozen distinct diseases have been attributed to expansion of trinucleotide repeats. These diseases fall into two basic categories, which we will consider in turn. Type I diseases are all neurodegenerative disorders that result from the expansion of the number of repeats of a CAG trinucleotide within the coding portion of the mutant gene (Figure 1). We can illustrate the nature of these disorders by considering the most prevalent and widely studied of the group, Huntington's disease (HD). HD is a fatal illness characterized by involuntary, uncoordinated movements; changes in personality, including depression and irritability; and gradual intellectual decline. Symptoms usually begin in the third to fifth decade of life and increase in severity until death.

The normal *HD* gene, which is stably transmitted, contains between 6 and 35 copies of the CAG trinucleotide. The protein en-

coded by the *HD* gene is called *huntingtin*, and its precise function remains unclear. CAG is a triplet that encodes the amino acid glutamine. Thus the normal huntingtin polypeptide contains a stretch of 6 to 35 glutamine residues—a polyglutamine tract—as part of its primary structure. We think of polypeptides as having a highly defined primary structure, and most of them do, but huntingtin is normally polymorphic with respect to the length of its polyglutamine tract. The protein appears to function normally as long as the tract length remains below approximately 35 glutamine residues. But if this number is exceeded, the protein takes on new properties, and the person is predisposed to developing HD.

HD exhibits a number of unusual characteristics. Unlike most inherited diseases, HD is a dominant genetic disorder, which means that a person with the mutant allele will develop the disease regardless of whether or not he or she has a normal *HD* allele. In fact, persons who are homozygous for the *HD* allele are no more seriously affected than heterozygotes. This observation indicates that the mutant huntingtin polypeptide causes the disease, not because it fails to carry out a particular function, but because it acquires some type of toxic property, which is referred to as a *gain-of-function* mutation. This interpretation is supported by studies with mice. Mice that are engineered to carry the mutant human *HD* allele (in addition to their own normal alleles) develop a neurodegenerative disease resembling that found in humans. The presence of the one abnormal allele is sufficient to cause disease. Another unusual characteristic of HD and the other CAG disorders is a phenomenon called genetic anticipation, which means that, as the disease is passed from generation to generation, its severity increases and/or it strikes at an increasingly earlier age. This was once a puzzling feature of HD but is now readily explained by the fact that the number of CAG repeats in a mutant allele (and the resulting consequences) often increases dramatically from one generation to the next.

The molecular basis of HD remains unclear, but there is no shortage of theories as to why an expanded glutamine tract may be

5' UTR	Intron	Coding exon	Intron	3' UTR

—5– 55
—60– 200
—200– >2000

—7– 22
—23– 200
— >200

—6– 35
—> 35

—5– 40
—45– 200
—200– >2000

CGG
Fragile
X syndrome

GAA
Friedreich's
ataxia

Type 1 diseases
CAG/polyglutamine
e.g. Huntington's disease

CTG
Myotonic
dystrophy
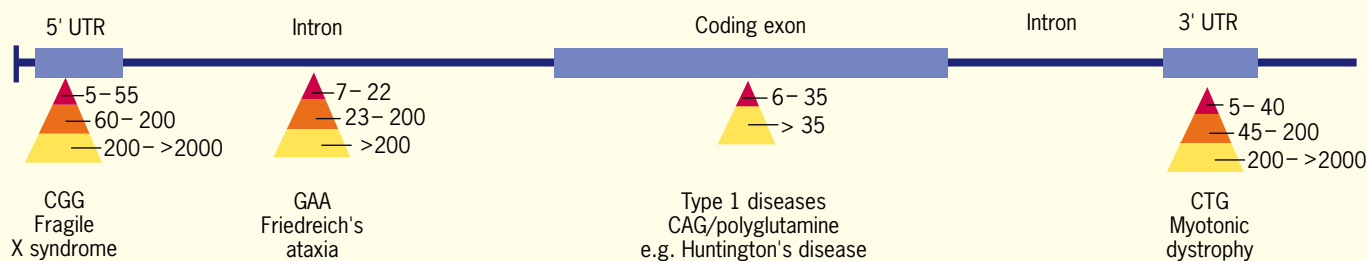
FIGURE 1 Trinucleotide repeat sequences and human disease. The top line shows a generalized gene that is transcribed into a messenger RNA with several distinct portions (purple boxes), including a 5′ noncoding portion called the 5′ UTR (5′ untranslated region), a coding exon that carries the information for the amino acid sequence of the polypeptide, and a 3′ noncoding portion (the 3′ UTR). The introns in the DNA (see Figure 11.29) are not represented in the mature messenger RNA. The general location of the trinucleotide responsible for each of four different diseases (fragile X syndrome, Friedreich's ataxia, Huntington's disease, and myotonic dystrophy) is indicated by the location of each pyramid. The number of repeats responsible for the normal (red), carrier (orange), and disease (yellow) conditions for each disease-causing gene is indicated. Genes responsible for Type I diseases, such as Huntington's, do not exhibit the intermediate "carrier" state in which an individual possesses an unstable allele but is not affected. (AFTER J.-L. MANDEL NATURE 386:768, 1997; © COPYRIGHT 1997, MACMILLAN MAGAZINES LTD.)

toxic to brain cells. One feature appears indisputable: when the polyglutamine tract of huntingtin exceeds 35 residues, the protein (or a fragment cleaved from it) undergoes abnormal folding to produce a misfolded molecule that (1) binds to other mutant huntingtin molecules to form insoluble aggregates, not unlike those seen in the brains of Alzheimer's victims, and (2) binds to a number of unrelated proteins that do not interact with normal, wild-type huntingtin molecules. Among the proteins bound by mutant huntingtin are several transcription factors, which are proteins involved in the regulation of gene expression. Several of the most important transcription factors present in cells, including TBP (see Figure 11.19) and CBP (see Figure 12.46), contain polyglutamine stretches themselves, which makes them particularly susceptible to aggregation by proteins with mutant, expanded polyglutamine tracts. In fact, the protein aggregates present in degenerating neurons of HD patients contain both of these transcription factors. These results suggest that mutant huntingtin sequesters transcription factors, disrupting the transcription of genes that are required for the health and survival of the affected neurons. This hypothesis has received support from a study in which mice were genetically engineered so that their brain cells lacked the ability to produce certain key transcription factors. These mice exhibited the same type of neurodegeneration as is seen in animals carrying a mutant *HD* gene. Other basic neuronal processes, including axonal transport, mitochondrial permeability, and protein degradation, are also disrupted by the *HD* mutation and represent possible causes of nerve-cell death. It should be noted that a number of HD researchers hold an alternate opinion about the cause of cell death. They argue that it is not the protein aggregates that are toxic, but the soluble mutant protein (or fragments of the mutant protein) itself. In fact, proponents of this view argue that the mutant protein aggregates protect the cell by sequestering the harmful molecules. It is important for practical reasons to distinguish between these possibilities because a number of proposed therapies are aimed at block-

ing formation of the aggregates, which could actually prove more harmful to a patient.

Type II trinucleotide repeat diseases differ from the Type I diseases in a number of ways. Type II diseases (1) arise from the expansion of a variety of trinucleotides, not only CAG, (2) the trinucleotides involved are present in a part of the gene that does not encode amino acids (Figure 1), (3) the trinucleotides are subject to massive expansion into thousands of repeats, and (4) the diseases affect numerous parts of the body, not only the brain. The best studied Type II disease is fragile X syndrome, so named because the mutant X chromosome is especially susceptible to damage. Fragile X syndrome is characterized by mental retardation as well as a number of physical abnormalities. The disease is caused by a dynamic mutation in a gene called *FMR1* that encodes an RNA-binding protein that regulates the translation of certain mRNAs involved in neuronal development and/or synaptic function. A normal allele of this gene contains anywhere from about 5 to 55 copies of a specific trinucleotide (CGG) that is repeated in a part of the gene that corresponds to the 5′ noncoding portion of the messenger RNA (Figure 1). However, once the number of copies rises above about 60, the locus becomes very unstable and the copy number tends to increase rapidly into the thousands. Females with an *FMR1* gene containing 60 to 200 copies of the triplet generally exhibit a normal phenotype but are carriers for the transmission of a highly unstable chromosome to their offspring. If the repeat number in the offspring rises above about 200, the individual is almost always mentally retarded. Unlike an abnormal *HD* allele, which causes disease as the result of a gain of function, an abnormal *FMR1* allele causes disease as the result of a loss of function; *FMR1* alleles containing an expanded CGG number are selectively inactivated so that the gene is not transcribed or translated. Although there is no effective treatment for any of the diseases caused by trinucleotide expansion, the risk of transmitting or possessing a mutant allele can be assessed through genetic screening.

Once it became apparent that eukaryotic genomes contain large numbers of copies of short DNA sequences, researchers sought to learn where these DNA sequences are located in the chromosomes. Are satellite DNA sequences, for example, clustered in particular parts of a chromosome, or scattered uniformly from one end to the other? It was noted above that the discovery of DNA reannealing led to the development of a vast nucleic acid hybridization methodology (discussed at length in Chapter 18). The ability to locate satellite DNA sequences illustrates the analytic power of nucleic acid hybridization.

In the reannealing experiments described on page 393, complementary DNA strands were allowed to bind to one another while colliding randomly in solution. An alternate experimental protocol called *in situ hybridization* was developed in 1969 by Mary Lou Pardue and Joseph Gall of Yale University and was used to determine the location of satellite DNA. The term *in situ* means "in place," which refers to the fact that the DNA of the chromosomes is kept in place while it is allowed to react with a particular preparation of labeled DNA. In early in situ hybridization studies, the DNA to be detected (the probe DNA) was radioactively labeled and was localized

by autoradiography. The resolution of the technique has been increased using probes that are labeled with fluorescent dyes and then localized with a fluorescence microscope (as in Figure 10.19). This latter technique is called **fluorescence in situ hybridization (FISH)** and has become so refined that it can be used to map the locations of different sequences along single DNA fibers.

To carry out nucleic acid hybridization, both interactants must be single-stranded. In the experiment depicted in Figure 10.19, the chromosomes from a mitotic cell are spread on a slide, and the DNA is made single-stranded by treating the chromosomes with a hot salt solution that causes the DNA strands to separate and remain apart. During the ensuing hybridization step, the denatured chromosomes are incubated with a solution of biotin-labeled, single-stranded satellite DNA, which binds selectively to complementary strands of immobilized satellite DNA located in the chromosomes. Following the incubation period, the soluble, unhybridized satellite DNA is washed away or digested, and the sites of the bound, labeled DNA fragments are revealed. As shown in Figure 10.19, and described in the accompanying legend, satellite DNA is localized in the centromeric regions of the

(a)

(b)

chromosome (see Figure 12.22). Another example using FISH technology is shown in Figure 10.20.

Moderately Repeated DNA Sequences The moderately repeated fraction of the genomes of plants and animals can vary from about 20 to more than 80 percent of the total DNA, depending on the organism. This fraction includes sequences that are repeated within the genome anywhere from a few times to tens of thousands of times. Included in the moderately repeated DNA fraction are sequences that code for known gene products, either RNAs or proteins, and those that lack a coding function.

1. ***Repeated DNA Sequences with Coding Functions.*** This fraction of the DNA includes the genes that code for ribosomal RNAs as well as those that code for an important group of chromosomal proteins, the histones. The repeated sequences that code for each of these products are typically identical to one another and located in tandem array. It is essential that the genes encoding ribosomal RNAs be present in multiple numbers because these

RNAs are required in large amounts and their synthesis does not benefit from the extra amplification step that occurs for protein-coding genes in which each mRNA acts as a template for the repeated synthesis of a polypeptide. Even though the production of histones does involve an intermediary messenger RNA, so many copies of these proteins are required during early development that several hundred DNA templates must be present.

2. ***Repeated DNAs that Lack Coding Functions.*** The bulk of the moderately repeated DNA fraction does not encode any type of product. Rather than occurring as clusters of tandem sequences, the members of these families are scattered (i.e., *interspersed*) throughout the genome as individual elements. Most of these repeated sequences can be grouped into two classes that are referred to as SINEs (*short interspersed elements*) or LINEs (*long interspersed elements*). SINEs and LINEs sequences are discussed on page 404.

Nonrepeated DNA Sequences As initially predicted by Mendel, classical studies on the inheritance patterns of visible traits led geneticists to conclude that each gene was present in one copy per single (haploid) set of chromosomes. When denatured eukaryotic DNA is allowed to reanneal, a significant fraction of the fragments are very slow to find partners, so slow in fact that they are presumed to be present in a single
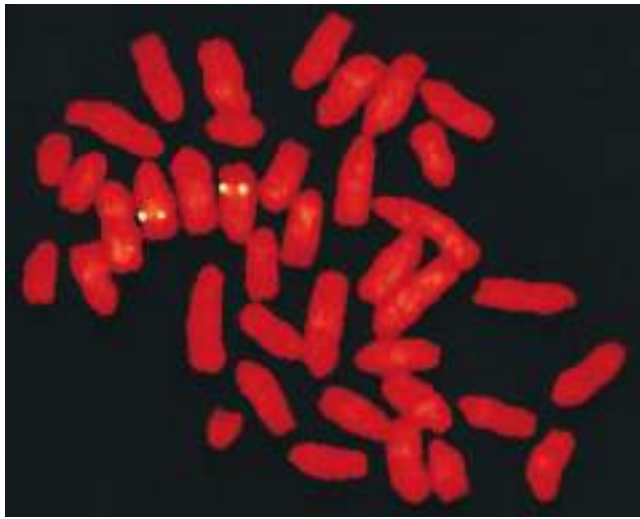
FIGURE 10.20 Chromosomal localization of a nonrepeated DNA sequence. These mitotic chromosomes were prepared from a dividing mouse cell and incubated with a purified preparation of biotin-labeled DNA encoding one of the nuclear lamin proteins (lamin B$_2$), which is encoded by a nonrepeated gene. The locations of the bound, labeled DNA appear as bright dots. The lamin gene is present on the homologues of chromosome 10. Each chromosome contains two copies of the gene because the DNA had been replicated prior to the cells entering mitosis. (FROM MONIKA ZEWE ET AL., COURTESY OF WERNER FRANKE, EUR. J. CELL BIOL. 56:349, 1991.)

copy per genome. This fraction comprises the *nonrepeated* (or *single-copy*) DNA sequences, which includes the genes that exhibit Mendelian patterns of inheritance. Because they are present in a single copy in the genome, nonrepeated sequences localize to a particular site on a particular chromosome (Figure 10.20).

Considering the variety of different sequences present, the nonrepeated fraction contains by far the greatest amount of genetic information. Included within the nonrepeated fraction are the DNA sequences that code for virtually all proteins other than histones. Even though these sequences are not present in multiple copies, genes that code for polypeptides are usually members of a family of related genes. This is true for the globins, actins, myosins, collagens, tubulins, integrins, and most other proteins in a eukaryotic cell. Each member of a multigene family is encoded by a different but related sequence. We will look into the origin of these multigene families in the following section.

Now that the human genome has been sequenced and analyzed, we finally have a relatively accurate measure of the DNA sequences that are responsible for encoding the amino acid sequences of our proteins, and it is remarkably small. If you would have suggested to a geneticist in 1960 that less than 1.5 percent of the human genome codes for the amino acids of our proteins, he or she would have considered the suggestion to be ridiculous. Yet, that is the reality that has emerged from the study of genome sequences. In the following section, we will try to better understand how the remaining 98+ percent of DNA sequences might have evolved.

REVIEW   **?**

1. What is a genome? How does the complexity of bacterial genomes differ from that of eukaryotic genomes?

2. What is meant by the term *DNA denaturation*? How does denaturation depend on the GC content of the DNA? How does this variable affect the T_m?

3. What is a microsatellite DNA sequence? What role do these sequences play in human disease?

4. Which fraction of the genome contains the most information? Why is this true?

10.5 THE STABILITY OF THE GENOME

Because DNA is the genetic material, we tend to think of it as a conservative molecule whose information content changes slowly over long periods of evolutionary time. In actual fact, the sequence organization of the genome is capable of rapid change, not only from one generation to the next, but within the lifetime of a given individual.

Whole-Genome Duplication (Polyploidization)

As discussed in the first sections of this chapter, peas and fruit flies have pairs of homologous chromosomes in each of their cells. These cells are said to have a **diploid** number of chromosomes. If a person were to compare the number of chromosomes present in the cells of closely related organisms, especially higher plants, they would find that some species have a much greater number of chromosomes than a close relative. Among animals, the widely studied amphibian *Xenopus laevis*, for example, has twice the number of chromosomes as its cousin *X. tropicalis*. These types of discrepancies can be explained by a process known as **polyploidization**, or **whole-genome duplication**. Polyploidization is an event in which offspring are produced that have twice the number of chromosomes in each cell as their diploid parents; the offspring have four homologues of each chromosome rather than two. Polyploidization is thought to occur in either of two ways: two related species mate to form a hybrid organism that contains the combined chromosomes from both parents, or, alternatively, a single-celled embryo undergoes chromosome duplication but the duplicates, rather than being separated into separate cells, are retained in a single cell that develops into a viable embryo. The first mechanism occurs most often in plants, and the second most often in animals. Polyploidization is particularly common in flowering plants, including numerous crop species (e.g., wheat, bananas, and coffee) depicted in Figure 10.21.

A "sudden" doubling of the number of chromosomes is a dramatic event, one that gives an organism remarkable evolutionary potential—assuming it can survive the increased number of chromosomes and reproduce. Depending on the circumstances, polyploidization may result in the production of a new species that has a great deal of "extra" genetic infor-

FIGURE 10.21 **A sample of agricultural crops that are polyploid.** Pictured are oil from oilseed rape, bread from bread wheat, rope from sisal, coffee beans, banana, cotton, potatoes, and maize. (FROM A. R. LEITCH AND I. J. LEITCH, SCIENCE 320:481, 2008; COPYRIGHT © 2008, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

mation. Several different fates can befall extra copies of a gene; they can be lost by deletion, rendered inactive by deleterious mutations, or, most importantly, they can evolve into new genes that possess new functions. Viewed in this way, extra genetic information is the raw material for evolutionary diversification. In 1971, Susumu Ohno of the City of Hope Cancer Center in Los Angeles put forward the "2R" hypothesis, in which he proposed that the evolution of vertebrates from a much simpler invertebrate ancestor was made possible by two separate rounds of whole-genome duplication during an early evolutionary period. Ohno suggested that the thousands of extra genes that would be generated by genome duplication could be molded by natural selection into new genes that were required to encode the more complex vertebrate body. Over the past three and a half decades, Ohno's proposal has been hotly debated as geneticists have tried to find evidence that either supports or refutes the notion.

The problem facing genome analysts is the vast stretch of time that has passed—hundreds of millions of years—since the origin of our earliest vertebrate ancestors. Just as a river or sea slowly wears away the face of the Earth, chromosomal rearrangement and mutation slowly wear away the face of an ancestral genome. Even with the complete sequence of a number of invertebrate and vertebrate genomes in hand, it has still proven a formidable challenge to identify the origin of many of our genes. The strongest evidence for the 2R hypothesis comes from the recent analysis of the amphioxus genome. Amphioxus lacks a backbone, which makes it an invertebrate, but it has a number of features (e.g., a notochord, dorsal tubular nerve cord, and segmental body musculature) that clearly identifies it as a member of the phylum Chordata to which vertebrates belong. The lineages leading to modern vertebrates and amphioxus are thought to have separated about 550 million years ago, yet the two groups share a remarkably similar collection of genes.

However, when researchers looked more closely at certain groups of genes, the genomes of vertebrates typically contained four times the number of such genes when compared to the homologous sequences in the amphioxus genome. This finding provides strong support for Ohno's hypothesis of two rounds of whole-genome duplication in the ancestral vertebrate lineage.

Duplication and Modification of DNA Sequences

Polyploidization is an extreme case of genome duplication and occurs only rarely during evolution. In contrast, **gene duplication**, which refers to the duplication of a small portion of a single chromosome, happens with surprisingly high frequency, and its occurrence is readily documented by genome analysis.[3] According to one estimate, each gene in the genome has about a 1 percent chance of being duplicated every million years. Duplication of a gene can probably occur by several different mechanisms but is most often thought to be produced by a process of *unequal crossing over*, as depicted in Figure 10.22. Unequal crossing over occurs when a pair of homologous chromosomes comes together during meiosis in such a way that they are not perfectly aligned. As a result of misalignment, genetic exchange between the homologues causes one chromosome to acquire an extra segment of DNA (a duplication) and the other chromosome to lose a DNA segment (a deletion). If the duplication of a particular sequence is repeated in subsequent generations, a cluster of tandemly repeated segments is generated at a localized site within that chromosome (see Figure 10.29).

The vast majority of gene duplicates are either lost during evolution through deletion or rendered nonfunctional by unfavorable mutation. However, in a small percentage of cases, the "extra" copy accumulates favorable mutations and acquires a new function. More often, both copies of the gene undergo mutation so that each evolves a more specialized function than that performed by the original gene. In either case, the two genes will have closely related sequences and encode similar polypeptides, which is to say that they encode different isoforms of a particular protein, such as α- and β-tubulin (page 324). Subsequent duplications of one of the genes can lead to the formation of additional isoforms (e.g., γ-tubulin), and so forth. It becomes apparent from this example that successive gene duplication can generate families of genes that encode polypeptides with related amino acid sequences. The production of a multigene family is illustrated by the evolution of the globin genes.

Evolution of Globin Genes Hemoglobin is a tetramer composed of four globin polypeptides (see Figure 2.38*b*). Examination of globin genes, whether from a mammal or a fish, reveals a characteristic organization. Each of these genes is constructed of three exons and two introns. Exons are parts of genes that code for amino acids in the encoded polypeptide, whereas introns do not; they are noncoding intervening sequences. The

---

[3]Actually, three categories of duplication can be distinguished: whole-genome, gene, and segmental duplication. The last category, which refers to the duplication of a large block of chromosomal material (from a few kilobases to hundreds of kilobases in length) is not discussed here, but has a significant impact in genome evolution. Approximately 5 percent of the present human genome consists of segmental duplications that have arisen during the past 35 million years.
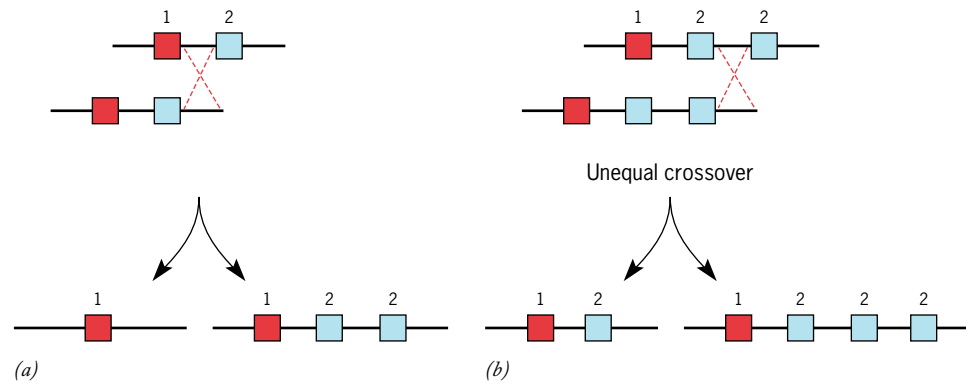
FIGURE 10.22 Unequal crossing over between duplicated genes provides a mechanism for generating changes in gene number. (*a*) The initial state shown has two related genes (1 and 2). In a diploid individual, gene 1 on one homologue can align with gene 2 on the other homologue during meiosis. If a crossover occurs during this misalignment, half the gametes will be missing gene 2 and half will have an extra gene 2. (*b*) As unequal crossing over continues to occur during meiotic divisions in subsequent generations, a tandemly repeated array of DNA sequences will gradually evolve.

subject of exons and introns is discussed in detail in Section 11.4 (see Figure 11.24). For the present purposes, we will simply use these gene parts as landmarks of evolution. Examination of genes encoding certain globin-like polypeptides, such as the plant protein leghemoglobin and the muscle protein myoglobin, reveals the presence of four exons and three introns. This is proposed to represent the ancestral form of the globin gene. It is thought that the modern globin polypeptide arose from the ancestral form as the result of the fusion of two of the globin exons (step 1, Figure 10.23) some 800 million years ago.

A number of primitive fish are known that have only one globin gene (step 2), suggesting that these fish diverged from other vertebrates prior to the first duplication of the globin gene (step 3). Following this duplication approximately 500 million years ago, the two copies diverged by mutation (step 4) to form two distinct globin types, an α type and a β type, located on a single chromosome. This is the present arrangement in the amphibian *Xenopus* and in zebra fish. In subsequent steps, the α and β forms are thought to have become separated from one another by a process of rearrangement that moved them to separate chromosomes (step 5). Each gene then underwent subsequent duplications and divergence (step 6), generating the arrangement of globin genes that exists today in humans (step 7). The evolution of vertebrate globin genes illustrates how gene duplication typically leads to

FIGURE 10.23 A pathway for the evolution of globin genes. Exons are shown in red, introns in yellow. The evolutionary steps depicted in the diagram are discussed in the text. The arrangement of α- and β-globin genes on human chromosomes 16 and 11 (shown without their introns on the right) are the products of several hundred million years of evolution. As discussed in Chapter 2, hemoglobin molecules consist of two pairs of polypeptide chains—one pair is always a member of the α-globin subfamily, and the other pair is always a member of the β-globin subfamily. Specific combinations of α- and β-globins are found at different stages of development. The α- and β-globin chains that are observed in embryonic, fetal, and adult hemoglobins are indicated.
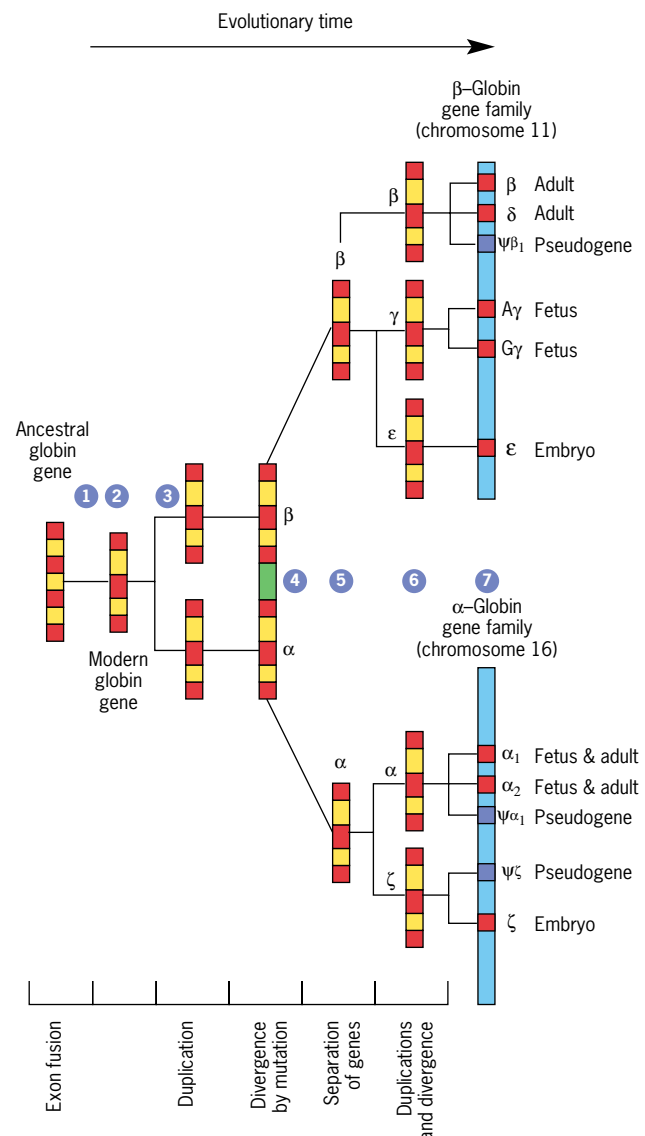
the generation of a family of genes whose individual members have specialized functions (in this case, embryonic, fetal, and adult forms) when compared to the single founding gene.

When the DNA sequences of globin gene clusters were analyzed, researchers found "genes" whose sequences are homologous to those of functional globin genes, but which have accumulated severe mutations that render them nonfunctional. Genes of this type, which are evolutionary relics, are known as **pseudogenes**. Examples of pseudogenes are found in both the human α- and β-globin gene clusters of Figure 10.23. Another point that is evident from examination of the two globin clusters of human chromosomes is how much of the DNA consists of noncoding sequences, either as introns within genes or as spacers between genes. In fact, the globin regions contain a much higher fraction of coding sequences than most other regions of the genome.

"Jumping Genes" and the Dynamic Nature of the Genome

If one looks at repeated sequences that have arisen during the normal course of evolution, one finds that repeats are sometimes present in tandem arrays, sometimes present on two or a few chromosomes (as in the case of the globin genes of Figure 10.23), and sometimes dispersed throughout the genome. If we assume that all members of a family of repeated sequences arose from a single copy, then how can individual members become dispersed among different chromosomes?

The first person to suggest that genetic elements were capable of moving around the genome was Barbara McClintock, a geneticist working with maize (corn) at the Cold Spring Harbor Laboratories in New York. Genetic traits in maize are often revealed as changes in the patterns and markings in leaf and kernel coloration (Figure 10.24). In the late 1940s, McClintock found that certain mutations were very unstable, appearing and disappearing from one generation to the next or even during the lifetime of an individual plant. After several years of careful study, she concluded that certain genetic elements were moving from one place in a chromosome to an entirely different site. She called this genetic rearrangement **transposition**, and the mobile genetic elements **transposable elements**. Meanwhile, molecular biologists working with bacteria were finding no evidence of "jumping genes." In their studies, genes appeared as stable elements situated in a linear array on the chromosome that remained constant from one individual to another and from one generation to the next. McClintock's findings were largely ignored.

Then, in the late 1960s, several laboratories discovered that certain DNA sequences in bacteria moved on rare occasion from one place in the genome to another. These bacterial transposable elements were called **transposons**. Most transposons encode a protein, or *transposase*, that single-handedly catalyzes the excision of a transposon from a donor DNA site and its subsequent insertion at a target DNA site. This "cut-and-paste" mechanism is mediated by two separate transposase subunits that bind to specific sequences at the two ends of the transposon (Figure 10.25, step 1). The two subunits then come together to form an active dimer (step 2) that catalyzes a series

FIGURE 10.24 Visible manifestations of transposition in maize. Kernels of corn are typically uniform in color. The spots on this kernel result from a mutation in a gene that codes for an enzyme involved in pigment production. Mutations of this type can be very unstable, arising or disappearing during the period in which a single kernel develops. These unstable mutations appear and disappear as the result of the movement of transposable elements into and out of these genes during the period of development. (COURTESY OF VENKATESAN SUNDARESAN, COLD SPRING HARBOR LABORATORY.)
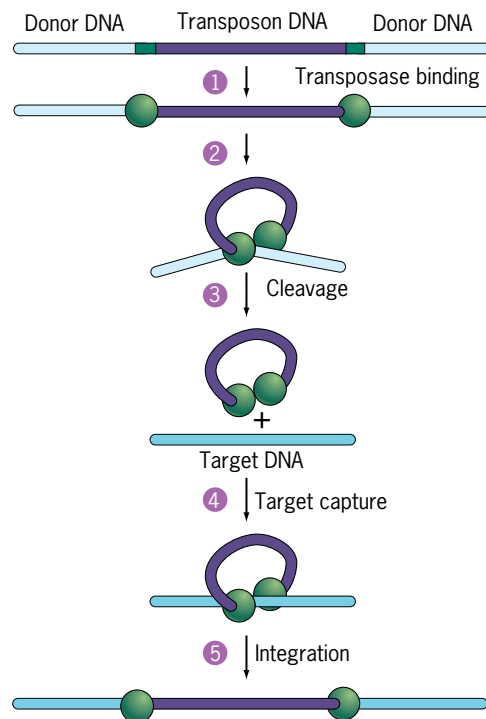
FIGURE 10.25 Transposition of a bacterial transposon by a "cut-and-paste" mechanism. As discussed in the text, the two ends of this bacterial Tn5 transposon are brought together by the dimerization of a pair of subunits of the transposase. Both strands of the double helix are cleaved at each end, which excises the transposon as part of a complex with the transposase. The transposon–transposase complex is "captured" by a target DNA, and the transposon is inserted in such a way as to produce a small duplication that flanks the transposed element. (*Note:* Not all DNA transposons move by this mechanism.) (AFTER D. R. DAVIES ET AL., SCIENCE 289:77, 2000; COPYRIGHT © 2000, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)
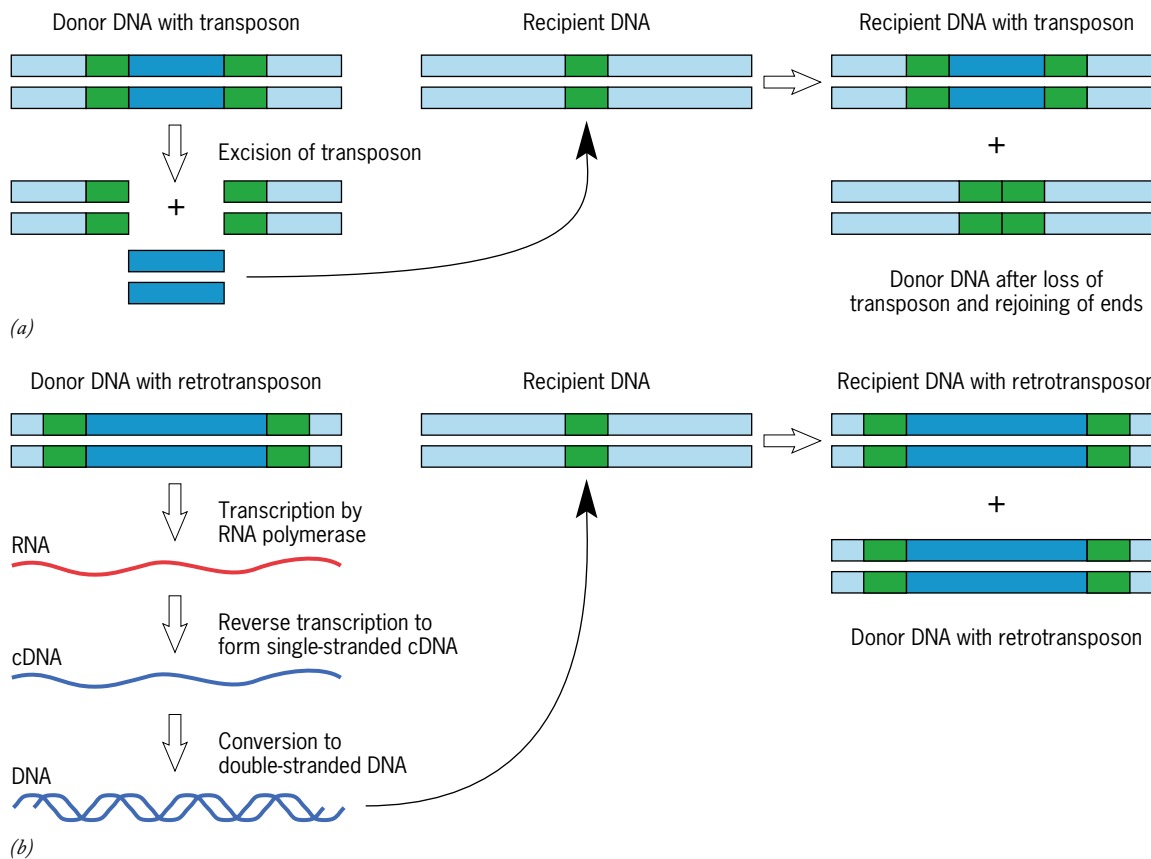
FIGURE 10.26 Schematic pathways in the movement of transposable elements. (*a*) DNA transposons move by a cut-and-paste pathway, whose mechanism is depicted in Figure 10.25. Approximately 3 percent of the human genome consists of DNA transposons, none of which are capable of transposition (i.e., all are relics left in the genome as a result of ancestral activity). (*b*) Retrotransposons move by a copy-and-paste pathway. The steps involved in retrotransposition take place both in the nucleus and cytoplasm and require numerous proteins, including those of the host. More than 40 percent of the human genome consists of retrotransposons, only a few of which (e.g., 40–100) are thought to be capable of transposition. More than one mechanism of retrotransposition is known.

of reactions leading to the excision of the transposon (step 3). The transposase–transposon complex then binds to a target DNA (step 4) where the transposase catalyzes the reactions required to integrate the transposon into its new residence (step 5). Integration of the element typically creates a small duplication in the target DNA that flanks the transposed element at the site of insertion (green segments in Figure 10.26). Target site duplications serve as "footprints" to identify sites in the genome that are occupied by transposable elements.

As originally demonstrated by McClintock, eukaryotic genomes contain large numbers of transposable elements. In fact, *at least* 45 percent of the DNA in a human cell nucleus has been derived from transposable elements! The vast majority (>99 percent) of transposable elements are incapable of moving from one place to another; they have either been crippled by mutation or their movement is suppressed by the cell. However, when transposable elements do change position, they insert widely throughout the target DNA. In fact, many transposable elements can insert themselves within the center of a protein-coding gene. Several examples of such an occurrence have been documented in humans, including a number of cases of hemophilia caused by a mobile genetic element that had "jumped" into the middle of one of the key blood clotting genes. It is estimated that approximately 1 out of 500 disease-causing mutations in humans is the result of the insertion of a transposable element.

Figure 10.26 illustrates two major types of eukaryotic transposable elements, DNA transposons and retrotransposons, and their differing mechanism of transposition. As described above for prokaryotes, most eukaryotic DNA transposons are excised from the DNA at the donor site and inserted at a distant target site (Figure 10.26*a*). This "cut-and-paste" mechanism is utilized, for example, by members of the *mariner* family of transposons, which are found throughout the plant and animal kingdoms. **Retrotransposons**, in contrast, operate by means of a "copy-and-paste" mechanism that involves an RNA intermediate (Figure 10.26*b*). The DNA of the transposable element is transcribed, producing an RNA, which is then "reverse transcribed" by an enzyme called **reverse transcriptase**, producing a complementary DNA. The DNA copy is made double-stranded and then integrated into a target DNA site. In most cases, the retrotransposon itself contains the sequence that codes for a reverse transcriptase. Retroviruses, such as the virus responsible for AIDS, use a very similar mechanism to replicate their RNA genome and integrate a DNA copy into a host chromosome.

The Role of Mobile Genetic Elements in Evolution It was noted on page 398 that moderately repeated DNA sequences constitute a significant portion of eukaryotic genomes. Unlike the highly repeated fraction of the genome (satellite, mini-satellite, and microsatellite DNA), whose sequences reside in tandem and arise by DNA duplication, most of the moderately repeated sequences of the genome are interspersed and arise by transposition of mobile genetic elements. In fact, the two most common families of moderately repeated sequences in human DNA—the *Alu* and L1 families—are retrotransposons. Recall from page 398 that there are two classes of interspersed elements, SINEs and LINEs. *Alu* is an example of the former and L1 is an example of the latter. A full-length, transposable L1 sequence (at least 6000 base pairs in length) encodes a unique protein with two catalytic activities: a reverse transcriptase that makes a DNA copy of the RNA that encoded it and an endonuclease that nicks the target DNA prior to insertion. The human genome is estimated to contain about 500,000 copies of L1, but the vast majority of these are incomplete, immobile elements. It is estimated that approximately 2 percent of humans possess a new genomic L1 insert, that is, an L1 element that was not present in the genome of either parent.

Even more abundant than L1 are the *Alu* sequences, which are interspersed at more than one million different sites throughout the human genome. *Alu* is a family of short, related sequences about 300 base pairs in length. The *Alu* sequence closely resembles that of the small RNA present in the signal recognition particles found in conjunction with membrane-bound ribosomes (page 277). It is presumed that during the course of evolution, this cytoplasmic RNA was copied into a DNA sequence by reverse transcriptase and integrated into the genome. The tremendous amplification of the *Alu* sequence is thought to have occurred by retrotransposition using the reverse transcriptase and endonuclease encoded by L1 sequences.

Given its prevalence in the human genome, one might expect the *Alu* sequence to be repeated in genomes throughout the rest of the animal kingdom, but this is not the case. Comparative genomic studies indicate that the *Alu* sequence first appeared as a transposable element in the genome of primates about 60 million years ago and has been increasing in copy number ever since. The rate of *Alu* transposition has slowed dramatically over the course of primate evolution to its current estimated rate in humans of approximately once in every 200 births. These transposition events generate differences in the locations of *Alu* sequences from one person to another and thus contribute to the genetic diversity in the human population (page 408).

When something new like a transposable element is discovered in an organism, a biologist's first question is usually: What is its function? Many researchers who study transposition believe that transposable elements are primarily "junk." According to this view, a transposable element is a type of genetic parasite that can invade a host genome from the outside world, spread within that genome, and be transmitted to offspring—so long as it doesn't have serious adverse effects on the ability of the host to survive and reproduce. Even if this is the case, it doesn't mean that transposable elements cannot make positive contributions to eukaryotic genomes. Keep in mind that evolution is an opportunistic process—there is no

preset path to be followed. Regardless of its origin, once a DNA sequence is present in a genome, it has the *potential* to be "put to use" in some beneficial manner during the course of evolution. For this reason, the portion of the genome formed by transposable elements has been referred to as a "genetic scrap yard." There are several ways that transposable elements appear to have been involved in adaptive evolution:

1. Transposable elements can, on occasion, carry adjacent parts of the host genome with them as they move from one site to another. Theoretically, two unlinked segments of the host genome could be brought together to form a new, composite segment. This may be a primary mechanism in the evolution of proteins that are composed of domains derived from different ancestral genes (as in Figure 2.36).

2. DNA sequences that were originally derived from transposable elements are found as parts of eukaryotic genes as well as the DNA segments that regulate gene expression. For example, several transcription factors, which are the proteins that regulate gene expression, bind to sites in the DNA that arose originally from transposable elements. Even when there is no direct evidence of a function, many transposable elements are very similar in position and sequence to elements in the genomes of distant vertebrate relatives. This type of evolutionary conservation suggests that these sequences carry out some beneficial role in the lives of their hosts (page 406).

3. In some cases, transposable elements themselves appear to have given rise to genes. The enzyme telomerase, which plays a key role in replicating the DNA at the ends of chromosomes (see Figure 12.20*c*), may be derived from a reverse transcriptase encoded by an ancient retrotransposon. The enzymes involved in the rearrangement of antibody genes (see Figure 17.18) are thought to be derived from a transposase encoded by an ancient DNA transposon. If this is in fact the case, our ability to ward off infectious diseases is a direct consequence of transposition.

One point is clear: transposition has had a profound impact on the genetic composition of organisms. It is interesting to note that, only a couple of decades ago, molecular biologists considered the genome to be a stable repository of genetic information. Now it seems remarkable that organisms can maintain themselves from one day to the next in the face of this large-scale disruption by genetic rearrangement. For her discovery of transposition, Barbara McClintock was the sole recipient of a Nobel Prize in 1983, at age 81, approximately 35 years after her initial discovery.

REVIEW ?

1. Describe the course of evolutionary events that are thought to give rise to multiple-gene families, such as those that encode the globins. How could these events give rise to pseudogenes? How could they give rise to proteins having entirely different functions?

2. Describe two mechanisms by which genetic elements are able to move from one site in the genome to another.

3. Describe the impact that transposable elements have had on the structure of the human genome over the past 50 million years.

10.6 SEQUENCING GENOMES: THE FOOTPRINTS OF BIOLOGICAL EVOLUTION

Determining the nucleotide sequence of all the DNA in a genome is a formidable task. During the 1980s and 1990s, the technology to accomplish this effort gradually improved, as researchers developed new vectors to clone large segments of DNA and increasingly automated procedures to determine the nucleotide sequences of these large fragments (discussed in Section 18.15). The first complete sequence of a prokaryotic organism was reported in 1995, and the first complete sequence of a eukaryote, the budding yeast *S. cerevisiae*, was published the following year. Over the next few years—as the scientific community awaited the results of work on the human genome—the genomic sequences of numerous prokaryotic and eukaryotic organisms (including a fruit fly, nematode, and flowering plant) were reported. Researchers were able to sequence these genomes with relative speed because they are considerably smaller than the human genome, which contains approximately 3.2 billion base pairs. To put this number into perspective, if each base pair in the DNA were equivalent to a single letter on this page, the information contained in the human genome would produce a book approximately 1 million pages long.

By 2001, the rough draft of the nucleotide sequence of the human genome had been published. The sequence was described as "rough" because each segment was sequenced an average of about four times, which is not enough to ensure complete accuracy, and many regions that proved difficult to sequence were excluded. The first attempts to *annotate* the human genomic sequence, that is, to interpret the sequence in terms of the numbers and types of genes it encoded, led to a striking observation concerning gene number. Researchers concluded that the human genome probably contained in the neighborhood of 30,000 protein-coding genes. Up until it had been sequenced, it had been widely assumed that the human genome contained at least 50,000 and maybe as many as 150,000 different genes.

The "finished" version of the human genome sequence was reported in 2004, which meant that (1) each site had been sequenced 7–10 times to ensure a very high degree of accuracy (at least 99.99 percent), and (2) the sequence contained a minimal number of gaps. The gaps that persisted contain regions of the chromosomes—often referred to as "dark matter"—that consist largely of long stretches of highly repeated DNA, especially those in and around the centromeres of each chromosome. Despite exhaustive efforts, these regions have proven impossible to clone, or their sequences have proven impossible to order correctly using current technology.

Despite this remarkable achievement in nucleotide sequencing, the actual number of protein-coding genes in the
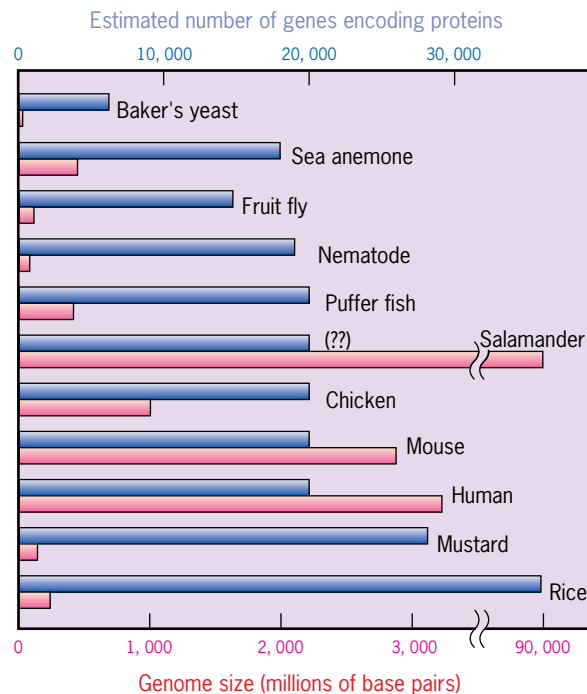
FIGURE 10.27 Genome comparisons. Among eukaryotes whose genomes have been sequenced, the number of protein-coding genes (blue bars) varies from about 6,200 in yeast to 37,000 in rice, with vertebrates thought to possess about 20,000. Whereas the number of estimated protein-coding genes varies over a modest range among eukaryotes, the amount of DNA in a genome (red bars) varies widely, reaching values of 90 billion base pairs in some salamanders (the actual gene number for these amphibians is unknown). Many of the organisms whose genomes have been sequenced (e.g., the mustard plant *Arabidopsis* and the puffer fish *Fugu*) were selected because they possess particularly compact genomes. (See *Nature Revs. Gen.* 9:689, 2008, for a discussion of gene numbers in animals.)

human genome remains uncertain. Identification of genes using various computer software programs (algorithms) has been plagued with difficulties, and, in fact, the earlier estimate of 30,000 protein-coding human genes has been revised steadily downward. Although it has come as a shock to most biologists, current estimates place the number in the neighborhood of 20,000! This means that humans have roughly the same number of protein-coding genes as a microscopic worm or a mustard plant, as well as most other vertebrates, including a puffer fish (*Fugu*), chicken, or mouse (Figure 10.27).[4]

If the differences in complexity between organisms cannot be explained by the number of protein-coding genes in their genomes, how can they be explained? We really don't

---

[4]It can also be seen from Figure 10.27 that there is very little correlation between the number of protein-coding genes and the total amount of DNA in the genome. The puffer fish, for example, which has about the same number of genes as other vertebrates, has a genome that is about one-eighth the size of its human counterpart. At the other end of the spectrum, the genome of certain salamanders is roughly 30 times the size of the human genome. The contrasts in genome size among vertebrates reflects a striking difference in the content of noncoding, largely repetitive DNA. The evolutionary significance of these differences is unclear.

have a very good answer to that question, but we can list a few possibilities to consider.

1. As will be described in Chapter 12, a single gene can encode a number of related proteins as a result of a process termed *alternative splicing* (see Section 12.5). Several recent studies suggest that over 90 percent of human genes might engage in alternative splicing, so that the actual number of proteins encoded by the human genome is at least several times greater than the number of genes it contains. It is likely that greater differences between organisms will emerge when this and other "gene-enhancing" mechanisms are explored further. This expectation is consistent with the observations that vertebrate genes tend to be more complex (i.e., contain more exons) than those of flies and worms and exhibit a higher incidence of alternative splicing.

2. Molecular biologists have devoted an enormous effort to studying the mechanisms that regulate gene expression. Despite this effort, our understanding of these mechanisms is very limited. We have learned, for example, that a large portion of the genome is transcribed into a bewildering array of RNAs, but we have very little information as to what most of these RNAs are doing within the cell. There is a growing belief that many of these RNAs have a gene regulatory function. There is also growing evidence that the number and complexity of these noncoding RNAs can be correlated with the levels of complexity of diverse organisms. In one recent study, for example, researchers sought to identify the number of microRNAs produced by various organisms. As discussed in the next chapter, microRNAs are one of the best studied regulatory RNAs. Researchers found that sponges expressed about 10 different microRNAs, and sea anemones approximately 40. This compares to approximately 150 identified in worms and fruit flies and at least one thousand in humans. This is not to imply that the number of microRNAs is the primary determinant of morphological complexity. Rather, it suggests that we will have to learn a great deal more about gene regulation if we are going to understand the basis of biological diversity.

3. Over the past decade or so, a new area of biological study (called systems biology) has emerged that focuses on the ways that proteins work together as complex networks rather than as individual actors. A very simple example of a protein network was presented in Figure 2.41. Given that cells produce thousands of different proteins, with varying degrees of interaction, these networks can become very complex and dynamic. A relatively small increase in the number of elements that make up a network, or an increase in the size and complexity of those elements, could dramatically increase the complexity of the network and thus the complexity of the entire organism.

Numerous other factors could be added to this discussion, but the general point is clear: The apparent difference in complexity between different groups of multicellular organisms is less a matter of the amount of genetic information in an organism's genome than that of the manner in which that information is put to use.

Comparative Genomics: "If It's Conserved, It Must Be Important"

Consider the following facts: (1) The majority of the genome consists of DNA that resides between the genes, and thus represents intergenic DNA, and (2) each of the roughly 20,000 or so protein-coding human genes consists largely of noncoding portions (intronic DNA). Taken together, these facts indicate that the protein-coding portion of the genome represents a remarkably small percentage of the total DNA (estimated at about 1.5 percent). Most of the intergenic and intronic DNA of the genome does not contribute to the survival and reproductive capability of an individual, so there is no selective pressure to maintain its sequence unchanged. As a result, most intergenic and intronic sequences tend to change rapidly as organisms evolve. In other words, these sequences tend not to be conserved. In contrast, those segments of the genome that encode protein sequences or contain regulatory sequences that control gene expression (see Figure 12.40) are subject to natural selection. Natural selection tends to eliminate individuals whose genome contains mutations in these functional elements.[5] As a result, these sequences tend to be conserved. It follows from these comments that the best way to identify functional sequences is to compare the genomes of different types of organisms.

Despite the fact that humans and mice have not shared a common ancestor for about 75 million years, the two species share similar genes, which tend to be clustered in a remarkably similar pattern. For example, the number and order of human globin genes depicted in Figure 10.23 is basically similar to that found on a comparable segment of DNA in the mouse genome. As a result, it becomes possible to align corresponding regions of the mouse and human genomes. Human chromosome number 12, for example, is composed of a series of segments, in which each segment corresponds roughly to a block of DNA from a mouse chromosome. The number of the mouse chromosome containing each block is indicated.

Human chromosome 12

Even a quick examination of this drawing of a human chromosome reveals the dramatic changes in chromosome structure that occur as evolution generates new species over time. Blocks of genes that are present on the same chromosome in one species can become separated into different chromosomes in a subsequent species. Over time, chromosome numbers can increase or decrease as chromosomes split apart or fuse together. Taken by themselves, these changes in the positions of genes have very little effect on the phenotypes of the organisms, but they provide a clear visual footprint of the evolutionary process.

---

[5]One can recognize two opposing sides of natural selection. Negative or purifying selection acts to maintain (conserve) sequences with important functions as they are because changes reduce the likelihood that the individual will survive and reproduce. These regions of the genome evolve more slowly than nonfunctional sequences whose change has no effect on the fitness of the individual and are not subject to natural selection (such sequences are said to evolve neutrally). In contrast, positive or Darwinian selection promotes speciation and evolution by selecting for changes in sequence that cause individuals to be better adapted to their environment and thus more likely to survive and reproduce. These regions evolve more rapidly than nonfunctional segments.

When homologous segments of the mouse and human genome are aligned on the basis of their nucleotide sequences, it is found that approximately 5 percent of DNA sequences are highly conserved between the two species. This is a considerably higher percentage than would have been expected by combining the known protein-coding regions and gene regulatory regions (together accounting for roughly 2 percent of the genome). If we adhere to one of the foremost principles of molecular evolution: "if it's conserved, it must be important," then these studies are telling us that parts of the genome that were presumed to consist of "useless" sequences actually have important, unidentified functions. Some of these regions undoubtedly encode RNAs that have various regulatory functions (discussed in Section 11.5). Others likely have "chromosomal functions" rather than "genetic functions." For example, these conserved sequences could be important for chromosome pairing prior to cell division. Whatever the function, these genomic elements are often located at great distances from the closest gene and include some of the most conserved sequences ever discovered, exhibiting virtual identity between human, rat, and mouse genomes.

By comparing stretches of the genome between two distantly related species, such as the human and mouse, one can identify those regions that have been conserved for tens of millions of years. However, this approach won't identify functional sequences in the human genome that have a more recent evolutionary origin. This could include, for example, genes that are present in humans and absent from mice, or regulatory regions that have changed their sequence over the course of primate evolution to allow them to bind new regulatory proteins. Regions of the genome that fall into one of these categories are best identified by comparing sequences in closely related species. In one study, for example, several corresponding segments of the genome from 17 different primate species were compared. These efforts produced the type of data illustrated in Figure 10.28. Those parts that are highly conserved (indicated by the arrows at the bottom of the figure) correspond to protein-coding segments (exons) as well as smaller regions that are likely to bind gene regulatory proteins. This figure shows that a comparison of sequences from just two or three different species would not have been sufficient to identify the most highly conserved sequences within these genomes. A concerted effort (called the ENCODE Project) is underway to identify all of the functional elements present in the human genome. Unfortunately, we lack the knowledge necessary to recognize many of these functional elements, which obviously confounds the task. One point has become clear from recent studies: a significant proportion of functional DNA sequences are constantly evolving and thus are not highly conserved. In other words, if we restrict our search to broadly conserved sequences, we run the risk of missing many of the most important functional elements in the genome.

The Genetic Basis of "Being Human"

By focusing on conserved sequences, we tend to learn about characteristics that we share with other organisms. If we want to better understand our own unique biological evolution, we
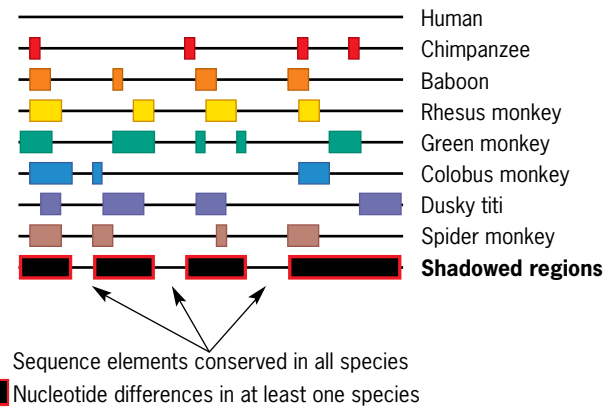
FIGURE 10.28 **Small segments of DNA are highly conserved between humans and related species.** The top line represents the nucleotide sequence of a segment of human DNA. Each line below it represents the nucleotide sequence from another primate, as indicated on the right. Those blocks on each line represent segments whose nucleotide sequence varies from that of the corresponding human segment. The bottom line shows those parts of this DNA segment that are highly conserved in all of the species shown, that is, those parts whose sequence does not vary in any of the eight species represented. (FROM R. A. GIBBS AND D. L. NELSON, SCIENCE 299:1333, 2003; COPYRIGHT © 2003, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

have to look more closely at those parts of the genome that distinguish us from other organisms. Chimpanzees are our closest living relative, having shared a common ancestor as recently as 5–7 million years ago. It was thought that a detailed analysis of the differences in DNA sequence and gene organization between chimpanzees and humans might tell us a great deal about the genetic basis for recently evolved features that make us uniquely human, such as our upright walk and advanced use of tools and language. These latter characteristics can be traced to our brain, which has a volume of about 1300 cm^3—nearly four times that of a chimpanzee.

The rough draft of the chimpanzee genome was reported in 2005. Overall, the chimpanzee and human genomes differ by roughly 4 percent, which amounts to tens of millions of differences—a level of divergence that is considerably greater than expected from preliminary studies. While some of this divergence is due to single nucleotide changes between the two genomes, most is attributed to larger differences, such as deletions and segmental duplications (see footnote, page 400).

Researchers have been able to identify hundreds of genes in the human lineage that are evolving more quickly than the background (or neutral) rate, presumably in response to natural selection. However, it remains unclear which, if any, of these genes contribute to "making us human." Some of the fastest evolving genes encode proteins involved in regulation of gene expression (i.e., transcription factors). These are precisely the types of genes that would be expected to generate major phenotypic differences because they can affect the expression of large numbers of other genes. In fact, differences between chimp and human transcription factors are presumably responsible for the differences in expression of brain proteins seen in Figure 2.48. Many other differences have also been described,

including alterations in certain RNA-encoding genes that appear to affect brain development. One of the genes in this latter category, called *HAR1*, stretches for a mere 118 base pairs but contains 18 base substitutions that distinguish it from the corresponding region in the chimp genome. This same region is highly conserved among other vertebrates, so it is only during the evolution of humans that it has experienced such pronounced effects of natural selection (*HAR* stands for *H*uman *A*ccelerated *R*egion). Although the function of *HAR1* is unknown, the fact that it is expressed primarily in the developing fetal brain makes it a gene of interest to researchers trying to unravel the genomic basis of human brain expansion.

Another gene of interest encodes the starch-digesting enzyme amylase that is present in saliva. Chimps have a relatively low-starch diet and have one copy of the amylase 1 (*AMY1*) gene in their genome (Figure 10.29*a*). During the evolution of humans, there has been an apparent selection for an increased number of copies of the *AMY1* gene, which has resulted in an increased concentration of the enzyme in human saliva. One might have predicted that an increased amylase level would have been accomplished during evolution by an increase in the level of expression of the single existing *AMY1* gene in the genome, but in this particular case it has resulted from gene duplication. As seen in Figure 10.29*b* and discussed in the next section, the number of copies of the *AMY1* gene is variable. In fact, the number of copies of this gene tends to be higher in human populations that ingest greater quantities of starch in their diet.

To date, the greatest interest in this field has focused on a gene encoding a transcription factor named *FOXP2*. A comparison of the FOXP2 protein between humans and chimps shows two amino acid differences that have appeared in the human lineage since the time of separation from our last common ancestor. What makes this gene so interesting is that persons with mutations in the *FOXP2* gene suffer from a severe speech and language disorder. Among other deficits, persons with this disorder are unable to perform the fine muscular movements of the lips and tongue that are required to engage in vocal communication. Calculations suggest that the changes in this "speech gene" that distinguish it from the chimp version became "fixed" in the human genome within the past 120,000 to 200,000 years, a time when modern humans are thought to have emerged. These findings suggest that changes in the *FOXP2* gene may have played an important role in human evolution.

We have focused in this section on genomic differences between humans and other mammals because this is the subject of the present chapter. It is probably evident from the limited nature of this discussion that researchers have not yet made much progress in identifying the genetic basis of "being human." Whether that will change in the next few years remains to be seen. Keep in mind that, even at the molecular level, many nongenomic factors that enter into the topic of "what makes us human" have not been considered.

Genetic Variation Within the Human Species Population

No one else in the world looks exactly like you because no one else, other than an identical twin, has the same DNA sequence throughout their genome. The human genome that was sequenced by the original Human Genome Project was derived largely from a single human male. Since the completion of that sequence, a great deal of attention has been focused on how DNA sequence varies within the human population. **Genetic polymorphisms** are sites in the genome that vary among different individuals. The term usually refers to a genetic variant that occurs in at least 1 percent of a species population. The concept of genetic polymorphisms began with the discovery in 1900 by an Austrian physician, Karl Landsteiner, that people could have at least three alternate types of blood, A, B, or O. As discussed on page 126, the ABO blood

(a)

(b)
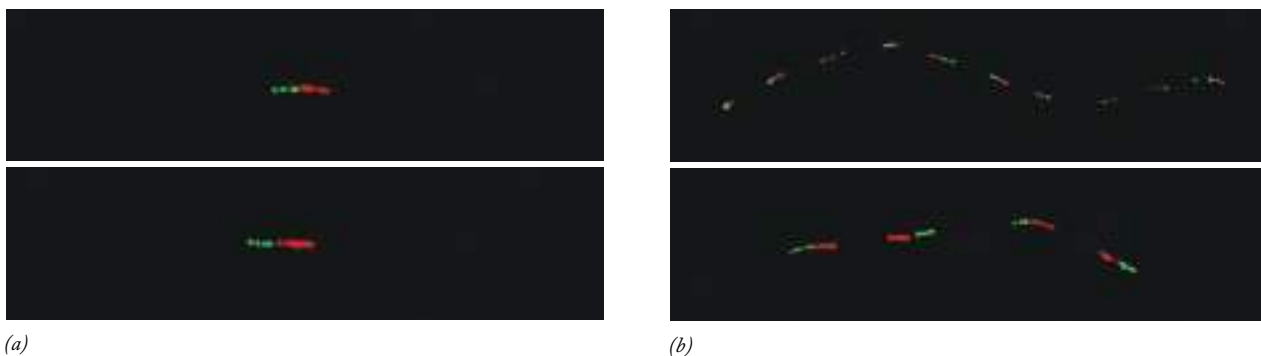
FIGURE 10.29 Duplication of the amylase gene during human evolution. In the results depicted here, a pair of homologous chromosomes from a chimp (*a*) or a human (*b*) have been hybridized to red and green colored probes that bind to different portions of the *AMY1* gene. This gene encodes the starch-digesting enzyme salivary amylase. The number of copies of the *AMY1* gene on each chromosome is revealed by the number of times the fluorescent probes are repeated. The chimp has one copy of the *AMY1* gene on each chromosome (i.e., one copy per genome), whereas humans have a number of copies. The number of copies of the gene varies within the human genome as illustrated here by the fact that one of the two homologous chromosomes from this individual has 4 copies of the gene and the other homologous chromosome has 10 copies. This is an example of a *copy number variation* (page 409). Moreover, the number of copies of the *AMY1* gene in the genomes of a given human population tends to correlate with the amount of starch in the diet of that population. This correlation strongly suggests that the copy number of the *AMY1* gene has been influenced by natural selection. (Reprinted from George H. Perry et al., courtesy of Nathaniel J. Dominy, Nature Gen. 39:1257, 2007 © copyright 2007 by Macmillan Magazines Limited.)

group is a result of members of the population having different alleles of a gene encoding a sugar-transferring enzyme. Now that we are in possession of our own genomic sequence, we are in a position to search for new types of genetic variation that would never have been revealed in the "pregenomic era."

DNA Sequence Variation   The most common type of genetic variability in humans occurs at sites in the genome where single nucleotide differences are found among different members of the population. These sites are called **single nucleotide polymorphisms (SNPs)**. The vast majority of SNPs (pronounced "snips") occur as two alternate alleles, such as A or G. Most SNPs are thought to have arisen by mutation only once during the course of human evolution and are shared by individuals through common descent. Commonly occurring SNPs have been identified by comparing DNA sequences from the genomes of hundreds of different individuals from diverse ethnic populations. On average, two randomly selected human genomes have about 3 million single nucleotide differences between them, or one every thousand base pairs. Although this might appear to be a large number, it means that humans are, on average, 99.9 percent similar to one another with respect to nucleotide sequence, which is probably much greater than most mammalian species. This sequence similarity reflects the fact that we are apparently a young species, in evolutionary terms, and our prehistoric population size was relatively small. Of the 15 million or so *common* SNPs in the genome, it is esti-

mated that approximately 60,000 are found within protein-coding sequences and lead to amino acid replacements within the encoded protein. This corresponds to about 2–3 amino acid substitutions per gene. It is this body of genetic variation that is thought to be largely responsible for the phenotypic variability observed among humans. The impact of human genome sequence variation on the practice of medicine is discussed in the accompanying Human Perspective.

Structural Variation   As illustrated in Figure 10.30*a*, segments of the genome can change as the result of duplications, deletions, insertions, inversions (when a piece of DNA exists in a reversed orientation), and other events. These types of changes typically involve segments of DNA ranging from thousands to millions of base pairs in length. Because of their relatively large size, these types of polymorphisms are called *structural variants*, and we are just beginning to grasp the extent and importance of their presence.

It has been known since the early days of microscopic chromosome analysis that not all healthy people have identical chromosome structures. Figure 10.30*b* shows an example of a common chromosome inversion whose existence has been known for over 30 years. Recent studies have revealed that intermediate-sized structural variants—those that are too small to be seen under the microscope and too large to be readily detected by conventional sequence analysis—are much more common than previously thought. In one study, for example, researchers compared the genome of a "normal" human with the original sequence generated by the Human Genome Project and found that they differed by 297 intermediate-sized (>8 kilobases) structural variants: 139 insertions, 102 deletions, and 56 inversions. This is a remarkable and unexpected level of large-scale genomic variation.

Copy Number Variation   As discussed on page 395, the technique of DNA fingerprinting depends on differences in the

(a)

(b)
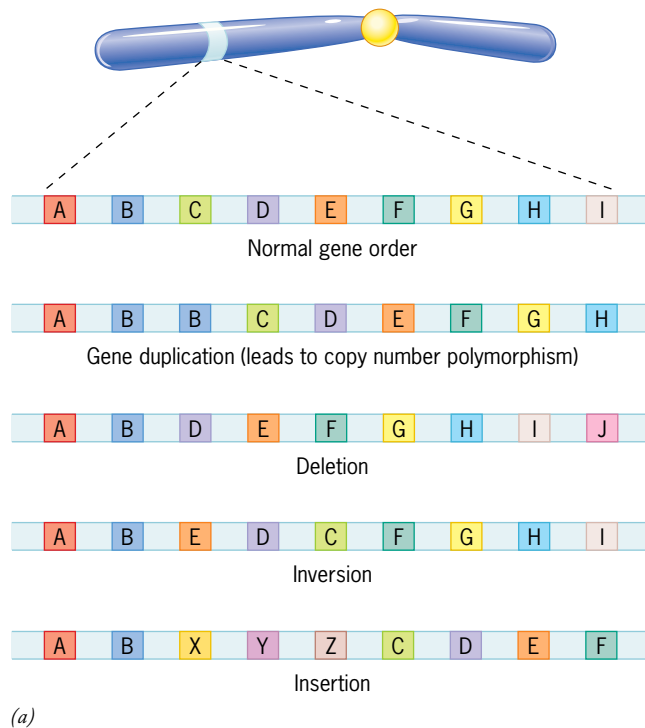
FIGURE 10.30   Structural variants. (*a*) Schematic representation of the major types of genomic polymorphisms that involve a significant segment of a chromosome (e.g., thousands of base pairs). Most of these polymorphisms are too small to be detected by microscopic examination of chromosomes but are readily detected when the number and/or chromosomal location of individual genes are determined. (*b*) On the left is a normal human chromosome #9, and on the right is a chromosome #9 containing a large inversion that includes the chromosome's centromere (arrow). This inversion, which is clearly visible microscopically, is present in 1–3 percent of the population. (B: REPRINTED WITH PERMISSION FROM C. LEE, NATURE GEN. 37:661, 2005 © COPYRIGHT 2005 BY MACMILLAN MAGAZINES LIMITED.)
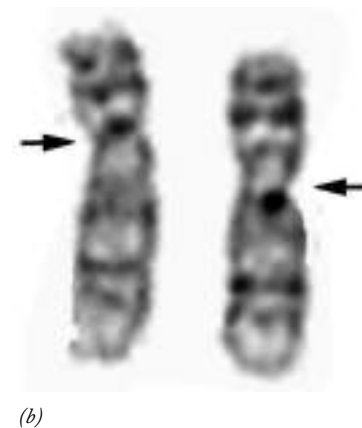
lengths of minisatellite sequences, which in turn depend on the number of copies of the sequence that are present at particular sites in the chromosomes. This is an example of a *copy number variation* (or *CNV*). Recent findings have revealed that much larger-sized CNVs (>1 kb) are also common in the human population, affecting thousands of different regions of the genome, including large numbers of protein-coding genes. These large-sized CNVs are a type of structural variation that results from a duplication or deletion of a DNA segment. Because of such CNVs, many of us carry extra copies of one or more genes that encode important physiological proteins. Extra copies of genes are generally associated with overproduction of a protein, which can disrupt the delicate biochemical balance

that exists within a cell. It is found, for example, that a significant number of persons who develop early-onset Alzheimer's disease possess extra copies of the *APP* gene, which encodes the protein thought to be responsible for the disease (page 66). In some cases, extra copies of a gene can be beneficial, as illustrated by the repeated duplication of the amylase gene (*AMY1*), which has been found in certain human populations that have a high starch content in their diet (Figure 10.29*b*). Figure 10.29*b* also provides an excellent example of a CNV in that it shows the number of *AMY1* genes on both copies of one person's homologous chromosomes. It can be seen that one chromosome contains only 4 copies of the *AMY1* gene, whereas the homologous chromosome contains 10 copies of the gene.

THE HUMAN PERSPECTIVE

Application of Genomic Analyses to Medicine

Over the past several decades, more than a thousand genes have been identified as the cause of rare, heritable diseases. In the vast majority of cases, the gene responsible for the disease being investigated was identified through traditional genetic linkage studies. Such studies begin by locating a number of families whose members exhibit a disproportionately high frequency of a particular disease. The initial challenge is to determine which altered region of the genome is shared by all affected members of the family. Once a region linked to the disease gene is identified, the DNA of that segment can be isolated and the mutant gene pinpointed. This type of genetic approach is well suited for the identification of genes that have very high penetrance, that is, genes whose mutant form virtually guarantees that the individual will be afflicted with the disorder. The mutated gene responsible for Huntington's disease (page 396), for example, exhibits very high penetrance. Furthermore, no other gene defect in the genome causes this disease.

Most of the common diseases that afflict our species, such as cancer, heart failure, Alzheimer's disease, diabetes, asthma, depression, and various age-related diseases, have a genetic component, which is to say that they show at least some tendency to run in families. But unlike inherited disorders such as Huntington's disease, there is no single gene that is clearly linked with the condition. Instead, numerous genes are likely to affect the disease risk, with each of them making only a small contribution to the overall likelihood of developing the disorder. In addition, the development of the disorder is often influenced by nongenetic (i.e., environmental) factors. The likelihood of developing diabetes, for example, is greatly increased in individuals who are seriously overweight, and the likelihood of developing lung cancer is greatly increased by smoking. One of the goals of the medical research community is to identify genes that contribute to the development of these common, but genetically complex, diseases.

Because of their low penetrance, most genes that increase the risk of developing common diseases cannot be identified through family linkage studies.[a] Instead, such genes are best identified by analyzing

disease occurrence in large populations. To carry out this type of investigation, researchers compare the genotypes of individuals who have the disease in question with individuals of a similar ethnic background who are not afflicted. The goal is to discover an *association* (or correlation) between a particular condition, such as diabetes, and common genetic polymorphisms. The potential value of this approach is illustrated by the discovery in the early 1990s of a strong association between one common allele of the gene encoding the lipoprotein ApoE and the likelihood of developing Alzheimer's disease. It was found in these studies that individuals who have at least one copy of the *APOE4* allele of the gene are much more likely to develop this disabling neurodegenerative disease than persons lacking the allele. These findings have opened a major branch of investigation into the relationship between cholesterol metabolism, cardiovascular health, and Alzheimer's disease.

As the name suggests, genome-wide association studies look for associations between a disease state and polymorphisms that may be located anywhere in the genome. This requires determining the nucleotide sequence of large parts of the genome of all the subjects participating in the study. As noted earlier, the most common types of genetic variability in humans are single nucleotide differences (SNPs), which are spread almost uniformly throughout the genome. It is likely that many of the SNPs that change the coding specificity of a gene, or the regulation of a gene's expression, play an important role in our susceptibility to complex diseases. As the cost of genotyping human DNA samples has decreased dramatically, researchers have begun scanning the genomes of large numbers of people to identify SNPs that occur more often in people afflicted with a particular disease than in healthy individuals. Even if the identified SNPs are not directly responsible for the disease, they serve as genetic markers for a nearby locus that might be responsible. These principles are well illustrated by a landmark genome-wide association study published in 2005 on age-related macular degeneration (AMD), which is the leading cause of blindness in the elderly. Initially, the researchers compared 96 AMD cases with 50 controls, looking for an association of the disease with over 100,000 SNPs that had been genotyped in the subjects. They found a strong association between individuals with the disease and a common SNP present within the intron (noncoding section) of a gene called *CFH* that is involved in immunity and inflammation. Persons homozygous for this SNP had a 7.4-fold greater risk of having the disease. Once they

---

[a]Not all genes involved in common diseases have a low penetrance. Alzheimer's disease, for example, occurs with high frequency in persons having certain *APP* alleles (page 66) and breast cancer occurs with high frequency in persons having certain *BRCA* alleles (Section 16.3). But these types of rare, high-penetrance polymorphisms are not major factors in the development of the large majority of cases for either of these diseases.

had pinned down this region of the genome as being associated with AMD, they sequenced the entire *CFH* gene in 96 individuals. They found that the SNP they had identified was tightly linked to a polymorphism within the coding region of the *CFH* gene that placed a histidine at a particular site in the protein. Other alleles of the CFH protein that were not associated with a risk for AMD had a tyrosine at this position. These findings provide additional evidence that inflammation is an underlying factor in the development of AMD and point out a clear target in the search for treatments of the disease.

A large number of genome-wide association studies have been conducted over the past few years. In many of these studies, the genomes of thousands of individuals have been scrutinized, and the results have been confirmed in independent analyses of separate case and control populations. These studies have become practical with the availability of commercially produced chips that house hundreds of thousands of DNA snippets containing the most common SNPs in the human population. Researchers can incubate the chip with a DNA sample to be tested and quickly determine which of the possible SNPs is present at each genomic location in that person's DNA. As a result of these studies, we now have lists of genes (and noncoding DNA regions) for which certain alleles or variants increase the likelihood that an individual will develop various diseases, including heart disease, Crohn's disease, types 1 and 2 diabetes, and several different types of cancer.

When researchers began these association studies, it was hoped that the studies would reveal new insights into the underlying basis of common diseases by turning up genes not previously suspected of being involved in particular diseases. The products of such genes may ultimately become targets of new avenues of drug therapy, just as identification of the importance of the HMG CoA reductase gene and LDL receptor gene in the disease familial hypercholesterolemia led to the development of the cholesterol-lowering statins (page 312). Although these genetic association studies have not yet led to the creation of new drugs, they have uncovered important mechanisms and pathways involved in the development of numerous common diseases. To cite just one example, several of the susceptibility alleles that contribute to type 2 diabetes encode proteins that function in the insulin-secreting cells of the pancreas, which has focused renewed attention on the dysfunction of these cells in the development of this disease. (Readers interested in examining the results of these studies can find them at the user-friendly site, www.snpedia.com.)

Despite the fact that these genome-wide association studies have proven successful in turning up genes that contribute to common diseases, they have also left many clinical geneticists disappointed. For many years, we have known the degree to which heredity contributes to the overall susceptibility of most common diseases. This measure has been obtained primarily by comparing the incidence of a given disease in members of the same family, as compared to members of the general population. When the results of the recent genome-wide association studies are examined, the vast majority of the "susceptibility alleles" that have been identified lead to only a small increase in the risk of developing a particular disease, typically less than 1.5 times the odds without the allele. If one adds up the increased risk a person would bear if he or she carried all of the identified susceptibility alleles for a particular disease, such as Alzheimer's disease or heart failure, it does not come close to accounting for the contribution that genetics is known to play in the development of that disease. In other words, there is a great deal of "missing heritability" that is yet to be discovered. Many factors could be responsible for the fact that most of the genetic risk factors for common diseases have yet to be identified. For example, a very large number of common alleles may contribute such a small risk (i.e., have a very low penetrance) that they have not been de-

tected in these association studies. If this turns out to be the case, then identification of such low-penetrant variants may be of little practical value. On the opposite end of the scale, there may be a large number of very rare alleles that significantly raise the risk of disease (i.e., they are moderately penetrant). However, because they are rare, this group of alleles may also have been overlooked in the current genome-wide association studies. Another strong possibility is that much of the risk for disease is conveyed by structural variations (page 409) that are not subject to detection in association studies that rely on SNPs.

Despite the present difficulties, it may be possible, one day, to determine whether a person is genetically predisposed to develop a particular disease simply by identifying which nucleotides are present at key positions within that person's genome. Several companies have already begun to put this concept into practice; send them a sample of your DNA and for a thousand dollars or more, they will provide you with a survey of the SNPs you carry that increase your risk of developing a number of common diseases. It is argued that by knowing this information you can modify your lifestyle with the goal of preventing the subsequent development of a particular disorder. Given the limitations of the present data, however, many geneticists have been highly critical of these services.

Information about genetic variation may also provide an indication as to how persons will react to a particular drug, whether they are likely to be helped by it and/or whether they may develop serious side effects. To cite just one example, individuals who carry an allele with two specific SNPs in a gene encoding the enzyme TPMT are unable to metabolize a class of thiopurine drugs that are commonly used to treat a type of childhood leukemia. Persons who are homozygous for this allele are at very high risk of developing a life-threatening suppression of their bone marrow when treated with normal doses of thiopurine drugs. Most diseases can be treated by a number of alternate medications, so that proper drug selection is an important aspect of the practice of medicine. The pharmaceutical industry is hoping that SNP data will eventually lead to an era of "customized drug therapy," allowing physicians to prescribe specific drugs at specific doses that are tailored to each individual patient based on their genetic profile. This era may have begun with the FDA approval of a DNA-containing chip that allows physicians to screen a patient's DNA to determine which variants of two different cytochrome P-450 genes they possess. These genes help determine how efficiently a person is able to metabolize various drugs, ranging from painkillers and antidepressants to over-the-counter heartburn medications.

The use of SNP data in association studies presents a formidable challenge because of the sheer number of these polymorphisms within the genome. As researchers began to study the distribution of SNPs in different human populations, they made a striking discovery that has made association studies much simpler: large blocks of SNPs have been inherited together as a unit over the course of recent human evolution. To illustrate: if you were told the particular base that resided at each of a handful of polymorphic sites in a particular region of a chromosome, then the bases at all of the other polymorphic sites in the region could be predicted with reasonably high (e.g., 90 percent) accuracy. Stretches of SNPs are thought to have remained together in the genome over a large number of generations because genetic recombination (i.e., crossing over) does not occur randomly along the DNA, as was suggested in the discussion of gene mapping on page 384. Instead, there are short segments of DNA (1–2 kb) where recombination is likely to occur and blocks of DNA between these "hot spots" that have a low frequency of recombination. As a result, certain blocks of DNA (typically about 20 kilobases long) tend to remain intact as they are transmitted from generation
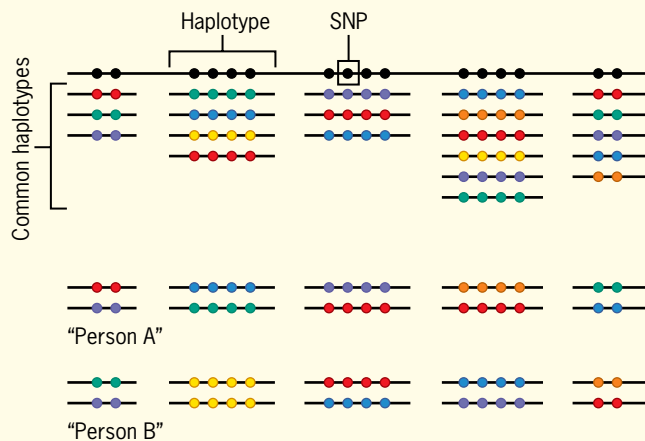
FIGURE 1 **The genome is divided into blocks (haplotypes).** The top line shows a hypothetical segment of DNA containing a number of SNPs (each SNP is indicated by a filled circle). This particular segment consists of five haplotypes separated by short stretches of DNA that are highly variable. Each haplotype occurs as a small number of variants. Each haplotype shown here exists as three to six different variants. Each haplotype variant is characterized by a specific set of SNPs, indicated by the colored circles. All of the SNPs of a particular haplotype variant are drawn in the same color to indicate that they are inherited as a group and are found together in different members of the population. Each person represented at the bottom of the illustration has a particular combination of haplotypes on each of their two chromosomes. Some haplotype variants are found in many different ethnic groups; others have a more restricted distribution.

to generation. These blocks are called **haplotypes**. Haplotypes are like "giant multigenic alleles"; if you select a particular site on a particular chromosome, only a limited number of alternate haplotypes can be found in that region (Figure 1). Each of the alternate haplotypes is defined by a small number of SNPs (called "tag SNPs") in that region of the genome. Once the identity of a handful of tag SNPs within a haplotype has been determined, the identity of the entire haplotype is known.

The average length of haplotypes and the number of alternate versions of each haplotype vary among different ethnic populations. For example, persons of Northern European descent have longer haplotypes, with fewer alternate versions, than persons of African descent. This finding suggests that the former group has arisen from a relatively small population in which many of the haplotypes of their ancestors had been lost. According to one "calculated guess," as few as 50 individuals living 27,000 to 50,000 years ago gave rise to all present-day Northern Europeans. African populations exhibit the greatest variety of haplotypes, in keeping with other studies suggesting that the modern human species arose in Africa (page 395).

In 2002 about 25 groups of investigators began a collaboration, called the International HapMap Project, aimed at identifying and mapping the various haplotypes that exist within the human population. The HapMap would contain all of the common haplotypes (i.e., haplotypes present in at least 5 percent of the population) that were present along the length of each of the 24 different human chromosomes in 270 members of four ethnically different populations. The populations to be examined were the Yoruba of Africa, Han Chinese, Japanese, and Western European (descendants who had settled in Utah). The project was completed in 2006 and resulted in publication of a HapMap built on more than four million tag SNPs spaced evenly throughout the genome.

Now that the HapMap is available, researchers are able to identify associations between a particular disease and a particular haplotype. Once such an association is made, the region of the genome containing the haplotype can be scrutinized for the particular gene(s) that contribute to disease susceptibility. The analysis of HapMaps has also provided data on the origins and migrations of human populations and has yielded valuable insights into the factors that have shaped the human genome. This is illustrated by the following example. Whether or not we are able to drink milk as an adult without upsetting our stomach (i.e., whether or not we are lactose tolerant) depends on which alleles of the lactase gene we carry in our genome. Lactose tolerance is associated with an unusually long haplotype that contains an allele of the lactase gene that is persistently expressed into adulthood. This particular haplotype is present at high frequency in European populations, which have had a long history of raising dairy animals, and is rare in most sub-Saharan African and Southeast Asian populations, which historically did not raise such animals. Its high frequency among Europeans suggests that this particular haplotype was under strong positive selective pressure in populations that depended on dairy products for nutrition. This single haplotype is over 1 million bases long, which indicates that it has not been together as a block long enough for crossing over to have had a chance to break it into smaller segments. In fact, it is estimated that this haplotype was subjected to strong selective pressure between 5,000 and 10,000 years ago, about the time that dairy farming is thought to have arisen. It is interesting to note that several East African populations that engage in dairy farming also exhibit persistent lactase expression. The mutations responsible for the lactose-tolerant phenotype in these Africans are different from the mutation found in Europeans, indicating that the trait evolved independently in the two groups and represents a good example of convergent evolution within the human lineage.

If we look ahead to the future, the greatest advances in "genomic medicine" are likely to spring from the current progress being made in DNA-sequencing technology. The Human Genome Project, which involved the efforts of more than a thousand researchers working for several years and led to the first sequence of the human genome, is estimated to have cost in the neighborhood of one billion dollars. By 2009, several companies have claimed to be able to sequence an entire human genome for a cost of $10,000 to $20,000. This dramatic reduction in cost and effort has been made possible by the development of a new generation of methods to determine DNA sequences. These efforts are expected to continue over the next few years and to culminate in the development of routine laboratory procedures for sequencing an individual's genome for a few thousand dollars or less. If this goal is achieved, we will all be able to carry around a CD containing a list of the letters that make up our complete genotype. Even if no one is interested in learning about his or her unique genetic identity, researchers should be able to use the information obtained to identify virtually all of the sites in the genome that contribute to human disease. In fact, an international collaboration began in 2008 to compare the complete genome sequences of at least one thousand individuals. This "1000 Genome Project" should allow researchers to make direct associations between a particular genomic variant and a particular trait instead of having to rely on the indirect genome-wide association studies that identify only the tag SNPs of the HapMap (which are not the actual causative defect). Moreover, this type of direct sequencing project will allow researchers to study the contribution to disease susceptibility of both structural variants (page 409) and rare SNPs (those that occur in less than 5 percent of the population). Neither of these types of variants is being included in current genome-wide association studies.

EXPERIMENTAL PATHWAYS

The Chemical Nature of the Gene

Three years after Gregor Mendel presented the results of his work on inheritance in pea plants, Friedrich Miescher graduated from a Swiss medical school and traveled to Tübingen, Germany, to spend a year studying under Ernst Hoppe-Seyler, one of the foremost chemists (and possibly the first biochemist) of the period. Miescher was interested in the chemical contents of the cell nucleus. To isolate material from cell nuclei with a minimum of contamination from cytoplasmic components, Miescher needed cells that had large nuclei and were easy to obtain in quantity. He chose white blood cells, which he obtained from the pus in surgical bandages that were discarded by a local clinic. Miescher treated the cells with dilute hydrochloric acid to which he added an extract from pig's stomach that removed protein (the stomach extract contained the proteolytic enzyme pepsin). The residue from this treatment was composed primarily of isolated cell nuclei that settled to the bottom of the vessel. Miescher then extracted the nuclei with dilute alkali. The alkali-soluble material was further purified by precipitation with dilute acid and reextraction with dilute alkali. Miescher found that the alkaline extract contained a substance that had properties unlike any previously discovered: the molecule was very large, acidic, and rich in phosphorus. He called the material "nuclein." His year up, Miescher returned home to Switzerland, while Hoppe-Seyler, who was cautious about the findings, repeated the work. With the results confirmed, Miescher published the findings in 1871.[1]

Back in Switzerland, Miescher continued his studies of the chemistry of the cell nucleus. Living near the Rhine River, Miescher had ready access to salmon that had swum upstream and were ripe with eggs or sperm. Sperm were ideal cells to study nuclei. Like white blood cells, they could be obtained in large quantity, and 90 percent of their volume was occupied by nuclei. Miescher's nuclein preparations from sperm cells contained a higher percentage of phosphorus (almost 10 percent by weight) than those from white blood cells, indicating they had less contaminating protein. In fact, they were the first preparations of relatively pure DNA. The term *nucleic acid* was coined in 1889 by Richard Altmann, one of Miescher's students, who worked out the methods of purifying protein-free DNA from various animal tissues and yeast.[2]

During the last two decades of the nineteenth century, numerous biologists focused on the chromosomes, describing their behavior during and between cell division (page 381). One way to observe chromosomes was to stain these cellular structures with dyes. A botanist named E. Zacharias discovered that the same stains that made the chromosomes visible also stained a preparation of nuclein that had been extracted using Miescher's procedure of digestion with pepsin in an HCl medium. Furthermore, when the pepsin–HCl-extracted cells were subsequently extracted with dilute alkali, a procedure that was known to remove nuclein, the cell residue (which included the chromosome remnants) no longer contained stainable material. These and other results pointed strongly to nuclein as a component of the chromosomes. In one remarkably farsighted proposal, Otto Hertwig, who had been studying the behavior of the chromosomes during fertilization, stated in 1884: "I believe that I have at least made it highly probable that nuclein is the substance that is responsible not only for fertilization but also for the transmission of hereditary characteristics."[3] Ironically, as more was learned about the properties of nuclein, the less it was considered as a candidate for the genetic material.

During the 50 years that followed Miescher's discovery of DNA, the chemistry of the molecule and the nature of its components were described. Some of the most important contributions in this pursuit were made by Phoebus Aaron Levene, who immigrated from Russia to the United States in 1891 and eventually settled in a position at the Rockefeller Institute in New York. It was Levene who finally solved one of the most resistant problems of DNA chemistry when he determined in 1929 that the sugar of the nucleotides was 2-deoxyribose.[4] To isolate the sugar, Levene and E. S. London placed the DNA into the stomach of a dog through a surgical opening and then collected the sample from the animal's intestine. As it passed through the stomach and intestine, various enzymes of the animal's digestive tract acted on the DNA, carving the nucleotides into their component parts, which could then be isolated and analyzed. Levene summarized the state of knowledge on nucleic acids in a monograph published in 1931.[5]

While Levene is credited for his work in determining the structure of the building blocks of DNA, he is also credited as having been the major stumbling block in the search for the chemical nature of the gene. Through this period, it became increasingly evident that proteins were very complex and exhibited great specificity in catalyzing a remarkable variety of chemical reactions. DNA, on the other hand, was thought to be composed of a monotonous repeat of its four nucleotide building blocks. The major proponent of this view of DNA, which was called the tetranucleotide theory, was Phoebus Levene. Because chromosomes consisted of only two components— DNA and protein—there seemed little doubt at that time that protein was the genetic material.

Meanwhile, as the structure of DNA was being worked out, a seemingly independent line of research was being carried out in the field of bacteriology. During the early 1920s, it was found that a number of species of pathogenic bacteria could be grown in the laboratory in two different forms. Virulent bacteria, that is, bacterial cells capable of causing disease, formed colonies that were smooth, dome shaped, and regular. In contrast, nonvirulent bacterial cells grew into colonies that were rough, flat, and irregular (Figure 1).[6] The British microbiologist J. A. Arkwright introduced the terms smooth (S) and rough (R) to describe these two types. Under the microscope, cells that formed S colonies were seen to be surrounded by a gelatinous capsule, whereas the capsule was absent from cells in the R colonies. The bacterial capsule helps protect a bacterium from its host's defenses, which explains why the R cells, which lacked these structures, were unable to cause infections in laboratory animals.

Because of its widespread impact on human health, the bacterium responsible for causing pneumonia (*Streptococcus pneumoniae*, or simply pneumococcus) has long been a focus of attention among microbiologists. In 1923, Frederick Griffith, a medical officer at the British Ministry of Health, demonstrated that pneumococcus also grew as either S or R colonies and, furthermore, that the two forms were interconvertible; that is, on occasion an R bacterium could revert to an S form, or vice versa.[7] Griffith found, for example, if he injected exceptionally large numbers of R bacteria into a mouse, the animal frequently developed pneumonia and produced bacteria that formed colonies of the S form.

It had been shown earlier that pneumococcus occurred as several distinct types (types I, II, and III) that could be distinguished

FIGURE 1 The large glistening colonies on the right are virulent S-type pneumococci, whereas the smaller colonies on the left are nonvirulent R-type pneumococci. As discussed below, the cells in these particular S colonies are the result of transformation of R bacteria by DNA from heat-killed S pneumococci. (FROM O. T. AVERY, C. M. MACLEOD, AND M. MCCARTY, J. EXP. MED. 79:153, 1944; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

from one another immunologically. In other words, antibodies could be obtained from infected animals that would react with only one of the three types. Moreover, a bacterium of one type never gave rise to cells of another type. Each of the three types of pneumococcus could occur as either the R or S form.

In 1928, Griffith made a surprising discovery while injecting various bacterial preparations into mice.[8] Injections of large numbers of heat-killed S bacteria or small numbers of living R bacteria, by themselves, were harmless to the mouse. However, if he injected both of these preparations together into the same mouse, it contracted pneumonia and died. Virulent bacteria could be isolated from the mouse and cultured. To extend the findings, he injected combinations of bacteria of different types (Figure 2). Initially, eight mice were injected with heat-killed, type I, S bacteria together with a small inoculum of a live, type II, R strain. Two of the eight animals contracted pneumonia, and Griffith was able to isolate and culture virulent, type I, S bacterial cells from the infected mice. Because there was no possibility that the heat-killed bacteria had been brought back to life, Griffith concluded that the dead type I cells had provided something to the live, nonencapsulated, type II cells that *transformed* them into an encapsulated type I form. The transformed
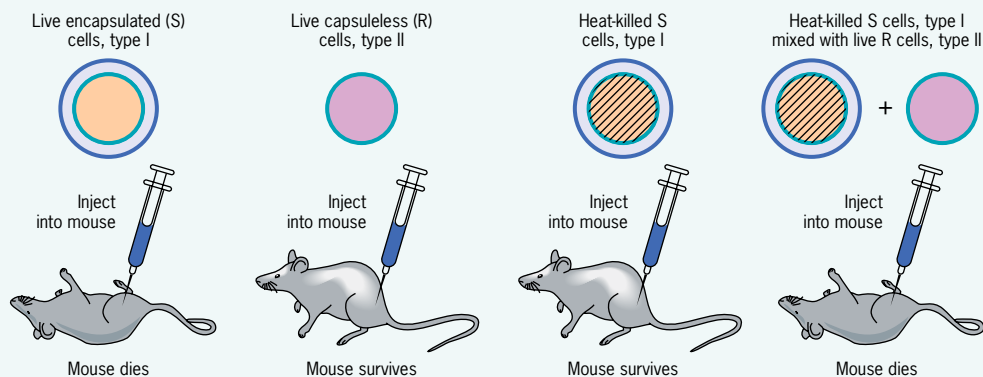
bacteria continued to produce type I cells when grown in culture; thus the change was stable and permanent.

Griffith's finding of transformation was rapidly confirmed by several laboratories around the world, including that of Oswald Avery, an immunologist at the Rockefeller Institute, the same institution where Levene was working. Avery had initially been skeptical of the idea that a substance released by a dead cell could alter the appearance of a living cell, but he was convinced when Martin Dawson, a young associate in his lab, confirmed the results.[9] Dawson went on to show that transformation need not occur within a living animal host. A crude extract of the dead S bacteria, when mixed in bacterial culture with a small number of the nonvirulent (R) cells in the presence of anti-R serum, was capable of converting the R cells into the virulent S form. The transformed cells were always of the type (I, II, or III) characteristic of the dead S cells.[10]

The next major step was taken by J. Lionel Alloway, another member of Avery's lab, who was able to solubilize the transforming agent. This was accomplished by rapidly freezing and thawing the killed donor cells, then heating the disrupted cells, centrifuging the suspension, and forcing the supernatant through a porcelain filter whose pores blocked the passage of bacteria. The soluble, filtered extract possessed the same transforming capacity as the original heat-killed cells.[11]

For the next decade, Avery and his colleagues focused their attention on purifying the substance responsible for transformation and determining its identity. As remarkable as it may seem today, no other laboratory in the world was pursuing the identity of the "transforming principle," as Avery called it. Progress on the problem was slow.[12] Eventually, Avery and his co-workers, Colin MacLeod and Maclyn McCarty, succeeded in isolating a substance from the soluble extract that was active in causing transformation when present at only 1 part per 600 million. All the evidence suggested that the active substance was DNA: (1) it exhibited a host of chemical properties that were characteristic of DNA, (2) no other type of material could be detected in the preparation, and (3) tests of various enzymes indicated that only those enzymes that digested DNA inactivated the transforming principle.

The paper published in 1944 was written with scrupulous caution and made no dramatic statements that genes were made of DNA rather than protein.[13] The paper drew remarkably little attention. Maclyn McCarty, one of the three authors, describes an incident in 1949 when he was asked to speak at Johns Hopkins University along with Leslie Gay, who had been testing the effects of the new drug Dramamine for the treatment of sea sickness. The large hall was packed with people, and "after a short period of ques-

FIGURE 2 Outline of the experiment by Griffith of the discovery of bacterial transformation.

tions and discussion following [Gay's] paper, the president of the Society got up to introduce me as the second speaker. Very little that he said could be heard because of the noise created by people streaming out of the hall. When the exodus was complete, after I had given the first few minutes of my talk, I counted approximately thirty-five hardy souls who remained in the audience because they wanted to hear about pneumococcal transformation or because they felt they had to remain out of courtesy." But Avery's awareness of the potential of his discovery was revealed in a letter he wrote in 1943 to his brother Roy, also a bacteriologist:

> If we are right, & of course that's not yet proven, then it means that nucleic acids are not merely structurally important but functionally active substances in determining the biochemical activities and specific characteristics of cells—& that by means of a known chemical substance it is possible to induce predictable and hereditary changes in cells. This is something that has long been the dream of geneticists. . . . Sounds like a virus—may be a gene. But with mechanisms I am not now concerned—one step at a time. . . . Of course the problem bristles with implications. . . . It touches genetics, enzyme chemistry, cell metabolism & carbohydrate synthesis—etc. But today it takes a lot of well documented evidence to convince anyone that the sodium salt of deoxyribose nucleic acid, protein free, could possibly be endowed with such biologically active & specific properties & that evidence we are now trying to get. It's lots of fun to blow bubbles,—but it's wiser to prick them yourself before someone else tries to.

Many articles and passages in books have dealt with the reasons why Avery's findings were not met with greater acclaim. Part of the reason may be due to the subdued manner in which the paper was written and the fact that Avery was a bacteriologist, not a geneticist. Some biologists were persuaded that Avery's preparation must have been contaminated with minuscule amounts of protein and that the contaminant, not the DNA, was the active transforming agent. Others questioned whether studies on transformation in bacteria had any relevance to the field of genetics.

During the years following the publication of Avery's paper, the climate in genetics changed in an important way. The existence of the bacterial chromosome was recognized, and a number of prominent geneticists turned their attention to these prokaryotes. These scientists believed that knowledge gained from the study of the simplest cellular organisms would shed light on the mechanisms that operate in the most complex plants and animals. In addition, the work of Erwin Chargaff and his colleagues on the base composition of DNA shattered the notion that DNA was a molecule consisting of a simple repetitive series of nucleotides (discussed on page 413).[14] This finding awakened researchers to the possibility that DNA might have the properties necessary to fulfill a role in information storage.

Seven years after the publication of Avery's paper on bacterial transformation, Alfred Hershey and Martha Chase of the Cold Spring Harbor Laboratories in New York turned their attention to an even simpler system—bacteriophages, or viruses that infect bacterial cells. By 1950, researchers recognized that viruses also had a genetic program. Viruses injected their genetic material into the host bacterium, where it directed the formation of new virus particles inside the infected cell. Within a matter of minutes, the infected cell broke open, releasing new bacteriophage particles, which then infected neighboring host cells.

It was clear that the genetic material directing the formation of viral progeny had to be either DNA or protein because these were the only two molecules the virus contained. Electron microscopic observations showed that, during the infection, the bulk of the bacteriophage remains outside the cell, attached to the cell surface by tail fibers (Figure 3). Hershey and Chase reasoned that the virus's genetic material must possess two properties. First, if the material were to direct the development of new bacteriophages during infection, it must pass into the infected cell. Second, if the material carries inherited information, it must be passed on to the next generation of bacteriophages. Hershey and Chase prepared two batches of bacteriophages to use for infection (Figure 4). One batch contained radioactively labeled DNA ($[^{32}P]$DNA); the other batch contained radioactively labeled protein ($[^{35}S]$protein). Because DNA lacks sulfur (S) atoms, and protein usually lacks phosphorus (P) atoms, these two radioisotopes provided distinguishing labels for the two types of macromolecules. Their experimental plan was to allow one or the other type of bacteriophage to infect a population of bacterial cells, wait a few minutes, and then strip the empty viruses from the surfaces of the cells. After trying several methods to separate the bacteria from the attached phage coats, they found this was best accomplished by subjecting the infected suspension to the spinning blades of a Waring blender. Once the virus particles had been detached from the cells, the bacteria could be centrifuged to the bottom of the tube, leaving the empty viral coats in suspension.

By following this procedure, Hershey and Chase determined the amount of radioactivity that entered the cells versus that which remained behind in the empty coats. They found that when cells were infected with protein-labeled bacteriophages, the bulk of the radioactivity remained in the empty coats. In contrast, when cells were infected with DNA-labeled bacteriophages, the bulk of the radioactivity passed inside the host cell. When they monitored the radioactivity passed onto the next generation, they found that less than 1 percent of the labeled protein could be detected in the progeny, whereas approximately 30 percent of the labeled DNA could be accounted for in the next generation.
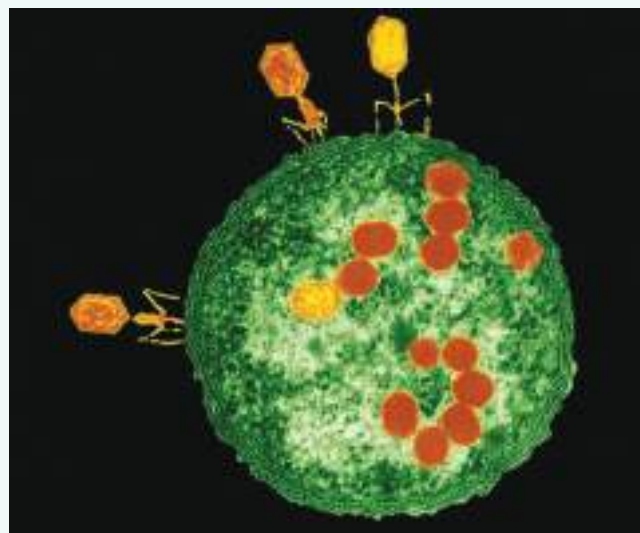
FIGURE 3 Electron micrograph of a bacterial cell infected by T4 bacteriophage. Each phage is seen attached by its tail fibers to the outer surface of the bacterium, while new phage heads are being assembled in the host cell's cytoplasm. (LEE D. SIMON/PHOTO RESEARCHERS, INC.)

Shear off adsorbed
particles in blender

~ 80% of
radioactivity

~ 20% of
radioactivity

Phage progeny
produced from bacteria
despite loss of
adsorbed phage coats.
Considerable % of
original radioactivity
present in progeny

(a)

Shear off adsorbed
particles in blender

~ 20% of
radioactivity

~ 80% of
radioactivity

Less than 1% of
radioactivity
transferred
to progeny

Unlabeled protein
Labeled protein
Unlabeled DNA
Labeled DNA

(b)

FIGURE 4 The Hershey-Chase experiment showing that bacterial cells infected with phage containing ^{32}P-labeled DNA (red DNA molecules) became radioactively labeled and produced labeled phage progeny. In contrast, bacterial cells infected with phage containing ^{35}S-labeled protein (red phage coats) did not become radioactively labeled and produced only nonlabeled progeny.

The publication of the Hershey-Chase experiments in 1952,[15] together with the abandonment of the tetranucleotide theory, removed any remaining obstacles to the acceptance of DNA as the genetic material. Suddenly, tremendous new interest was generated in a molecule that had largely been ignored. The stage was set for the discovery of the double helix.

References

1. MIESCHER, J. F. 1871. *Hoppe-Seyler's Med. Chem. Untersuchungen.* 4:441.
2. ALTMANN, R. 1889. *Anat. u. Physiol., Physiol. Abt.* 524.
3. Taken from MIRSKY, A. E. 1968. The discovery of DNA. *Sci. Am.* 218: 78–88. (June)
4. LEVENE, P. A. & LONDON, E. S. 1929. The structure of thymonucleic acid. *J. Biol. Chem.* 83:793–802.
5. LEVENE, P. A. & BASS, L. W. 1931. *Nucleic Acids.* The Chemical Catalog Co.
6. ARKWRIGHT, J. A. 1921. Variation in bacteria in relation to agglutination both by salts and by specific serum. *J. Path. Bact.* 24:36–60.
7. GRIFFITH, F. 1923. The influence of immune serum on the biological properties of pneumococci. *Rep. Public Health Med. Subj.* 18:1–13.
8. GRIFFITH, F. 1928. The significance of pneumococcal types. *J. Hygiene* 27:113–159.
9. DAWSON, M. H. 1930. The transformation of pneumoccal types. *J. Exp. Med.* 51:123–147.
10. DAWSON, M. H. & SIA, R.H.P. 1931. In vitro transformation of pneumococcal types. *J. Exp. Med.* 54:701–710.
11. ALLOWAY, J. L. 1932. The transformation in vitro of R pneumococci into S forms of different specific types by use of filtered pneumococcus extracts. *J. Exp. Med.* 55:91–99.
12. McCARTY, M. 1985. *The Transforming Principle: Discovering That Genes Are Made of DNA.* Norton.
13. AVERY, O. T., MACLEOD, C. M., & McCARTY, M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* 79:137–158.
14. CHARGAFF, E. 1950. Chemical specificity of nucleic acids and mechanism of their enzymic degradation. *Experentia* 6:201–209.
15. HERSHEY, A. D. & CHASE, M. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* 36:39–56.

SYNOPSIS

Chromosomes are the carriers of genetic information. A number of early observations led biologists to consider the genetic role of chromosomes. These included observations of the precision with which chromosomes are divided between daughter cells during cell division; the realization that the chromosomes of a species remained constant in shape and number from one division to the next; the finding that embryonic development required a particular complement of chromosomes; and the observation that the chromosome number is divided in half prior to formation of the gametes and is doubled following union of a sperm and egg at fertilization. Mendel's findings provided biologists with a new set of criteria for identifying the carriers of the genes. Sutton's studies of gamete formation in the grasshopper revealed the existence of homologous chromosomes, the association of homologues during the cell divisions that preceded gamete formation, and the separation of the homologues during the first of these meiotic divisions. (p. 380)

If genes are packaged together on chromosomes that are passed from parents to offspring, then genes on the same chromosome should be linked to one another to form a linkage group. The existence of linkage groups was confirmed in various systems, particularly in fruit flies where dozens of mutations were found to assort into four linkage groups that correspond in size and number to the chromosomes present in the cells of these insects. At the same time, it was discovered that linkage was incomplete; that is, the alleles originally present on a given chromosome did not necessarily remain together during formation of the gametes, but could be reshuffled between maternal and paternal homologues. This phenomenon, which Morgan called crossing over, was proposed to be the result of breakage and reunion of segments of homologous chromosomes that occurred while homologues were physically associated during the first meiotic division. Analyses of offspring from matings between adults carrying a variety of mutations on the same chromosome indicated that the frequency of recombination between two genes provided a measure of the relative distance that separates those two genes. Thus, recombination frequencies could be used to prepare detailed maps of the serial order of genes along each chromosome of a species. Genetic maps of the fruit fly based on recombination frequencies were verified independently by examination of the locations of various bands in the giant polytene chromosomes found in certain larval tissues of these insects. (p. 382)

Experiments discussed in the Experimental Pathways provided conclusive evidence that DNA was the genetic material. DNA is a helical molecule consisting of two chains of nucleotides running in opposite directions with their backbones on the outside and the nitrogenous bases facing inward. Adenine-containing nucleotides on one strand always pair with thymine-containing nucleotides on the other strand, likewise for guanine- and cytosine-containing nucleotides. As a result, the two strands of a DNA molecule are complementary to one another. Genetic information is encoded in the specific linear sequence of nucleotides that makes up the strands. The Watson-Crick model of DNA structure suggested a mechanism of replication that included strand separation and the use of each strand as a template that directed the order of nucleotide assembly during construction of the complementary strand. The mechanism by which DNA governed the assembly of a specific protein remained a total mystery. The DNA molecule depicted in Figure 10.10 is in a relaxed state having 10 base pairs per turn of the helix. DNA within a cell tends to be underwound (contains a greater number of base pairs per turn) and is said to be negatively supercoiled, a condition that facilitates the separation of strands during replication and tran-

scription. The supercoiled state of DNA is altered by topoisomerases, enzymes that are able to cut, rearrange, and reseal DNA strands. (p. 386)

All of the genetic information present in a single haploid set of chromosomes of an organism constitutes that organism's genome. The variety of DNA sequences that make up the genome and the number of copies of these various sequences describe the complexity of the genome. Understanding the complexity of a genome has grown out of early studies showing that the two strands that make up a DNA molecule can be separated by heat; when the temperature of the solution is lowered, complementary single strands are capable of reassociating to form stable, double-stranded DNA molecules. Analysis of the kinetics of this reassociation process provides a measure of the concentration of complementary sequences, which in turn provides a measure of the variety of sequences that are present within a given quantity of DNA. The greater the number of copies of a particular sequence in the genome, the greater is its concentration and the faster it reanneals. (p. 393)

When DNA fragments from eukaryotic cells are allowed to reanneal, the curve typically shows three steps, which correspond to the reannealing of three different classes of DNA sequences. The highly repeated fraction consists of short DNA sequences that are repeated in great number; these include satellite DNAs situated at the centromeres of the chromosomes, minisatellite DNAs, and microsatellite DNAs. The latter two groups tend to be highly variable, causing certain inherited diseases and forming the basis of DNA fingerprint techniques. The moderately repeated fraction includes DNA sequences that encode ribosomal and transfer RNAs and histone proteins, as well as various sequences with noncoding functions. The nonrepeated fraction contains protein-coding genes, which are present in one copy per haploid set of chromosomes. (p. 394)

The sequence organization of the genome is capable of change, either slowly over the course of evolution or rapidly as the result of transposition. The genes encoding eukaryotic proteins are often members of multigene families whose members show evidence that they have evolved from a common ancestral gene. The first step in this process is thought to be the duplication of a gene, which probably occurs primarily by unequal crossing over. Once duplication has occurred, nucleotide substitutions would be expected to modify various members in different ways, producing a family of repeated sequences of similar but nonidentical structure. The globin genes, for example, consist of clusters of genes located on two different chromosomes. Each cluster contains a number of related genes that code for globin polypeptides produced at different stages in the life of an animal. The clusters also contain pseudogenes, which are homologous to the globin genes, but are nonfunctional. (p. 399)

Certain DNA sequences are capable of moving rapidly from place to place in the genome by transposition. These transposable elements are called transposons, and they are capable of integrating themselves randomly throughout the genome. The best studied transposons occur in bacteria. They are characterized by inverted repeats at their termini, an internal segment that codes for a transposase required for their integration, and the formation of short repeated sequences in the target DNA that flank the element at the site of integration. Eukaryotic transposable elements are capable of moving by several mechanisms. In most cases, the element is transcribed into RNA, which is copied by a reverse transcriptase encoded by the element, and the DNA copy is integrated into the target site. The moderately repeated fraction of human DNA con-

tains two large families of transposable elements, the *Alu* and L1 families. (p. 402)

Over the past decade or so, large numbers of genomes, both prokaryotic and eurkaryotic, have been sequenced. These efforts have provided remarkable insight into the arrangement and evolution of both protein-coding and noncoding components of the genome. The human genome contains only about 20,000 protein-coding genes, but many more polypeptides can be synthesized as the result of gene-enhancing activities, such as alternative splicing. Information

about the functional elements of the human genome can be gained by comparing the sequence with homologous sections of other genomes to determine those portions that are conserved and those that tend to undergo sequence diversification. Conserved sequences are presumed to have a function, either in a coding or regulatory capacity. Chimpanzee and human genomes differ by about 4 percent. Comparison of sequences in the two genomes provides information on genes that have been subjected to positive selection in the human lineage and underlie some of our unique human characteristics. (p. 405).

ANALYTIC QUESTIONS

1. How would Mendel's results have differed if two of the seven traits he studied were encoded by genes that resided close to one another on one of the chromosomes of a pea plant?

2. Sutton was able to provide visual evidence for Mendel's law of segregation. Why would it have been impossible for him visually to confirm or refute Mendel's law of independent assortment?

3. Genes X, Y, and Z are all located on one chromosome. Draw a simple map showing the gene order and relative distances among the genes X, Y, and Z using the data below:

Crossover Frequency	Between These Genes
36%	X-Z
10%	Y-Z
26%	X-Y

4. Alleles on opposite ends of a chromosome are so likely to be separated by crossing over between them that they segregate independently. How would one be able to determine, using genetic crosses, that these two genes belong to the same linkage group? How might this be established using nucleic acid hybridization?

5. Where in the curve of Figure 10.17 would you expect to see the reannealing of the DNA that codes for ribosomal RNA? Suppose you had a pure preparation of DNA fragments whose sequences code for ribosomal RNA. Draw the reannealing of this DNA superimposed over the curve for total DNA shown in Figure 10.17.

6. Approximately 5 percent of the present human genome consists of segmental duplications that have arisen during the past 35 million years. How do you suppose researchers are able to estimate how long it has been since a particular region of a chromosome was duplicated?

7. It was noted on page 403 that at least 45 percent of the human genome is derived from transposable elements. The actual number could be much higher, but it is impossible to make a determination about the origin of many other sequences. Why do you suppose it is difficult to make assignments about the origin of many of the sequences in the human genome?

8. Suppose you had two solutions of DNA, one single stranded and the other double stranded, with equivalent absorbance of

ultraviolet light. How would the concentrations of DNA compare in these two solutions?

9. What type of labeling pattern would you expect after in situ hybridization between mitotic chromosomes and labeled myoglobin mRNA? Between the same chromosomes and labeled histone DNA?

10. According to Chargaff's determination of base composition, which of the following would characterize any sample of DNA? (1) [A] + [T] = [G] + [C]; (2) [A]/[T] = 1; (3) [G] = [C]; (4) [A] + [G] = [T] + [C]. If the C content of a preparation of double-stranded DNA is 15 percent, what is the A content?

11. If 30 percent of the bases on a single strand of DNA are T, then 30 percent of the bases on that strand are A. True or false? Why?

12. Would you agree with the statement that the T_m represents the temperature at which 50 percent of the DNA molecules in a solution are single stranded?

13. Suppose you were to find a primate that had a β-globin gene with only one intervening sequence. Do you think this animal may have evolved from a primitive ancestor that split away from other animals prior to the appearance of the second intron?

14. In 1996, a report was published in the journal *Lancet* on the level of trinucleotide expansion in the DNA of blood donors. These researchers found that the level of trinucleotide expansion decreased significantly with age. Can you provide any explanation for these findings?

15. Suppose you were looking at a particular SNP in the genome where half the population had an adenine (A) and the other half had a guanine (G). Is there any way that you might be able to determine which of these variants was present in ancestral humans and which became frequent during human evolution?

16. Look over Figures 10.23 and 10.29. The arrangement of genes depicted in the former figure has evolved over hundreds of millions of years, whereas that in the latter figure has evolved over thousands of years. Discuss the likely possibilities for the future evolution of the human amylase genes.

11
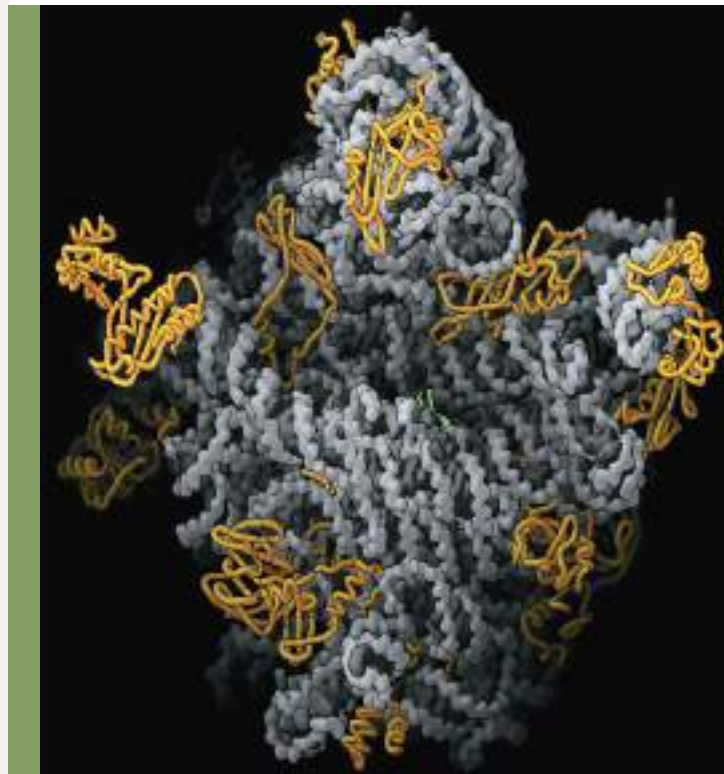
Gene Expression:
From Transcription to Translation

I n many ways, progress in biology over the past century is reflected in our changing concept of the gene. As the result of Mendel's work, biologists learned that genes are discrete elements that govern the appearance of specific traits. Given the opportunity, Mendel might have argued for the concept that one gene determines one trait. Boveri, Weismann, Sutton, and their contemporaries discovered that genes have a physical embodiment as parts of the chromosome. Morgan, Sturtevant, and colleagues demonstrated that genes have specific addresses—they reside at particular locations on particular chromosomes, and these addresses remain constant from one individual of a species to the next. Griffith, Avery, Hershey, and Chase demonstrated that genes were composed of DNA, and Watson and Crick solved the puzzle of DNA structure, which explained how this remarkable macromolecule could encode heritable information.

Although the formulations of these concepts were important steps along the path to genetic understanding, none of them addressed the mechanism by which the information stored in a gene is put to work in governing cellular activities. This is the major topic to be discussed in the present chapter. We will begin with additional insights into the nature of a gene, which brings us closer to its role in the expression of inherited traits. ■

A model of the large subunit of a prokaryotic ribosome at 2.4 Å resolution as determined by X-ray crystallography. This view looks into the subunit's active site cleft, which consists entirely of RNA. RNA is shown in gray, protein in gold, and the active site is revealed by a bound inhibitor (green). (COURTESY OF THOMAS A. STEITZ, YALE UNIVERSITY.)

419

11.1 THE RELATIONSHIP BETWEEN GENES AND PROTEINS

The first meaningful insight into gene function was gained by Archibald Garrod, a Scottish physician who reported in 1908 that the symptoms exhibited by persons with certain rare inherited diseases were caused by the absence of specific enzymes. One of the diseases investigated by Garrod was *alcaptonuria*, a condition readily diagnosed because the urine becomes dark on exposure to air. Garrod found that persons with alcaptonuria lacked an enzyme in their blood that oxidized homogentisic acid, a compound formed during the breakdown of the amino acids phenylalanine and tyrosine. As homogentisic acid accumulates, it is excreted in the urine and darkens in color when oxidized by air. Garrod had discovered the relationship between a genetic defect, a specific enzyme, and a specific metabolic condition. He called such diseases "inborn errors of metabolism." As seems to have happened with other early observations of basic importance in genetics, Garrod's findings went unappreciated for decades.

The idea that genes direct the production of enzymes was resurrected in the 1940s by George Beadle and Edward Tatum of the California Institute of Technology. They studied *Neurospora*, a tropical bread mold that normally grows in a very simple medium containing a single organic carbon source (e.g., a sugar), inorganic salts, and biotin (a B vitamin). Because it needs so little to live on, *Neurospora* was presumed to synthesize all of its required metabolites. Beadle and Tatum reasoned that an organism with such broad synthetic capacity should be very sensitive to enzymatic deficiencies, which should be easily detected using the proper experimental protocol. A simplified outline of their protocol is given in Figure 11.1.

Beadle and Tatum's plan was to irradiate mold spores and screen them for mutations that caused cells to lack a particular enzyme. To screen for such mutations, irradiated spores were tested for their ability to grow in a *minimal medium* that
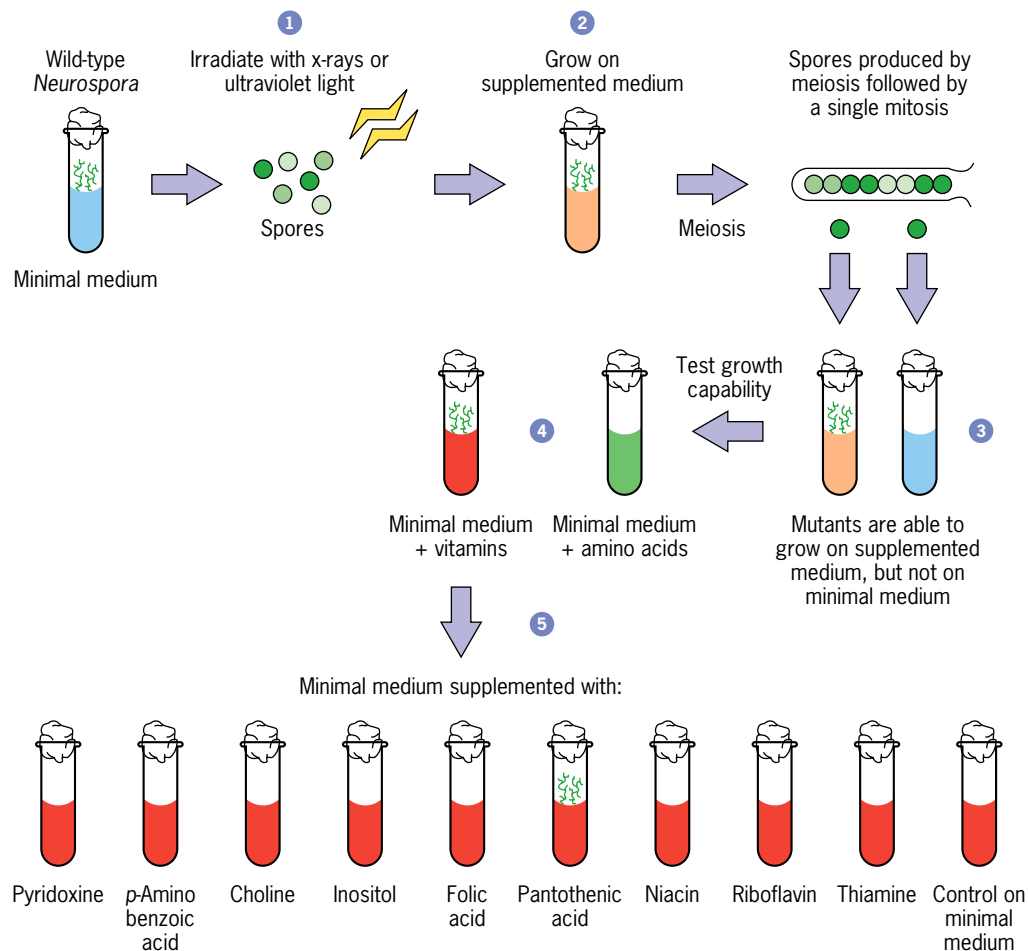
FIGURE 11.1 **The Beadle-Tatum experiment for the isolation of genetic mutants in *Neurospora*.** Spores were irradiated to induce mutations (step 1) and allowed to grow into colonies in tubes containing supplemented medium (step 2). Individual spores produced by the colonies were then tested for their ability to grow on the minimal medium (step 3). Those that failed to grow were mutants, and the task was to identify the mutant gene. In the example shown in step 4, a sample of mutant cells was found to grow in the minimal medium supplemented with vitamins, but would not grow in medium supplemented with amino acids. This observation indicates a deficiency in an enzyme leading to the formation of a vitamin. In step 5, growth of these same cells in minimal medium supplemented with one or another of the vitamins indicates the deficiency resides in a gene involved in the formation of pantothenic acid (part of coenzyme A).

lacked the essential compounds known to be synthesized by this organism (Figure 11.1). If a spore is unable to grow in minimal medium but can grow in a medium supplemented with a particular coenzyme (e.g., pantothenic acid of coenzyme A), then the researchers could conclude that the cells have an enzymatic deficiency that prevents them from synthesizing this essential compound.

Beadle and Tatum began by irradiating over a thousand cells. Two of the cells proved unable to grow on the minimal medium: one needed pyridoxine (vitamin B_6), and the other needed thiamine (vitamin B_1). Eventually, the progeny of about 100,000 irradiated spores were tested, and dozens of mutants were isolated. Each mutant had a gene defect that resulted in an enzyme deficiency that prevented the cells from catalyzing a particular metabolic reaction. The results were clear-cut: a gene carries the information for the construction of a particular enzyme. This conclusion became known as the "one gene–one enzyme" hypothesis. Once it was learned that enzymes are often composed of more than one polypeptide chain, each of which is encoded by its own gene, the concept became modified to "one gene–one polypeptide." Although this relationship remains a close approximation of the basic function of a gene, it too has had to be modified owing to the discovery that a single gene often generates a variety of polypeptides, primarily as the result of alternative splicing (discussed in Section 12.5).

What is the molecular nature of the defect in a protein caused by a genetic mutation? The answer to this question came in 1956, when Vernon Ingram of Cambridge University reported on the molecular consequence of the mutation that causes sickle cell anemia. Hemoglobin consists of four large polypeptides. Although they were too large to be sequenced by the technology of that period, Ingram devised a shortcut. He cleaved preparations of both the normal and sickle cell forms of hemoglobin into a number of specific pieces using the proteolytic enzyme trypsin. He then analyzed these peptide fragments by paper chromatography to determine if any of the products of trypsin digestion of the normal and sickled

hemoglobin types were different. Of the 30 or so peptides in the mixture, one migrated differently in the two preparations; this one difference was apparently responsible for all of the symptoms associated with the disease. Once the peptides had been separated, Ingram had only one small fragment to sequence rather than an entire protein. The difference in the proteins proved to be the substitution of a single amino acid, valine, in the sickle cell hemoglobin for a glutamic acid in the normal molecule. Ingram had demonstrated that a mutation in a single gene had caused a single substitution in the amino acid sequence of a single polypeptide.

An Overview of the Flow of Information through the Cell

We have seen how the relationship between genetic information and amino acid sequence was established, but this knowledge by itself provided no clue as to the mechanism by which the specific polypeptide chain is generated. As we now know, there is an intermediate between a gene and its polypeptide; the intermediate is **messenger RNA (mRNA)**. The momentous discovery of mRNA was made in 1961 by François Jacob and Jacques Monod of the Pasteur Institute in Paris, Sydney Brenner of the University of Cambridge, and Matthew Meselson of the California Institute of Technology. A messenger RNA is assembled as a complementary copy of one of the two DNA strands that make up a gene. The synthesis of an RNA from a DNA template is called **transcription**. Because its nucleotide sequence is complementary to that of the gene from which it is transcribed, the mRNA retains the same information as the gene itself. An overview of the role of mRNA in the flow of information through a eukaryotic cell is illustrated in Figure 11.2.

The use of messenger RNA allows the cell to separate information storage from information utilization. While the gene remains stored away in the nucleus as part of a huge, stationary DNA molecule, its information can be imparted to a much smaller, mobile nucleic acid that passes into the cyto-

DNA of chromosome   Pre-mRNAs   Segment of DNA being transcribed

mRNA being translated   Protein

Cytoplasm   mRNAs
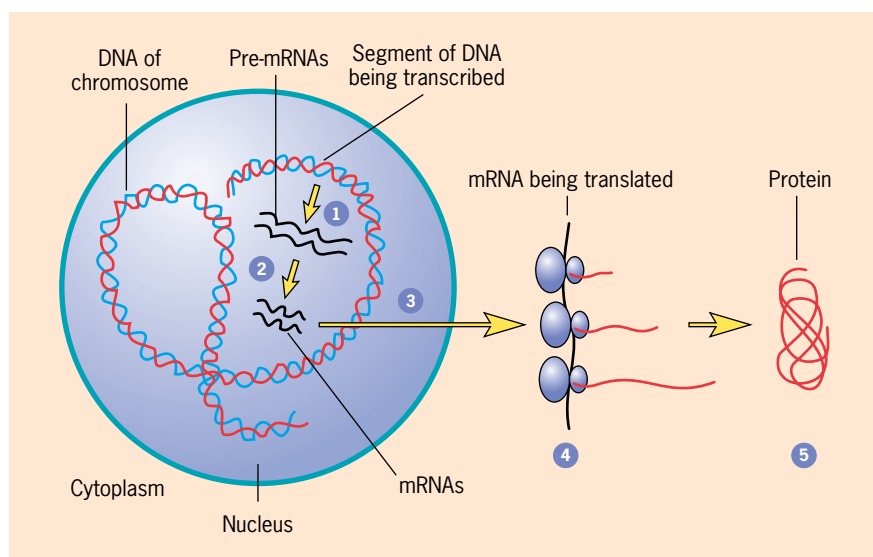
Nucleus

FIGURE 11.2 An overview of the flow of information in a eukaryotic cell. The DNA of the chromosomes located within the nucleus contains the entire store of genetic information. Selected sites on the DNA are transcribed into pre-mRNAs (step 1), which are processed into messenger RNAs (step 2). The messenger RNAs are transported out of the nucleus (step 3) into the cytoplasm, where they are translated into polypeptides by ribosomes that move along the mRNA (step 4). Following translation, the polypeptide folds to assume its native conformation (step 5).

plasm. Once in the cytoplasm, the mRNA can serve as a template to direct the incorporation of amino acids in a particular order encoded by the nucleotide sequence of the DNA and mRNA. The use of a messenger RNA also allows a cell to greatly amplify its synthetic output. One DNA molecule can serve as the template in the formation of many mRNA molecules, each of which can be used in the formation of a large number of polypeptide chains.

Proteins are synthesized in the cytoplasm by a complex process called **translation**. Translation requires the participation of dozens of different components, including ribosomes. Ribosomes are nonspecific components of the translation machinery. These complex, cytoplasmic "machines" can be programmed, like a computer, to translate the information encoded by any mRNA. Ribosomes contain both protein and RNA. The RNAs of a ribosome are called **ribosomal RNAs** (or **rRNAs**), and like mRNAs, each is transcribed from one of the DNA strands of a gene. Rather than functioning in an informational capacity, rRNAs provide structural support and catalyze the chemical reaction in which amino acids are covalently linked to one another. **Transfer RNAs** (or **tRNAs**) constitute a third major class of RNA that is required during protein synthesis. Transfer RNAs are required to translate the information in the mRNA nucleotide code into the amino acid "alphabet" of a polypeptide.
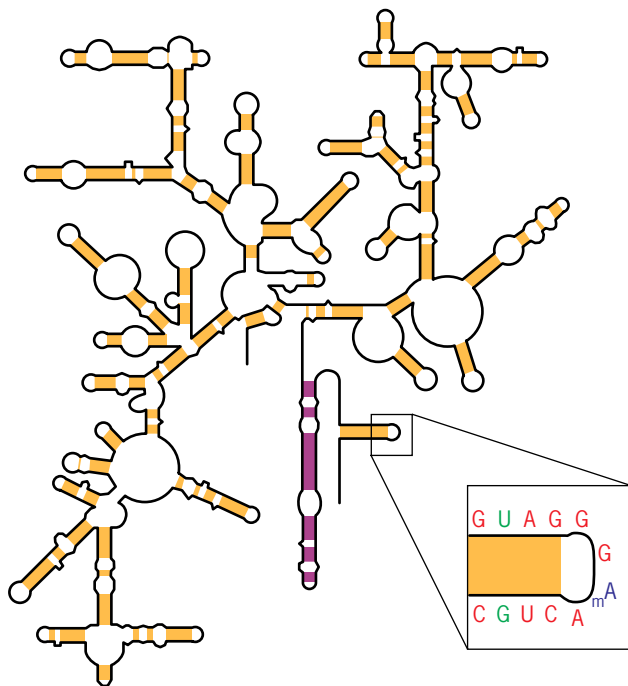
FIGURE 11.3 **Two-dimensional structure of a bacterial ribosomal RNA** showing the extensive base-pairing between different regions of the single strand. The expanded section shows the base sequence of a stem and loop, including a nonstandard base pair (G-U) and a modified nucleotide, methyladenosine. One of the helices is shaded differently because it plays an important role in ribosome function as discussed on page 463. An example of the three-dimensional structure of an RNA is shown in Figure 11.43*b*.

Both rRNAs and tRNAs owe their activity to their complex secondary and tertiary structures. Unlike DNA, which has a similar double-stranded, helical structure regardless of the source, many RNAs fold into a complex three-dimensional shape, which is markedly different from one type of RNA to another. Thus, like proteins, RNAs carry out a diverse array of functions because of their different shapes. As with proteins, the folding of RNA molecules follows certain rules. Whereas protein folding is driven by the withdrawal of hydrophobic residues into the interior, RNA folding is driven by the formation of regions having complementary base pairs (Figure 11.3). As seen in Figure 11.3, base-paired regions typically form double-stranded (and double-helical) "stems," which are connected to single-stranded "loops." Unlike DNA, which consists exclusively of standard Watson-Crick (G-C, A-T) base pairs, RNAs often contain nonstandard base pairs (inset, Figure 11.3) and modified nitrogenous bases. These unorthodox regions of the molecule serve as recognition sites for proteins and other RNAs, promote RNA folding, and help stabilize the structure of the molecule. The importance of complementary base-pairing extends far beyond tRNA and rRNA structure. As we will see throughout this chapter, base pairing *between* RNA molecules plays a central role in most of the activities in which RNAs are engaged.

The roles of mRNAs, rRNAs, and tRNAs are explored in detail in the following sections of this chapter. Eukaryotic cells make a host of other RNAs, which also play vital roles in cellular metabolism; these include small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), small interfering RNAs (siRNAs) and microRNAs (miRNAs). This last class of RNAs has burst onto the scene in the past few years, reminding us once again that there are many hidden cellular activities waiting to be discovered. Our genome encodes hundreds of these tiny microRNAs, yet we have very little idea what any of them do.

For background material on the structure of RNA, you might review page 74.

REVIEW ?

1. How were Beadle and Tatum able to conclude that a gene encoded a specific enzyme?
2. How was Ingram able to determine the amino acid substitution in the hemoglobin of persons with sickle cell anemia without sequencing the protein?

11.2 AN OVERVIEW OF TRANSCRIPTION IN BOTH PROKARYOTIC AND EUKARYOTIC CELLS

Transcription is a process in which a DNA strand provides the information for the synthesis of an RNA strand. The enzymes responsible for transcription in both prokaryotic and eukaryotic cells are called **DNA-dependent RNA polymerases**, or simply **RNA polymerases**. These enzymes are able to incor-

porate nucleotides, one at a time, into a strand of RNA whose sequence is complementary to one of the DNA strands, which serves as the *template.*

The first step in the synthesis of an RNA is the association of the polymerase with the DNA template. This brings up a matter of more general interest, namely, the specific interactions of two very different macromolecules, proteins and nucleic acids. Just as different proteins have evolved to bind different types of substrates and catalyze different types of reactions, so too have some of them evolved to recognize and bind to specific sequences of nucleotides in a strand of nucleic acid. The site on the DNA to which an RNA polymerase molecule binds prior to initiating transcription is called the **promoter**. Cellular RNA polymerases are not capable of recognizing promoters on their own but require the help of additional proteins called **transcription factors**. In addition to providing a binding site for the polymerase, the promoter contains the information that determines which of the two DNA strands is transcribed and the site at which transcription begins.

RNA polymerase moves along the template DNA strand toward its $5'$ end (i.e., in a $3' \rightarrow 5'$ direction). As the polymerase progresses, the DNA is temporarily unwound, and the polymerase assembles a complementary strand of RNA that grows from its $5'$ terminus in a $3'$ direction (Figure 11.4*a,b*). As indicated in Figure 11.4*c*, RNA polymerase catalyzes the reaction

$$RNA_n + NTP \rightarrow RNA_{n+1} + PP_i$$
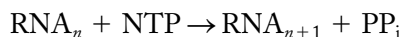
in which ribonucleoside triphosphate substrates (NTPs) are cleaved into nucleoside monophosphates as they are polymerized into a covalent chain. Reactions leading to the synthesis of nucleic acids (and proteins) are inherently different from those of intermediary metabolism discussed in Chapter 3. Whereas some of the reactions leading to the formation of small molecules, such as amino acids, may be close enough to equilibrium that a considerable reverse reaction can be measured, those reactions leading to the synthesis of nucleic acids and proteins must occur under conditions in which there is virtually no reverse reaction. This condition is met during transcription with the aid of a second reaction

$$PP_i \rightarrow 2\ P_i$$
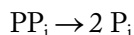
catalyzed by a different enzyme, a pyrophosphatase. In this case, the pyrophosphate (PP_i) produced in the first reaction is hydrolyzed to inorganic phosphate (P_i). The hydrolysis of pyrophosphate releases a large amount of free energy and makes the incorporation of nucleotides essentially irreversible.

As the polymerase moves along the DNA template, it incorporates complementary nucleotides into the growing RNA chain. A nucleotide is incorporated into the RNA strand if it is able to form a proper (Watson-Crick) base pair with the nucleotide in the DNA strand being transcribed. This can be seen in Figure 11.4*c* where the incoming adenosine $5'$-triphosphate pairs with the thymine-containing nucleotide of the template. Once the polymerase has moved past a particular stretch of DNA, the DNA double helix reforms (as in Figure 11.4*a,b*). Consequently, the RNA chain does not remain

associated with its template as a DNA–RNA hybrid (except for about nine nucleotides just behind the site where the polymerase is operating). RNA polymerases are capable of incorporating from about 20 to 50 nucleotides into a growing RNA molecule per second, and many genes in a cell are transcribed simultaneously by a hundred or more polymerases (as seen in Figure 11.11*c*). The electron micrograph of Figure 11.4*d* shows a molecule of phage DNA with a number of bound RNA polymerase molecules.

RNA polymerases are capable of forming prodigiously long RNAs. Consequently, the enzyme must remain attached to the DNA over long stretches of template (the enzyme is said to be *processive*). At the same time, the enzyme must be associated loosely enough that it can move from nucleotide to nucleotide of the template. It is difficult to study certain properties of RNA polymerases, such as processivity, using biochemical methodologies that tend to average out differences between individual protein molecules. Consequently, researchers have developed techniques to follow the activities of single RNA polymerase molecules similar to those used to study individual cytoskeletal motors. Two examples of such studies are depicted in Figure 11.5. In both of these examples, a single RNA polymerase is attached to the surface of a glass coverslip and allowed to transcribe a DNA molecule containing a fluorescent bead covalently linked to one of its ends. The movement of the fluorescent bead can be monitored under a fluorescence microscope.

In Figure 11.5*a*, the bead is free to move in the medium and its range of movement is proportional to the length of the DNA between the polymerase and the bead. As the polymerase transcribes the template, the connecting DNA strand is elongated, and the movement of the bead is increased. This type of system allows investigators to study the rate of transcription of an individual polymerase and to determine if the polymerase transcribes the DNA in a steady or discontinuous movement. In Figure 11.5*b*, the bead at the end of the DNA molecule being transcribed is trapped by a focused laser beam (page 139). The minute force exerted by the laser trap can be varied, until it is just sufficient to stop the polymerase from continuing to transcribe the DNA. Measurements carried out on single RNA polymerase molecules in the act of transcription indicate that these enzymes can move over the template, one base (3.4 Å) at a time, with a force more than twice that of a myosin molecule.

Even though polymerases are relatively powerful motors, these enzymes do not necessarily move in a steady, continuous fashion but may pause at certain locations along the template or even backtrack before resuming their forward progress. In some cases, as occurs following the incorporation of an incorrect nucleotide, a stalled polymerase must digest away the $3'$ end of the newly synthesized transcript and resynthesize the missing portion before continuing its movement. A number of elongation factors have been identified that enhance the enzyme's ability to traverse these various roadblocks.

At this point in the discussion, it is useful to examine the differences in the process of transcription between bacterial and eukaryotic cells.

(a)

(b)

(c)

(d)
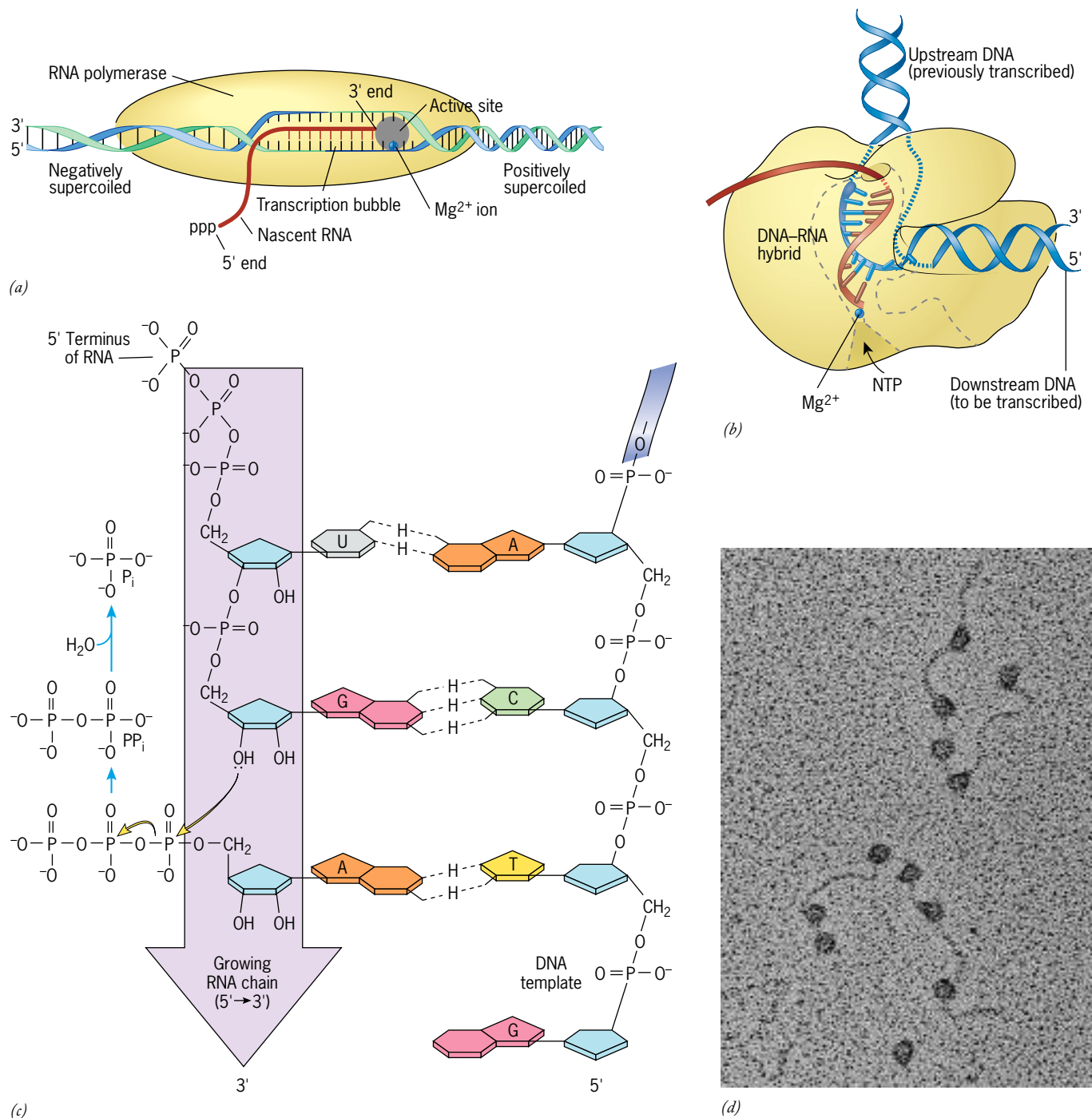
FIGURE 11.4 Chain elongation during transcription. (*a*) A schematic model of the elongation of a newly synthesized RNA molecule during transcription. The polymerase covers approximately 35 base pairs of DNA, the transcription bubble composed of single-stranded (melted) DNA contains about 15 base pairs, and the segment present in a DNA–RNA hybrid includes about 9 base pairs. The enzyme generates overwound (positively supercoiled) DNA ahead of itself and underwound (negatively supercoiled) DNA behind itself (page 391). (*b*) Schematic drawing of an RNA polymerase in the act of transcription elongation. The downstream DNA lies in a groove within the polymerase, clamped by a pair of jaws formed by the two largest subunits of the enzyme. The DNA makes a sharp turn in the region of the active site, so that the upstream DNA extends upward in this drawing. The nascent RNA exits from the enzyme's active site through a separate channel. (*c*) Chain elongation results following an attack by the 3′ OH of the nucleotide at the end of the growing strand on the 5′ α-phosphate of the incoming nucleoside triphosphate. The pyrophosphate released is subsequently cleaved, which drives the reaction in the direction of polymerization. The geometry of base-pairing between the nucleotide of the template strand and the incoming nucleotide determines which of the four possible nucleoside triphosphates is incorporated into the growing RNA chain at each site. (*d*) Electron micrograph of several RNA polymerase molecules bound to a phage DNA template. (D: COURTESY OF ROBLEY C. WILLIAMS.)

Transcription in Bacteria

Bacteria, such as *E. coli*, contain a single type of RNA polymerase composed of five subunits that are tightly associated to form a *core enzyme*. If the core enzyme is purified from bacterial cells and added to a solution of bacterial DNA molecules and ribonucleoside triphosphates, the enzyme binds to the DNA and synthesizes RNA. The RNA molecules produced by a purified polymerase, however, are not the same as those found within cells because the core enzyme has attached to random sites in the DNA, sites that it would normally have ignored in vivo. If, however, a purified accessory polypeptide called *sigma factor* (σ) is added to the RNA polymerase before it attaches to DNA, transcription begins at selected locations (Figure 11.6a,b). Attachment of σ factor to the core enzyme increases the enzyme's affinity for promoter sites in DNA and decreases its affinity for DNA in general. As a result, the complete enzyme is thought to slide freely along the DNA until it recognizes and binds to a suitable promoter region.

X-ray crystallographic analysis of the bacterial RNA polymerase (see Figure 11.8) reveals a molecule shaped like a "crab claw" with a pair of mobile pincers (or jaws) enclosing a positively charged internal channel. As the sigma factor interacts with the promoter, the jaws of the enzyme grip the downstream DNA duplex, which resides within the channel (as in

(a)

(b)
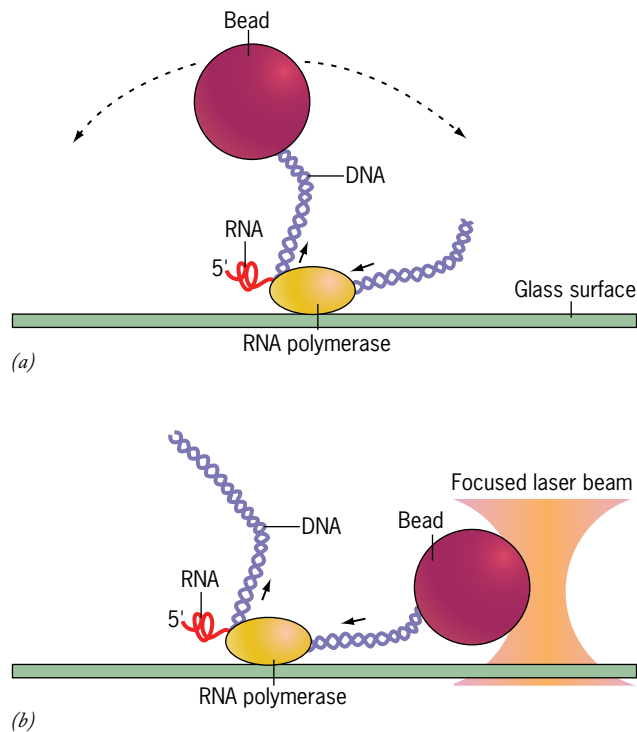
FIGURE 11.5 Examples of experimental techniques to follow the activities of single RNA polymerase molecules. (*a*) In this protocol, the RNA polymerase molecule is attached to the underlying coverslip and allowed to transcribe a DNA molecule containing a fluorescent bead at the upstream end. The arrows indicate the movement of the DNA through the polymerase. The rate of movement and progression of the polymerase can be followed by observing the position of the bead over time using a fluorescence microscope. (*b*) In this protocol, the attached polymerase is transcribing a DNA molecule with a bead at its downstream end. As in *a*, the arrows indicate direction of DNA movement. The bead is caught in an optical (laser) trap, which delivers a known force that can be varied by adjusting the laser beam. This type of apparatus can measure the forces generated by a transcribing polymerase. (REPRINTED FROM J. GELLES AND R. LANDICK, CELL 93:15, 1998, COPYRIGHT 1998, WITH PERMISSION FROM ELSEVIER SCIENCE.)

(a)

(b)

(c)

(d)

FIGURE 11.6 Schematic representation of the initiation of transcription in bacteria. (*a*) In the absence of the σ factor, the core enzyme does not interact with the DNA at specific initiation sites. (*b–d*) When the core enzyme is associated with the σ factor, the complete enzyme (or holoenzyme) is able to recognize and bind to the promoter regions of the DNA, separate the strands of the DNA double helix, and initiate transcription at the proper start sites (see Figure 11.7). In the traditional model shown here, the σ factor dissociates from the core enzyme, which is capable of transcription elongation. Recent studies suggest that, in at least some cases, σ may remain with the polymerase.

FIGURE 11.7 **The basic elements of a promoter region in the DNA of the bacterium *E. coli*.** The key regulatory sequences required for initiation of transcription are found in regions located at −35 and −10 base pairs from the site at which transcription is initiated. The initiation site marks the boundary between the + and − sides of the gene.
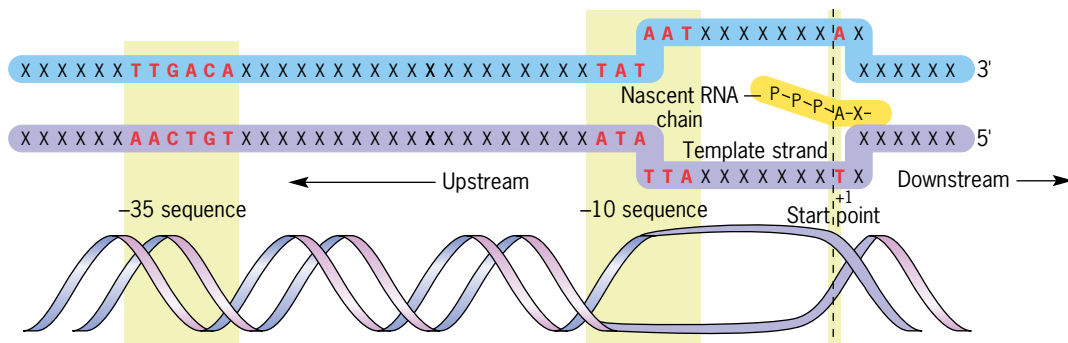
Figure 11.4*b*). The enzyme then separates (or *melts*) the two DNA strands in the region surrounding the start site (Figure 11.6*c*). Strand separation makes the template strand accessible to the enzyme's active site, which resides at the back wall of the channel. Initiation of transcription appears to be a difficult undertaking because an RNA polymerase typically makes several unsuccessful attempts to assemble an RNA transcript. Once 10–12 nucleotides have been successfully incorporated into a growing transcript, the enzyme undergoes a major change in conformation and is transformed into a *transcriptional elongation complex* that can move processively along the DNA. In the model shown in Figure 11.6*d*, the formation of an elongation complex is followed by release of the sigma factor.

As noted above, promoters are the sites in DNA that bind RNA polymerase. Bacterial promoters are located in the region of a DNA strand just preceding the initiation site of RNA synthesis (Figure 11.7). The nucleotide at which transcription is initiated is denoted as +1 and the preceding nucleotide as −1. Those portions of the DNA preceding the initiation site (toward the 3′ end of the template) are said to be *upstream* from that site. Those portions of the DNA succeeding it (toward the 5′ end of the template) are said to be *downstream* from that site. Analysis of the DNA sequences just upstream from a large number of bacterial genes reveals that two short stretches of DNA are similar from one gene to another. One of these stretches is centered at approximately 35 bases upstream from the initiation site and typically occurs as the sequence TTGACA (Figure 11.7). This TTGACA sequence (known as the −35 element) is called a **consensus sequence**, which indicates that it is the most common version of a conserved sequence, but that some variation occurs from one gene to another. The second conserved sequence is found approximately 10 bases upstream from the initiation site and occurs at the consensus sequence TATAAT (Figure 11.7). This site in the promoter, named the *Pribnow box* after its discoverer, is responsible for identifying the precise nucleotide at which transcription begins.

Bacterial cells possess a variety of different σ factors that recognize different versions of the promoter sequence. The σ^{70} is known as the "housekeeping" σ factor, because it initiates transcription of most genes. Alternative σ factors initiate transcription of a small number of specific genes that participate in a common response. For example, when *E. coli* cells are subjected to a sudden rise in temperature, a new σ factor is synthesized that recognizes a different promoter sequence and leads to the coordinated transcription of a battery of *heat-shock*

genes. The products of these genes protect the proteins of the cell from thermal damage (page 78).

Just as transcription is initiated at specific points in the chromosome, it also terminates when a specific nucleotide sequence is reached. In roughly half of the cases, a ring-shaped protein called *rho* is required for termination of bacterial transcription. Rho encircles the newly synthesized RNA and moves along the strand in a 3′ direction to the polymerase, where it separates the RNA transcript from the DNA to which it is bound. In other cases, the polymerase stops transcription when it reaches a *terminator sequence* and releases the completed RNA chain without requiring additional factors.

Transcription and RNA Processing in Eukaryotic Cells

As discovered in 1969 by Robert Roeder at the University of Washington, eukaryotic cells have three distinct transcribing enzymes in their cell nuclei. Each of these enzymes is responsible for synthesizing a different group of RNAs (Table 11.1). Plants have two additional RNA polymerases that are not essential for life. No prokaryote has been found with multiple RNA polymerases, whereas the simplest eukaryotes (yeast) have the same three nuclear types that are present in mammalian cells. This difference in number of RNA polymerases is another sharp distinction between prokaryotic and eukaryotic cells.

Figure 11.8*a* shows the surface structures of RNA polymerases from each of the three domains of life: archaea, bacteria, and eukaryotes. Several features are apparent from a close examination of this illustration. It is evident that the archaeal and eukaryotic enzymes are more similar in structure to one another than are the bacterial and eukaryotic enzymes. This feature re-

TABLE 11.1 **Eukaryotic <u>Nuclear</u> RNA Polymerases**

Enzyme	RNAs Synthesized
RNA polymerase I	larger rRNAs (28S, 18S, 5.8S)
RNA polymerase II	mRNAs, most small nuclear RNAs (snRNAs and snoRNAs), most microRNAs, and telomerase RNA
RNA polymerase III	small RNAs, including tRNAs, 5S rRNA, and U6 snRNA
RNA polymerase IV, V (plants only)	siRNAs

(a)

(b)
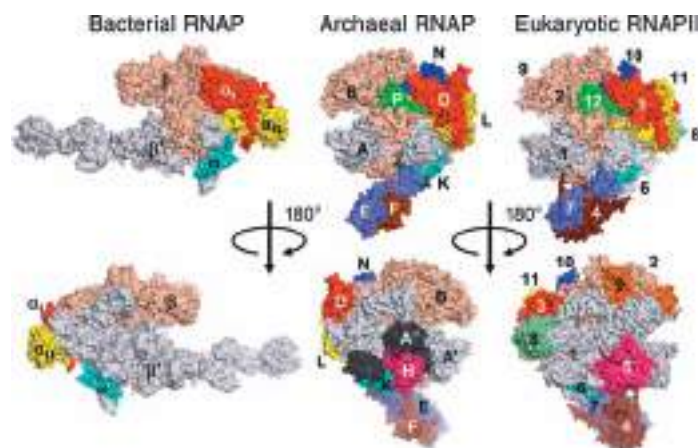
FIGURE 11.8 A comparison of prokaryotic and eukaryotic RNA polymerase structure. (*a*) RNA polymerases from the three domains of life. Each subunit of an enzyme is denoted by a different color and labeled according to conventional nomenclature for that enzyme. Homologous subunits are depicted by the same color. It can be seen that the archaeal and eukaryotic polymerases are more similar in structure to one another than are the bacterial and eukaryotic enzymes. RNA polymerase II (shown here) is only one of three major eukaryotic nuclear RNA polymerases. (*b*) Ribbon diagram of the core structure of yeast RNA polymerase II. Regions of the bacterial polymerase that are structurally homologous to the yeast enzyme are shown in green. The large channel that grips the downstream DNA is evident. The divalent metal situated at the end of the channel and within the active site is seen as a red sphere. (A: FROM AKIRA HIRATA, BRIANNA J. KLEIN, AND KATSUHIKO S. MURAKAMI, NATURE 451:852, 2008, © COPYRIGHT 2008, MACMILLAN MAGAZINES LTD.; B: FROM PATRICK CRAMER, DAVID A. BUSHNELL, AND ROGER D. KORNBERG, SCIENCE 292:1874, 2001; © COPYRIGHT 2001, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)
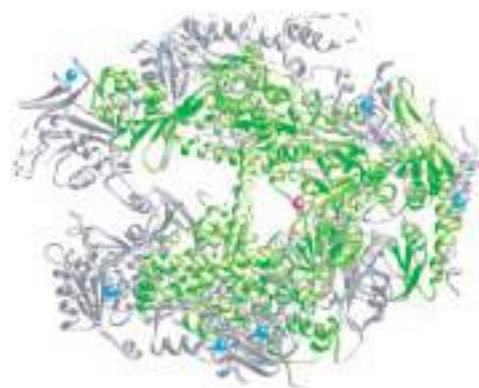
flects the evolutionary relationship between archaea and bacteria that was discussed back on page 28. The subunits that make up each of the proteins depicted in Figure 11.8*a* are indicated by a different color, and it is immediately evident that RNA polymerases are multisubunit enzymes. Those subunits between the different enzymes that are homologous to one another (i.e., are derived from a common ancestral polypeptide) are shown in the same color. Thus, it is also evident from Figure 11.8*a* that the RNA polymerases from the three domains share a number of subunits. Although the yeast enzyme shown in Figure 11.8*a* has a total of 12 subunits, 7 more than its bacterial counterpart, the fundamental core structure of the 2 enzymes is virtually identical. This evolutionary conservation in RNA polymerase structure is revealed in Figure 11.8*b*, which shows the homologous regions of the bacterial and eukaryotic enzymes at much higher resolution.

Our understanding of transcription in eukaryotes was greatly advanced with the 2001 publication of the X-ray crystallographic structure of yeast RNA polymerase II by Roger Kornberg and colleagues at Stanford University. As a result of these studies, and those of other laboratories in subsequent years, we now know a great deal about the mechanism of action of RNA polymerases as they move along the DNA, transcribing a complementary strand of RNA. A major distinction between transcription in prokaryotes and eukaryotes is the requirement in eukaryotes for a large variety of accessory proteins, or **transcription factors**. These proteins play a role in virtually every aspect of the transcription process, from the binding of the polymerase to the DNA template, to the initiation of transcription, to its elongation and termination. Although transcription factors are crucial for the operation of all three types of eukaryotic RNA polymerases, they will only be discussed in regard to the synthesis of mRNAs by RNA polymerase II (page 435).

All three major types of eukaryotic RNAs—mRNAs, rRNAs, and tRNAs—are derived from precursor RNA molecules that are considerably longer than the final RNA product. The initial precursor RNA is equivalent in length to the full length of the DNA transcribed and is called the **primary transcript**, or **pre-RNA**. The corresponding segment of DNA from which a primary transcript is transcribed is called a **transcription unit**. Primary transcripts do not exist within the cell as naked RNA but become associated with proteins even as they are synthesized. Primary transcripts typically have a fleeting existence, being *processed* into smaller, functional RNAs by a series of "cut-and-paste" reactions. RNA processing requires a variety of small RNAs (90 to 300 nucleotides long) and their associated proteins. In the following sections, we will examine the activities associated with the transcription and processing of each of the major eukaryotic RNAs.

REVIEW

1. What is the role of a promoter in gene expression? Where are the promoters for bacterial polymerases located?

2. Describe the steps during initiation of transcription in bacteria. What is the role of the σ factor? What is the nature of the reaction in which nucleotides are incorporated into a growing RNA strand? How is the specificity of nucleotide incorporation determined? What is the role of pyrophosphate hydrolysis?

3. How do the number of RNA polymerases distinguish prokaryotes and eukaryotes? What is the relationship between a pre-RNA and a mature RNA?

11.3 SYNTHESIS AND PROCESSING OF RIBOSOMAL AND TRANSFER RNAs

A eukaryotic cell may contain millions of ribosomes, each consisting of several molecules of rRNA together with dozens of ribosomal proteins. The composition of a mammalian ribosome is shown in Figure 11.9. Ribosomes are so numerous that more than 80 percent of the RNA in most cells consists of ribosomal RNA. To furnish the cell with such a large number of transcripts, the DNA sequences encoding rRNA are normally repeated hundreds of times. This DNA, called **rDNA**, is typically clustered in one or a few regions of the genome. The human genome has five rDNA clusters, each on a different chromosome. In a nondividing (interphase) cell, the clusters of rDNA are gathered together as part of one or more irregularly shaped nuclear structures, called **nucleoli** (singular, **nucleolus**), that function in producing ribosomes (Figure 11.10*a*).

The bulk of a nucleolus is composed of nascent ribosomal subunits that give the nucleolus a granular appearance (Figure 11.10*b*). Embedded within this granular mass are one or more rounded cores consisting primarily of fibrillar material.

(a)

(b)
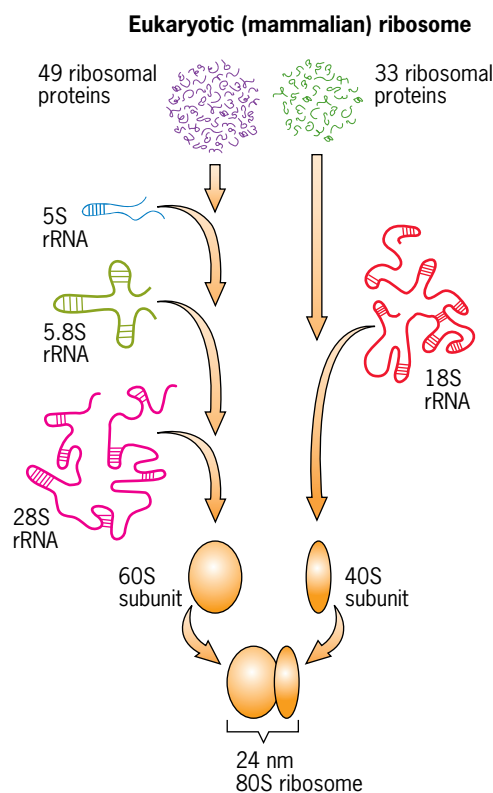
FIGURE 11.9 The macromolecular composition of a mammalian ribosome. This schematic drawing shows the components present in each of the subunits of a mammalian ribosome. The synthesis and processing of the rRNAs and the assembly of the ribosomal subunits are discussed in the following pages. (FROM D. P. SNUSTAD ET AL., PRINCIPLES OF GENETICS; COPYRIGHT © 1997, JOHN WILEY & SONS, INC. REPRINTED BY PERMISSION OF JOHN WILEY & SONS, INC.)
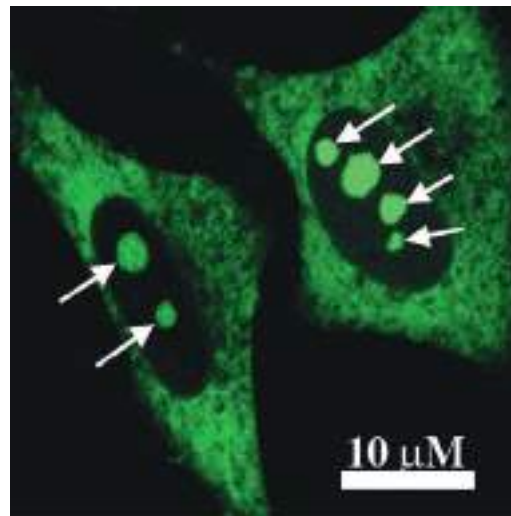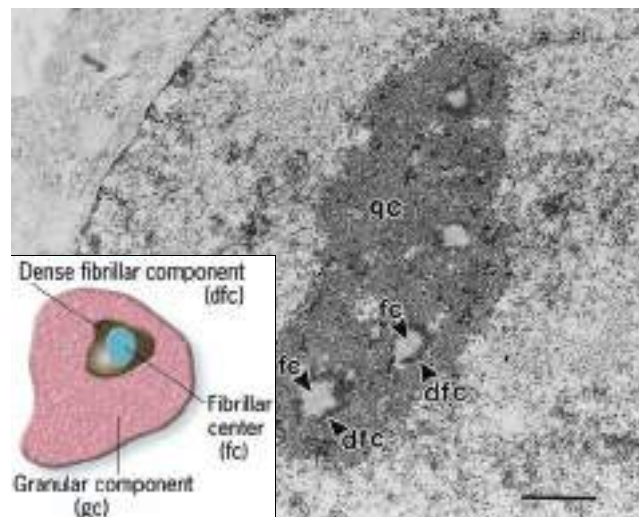
FIGURE 11.10 The nucleolus. (*a*) Light micrograph of two human HeLa cells transfected with a gene for a ribosomal protein fused to the green fluorescent protein (GFP). The fluorescent ribosomal protein can be seen in the cytoplasm where it is synthesized and ultimately functions, and in the nucleoli (white arrows), where it is assembled into ribosomes. (*b*) Electron micrograph of a section of part of a nucleus with a nucleolus. Three distinct nucleolar regions can be distinguished morphologically. The bulk of the nucleolus consists of a granular component (gc), which contains ribosomal subunits in various stages of assembly. Embedded within the granular regions are fibrillar centers (fc) that are surrounded by a more dense fibrillar component (dfc). The inset shows a schematic drawing of these parts of the nucleolus. According to one model, the fc contains the DNA that codes for ribosomal RNA, and the dfc contains the nascent pre-rRNA transcripts and associated proteins. According to this model, transcription of the pre-rRNA precursor takes place at the border between the fc and dfc. (*Note:* Nucleoli have other functions unrelated to ribosome biogenesis that are not discussed in this text.) Bar, 1 μm. (A: FROM C. E. LYON AND A. I. LAMOND, CURR. BIOL. 10:R323, 2000; B: FROM PAVEL HOZAK ET AL., J. CELL SCIENCE 107:641, 1994; BY PERMISSION OF THE COMPANY OF BIOLOGISTS, LTD.)

As discussed in the legend of Figure 11.10, the fibrillar material is thought to consist of rDNA templates and nascent rRNA transcripts. In the following sections, we will examine the process by which these rRNA transcripts are synthesized.

Synthesizing the rRNA Precursor

Oocytes are typically very large cells (e.g., 100 μm in diameter in mammals); those of amphibians are generally enormous (up to 2.5 mm in diameter). During the growth of amphibian oocytes, the amount of rDNA in the cell is greatly increased as are the number of nucleoli (Figure 11.11a). The selective amplification of rDNA is necessary to provide the large numbers of ribosomes that are required by the fertilized egg to begin embryonic development. Because these oocytes contain hundreds of nucleoli, each actively manufacturing rRNA, they are ideal subjects for the study of rRNA synthesis and processing.

Our understanding of rRNA synthesis was greatly advanced in the late 1960s with the development of techniques by Oscar Miller, Jr. of the University of Virginia to visualize "genes in action" with the electron microscope. To carry out these studies, the fibrillar cores of oocyte nucleoli were gently dispersed to reveal the presence of a large circular fiber. When one of these fibers was examined in the electron microscope, it was seen to resemble a chain of Christmas trees (Figure 11.11b,c). The electron micrographs of Figure 11.11 reveal a number of aspects of nucleolar activity and rRNA synthesis.

1. The micrograph in Figure 11.11b shows numerous genes for ribosomal RNA situated one after the other along a single DNA molecule, thus revealing the tandem arrangement of the repeated rRNA genes.

2. The micrograph in Figure 11.11b shows a static image of dynamic events that occur in the nucleolus. We can inter-

Nucleoli

(a)

(b)   2 μm

(c)   0.5 μm

FIGURE 11.11 The synthesis of ribosomal RNA. (a) Light micrograph of an isolated nucleus from a *Xenopus* oocyte stained to reveal the hundreds of nucleoli. (b) Electron micrograph of a segment of DNA isolated from one of the nucleoli from a *Xenopus* oocyte. The DNA (called rDNA) contains the genes that encode the two large ribosomal RNAs, which are carved from a single primary transcript. Numerous genes are shown, each in the process of being transcribed. Transcription is evident by the fibrils that are attached to the DNA. These fibrils consist of nascent RNA and associated protein. The stretches of DNA between the transcribed genes are nontranscribed spacers. The arrows indicate sites where transcription is initiated. (c) A closer view of two nucleolar genes being transcribed. The length of the nascent rRNA primary transcript increases with increasing distance from the point of initiation. The RNA polymerase molecules at the base of each fibril can be seen as dots. (A: FROM DONALD D. BROWN AND IGOR B. DAWID, SCIENCE 160:272, 1968; © COPYRIGHT 1968, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE. B–C: COURTESY OF OSCAR L. MILLER, JR., AND BARBARA R. BEATTY.)

FIGURE 11.12 The rRNA transcription unit. The top drawing depicts the appearance of a portion of the DNA from a nucleolus as it is transcribed into rRNA. The lower drawings illustrate one of the transcription units that codes for rRNA in *Xenopus* and the mouse. Those parts of the DNA that encode the mature rRNA products are shown in blue. Regions of transcribed spacer, that is, portions of the DNA that are tran- scribed, but whose corresponding RNAs are degraded during process- ing, are shown in yellow. The nontranscribed spacer, which lies between the transcription units, contains the promoter region at the 5′ side of the gene. (AFTER B. SOLLNER-WEBB AND E. B. MOUGEY, TRENDS BIOCHEM. SCI. 16:59, 1991.)

pret this photograph to provide a great deal of informa- tion about the process of rRNA transcription. Each of the 100 or so fibrils emerging from the DNA as a branch of a Christmas tree is a nascent rRNA transcript caught in the act of elongation. The dark granule at the base of each fibril, visible in the higher magnification photograph of Figure 11.11*c*, is the RNA polymerase I molecule synthe- sizing that transcript. The length of the fibrils increases gradually from one end of the Christmas tree trunk to the other. The shorter fibrils are RNA molecules of fewer nu- cleotides that are attached to polymerase molecules bound to the DNA closer to the transcription initiation site. The longer the fibril, the closer the transcript is to completion. The length of DNA between the shortest and longest RNA fibrils corresponds to a single transcription unit. The promoter lies just upstream from the transcription initiation site. The high density of RNA polymerase molecules along each transcription unit (about 1 every 100 base pairs of DNA) reflects the high rate of rRNA synthesis in the nucleoli of these oocytes.

3. It can be seen in Figure 11.11*c* that the nascent RNA transcripts contain associated particles. These particles consist of RNA and protein that work together to convert the rRNA precursors to their final rRNA products and assemble them into ribosomal subunits. These processing events occur as the RNA molecule is being synthesized.

4. It can be noted from Figure 11.11*b* that the region of the DNA fiber between adjacent transcription units is devoid of nascent RNA chains. Because this region of the riboso- mal gene cluster is not transcribed, it is referred to as the **nontranscribed spacer** (Figure 11.12). Nontranscribed spacers are present between various types of tandemly re- peated genes, including those of tRNAs and histones.

Processing the rRNA Precursor

Eukaryotic ribosomes have four distinct ribosomal RNAs, three in the large subunit and one in the small subunit. In humans, the large subunit contains a 28S, 5.8S, and 5S RNA molecule, and the small subunit contains an 18S RNA mole- cule.[1] Three of these rRNAs (the 28S, 18S, and 5.8S) are carved by various nucleases from a single primary transcript (called the **pre-rRNA**). The 5S rRNA is synthesized from a separate RNA precursor outside the nucleolus. We will begin this discussion with the pre-rRNA.

Two of the peculiarities of the pre-rRNA, as compared with other RNA transcripts, are the large number of methy- lated nucleotides and pseudouridine residues. By the time a human pre-rRNA precursor is first cleaved, over 100 methyl groups have been added to ribose groups in the molecule, and approximately 95 of its uridine residues have been enzymati- cally converted to pseudouridine (see Figure 11.15*a*). All of these modifications occur after the nucleotides are incorpo- rated into the nascent RNA, that is, *posttranscriptionally*. The altered nucleotides are located at specific positions and are clustered within portions of the molecule that have been conserved among organisms. All of the nucleotides in the pre- rRNA that are altered remain as part of the final products, while unaltered sections are discarded during processing. The functions of the methyl groups and pseudouridines are unclear. These modified nucleotides may protect parts of the pre-rRNA from enzymatic cleavage, promote folding of

---

[1]The S value (or Svedberg unit) refers to the sedimentation coefficient of the RNA; the larger the number, the more rapidly the molecule moves through a field of force during centrifugation, and, (for a group of chemically similar molecules) the larger the size of the molecule. The 28S, 18S, 5.8S, and 5S RNAs consist of nucleotide lengths of about 5000, 2000, 160, and 120 nucleotides, respectively.

rRNAs into their final three-dimensional structures, and/or promote interactions of rRNAs with other molecules.

Because rRNAs are so heavily methylated, their synthesis can be followed by incubating cells with radioactively labeled methionine, a compound that serves in most cells as a methyl group donor. The methyl group is transferred enzymatically from methionine to nucleotides in the pre-rRNA. When [^{14}C] methionine is provided to cultured mammalian cells for a short period of time, a considerable fraction of the incorporated radioactivity is present in a 45S RNA molecule, equal to a length of about 13,000 nucleotides. The 45S RNA is cleaved into smaller molecules that are then trimmed down to the 28S, 18S, and 5.8S rRNA molecules. The combined length of the three mature rRNAs is approximately 7000 nucleotides, or a little more than half of the primary transcript.

Some of the steps in the processing pathway from the 45S pre-rRNA to the mature rRNAs can be followed by incubat-ing mammalian cells very briefly with labeled methionine and then chasing the cells in nonlabeled medium for varying lengths of time (Figure 11.13). (This type of "pulse-chase" ex-periment was discussed on page 267). As noted above, the first species to become labeled in this type of experiment is the 45S primary transcript, which is seen as a peak of radioactivity (red dotted line) in the nucleolar RNA fraction after 10 minutes. After about an hour, the 45S RNA has disappeared from the nucleolus and is largely replaced by a 32S RNA, which is one of the two major products produced from the 45S primary transcript. The 32S RNA is seen as a distinct peak in the nu-cleolar fraction from 40 minutes to 150 minutes. The 32S RNA is a precursor to the mature 28S and 5.8S rRNAs. The other major product of the 45S pre-rRNA leaves the nucleo-lus quite rapidly and appears in the cytoplasm as the mature 18S rRNA (seen in the 40-minute cytoplasmic fraction). After a period of two or more hours, nearly all of the radio-
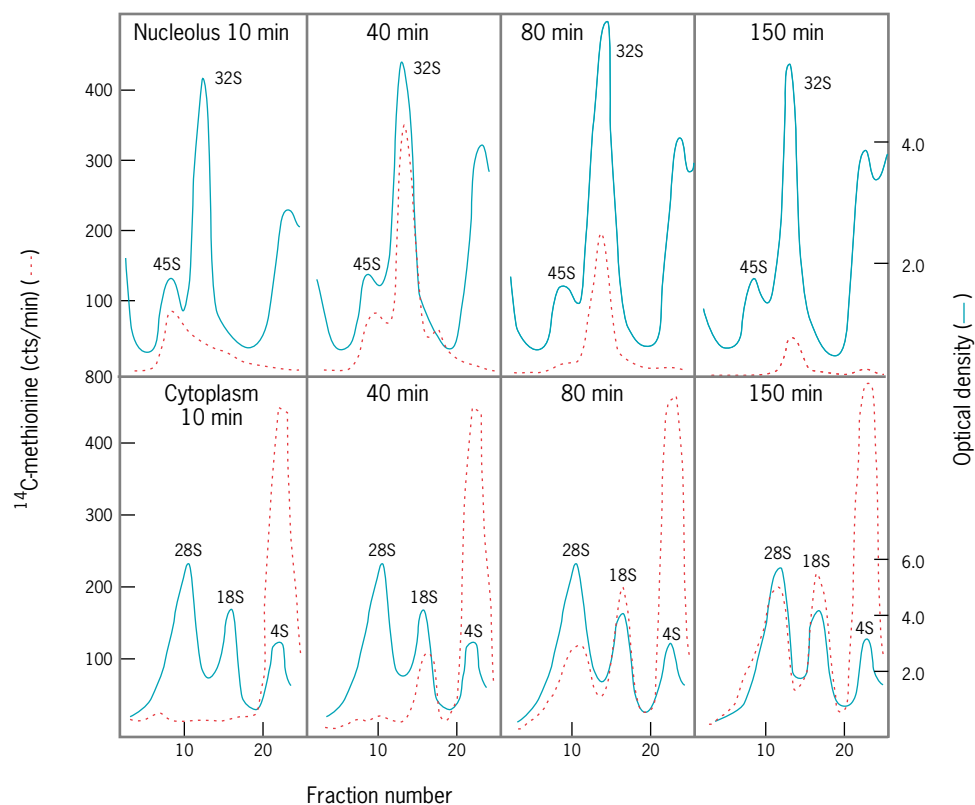
FIGURE 11.13 Kinetic analysis of rRNA synthesis and processing. A culture of mammalian cells was incubated for 10 minutes in [^{14}C]me-thionine and was then chased in unlabeled medium for various times as indicated in each panel. After the chase, cells were washed free of isotope and homogenized, and nucleolar and cytoplasmic fractions were pre-pared. The RNA was extracted from each fraction and analyzed by su-crose density-gradient sedimentation. As discussed in Section 18.10, this technique separates RNAs according to size (the larger the size, the closer the RNA is to the bottom of the tube, which corresponds to frac-tion 1). The continuous line represents the UV absorbance of each cellu-lar fraction, which provides a measure of the amount of RNA of each size class. This absorbance profile does not change with time. The dot-ted line shows the radioactivity at various times during the chase. The graphs of nucleolar RNA (upper profiles) show the synthesis of the 45S rRNA precursor and its subsequent conversion to a 32S molecule, which is a precursor to the 28S and 5.8S rRNAs. The other major product of the 45S precursor leaves the nucleus very rapidly and, therefore, does not appear prominently in the nucleolar RNA. The lower profiles show the time course of the appearance of the mature rRNA molecules in the cy-toplasm. The 18S rRNA appears in the cytoplasm well in advance of the larger 28S species, which correlates with the rapid exodus of the former from the nucleolus. (FROM H. GREENBERG AND S. PENMAN, J. MOL. BIOL. 21:531, 1966; COPYRIGHT 1966, BY PERMISSION OF THE PUBLISHER ACADE-MIC PRESS.)
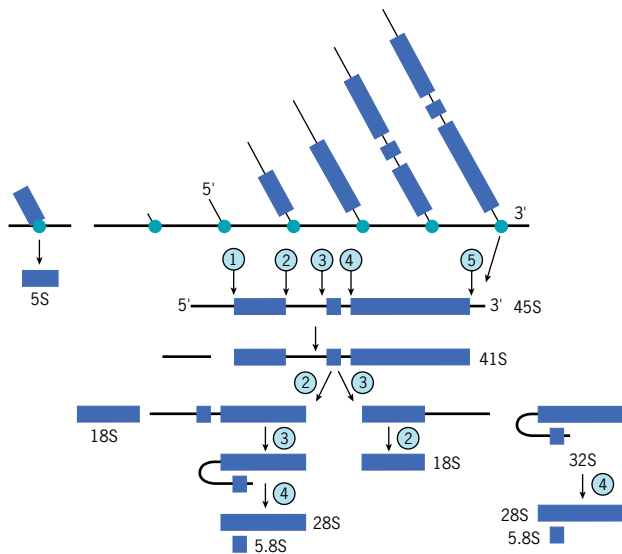
FIGURE 11.14 **A proposed scheme for the processing of mammalian ribosomal RNA.** The primary transcript for the rRNAs is a 45S molecule of about 13,000 bases. The principal cleavage events during processing of this pre-rRNA are indicated by the circled numbers. Cleavage of the primary transcript at sites 1 and 5 removes the 5′ and 3′ external transcribed sequences and produces a 41S intermediate. The second cleavages can occur at either site 2 or 3 depending on the type of cell. Cleavage at site 3 generates the 32S intermediate seen in curves of the previous figure. During the final processing steps, the 28S and 5.8S sections are separated from one another, and the ends of the various intermediates are trimmed down to their final mature size. (AFTER R. P. PERRY, J. CELL BIOL. 91:29S, 1981; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

activity has left the nucleolus, and most has accumulated in the 28S and 18S rRNAs of the cytoplasm. The radioactivity in the 4S RNA peak includes the 5.8S rRNA as well as methyl groups that have been transferred to small tRNA molecules. Figure 11.14 shows a likely pathway in the processing of an rRNA primary transcript.

The Role of the snoRNAs   The processing of the pre-rRNA is accomplished with the help of a large number of **small, nucleolar RNAs** (or **snoRNAs**) that are packaged with particular proteins to form particles called **snoRNPs (small, nucleolar ribonucleoproteins)**. Electron micrographs indicate that snoRNPs begin to associate with the rRNA precursor before it is fully transcribed. The first RNP particle to attach to a pre-rRNA transcript contains the U3 snoRNA and more than two dozen different proteins. This huge component of the rRNA-processing machinery can be seen as a ball bound to the outer end of each nascent RNA fibril in Figure 11.11c, where it catalyzes the removal of the 5′ end of the transcript (Figure 11.14). Some of the other enzymatic cleavages indicated in Figure 11.14 are thought to be catalyzed by the "exosome," which is an RNA degrading machine that consists of nearly a dozen different exonucleases.

U3 and several other snoRNAs were identified years ago because they are present in large quantities (about 10^6 copies per cell). Subsequently, a different class of snoRNAs present

at lower concentration (about 10^4 copies per cell) was discovered. These low-abundance snoRNAs can be divided into two groups based on their function and similarities in nucleotide sequence. Members of one group (called the *box C/D snoRNAs*) determine which nucleotides in the pre-rRNA will have their ribose moieties methylated, whereas members of the other group (called the *box H/ACA snoRNAs*) determine which uridines are converted to pseudouridines. The structures of nucleotides modified by these two reactions are shown in Figure 11.15a.

Both groups of snoRNAs contain relatively long stretches (10 to 21 nucleotides) that are complementary to sections of the rRNA transcript. These snoRNAs provide an excellent example of the principle that single-stranded nucleic acids having complementary nucleotide sequences are capable of forming double-stranded hybrids. In this case, each snoRNA binds to a specific portion of the pre-rRNA to form an RNA–RNA duplex. The bound snoRNA then guides an enzyme—either a methylase or a pseudouridylase—within the snoRNP to modify a particular nucleotide in the pre-rRNA. Taken together, there are roughly 200 different snoRNAs, one for each site in the pre-rRNA that is ribose-methylated or pseudouridylated. If the gene encoding one of these snoRNAs is deleted, one of the nucleotides of the pre-rRNA fails to be enzymatically modified. The mechanism of action of the two types of snoRNAs is illustrated in Figure 11.15b,c.

The nucleolus is the site not only of rRNA processing, but also of assembly of the two ribosomal subunits. Two types of proteins become associated with the RNA as it is processed: ribosomal proteins that remain in the subunits, and accessory proteins that have a transient interaction with the rRNA intermediates and are required only for processing. Among this latter group are more than a dozen RNA helicases, which are enzymes that unwind regions of double-stranded RNA. These enzymes are presumably involved in the many structural rearrangements that occur during ribosome formation, including the association and dissociation of the snoRNAs.

Synthesis and Processing of the 5S rRNA

A 5S rRNA, about 120 nucleotides long, is present as part of the large ribosomal subunit of both prokaryotes and eukaryotes. In eukaryotes, the 5S rRNA molecules are encoded by a large number of identical genes that are separate from the other rRNA genes and are located outside the nucleolus. Following its synthesis, the 5S rRNA is transported to the nucleolus to join the other components involved in the assembly of ribosomal subunits.

The 5S rRNA genes are transcribed by RNA polymerase III. RNA polymerase III is unusual among the three polymerases in that it can bind to a promoter site located within the transcribed portion of the target gene.[2] The inter-

---

[2]RNA polymerase III transcribes several different RNAs. It binds to an internal promoter when transcribing a pre-5S RNA or pre-tRNA, but binds to an upstream promoter when transcribing the precursors for several others, including U6 snRNA.

FIGURE 11.15 **Modifying the pre-rRNA.** (*a*) The most frequent modifications to nucleotides in a pre-rRNA are the conversion (isomerization) of a uridine to a pseudouridine and the methylation of a ribose at the 2′ site of the sugar. To convert uridine to pseudouridine, the N1—C1′ bond is cleaved, and the uracil ring is rotated through 120°, which brings the C5 of the ring into place to form the new bond with C1′ of ribose. These chemical modifications are catalyzed by a protein component of the snoRNPs called dyskerin. (*b*) The formation of an RNA–RNA duplex between the U20 snoRNA and a portion of the pre-rRNA that leads to 2′ ribose methylation. In each case, the methylated nucleotide in the rRNA is hydrogen bonded to a nucleotide of the snoRNA that is located five base pairs from box D. Box D, which contains the invariable sequence CUGA, is present in all snoRNAs that guide ribose methylation. (*c*) The formation of an RNA–RNA duplex between U68 snoRNA and a portion of the pre-rRNA that leads to conversion of a uridine to pseudouridine (ψ). Pseudouridylation occurs at a fixed site relative to a hairpin fold in the snoRNA. The snoRNAs that guide pseudouridylation have a common ACA sequence. (B: AFTER J.-P. BACHELLERIE AND J. CAVAILLÉ, TRENDS IN BIOCH. SCI. 22:258, 1997; COPYRIGHT 1997, WITH PERMISSION FROM ELSEVIER SCIENCE; C: AFTER P. GANOT ET AL., CELL 89:802, 1997.)

nal position of the promoter was first clearly demonstrated by introducing modified 5S rRNA genes into host cells and determining the ability of that DNA to serve as templates for the host polymerase III. It was found that the entire 5′ flanking region could be removed and the polymerase would still transcribe the DNA starting at the normal initiation site. If, however, the deletion included the central part of the gene, the polymerase would not transcribe the DNA or even bind to it. If the internal promoter from a 5S rRNA gene is inserted into another region of the genome, the new site becomes a template for transcription by RNA polymerase III.

Transfer RNAs

Plant and animal cells are estimated to have approximately 50 different species of transfer RNA, each encoded by a repeated DNA sequence. The degree of repetition varies with the organism: yeast cells have a *total* of about 275 tRNA genes, fruit flies about 850, and humans about 1300. Transfer RNAs are synthesized from genes that are found in small clusters scattered around the genome. A single cluster typically contains multiple copies of *different* tRNA genes, and conversely, the DNA sequence encoding a given tRNA is typically found in more than one cluster. The DNA within a cluster (*tDNA*) consists largely of nontranscribed spacer sequences with the tRNA coding sequences situated at irregular intervals in a tandemly repeated arrangement (Figure 11.16).

Like the 5S rRNA, tRNAs are transcribed by RNA polymerase III and the promoter sequence lies within the coding section of the gene rather than being located at its 5′ flank. The primary transcript of a transfer RNA molecule is larger than the final product, and pieces on both the 5′ and 3′ sides of the precursor tRNA (and a small interior piece in some cases) must be trimmed away. In addition, numerous bases must be modified (see Figure 11.42). One of the enzymes involved in pre-tRNA processing is an endonuclease called
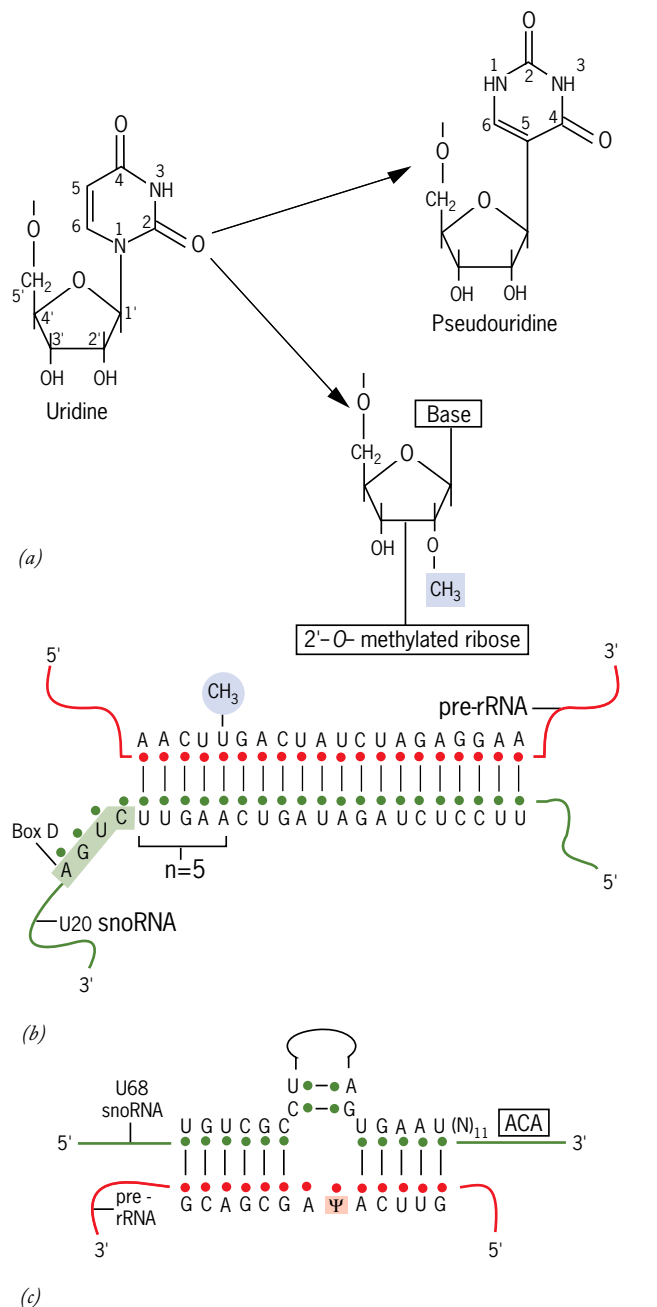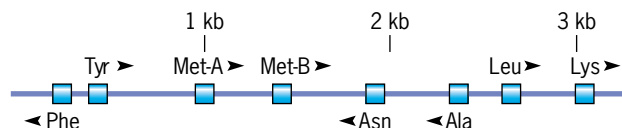
FIGURE 11.16 **The arrangement of genes that code for transfer RNAs in *Xenopus*.** A 3.18-kilobase fragment of DNA, showing the arrangement of various tRNA genes and spacers. (FROM S. G. CLARKSON ET AL., IN D. D. BROWN, ED., DEVELOPMENTAL BIOLOGY USING PURIFIED GENES, ACADEMIC PRESS, 1981.)

ribonuclease P, which is present in both bacterial and eukaryotic cells and consists of both RNA and protein subunits. It is the RNA subunit of ribonuclease P that catalyzes cleavage of the pre-tRNA substrate, a subject discussed in the Experimental Pathways, on page 469.

REVIEW ?

1. Describe the differences between a primary transcript, a transcription unit, a transcribed spacer, and a mature rRNA.
2. Draw a representation of an electron micrograph of rDNA being transcribed. Label the nontranscribed spacer, the transcribed spacer, the RNA polymerase molecules, the U3 snRNP, and the promoter.
3. Compare the organization of the genes that code for the large rRNAs and the tRNAs within the vertebrate genome.

11.4 SYNTHESIS AND PROCESSING OF MESSENGER RNAs

When eukaryotic cells are incubated for a short period in [^{3}H]uridine or [^{32}P]phosphate and immediately killed, most of the radioactivity is incorporated into a large group of RNA molecules that have the following properties: (1) they have large molecular weights (up to about 80S, or 50,000 nucleotides); (2) as a group, they are represented by RNAs of diverse (heterogeneous) nucleotide sequence; and (3) they are found only in the nucleus. Because of these properties, these RNAs are referred to as **heterogeneous nuclear RNAs** (**hnRNAs**), and they are indicated by the red radioactivity line in Figure 11.17a. When cells that have been incubated in [^{3}H]uridine or [^{32}P]phosphate for a brief pulse are placed into unlabeled medium, chased for several hours before they are killed and the RNA extracted, the amount of radioactivity in the large nuclear RNAs drops sharply and appears instead in much smaller mRNAs found in the cytoplasm (red line of Figure 11.17b). These early experiments, begun initially by James Darnell, Jr., Klaus Scherrer and colleagues, suggested that the large, rapidly labeled hnRNAs were primarily precursors to the smaller cytoplasmic mRNAs. This interpretation has been fully substantiated by a large body of research over the past 40 years.

It is important to note that the blue and red lines in Figure 11.17 follow a very different course. The blue lines, which indicate the optical density (i.e., the absorbance of UV light) of each fraction, provide information about the amount of RNA in each fraction following centrifugation. It is evident from the blue lines that most of the RNA in the cell is present as 18S and 28S rRNA (along with various small RNAs that stay near the top of the tube). The red lines, which indicate the radioactivity in each fraction, provide information about the number of radioactive nucleotides incorporated into different-sized RNAs during the brief pulse. It is evident from these graphs that neither the hnRNA (Figure 11.17a) nor the

(a)

(b)
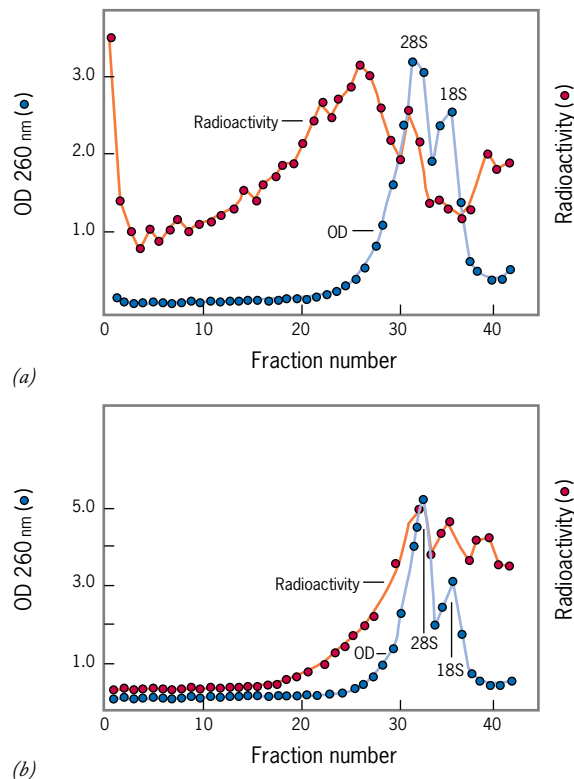
FIGURE 11.17 The formation of heterogeneous nuclear RNA (hnRNA) and its conversion into smaller mRNAs. (a) Curves showing the sedimentation pattern of total RNA extracted from duck blood cells after exposure to [^{32}P]phosphate for 30 minutes. The larger the RNA, the farther it travels during centrifugation and the closer to the bottom of the tube it ultimately lies. The absorbance (blue line) indicates the total amount of RNA in different regions of the centrifuge tube, whereas the red line indicates the corresponding radioactivity. It is evident that most of the newly synthesized RNA is very large, much larger than the stable 18S and 28S rRNAs. These large RNAs are the hnRNAs. (b) The absorbance and radioactivity profiles of the RNA were extracted from cells that were pulse-labeled for 30 minutes as in part a, but then chased for 3 hours in the presence of actinomycin D, which prevents the synthesis of additional RNA. It is evident that the large hnRNAs have been processed into smaller RNA products. (FROM G. ATTARDI ET AL., J. MOL. BIOL. 20:160, 1966; COPYRIGHT 1966, BY PERMISSION OF THE PUBLISHER ACADEMIC PRESS.)

mRNA (Figure 11.17b) constitute a significant fraction of the RNA in the cell. If they did, there would be a greater correspondence between the blue and red lines. Thus, even though the mRNAs (and their heterogeneous nuclear RNA precursors) constitute only a small percentage of the total RNA of most eukaryotic cells, they constitute a large percentage of the RNA that is being synthesized by that cell at any given moment (Figure 11.17a). The reason that there is little or no evidence of the hnRNAs and mRNAs in the optical density plots of Figure 11.17 is that these RNAs are degraded after relatively brief periods of time. This is particularly true of the hnRNAs, which are processed into mRNAs (or completely degraded) even as they are being synthesized. In contrast, the rRNAs and tRNAs have half-lives that are measured in days or weeks and thus gradually accumulate to become the

predominant species in the cell. Some accumulation of radioactivity in the mature 28S and 18S rRNAs can be seen after a 3-hour chase (Figure 11.17*b*). The half-lives of mRNAs vary depending on the particular species, ranging from about 15 minutes to a period of days.

The Machinery for mRNA Transcription

All eukaryotic mRNA precursors are synthesized by RNA polymerase II, an enzyme composed of a dozen different subunits that is remarkably conserved from yeast to mammals. RNA polymerase II binds the promoter with the cooperation of a number of **general transcription factors** (**GTFs**) to form a *preinitiation complex* (*PIC*). These proteins are referred to as general transcription factors because the same ones are required for accurate initiation of transcription of a diverse array of genes in a wide variety of different organisms. The promoter elements that nucleate PIC assembly lie to the 5′ side of each transcription unit, although the enzyme makes contacts with the DNA on both sides of the transcription start site (see Figure 11.19*b*). We will restrict the discussion to the best studied promoters, which are those associated with highly expressed, tissue-specific genes such as the ovalbumin gene, which encodes the white of a chicken egg, and the globin genes, which encode the polypeptides of hemoglobin. A critical portion of the promoter of such genes lies between 24 and 32 bases upstream from the site at which transcription is initiated (Figure 11.18*a*). This region often contains a consensus sequence that is either identical or very similar to the oligonucleotide 5′-TATAAA-3′ and is known as the TATA box.

The first step in assembly of the preinitiation complex is binding of a protein, called *TATA-binding protein* (*TBP*), that recognizes the TATA box of these promoters (Figure 11.18*b*). Thus, as in bacterial cells, a purified eukaryotic polymerase is

FIGURE 11.18 Initiation of transcription from a eukaryotic polymerase II promoter. (*a*) Nucleotide sequence of the region just upstream from the site where transcription is initiated in three different eukaryotic genes. The TATA box is indicated by the blue shading. Many eukaryotic promoters contain a second, conserved core promoter element called the initiator (Inr), which includes the site where transcription is initiated (shown in orange). Other promoter elements are depicted in Figure 12.40. It should be noted that (1) most eukaryotic promoters lack a recognizable TATA box and (2) numerous other promoter elements, for example, the DPE shown in part *b*, have been identified that lie downstream of the transcription start site. Different genes contain different combinations of promoter elements so that only a subset are required to nucleate PIC assembly. (*b*) A highly schematic model of the steps in the assembly of the preinitiation complex for RNA polymerase II. The polymerase itself is denoted RNAPII; the other components are the various general transcription factors required in assembly of the complete complex. TFIID includes the TBP subunit, which specifically binds to the TATA box, and a number of other subunits, which collectively are called TBP-associated factors (TAFs). TFIIB is thought to provide a binding site for RNA polymerase. TFIIF, which contains a subunit homologous to the bacterial σ factor, is bound to the entering polymerase. TFIIH contains 10 subunits, 3 of which possess enzymatic activities.

not able to recognize a promoter directly and cannot initiate accurate transcription on its own. TBP is present as a subunit of a much larger protein complex called TFIID (*t*ranscription *f*actor for polymerase *II*, fraction *D*).[3] X-ray crystallography has revealed that binding of TBP to a polymerase II promoter causes a dramatic distortion in conformation of the DNA. As shown in Figure 11.19*a*, TBP inserts itself into the minor groove of the double helix, bending the DNA molecule more

---

[3]TBP is actually a universal transcription factor that mediates the binding of all three eukaryotic RNA polymerases. TBP is present as one of the subunits of three different protein complexes. As a subunit of TFIID, TBP promotes the binding of RNA polymerase *II*. As subunits of the proteins SL1 or TFIIIB, TBP promotes the binding of RNA polymerases I and III, respectively. Polymerase I and III promoters lack a TATA box, as do many polymerase II promoters, yet all of these regions of the DNA bind TBP.

Chicken ovalbumin   GAGGCTATATATTCCCCAGGGCTCAGCCAGTGTCTGTACA

Rabbit β–globin   TTGGGCATAAAAGGCAGAGCAGGGCAGCTGCTGCTAACACT

Mouse β–globin major   GAGCATATAAGGTGAGGTAGGATCAGTTGCTCCTCACATTT

(a)

(b)

(a)   (b)

FIGURE 11.19   Structural models of the formation of the preinitiation complex. (*a*) Model of the complex formed by DNA and three of the GTFs, TBP of TFIID, TFIIA, and TFIIB. Interaction between the TATA box and TBP bends the DNA approximately 80° and allows TFIIB to bind to the DNA both upstream and downstream of the TATA box. (*b*) Top and bottom view of a model of the preinitiation complex. Unlike the schematic model of Figure 11.18*b*, DNA (shown in white) is thought to wrap around the preinitiation complex so that GTFs can contact the DNA on both sides of the transcription start site. (A: FROM GOURISANKAR GHOSH AND GREGORY D. VAN DUYNE, STRUCTURE 4:893, 1996; B: FROM M. DOUZIECH ET AL., MOL. CELL BIOL. 20:8175, 2000; COURTESY OF BENOIT COULOMBE.)

than 80° at the site of DNA–protein interaction. While TBP binds TATA, other subunits of the TFIID complex bind to other regions of the DNA, including elements that lie downstream of the transcription start site.

Binding of TFIID sets the stage for the assembly of the complete preinitiation complex, which is thought to occur in a stepwise manner as depicted in Figure 11.18*b*. Interaction of three of the GTFs (TBP of TFIID, TFIIA, and TFIIB) with DNA is shown in Figure 11.19*a*. The presence of these three GTFs bound to the promoter provides a platform for the subsequent binding of the huge, multisubunit RNA polymerase with its attached TFIIF (Figure 11.18*b*). Once the RNA polymerase–TFIIF is in position, another pair of GTFs (TFIIE and TFIIH) join the complex and transform the polymerase into an active, transcribing machine. A three-dimensional model of the preinitiation complex is shown in Figure 11.19*b*.

TFIIH is the only GTF known to possess enzymatic activities. One of the subunits of TFIIH functions as a protein kinase to phosphorylate RNA polymerase (discussed below), whereas two other subunits of this protein function as DNA unwinding enzymes (helicases). DNA helicase activity is required to separate the DNA strands of the promoter, allowing the polymerase access to the template strand. Once transcription begins, certain of the GTFs (including TFIID) may be left behind at the promoter, while others are released from the complex (Figure 11.20). As long as TFIID remains bound to the promoter, additional RNA polymerase molecules may be able to attach to the promoter site and initiate additional rounds of transcription without delay.

The carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II has an unusual structure; it consists of a sequence of seven amino acids (-Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7-) that is repeated over and over. In humans, the CTD consists of 52 repeats of this heptapeptide. All seven residues of the heptapeptide can be enzymatically modified in one way or another; we will limit the discussion to serines two and five, which are prime candidates for phosphorylation by protein kinases. The RNA polymerase that assem-

bles into the preinitiation complex is not phosphorylated, whereas the same enzyme engaged in transcription is heavily phosphorylated; all of the added phosphate groups are local-
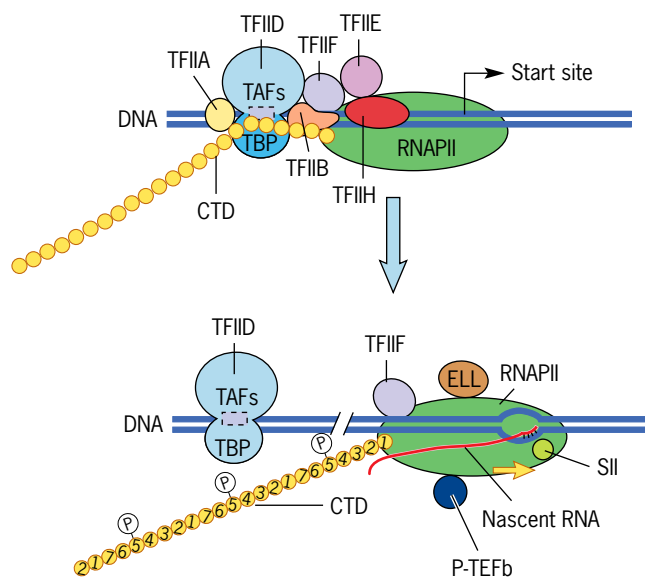
FIGURE 11.20   Initiation of transcription by RNA polymerase II is associated with phosphorylation of the C-terminal domain (CTD). The initiation of transcription is associated with TFIIH-catalyzed phosphorylation of serine residues at the 5 position of each heptad repeat of the CTD. Phosphorylation is thought to provide the trigger for the separation of the transcriptional machinery from the general transcription factors and/or the promoter DNA. Escape from the promoter is associated with major changes in the conformation of the polymerase, which converts it from an enzyme that "has trouble" synthesizing RNA to one that is highly processive. SII, ELL, and P-TEFb are three of a number of elongation factors that may become associated with the polymerase as it moves along the DNA. SII helps the polymerase get moving again after it pauses, whereas P-TEFb is a kinase that phosphorylates the #2 serine residues of the CTD after elongation begins. Phosphorylation of #2 serine residues is thought to promote the recruitment of RNA splicing and polyadenylation factors, whose activities are discussed in the following sections (see Figure 11.35).

ized in the CTD (Figure 11.20). CTD phosphorylation can be catalyzed by at least four different protein kinases, including TFIIH, which phosphorylates the serine residues at position #5. Phosphorylation of the polymerase by TFIIH may act as the trigger that uncouples the enzyme from the GTFs and/or the promoter DNA, allowing the enzyme to escape from the preinitiation complex and move down the DNA template. As the RNA polymerase moves along the gene being transcribed, another kinase (P-TEFb) phosphorylates the CTD on the serine residues at position #2 (see Figure 11.35). This change in phosphorylation pattern is thought to facilitate the recruitment of additional protein factors involved in RNA processing and transcription termination. In this way, the CTD acts as a platform for the dynamic gain and loss of factors required for the formation of a mature mRNA. According to some estimates, an elongating RNA polymerase II may contain over 50 components and constitute a total molecular mass of more than 3 million daltons.

Termination of transcription by RNA polymerase II is not well understood. There is no evidence that the DNA of protein-coding genes contains well-defined termination sequences as in bacterial cells. In fact, a transcribing RNA polymerase II can travel a variable and extensive distance past the point that will ultimately give rise to the 3′ terminus of the processed mRNA.

Together, RNA polymerase II and its GTFs are sufficient to promote a low, basal level of transcription from most promoters under in vitro conditions. As will be discussed at length in Chapter 12, a variety of *specific* transcription factors are able to bind at numerous sites in the regulatory regions of the DNA. These specific transcription factors can determine (1) whether or not a preinitiation complex assembles at a particular promoter and/or (2) the rate at which the polymerase initiates new rounds of transcription from that promoter. Before discussing the pathway by which mRNAs are produced, let us first describe the structure of mRNAs so that the reasons for some of the processing steps will be clear.

The Structure of mRNAs   Messenger RNAs share certain properties:

1. They contain a continuous sequence of nucleotides encoding a specific polypeptide.
2. They are found in the cytoplasm.
3. They are attached to ribosomes when they are translated.
4. Most mRNAs contain a significant noncoding segment, that is, a portion that does not direct the assembly of amino acids. For example, approximately 25 percent of each globin mRNA consists of noncoding, nontranslated regions (Figure 11.21). Noncoding portions are found at both the 5′ and 3′ ends of a messenger RNA and contain sequences that have important regulatory roles (Section 12.6).
5. Eukaryotic mRNAs have special modifications at their 5′ and 3′ termini that are not found on either bacterial mRNAs or on tRNAs or rRNAs. The 3′ end of nearly all eukaryotic mRNAs has a string of 50 to 250 adenosine residues that form a poly(A) tail, whereas the 5′ end has a methylated guanosine cap (Figure 11.21).

We will return shortly to describe how mRNAs are provided with their specialized 5′ and 3′ termini. First, however, it is necessary to take a short detour to understand how mRNAs are formed in the cell.

FIGURE 11.21 **Structure of the human β-globin mRNA.** The mRNA contains a 5′ methylguanosine cap, a 5′ and 3′ noncoding region that flanks the coding segment, and a 3′ poly(A) tail. The lengths of each segment are given in numbers of nucleotides. The length of the poly(A) tail is variable. It typically begins at a length of about 250 nucleotides and is gradually reduced in length, as discussed on page 526. The structure of the 5′ cap is shown.

Split Genes: An Unexpected Finding

Almost as soon as hnRNAs were discovered, it was proposed that this group of rapidly labeled nuclear RNAs were precursors to cytoplasmic mRNAs (page 434). The major sticking point was the difference in size between the two RNA populations: the hnRNAs were several times the size of the mRNAs (Figure 11.22). Why would cells synthesize huge molecules that were precursors to much smaller versions? Early studies on the processing of ribosomal RNA had shown that mature RNAs were carved from larger precursors. Recall that large segments were removed from both the 5′ and 3′ sides of various rRNA intermediates (Figure 11.14) to yield the final, mature rRNA products. It was thought that a similar pathway might account for the processing of hnRNAs to mRNAs. But mRNAs constitute such a diverse population
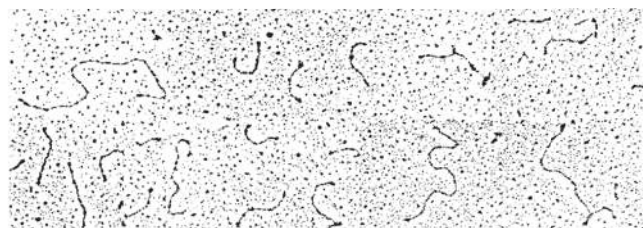
(a)
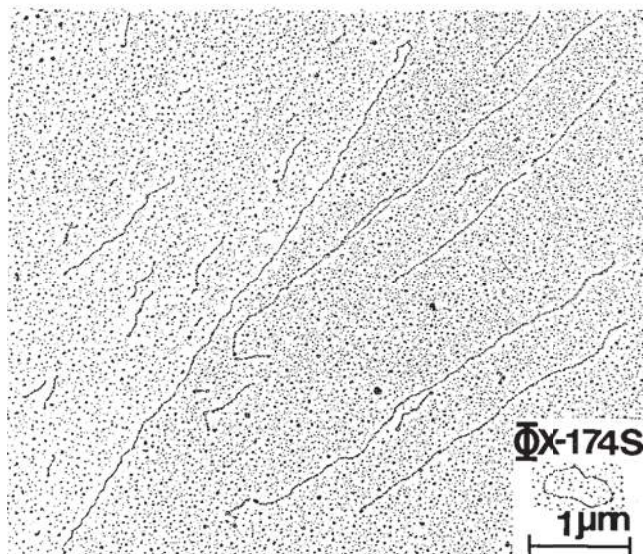
(b)
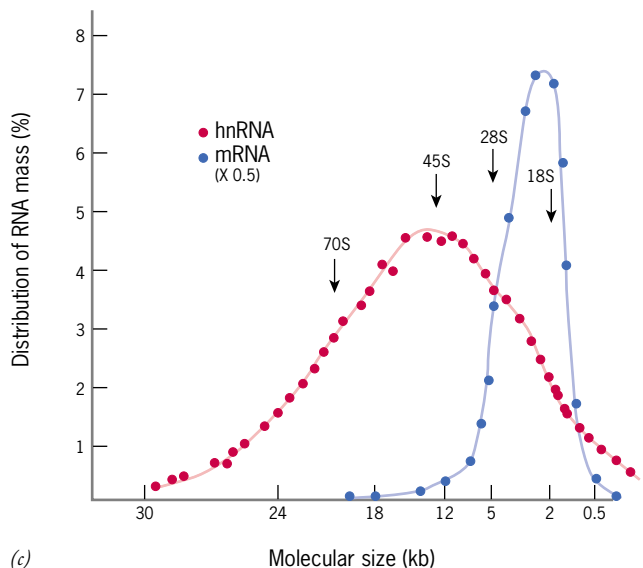
(c)

FIGURE 11.22 The difference in size between hnRNAs and mRNAs. (*a,b*) Electron micrographs of metal-shadowed preparations of poly(A)-mRNA (*a*) and poly(A)-hnRNA (*b*) molecules. Representative size classes of each type are shown. Reference molecule is φX-174 viral DNA. (*c*) Size distribution of hnRNA and mRNA from mouse L cells as determined by density-gradient sedimentation. The red line represents rapidly labeled hnRNA, whereas the purple line represents mRNA that was isolated from polyribosomes after a 4-hour labeling period. The abscissa has been converted from fraction number (indicated by the points) to molecular size by calibration of the gradients. (FROM JOHN A. BANTLE AND WILLIAM E. HAHN, CELL 8:145, 1976; BY PERMISSION OF CELL PRESS.)

that it was impossible to follow the steps in the processing of a single mRNA species as had been attempted for the rRNAs. The problem was solved by an unexpected discovery.

Until 1977, molecular biologists assumed that a continuous linear sequence of nucleotides in a messenger RNA is complementary to a continuous sequence of nucleotides in one strand of the DNA of a gene. Then in that year, a remarkable new finding was made by Phillip Sharp and his colleagues at MIT and Richard Roberts, Louise Chow, and their colleagues at the Cold Spring Harbor Laboratories in New York. These groups found that the mRNAs they were studying were transcribed from segments of DNA that were separated from one another along the template strand.

The first observations of major importance were made during the analysis of transcription of the adenovirus genome. Adenovirus is a pathogen capable of infecting a variety of mammalian cells. It was found that a number of different adenovirus mRNAs had the same 150- to 200-nucleotide 5′ terminus. One might expect that this leader sequence represents a repeated stretch of nucleotides located near the promoter region of each of the genes for these mRNAs. However, further analysis revealed that the 5′ leader sequence is not complementary to a repeated sequence and, moreover, is not even complementary to a continuous stretch of nucleotides in the template DNA. Instead, the leader is transcribed from three distinct and separate segments of DNA (represented by blocks x, y, and z in the top line of Figure 11.23). The regions of DNA between these blocks, referred to as **intervening sequences** (I_1 to I_3 in Figure 11.23), are somehow missing in the corresponding mRNA. It could have been argued that the presence of intervening sequences is a peculiarity of viral genomes, but the basic observation was soon extended to cellular genes themselves.

The presence of intervening sequences in nonviral, cellular genes was first reported in 1977 by Alec Jeffreys and Richard Flavell in The Netherlands and Pierre Chambon in France. Jeffreys and Flavell discovered an intervening sequence of approximately 600 bases located directly within a part of the globin gene that coded for the amino acid sequence of the globin polypeptide (Figure 11.24). The basis for this finding is discussed in the legend accompanying the figure. Intervening sequences were soon found in other genes, and it became apparent that the presence of genes with intervening sequences—called **split genes**—is the rule not the exception. Those parts of a split gene that contribute to the mature RNA product are called **exons**, whereas the intervening sequences are called **introns**. Split genes are widespread among eukaryotes, although the introns of simpler eukaryotes (e.g., yeast and nematodes) tend to be fewer in number and smaller in size than those of more complex plants and animals. Introns are found in all types of genes, including those that encode tRNAs, rRNAs, as well as mRNAs.

The discovery of genes with introns immediately raised the question as to how such genes were able to produce messenger RNAs lacking these sequences. One likely possibility was that cells produce a primary transcript that corresponds to the entire transcription unit, and that those portions
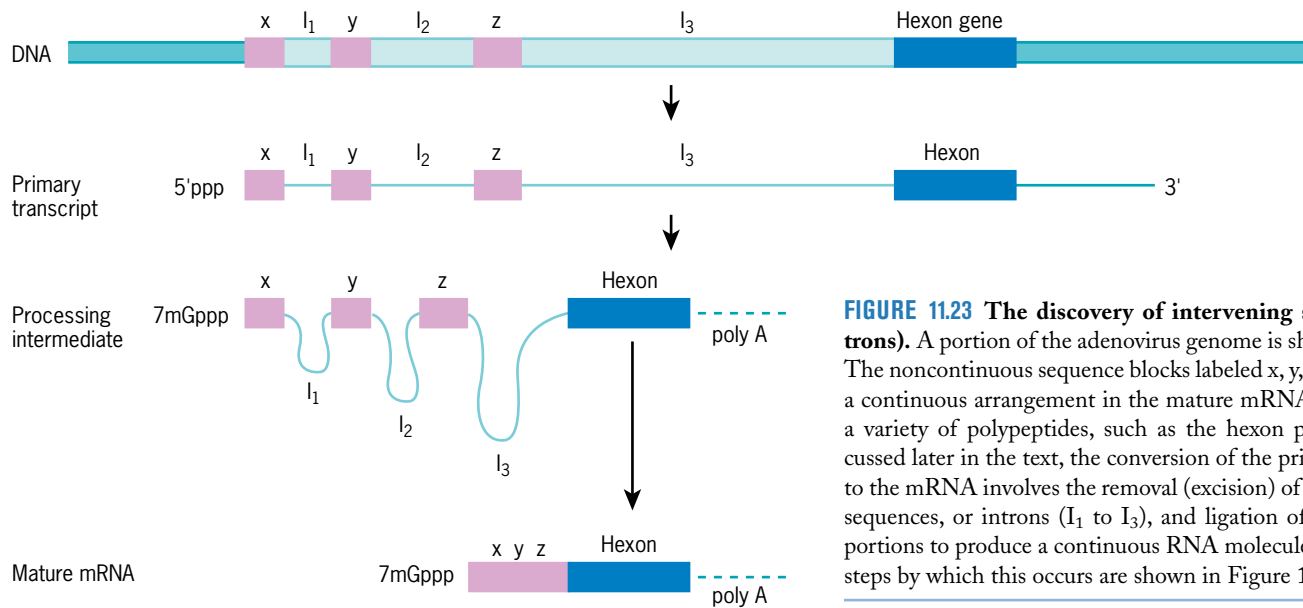
FIGURE 11.23 **The discovery of intervening sequences (introns).** A portion of the adenovirus genome is shown at the top. The noncontinuous sequence blocks labeled x, y, and z appear in a continuous arrangement in the mature mRNAs that code for a variety of polypeptides, such as the hexon protein. As discussed later in the text, the conversion of the primary transcript to the mRNA involves the removal (excision) of the intervening sequences, or introns (I_1 to I_3), and ligation of the remaining portions to produce a continuous RNA molecule (bottom). The steps by which this occurs are shown in Figure 11.32.

of the RNA corresponding to the introns in the DNA are somehow removed. If this were the case, then the segments corresponding to the introns should be present in the primary transcript. Such an explanation would also provide a reason why hnRNA molecules are so much larger than the mRNA molecules they ultimately produce.

Research on nuclear RNA had proceeded by this time to the point where the size of a few mRNA precursors (**pre-mRNAs**) had been determined. The globin sequence, for example, was found to be present within a nuclear RNA molecule that sediments at 15S, unlike the final globin mRNA, which has a sedimentation coefficient of 10S. An ingenious technique (known as R-loop formation) was employed by Shirley Tilghman, Philip Leder, and their co-workers at the

National Institutes of Health to determine the physical relationship between the 15S and 10S globin RNAs and provide information on the transcription of split genes.

Recall from page 393, that single-stranded, complementary DNA strands can bind specifically to one another. Single-stranded DNA and RNA molecules can also bind to one another as long as their nucleotide sequences are complementary; this is the basis of the technique of DNA–RNA hybridization discussed in Section 18.11 (the DNA–RNA complex is called a *hybrid*). Tilghman and her co-workers used the electron microscope to examine a fragment of DNA containing the globin gene that had been hybridized to the 15S globin RNA. The hybrid was seen to consist of a continuous, double-stranded, DNA–RNA complex (the red dotted
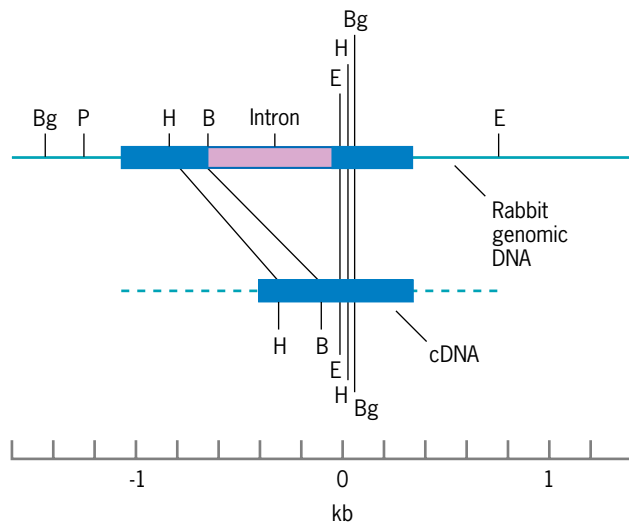
FIGURE 11.24 **The discovery of introns in a eukaryotic gene.** As discussed in Chapter 18, bacteria contain restriction enzymes that recognize and cleave DNA molecules at the site of certain nucleotide sequences. The drawing shows a map of restriction enzyme cleavage sites in the region of the rabbit β-globin gene (upper) and the corresponding map of a cDNA prepared from the β-globin mRNA (lower). (A cDNA is a DNA made in vitro by reverse transcriptase using the mRNA as a template. Thus, the cDNA has the complementary sequence to the mRNA. cDNAs had to be used for this experiment because restriction enzymes don't cleave RNAs.) The letters indicate the sites at which various restriction enzymes cleave the two DNAs. The upper map shows that the globin gene contains a restriction site for the enzyme *Bam*H1 (B) located approximately 700 base pairs from a restriction site for the enzyme *Eco*R1 (E). When the globin cDNA was treated with these same enzymes (lower map), the corresponding B and E sites were located only 67 nucleotides apart. It is evident that the DNA prepared from the genome has a sizeable region that is absent from the corresponding cDNA (and thus absent from the mRNA from which the cDNA was produced). Complete sequencing of the globin gene later showed it to contain a second smaller intron. (FROM A. J. JEFFREYS AND R. A. FLAVELL, CELL 12:1103, 1977; BY PERMISSION FROM CELL PRESS.)
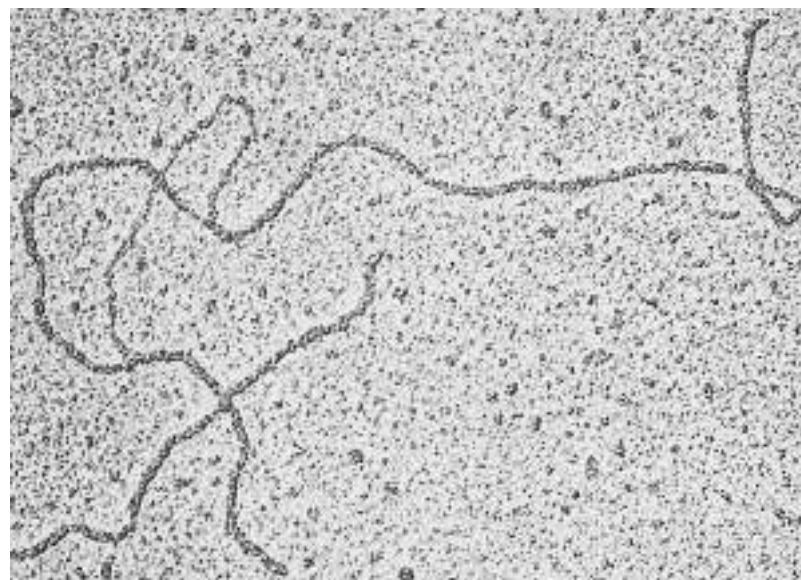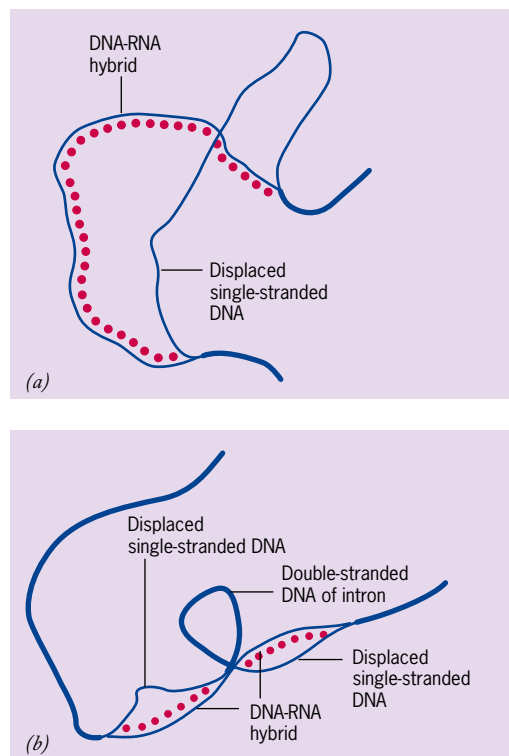
(a)
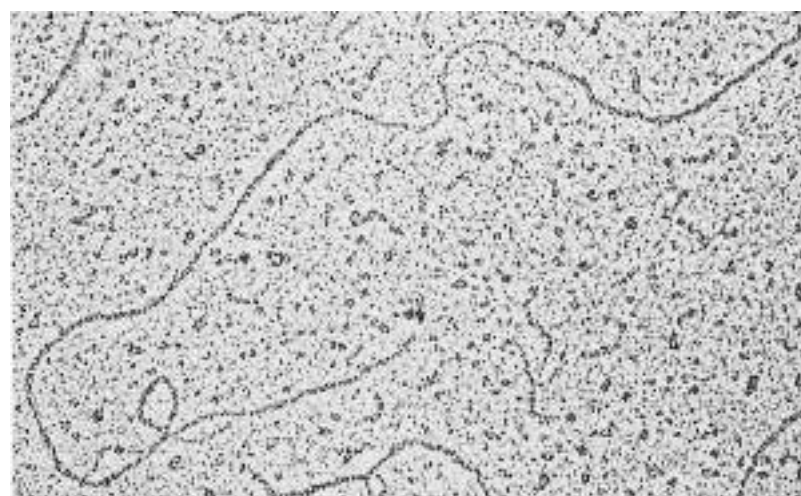
(b)

FIGURE 11.25 Visualizing an intron in the globin gene. Electron micrographs of hybrids formed between (a) 15S globin precursor RNA and the DNA of a globin gene and (b) 10S globin mRNA and the same DNA as in a. The red dotted lines indicate the positions of the RNA molecules. The precursor mRNA is equivalent in length and sequence to the DNA of the globin gene, but the 10S mRNA is missing a portion that is present in the DNA of the gene. These results suggest that the 15S RNA is processed by removing an internal RNA sequence and rejoining the flanking regions. (FROM SHIRLEY M. TILGHMAN ET AL., PROC. NAT'L. ACAD. SCI. U.S.A. 75:1312, 1978.)

(a)

(b)

and blue lines of Figure 11.25a). In contrast, when the same DNA fragment was incubated with the mature 10S globin mRNA, a large segment of the DNA in the center of the coding region was seen to bulge out to form a double-stranded loop (Figure 11.25b). The loop resulted from a large intron in the DNA that was not complementary to any part of the smaller globin message. It was apparent that the 15S RNA does indeed contain segments corresponding to the introns of the genes that are removed during formation of the 10S mRNA.

At about the same time, a similar type of hybridization experiment was performed between the DNA that codes for ovalbumin, a protein found in hen's eggs, and its corresponding mRNA. The ovalbumin DNA and mRNA hybrid contains seven distinct loops corresponding to seven introns (Figure 11.26). Taken together, the introns account for about three times as much DNA as that present in the eight combined coding portions (exons). Subsequent studies have revealed that individual exons average about 150 nucleotides. In contrast, individual introns average about 3500 nucleotides,

which is why hnRNA molecules are much longer than mRNAs. To cite two extreme cases, the human dystrophin gene extends for roughly 100 times the length needed to code for its corresponding message, and the type I collagen gene contains over 50 introns. The average human gene contains about 9 introns making up 90 percent of the transcription unit.

These and other findings provided strong evidence for the proposition that mRNA formation in eukaryotic cells occurs by the removal of internal sequences of ribonucleotides from a much larger pre-mRNA. Let us turn to the steps by which this occurs.

The Processing of Eukaryotic Messenger RNAs

RNA polymerase II assembles a primary transcript that is complementary to the DNA of the entire transcription unit. Electron microscopic examination of transcriptionally active genes indicates that RNA transcripts become associated with
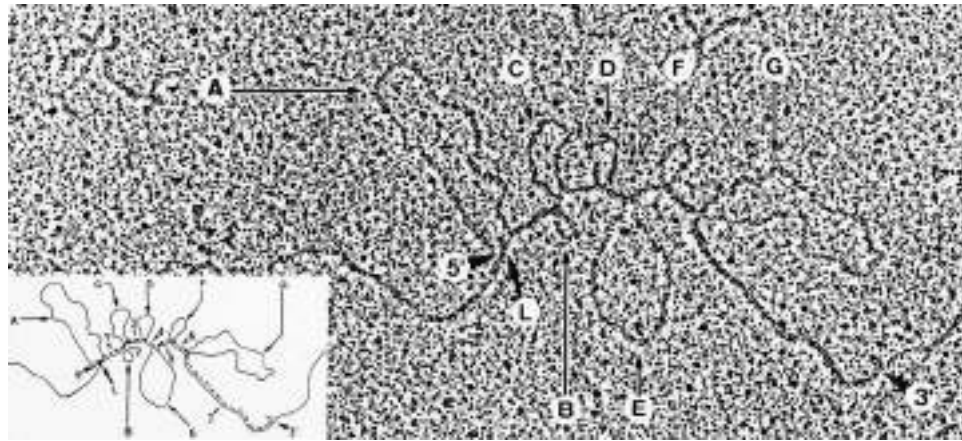
FIGURE 11.26 Visualizing introns in the ovalbumin gene. Electron micrograph of a hybrid formed between ovalbumin mRNA and a fragment of genomic chicken DNA containing the ovalbumin gene. The hybrid shown in this micrograph is similar in nature to that of Figure 11.25*b*. In both cases, the DNA contains the entire gene sequence because it was isolated directly from the genome. In contrast, the RNA has been completely processed, and those portions that were transcribed from the introns have been removed. When the genomic DNA and the mRNA are hybridized, those portions of the DNA that are not represented in the mRNA are thrown into loops. The loops of the seven introns (A–G) can be distinguished. (COURTESY OF PIERRE CHAMBON.)

proteins and larger particles while they are still in the process of being synthesized (Figure 11.27). These particles, which consist of proteins and ribonucleoproteins, include the agents responsible for converting the primary transcript into a mature messenger. This conversion process requires addition of a 5′ cap and 3′ poly(A) tail to the ends of the transcript, and removal of any intervening introns. Once processing is completed, the mRNP, which consists of mRNA and associated proteins, is ready for export from the nucleus.

5′ Caps and 3′ Poly(A) Tails   The 5′ ends of all RNAs initially possess a triphosphate derived from the first nucleoside triphosphate incorporated at the site of initiation of RNA synthesis. Once the 5′ end of an mRNA precursor has been synthesized, several enzyme activities act on this end of the molecule (Figure 11.28). In the first step, the last of the three phosphates is removed, converting the 5′ terminus to a diphosphate (step 1, Figure 11.28). Then, a GMP is added in an *inverted* orientation so that the 5′ end of the guanosine is facing
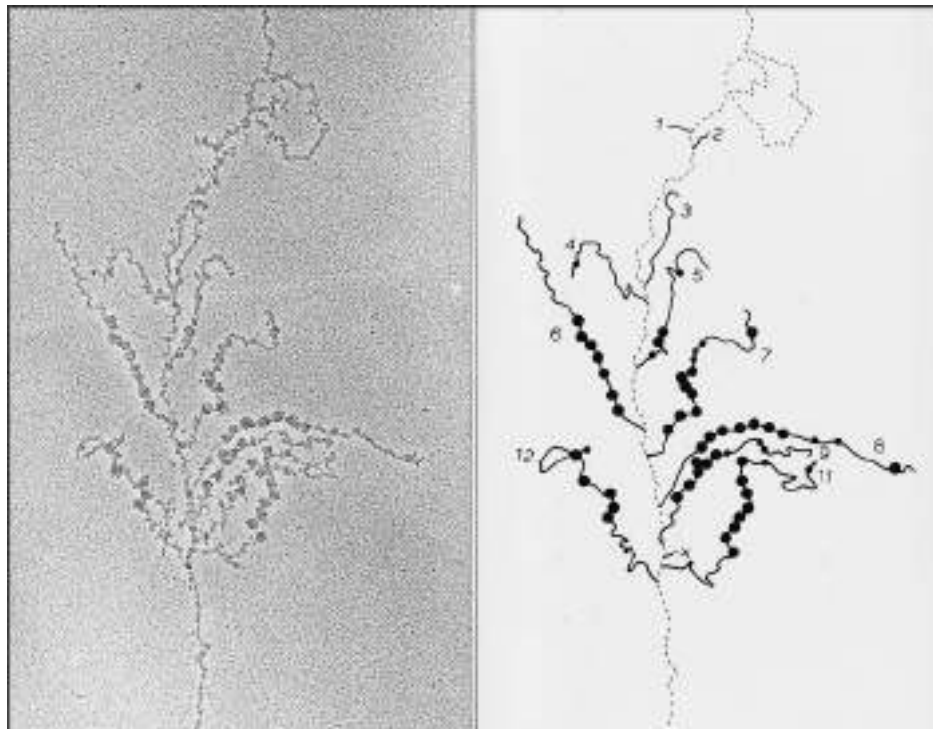
(a)   (b)

FIGURE 11.27 Pre-mRNA transcripts are processed as they are synthesized (i.e., cotranscriptionally). (*a*) Electron micrograph of a nonribosomal transcription unit showing the presence of ribonucleoprotein particles attached to the nascent RNA transcripts. (*b*) Interpretive tracing of the micrograph shown in part *a*. The dotted line represents the chromatin (DNA) strand, the solid lines represent ribonucleoprotein (RNP) fibrils, and solid circles represent RNP particles associated with the fibrils. Individual transcripts are numbered, beginning with 1, which is closest to the point of initiation. The RNP particles are not distributed randomly along the nascent transcript, but rather are bound at specific sites where RNA processing is taking place. (FROM ANN L. BEYER, OSCAR L. MILLER, JR., AND STEVEN L. MCKNIGHT, CELL 20:78, 1980; BY PERMISSION OF CELL PRESS.)

FIGURE 11.28 **Steps in the addition of a 5′ methylguanosine cap and a 3′ poly(A) tail to a pre-mRNA.** The 5′ end of the nascent pre-mRNA binds to a capping enzyme, which in mammals has two active sites that catalyze different reactions: a triphosphatase that removes the terminal phosphate group (step 1) and a guanylyltransferase that adds a guanine residue in a reverse orientation, by means of a 5′-to-5′ linkage (step 2). In step 3, different methyltransferases add a methyl group to the terminal guanosine cap and to the ribose of the nucleotide that had been at the end of the nascent RNA. A protein complex (called CBC) binds to the completed cap (not shown). A very different series of events occurs at the 3′ end of the pre-mRNA, where a large protein complex is assembled. First, an endonuclease cleaves the primary RNA transcript, generating a new 3′ end upstream from the original 3′ terminus. In steps a–c, poly(A) polymerase adds adenosine residues to the 3′ end without the involvement of a DNA template. A typical mammalian mRNA contains 200 to 250 adenosine residues in its completed poly(A) tail; the number is considerably less in lower eukaryotes. (AFTER D. A. MICKLOS AND G. A. FREYER, DNA SCIENCE, CAROLINA BIOLOGICAL SUPPLY CO.)
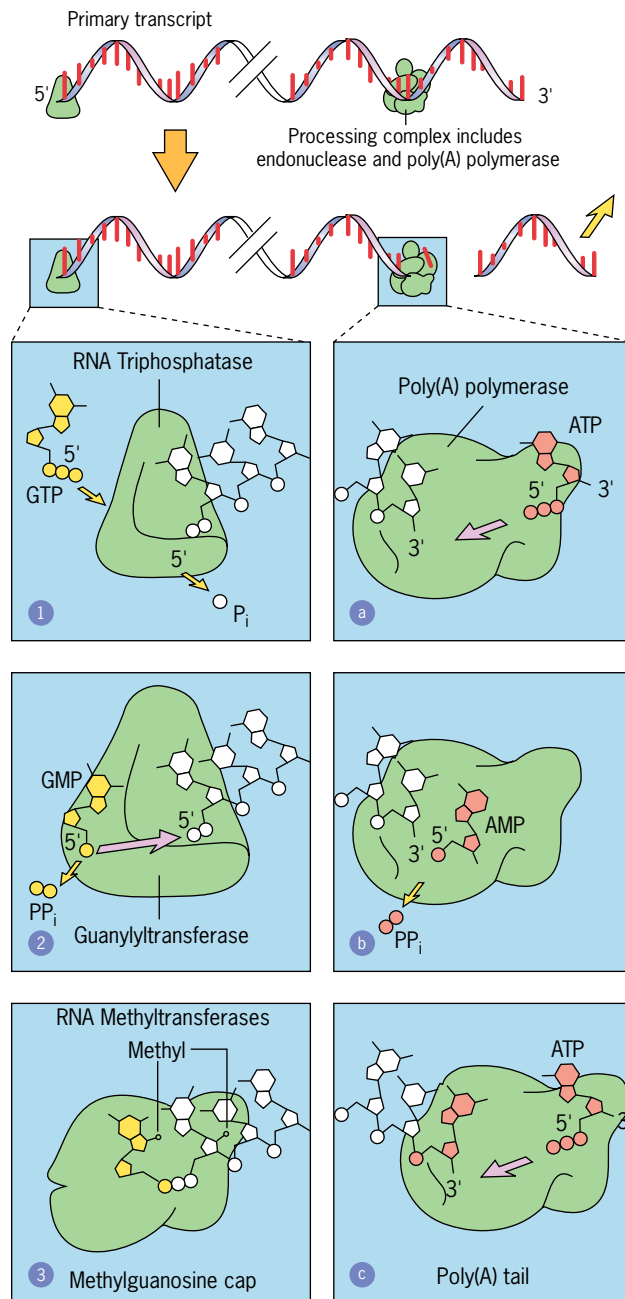
the 5′ end of the RNA chain (step 2, Figure 11.28). As a result, the first two nucleosides are joined by an unusual 5′–5′ triphosphate bridge. Finally, the terminal, inverted guanosine is methylated at the 7 position on its guanine base, while the nucleotide on the internal side of the triphosphate bridge is methylated at the 2′ position of the ribose (step 3, Figure 11.28). The 5′ end of the RNA now contains a **methylguanosine cap** (shown in greater detail in Figure 11.21). These enzymatic modifications at the 5′ end of the primary transcript occur very quickly, while the RNA molecule is still in its very early stages of synthesis. In fact, the capping enzymes are recruited by the CTD of the polymerase (see Figure 11.35). The methylguanosine cap at the 5′ end of an mRNA serves several functions: it prevents the 5′ end of the mRNA from being digested by exonucleases, it aids in transport of the mRNA out of the nucleus, and it plays an important role in the initiation of mRNA translation.

As noted above, the 3′ end of an mRNA contains a string of adenosine residues that forms a **poly(A) tail**. As a number of mRNAs were sequenced, it became evident that the poly(A) tail invariably begins approximately 20 nucleotides downstream from the sequence AAUAAA. This sequence in the primary transcript serves as a recognition site for the assembly of a complex of proteins that carry out the processing reactions at the 3′ end of the mRNA (Figure 11.28). The poly(A) processing complex is also physically associated with RNA polymerase II as it synthesizes the primary transcript (see Figure 11.35).

Included among the proteins of the processing complex is an endonuclease (Figure 11.28, top) that cleaves the pre-mRNA downstream from the recognition site. Following cleavage by the nuclease, an enzyme called *poly(A) polymerase* adds 250 or so adenosines without the need of a template (steps a–c, Figure 11.28). As discussed in Section 12.6, the poly(A) tail together with an associated protein protects the mRNA from premature degradation by exonucleases. Poly(A) tails have also proven useful to molecular biologists interested in isolating mRNA. When a mixture of cellular

RNAs is passed through a column containing bound synthetic poly(T), the mRNAs bind to the column and are removed from solution, while the more abundant tRNAs, rRNAs, and snRNAs that lack the poly(A) tail pass through the column along with the aqueous solvent.

RNA Splicing: Removal of Introns from a Pre-RNA The key steps in the processing of a pre-mRNA are shown in Figure 11.29. In addition to formation of the 5′ cap and poly(A) tail, which have already been discussed, those parts of a primary transcript that correspond to the intervening DNA sequences (the introns) must be removed by a complex process known as **RNA splicing**. To splice an RNA, breaks in the strand must be introduced at the 5′ and 3′ ends (the **splice sites**) of each intron, and the exons situated on either side of
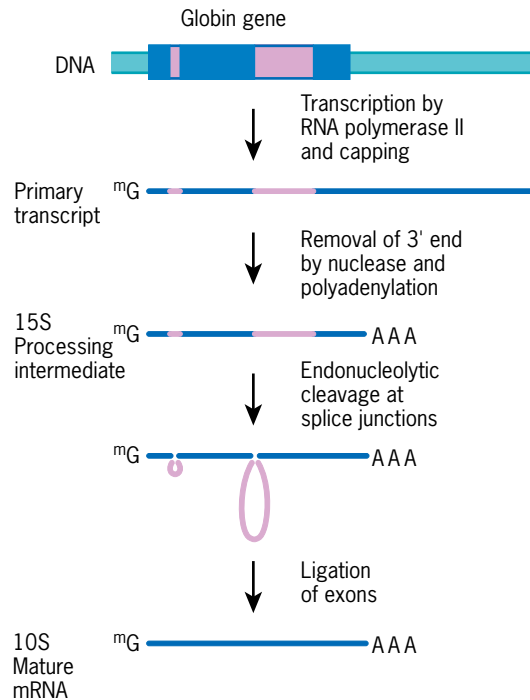
FIGURE 11.29 Overview of the steps during the processing of the globin mRNA. Introns are shown in violet, whereas dark blue portions of the gene indicate the positions of the exons, that is, the DNA sequences that are represented in the mature messenger RNA.

the splice sites must be covalently joined (ligated). It is imperative that the splicing process occur with absolute precision, because the addition or loss of a single nucleotide at any of the splice junctions would cause the resulting mRNA to be mistranslated.

How does the same basic splicing machinery recognize the exon–intron boundaries in thousands of different pre-mRNAs? Examination of hundreds of junctions between exons and introns in eukaryotes ranging from yeast to insects to mammals revealed the presence in splice sites of a conserved nucleotide sequence of ancient evolutionary origin. The sequence most commonly found at the exon–intron borders within mammalian pre-mRNA molecules is shown in Figure 11.30. The G/GU at the 5′ end of the intron (*the 5′ splice site*), the AG/G at the 3′ end of the intron (*the 3′ splice*

site), and the *polypyrimidine tract* near the 3′ splice site are present in the vast majority of eukaryotic pre-mRNAs.[4] In addition, the adjacent regions of the intron contain preferred nucleotides, as indicated in Figure 11.30, which play an important role in splice site recognition. The sequences depicted in Figure 11.30 are necessary for splice-site recognition, but they are not sufficient. Introns typically run for thousands of nucleotides and often contain internal segments that match the consensus sequence shown in Figure 11.30—but the cell doesn't recognize them as splicing signals, and consequently ignores them. The additional clues that allow the splicing machinery to distinguish between exons and introns are provided by specific sequences, most notably the *exonic splicing enhancers*, or *ESEs*, situated within exons (see Figure 11.32, inset A). Changes in DNA sequence within either a splice site or an ESE can lead to the inclusion of an intron or the exclusion of an exon. It is estimated that approximately 15 percent of inherited human disease results directly from mutations that alter pre-mRNA splicing. In addition, much of the "normal" genetic variation in susceptibility to common diseases that is present in the human population (page 410) may result from the effects of this variation on RNA splicing efficiency.

Understanding the mechanism of RNA splicing has come about through an appreciation of the remarkable capabilities of RNA molecules. The first evidence that RNA molecules were capable of catalyzing chemical reactions was obtained in 1982 by Thomas Cech and colleagues at the University of Colorado. As discussed at length in the Experimental Pathways for this chapter, these researchers found that the ciliated protozoan *Tetrahymena* synthesized an rRNA precursor (a pre-rRNA) that was capable of splicing itself. In addition to revealing the existence of RNA enzymes, or **ribozymes**, these experiments changed the thinking among biologists about the relative roles of RNA and protein in the mechanism of RNA splicing.

The intron in the *Tetrahymena* pre-rRNA is an example of a *group I intron*, which won't be discussed. Another type of self-splicing intron, called a *group II intron*, was subsequently discovered in fungal mitochondria, plant chloroplasts, and a

---

[4]Approximately 1 percent of introns have AT and AC dinucleotides at their 5′ and 3′ ends (rather than GU and AG). These AT/AC introns are processed by a different type of spliceosome, called a U12 spliceosome, because it contains a U12 snRNA in place of the U2 snRNA of the major spliceosome.

Polypyrimidine tract

FIGURE 11.30 Nucleotide sequences at the splice sites of pre-mRNAs. In addition to encoding the information to construct a polypeptide, a pre-mRNA must also contain information that directs the machinery responsible for RNA splicing. The nucleotide sequences shown in the regions of the splice sites are based on analysis of a large number of pre-mRNAs and are therefore referred to as *consensus* sequences. The bases shown in orange are virtually invariant; those in black represent the preferred base at that position. N represents any of the four nucleotides; Y represents a pyrimidine. The polypyrimidine tract near the 3′ splice site typically contains between 10 and 20 pyrimidines.

variety of bacteria and archaea. Group II introns fold into a complex structure shown two-dimensionally in Figure 11.31*a*. Group II introns undergo self-splicing by passing through an intermediate stage, called a *lariat* (Figure 11.31*b*) because it resembles the type of rope used by cowboys to catch runaway calves. The first step in group II intron splicing is the cleavage of the 5′ splice site (step 1, Figure 11.31*b*), followed by formation of a lariat by means of a covalent bond between the 5′ end of the intron and an adenosine residue near the 3′ end of the intron (step 2). The subsequent cleavage of the 3′ splice site releases the lariat and allows the cut ends of the exon to be covalently joined (ligated) (step 3).

(a)

(b)
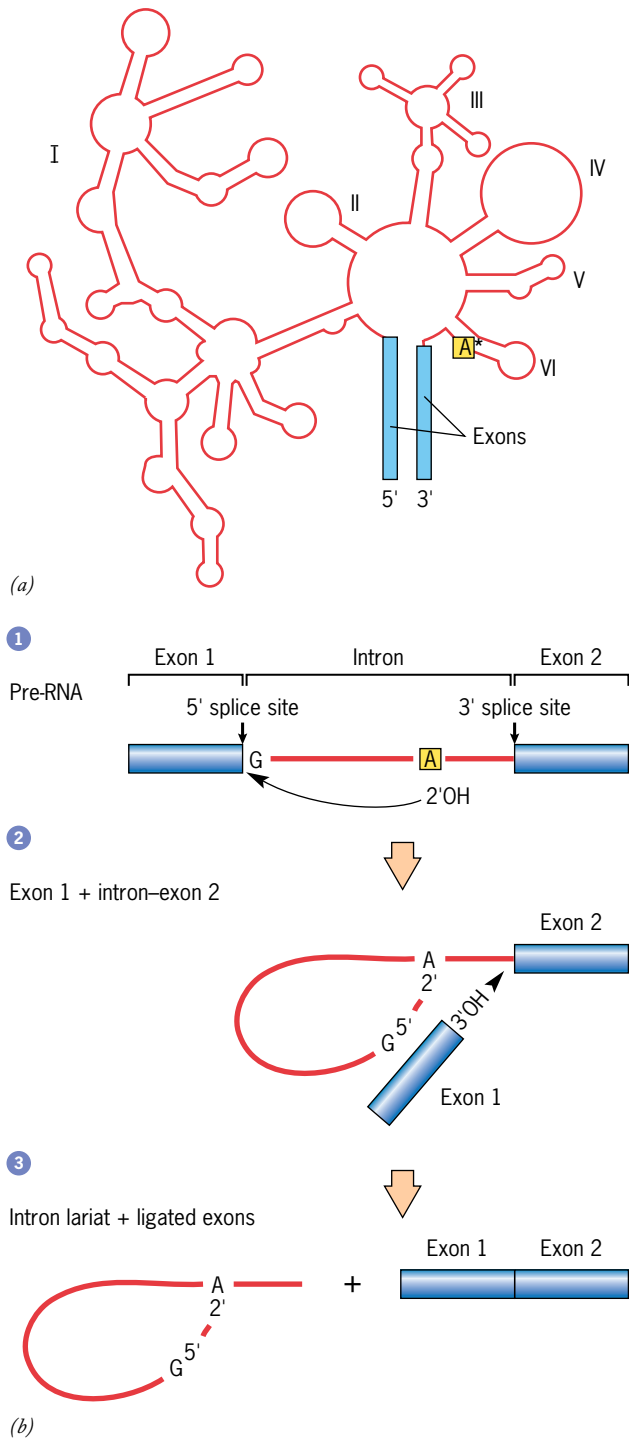
The steps that occur during the removal of introns from pre-mRNA molecules in eukaryotic cells are quite similar to those followed by group II introns. The primary difference is that the pre-mRNA is not able to splice itself, requiring instead a host of **small nuclear RNAs** (**snRNAs**) and their associated proteins. As each large hnRNA molecule is transcribed, it becomes associated with a variety of proteins to form an **hnRNP** (**heterogeneous nuclear ribonucleoprotein**), which represents the substrate for the processing reactions that follow. Processing occurs as each intron of the pre-mRNA becomes associated with a macromolecular complex called a **spliceosome**. Each spliceosome consists of a variety of proteins and a number of distinct ribonucleoprotein particles, called **snRNPs** (pronounced "snurps") because they are composed of snRNAs bound to specific proteins. Spliceosomes are not present within the nucleus in a prefabricated state, but rather are assembled as their component snRNPs bind to the pre-mRNA. Once the spliceosome machinery is assembled, the snRNPs carry out the reactions that cut the introns out of the transcript and paste the ends of the exons together. The excised introns, which constitute roughly 90 percent of the average mammalian pre-mRNA, are simply degraded within the nucleus.

Our understanding of the steps in RNA splicing has been achieved largely through studies of cell-free extracts that can accurately splice pre-mRNAs in vitro. Some of the major steps in the assembly of a spliceosome and removal of an intron are indicated in Figure 11.32 and described in some detail in the accompanying legend. Taken together, removal of an intron requires several snRNP particles: the U1 snRNP, U2 snRNP, U5 snRNP, and the U4/U6 snRNP, which contains the U4 and U6 snRNAs bound together. In addition to its snRNA, each snRNP contains a dozen or more proteins. One family, called *Sm proteins*, are present in all of the snRNPs. Sm proteins bind to one another and to a conserved site on each snRNA (except U6 snRNA) to form the core of the snRNP. Figure 11.33 shows a structural model of the U1 snRNP, with the locations of the snRNA, Sm proteins, and other proteins within the particle indicated. Sm proteins were first identified

FIGURE 11.31 The structure and self-splicing pathway of group II introns. (*a*) Two-dimensional structure of a group II intron (shown in red). The intron folds into six characteristic domains that radiate from a central structure. The asterisk indicates the adenosine nucleotide that bulges out of domain VI and forms the lariat structure as described in the text. The two ends of the intron become closely applied to one another as indicated by the proximity of the two intron–exon boundaries. (*b*) Steps in the self-splicing of group II introns. In step 1, the 2′ OH of an adenosine within the intron (asterisk in domain VI of part *a*) carries out a nucleophilic attack on the 5′ splice site, cleaving the RNA and forming an unusual 2′–5′ phosphodiester bond with the first nucleotide of the intron. This branched structure is described as a lariat. In step 2, the free 3′ OH of the displaced exon attacks the 3′ splice site, which cleaves the RNA at the other end of the intron. As a result of this reaction, the intron is released as a free lariat, and the 3′ and 5′ ends of the two flanking exons are ligated (step 3). A similar pathway is followed in the splicing of introns from pre-mRNAs, but rather than occurring by self-splicing, these steps require the aid of a number of additional factors.

FIGURE 11.32 **Schematic model of the assembly of the splicing machinery and some of the steps that occur during pre-mRNA splicing.** Step 1 shows the portion of the pre-mRNA to be spliced. In step 2, the first of the splicing components, U1 snRNP, has become attached at the 5′ splice site of the intron. The nucleotide sequence of U1 snRNA is complementary to the 5′ splice site of the pre-mRNA, and evidence indicates that U1 snRNP initially binds to the 5′ side of the intron by the formation of specific base pairs between the splice site and U1 snRNA (see inset A). The U2 snRNP is next to enter the splicing complex, binding to the pre-mRNA (as shown in inset A) in a way that causes a specific adenosine residue (dot) to bulge out of the surrounding helix (step 3). This is the site that later becomes the branch point of the lariat. U2 is thought to be recruited by the protein U2AF, which binds to the polypyrimidine tract near the 3′ splice site. U2AF also interacts with SR proteins that bind to the exonic splicing enhancers (ESEs). These interactions play an important role in recognizing intron/exon borders. The next step is the binding of the U4/U6 and U5 snRNPs to the pre-mRNA with the accompanying displacement of U1 (step 4). The assembly of a spliceosome involves a series of dynamic interactions between the pre-mRNA and specific snRNAs and among the snRNAs themselves. As they enter the complex with the pre-mRNA, the U4 and U6 snRNAs are extensively base-paired to one another (inset B). The U4 snRNA is subsequently stripped away from the duplex, and the regions of U6 that were paired with U4 become base-paired to a portion of the U2 snRNA (inset C). Another portion of the U6 snRNA is situated at the 5′ splice site (inset C), having displaced the U1 snRNA that was previously bound there (inset A). It is proposed that U6 is a ribozyme and that U4 is an inhibitor of its catalytic activity. According to this hypothesis, once the U1 and U4 snRNA have been displaced, the U6 snRNA is in position to catalyze the two chemical reactions required for intron removal. According to an alternate view, the reactions are catalyzed by the combined activity of U6 snRNA and a protein of the U5 snRNP. Regardless of the mechanism, the first reaction (indicated by the arrow in inset C) results in the cleavage of the 5′ splice site, forming a free 5′ exon and a lariat intron–3′ exon intermediate (step 5). The free 5′ exon is thought to be held in place by its association with the U5 snRNA of the spliceosome, which also interacts with the 3′ exon (step 5). The first cleavage reaction at the 5′ splice site is followed by a second cleavage reaction at the 3′ splice site (arrow, step 5), which excises the lariat intron and simultaneously joins the ends of the two neighboring exons (step 6). Following splicing, the snRNPs must be released from the pre-mRNA, the original associations between snRNAs must be restored, and the snRNPs must be reassembled at the sites of other introns.
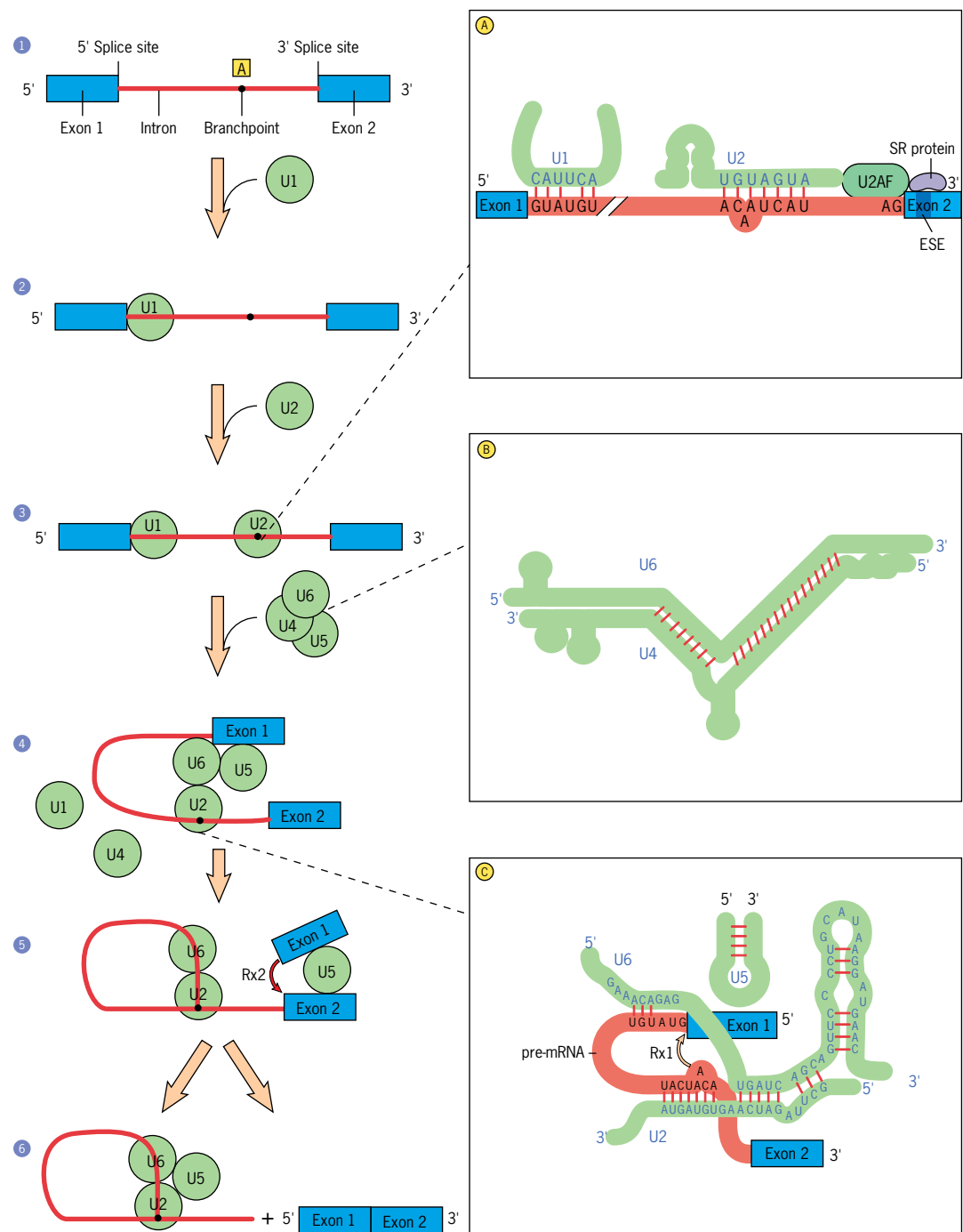
because they are the targets of antibodies produced by patients with the autoimmune disease systemic lupus erythematosus. The other proteins of the snRNPs are unique to each particle.

The events described in Figure 11.32 provide excellent examples of the complex and dynamic interactions that can occur between RNA molecules. The multiple rearrangements among

(a)   (b)
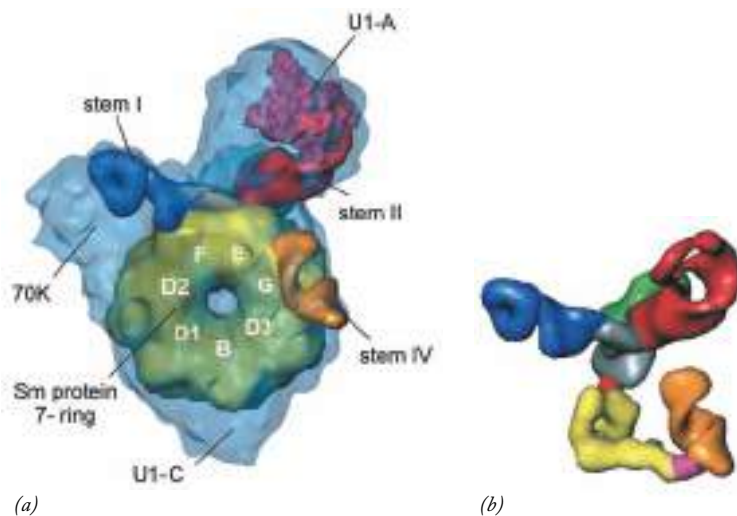
FIGURE 11.33 **The structure of an snRNP.** (*a*) Model of a U1 snRNP particle based on biochemical data and structural information obtained by cryoelectron microscopy. At the core of the particle is a ring-shaped protein complex composed of the seven different Sm proteins that are common to all U snRNPs. Three other proteins are unique to the U1 snRNP (named 70K, U1-A, and U1-C). Stems I, II and IV are parts of the 165-base U1 snRNA. The snRNP is assembled in the cytoplasm and imported into the nucleus, where it carries out its function. (*b*) A model of the U1 snRNA in the same orientation as in *a*. (REPRINTED WITH PERMISSION FROM HOLGER STARK ET AL., NATURE 409:541, 2001; © COPYRIGHT 2001, BY MACMILLAN MAGAZINES LIMITED.)

RNA molecules that occur during the assembly of a spliceosome are primarily mediated by ATP-consuming RNA helicases present within the snRNPs. RNA helicases can unwind double-stranded RNAs, such as the U4–U6 duplex shown in Figure 11.32, inset B, which allows the displaced RNAs to bind new partners. Spliceosomal helicases are also thought to strip RNAs from bound proteins, including the U2AF protein of Figure 11.32, inset A. At least eight different helicases have been implicated in splicing of pre-mRNAs in yeast.

The fact that (1) pre-mRNAs are spliced by the same pair of chemical reactions that occur as group II introns splice themselves, and (2) the snRNAs that are required for splicing pre-mRNAs closely resemble parts of the group II introns (Figure 11.34), suggested that the snRNAs are the catalytically active components of the snRNPs, not the proteins. According to this scenario, the spliceosome would act as a ribozyme and the proteins would serve various supplementary roles, such as maintaining the proper three-dimensional structure of the snRNA and selecting the splice sites to be used during the processing of a particular pre-mRNA. Of the various snRNAs that participate in RNA splicing, U6 is considered the most likely candidate for the catalytic species. However, recent studies have placed at least one of the proteins of the U5 snRNP (namely, a protein called Prp8) very close to the catalytic site of the spliceosome. In addition, Prp8 contains an RNase domain that might be well suited for cleaving the pre-mRNA. This finding has revived the proposal that the combined action of both an RNA and a protein component of the spliceosome is responsible for catalyzing the two chemical reactions required for RNA splicing.

It was mentioned above that sequences situated within exons, called exonic splicing enhancers (ESEs), play a key role in recognition of exons by the splicing machinery. ESEs serve as binding sites for a family of RNA-binding proteins, called *SR proteins*—so named because of their large number of arginine (R)–serine (S) dipeptides. SR proteins are thought to form interacting networks that span the intron/exon borders and help recruit snRNPs to the splice sites (see Figures 11.32, inset A and 12.52). Positively charged SR proteins may also bind electrostatically to the negatively charged phosphate

groups that are added to the CTD of the polymerase as transcription is initiated (Figure 11.20). As a result, the assembly of the splicing machinery at an intron occurs in conjunction with the synthesis of the intron by the polymerase. The CTD

(a)

(b)
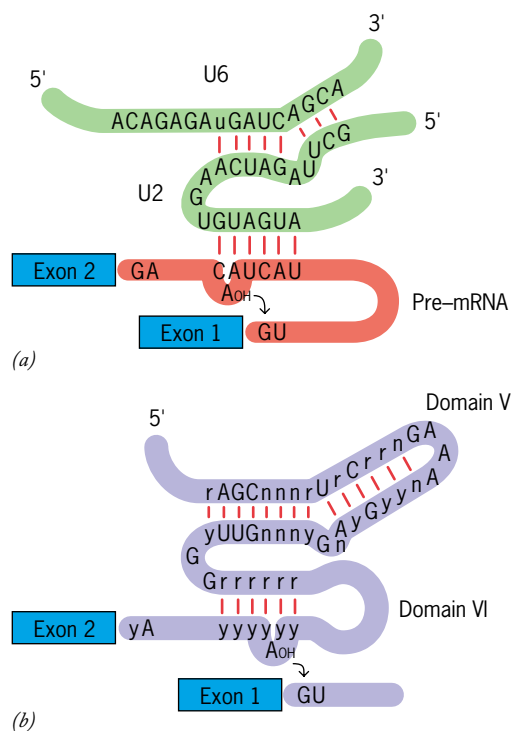
FIGURE 11.34 **Proposed structural similarity between the splicing reactions carried out by the spliceosome on pre-mRNAs and the self-splicing reactions of group II introns.** (*a*) Interactions between an intron of a pre-mRNA molecule (flanked by exons 1 and 2) and two snRNAs (U2 and U6) that are required for splicing. (*b*) The structure of a group II intron showing the arrangement of critical parts that become aligned during self-splicing (as indicated in Figure 11.31). The resemblance of the parts of the group II intron to the combined RNAs in part *a* is evident. Invariant residues are shown in upper case letters; conserved purines and pyrimidines are denoted by r and y, respectively. Variable residues are denoted by n. (AFTER A. M. WEINER, CELL 72:162, 1993; BY PERMISSION OF CELL PRESS.)
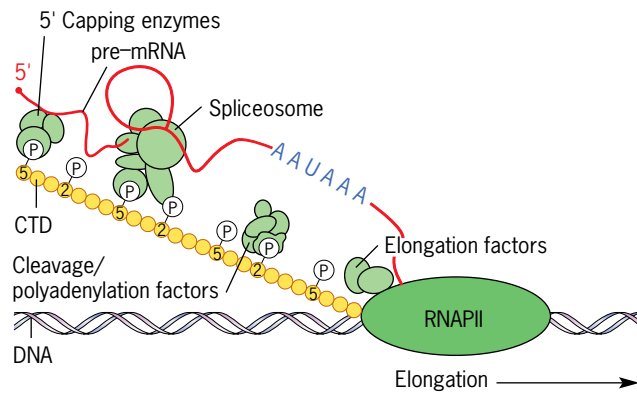
FIGURE 11.35 **Schematic representation of a mechanism for the coordination of transcription, capping, polyadenylation, and splicing.** In this simplified model, the C-terminal domain (CTD) of the large subunit of the RNA polymerase (page 436) serves as a flexible scaffold for the organization of factors involved in processing pre-mRNAs, including those for capping, polyadenylation, and intron removal. In addition to the proteins depicted here, the polymerase is probably associated with a host of transcription factors, as well as enzymes that modify the chromatin template. The proteins bound to the polymerase at any particular time may depend on which of the serine residues of the CTD are phosphorylated. The pattern of phosphorylated serine residues changes as the polymerase proceeds from the beginning to the end of the gene being transcribed (compare to Figure 11.20). The phosphate groups linked to the #5 residues are largely lost by the time the polymerase has transcribed the 3′ end of the RNA. (AFTER E. J. STEINMETZ, CELL 89:493, 1997; BY PERMISSION OF CELL PRESS.)

is thought to recruit a wide variety of processing factors. In fact, most of the machinery required for mRNA processing and export to the cytoplasm travels with the polymerase as part of a giant "mRNA factory" (Figure 11.35).

Because most genes contain a number of introns, the splicing reactions depicted in Figure 11.32 must occur repeatedly on a single primary transcript. Evidence suggests that introns are removed in a preferred order, generating specific processing intermediates whose size lies between that of the primary transcript and the mature mRNA. An example of the intermediates that form during the nuclear processing of the ovomucoid mRNA in cells of the hen's oviduct is shown in Figure 11.36.

Evolutionary Implications of Split Genes and RNA Splicing

The discovery that RNAs are capable of catalyzing chemical reactions has had an enormous impact on our view of biological evolution. Ever since the discovery of DNA as the genetic material, biologists have wondered which came first, protein or DNA. The dilemma arose from the seemingly nonoverlapping functions of these two types of macromolecules. Nucleic acids store information, whereas proteins catalyze reactions. With the discovery of ribozymes in the early 1980s, it became apparent that one type of molecule—RNA—could do both.

These findings have fueled the belief that both DNA and protein were absent at an early stage in the evolution of life.
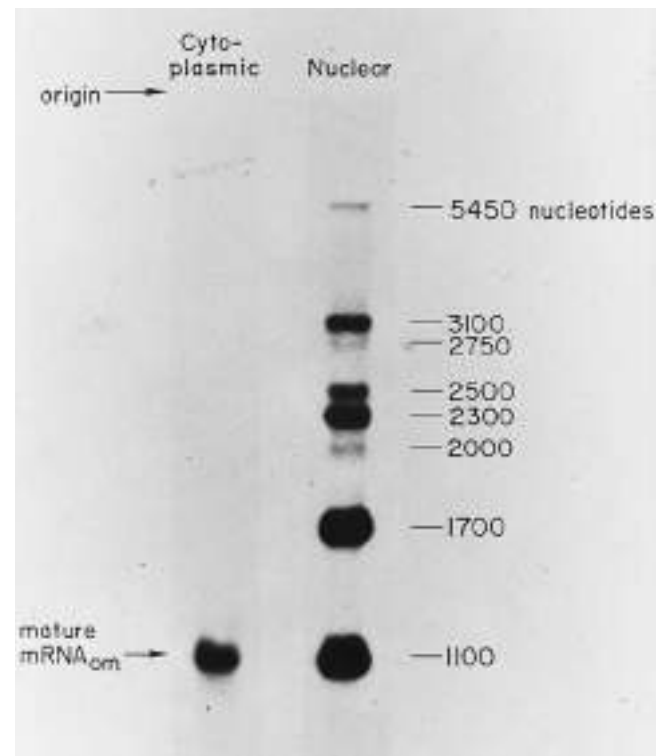
FIGURE 11.36 **Processing the ovomucoid pre-mRNA.** The photograph shows a technique called a Northern blot in which extracted RNA (in this case from the nuclei of hen oviduct cells) is fractionated by gel electrophoresis and blotted onto a membrane filter. The immobilized RNA on the filter is then incubated with a radioactively labeled cDNA (in this case a cDNA made from the ovomucoid mRNA) to produce bands that reveal the position of RNAs containing the complementary sequence. The mature mRNA encoding the ovomucoid protein is 1100 nucleotides long and is shown at the bottom of the blot. It is evident that the nucleus contains a number of larger-sized RNAs that also contain the nucleotide sequence of the ovomucoid mRNA. The largest RNA on the blot has a length of 5450 nucleotides, which corresponds to the size of the ovomucoid transcription unit; this RNA is presumably the primary transcript from which the mRNA is ultimately carved. Other prominent bands contain RNAs with lengths of 3100 nucleotides (which corresponds to a transcript lacking introns 5 and 6), 2300 nucleotides (a transcript lacking introns 4, 5, 6, and 7), and 1700 nucleotides (a transcript lacking all introns except number 3). (COURTESY OF BERT O'MALLEY.)

During this period, RNA molecules performed double duty: they served as genetic material, and they catalyzed chemical reactions including those required for RNA replication. Life at this stage is described as an **"RNA world."** Only at a later stage in evolution were the functions of catalysis and information storage taken over by protein and DNA, respectively, leaving RNA to function primarily as a go-between in the flow of genetic information. Many researchers believe that splicing provides an example of a legacy from an ancient RNA world.

Although the presence of introns creates an added burden for cells, because they have to remove these intervening sequences from their transcripts, introns are not without their virtues. As we will see in the following chapter, RNA splicing

is one of the steps along the path to mRNA formation that is subject to cellular regulation. Many primary transcripts can be processed by two or more pathways so that a sequence that acts as an intron in one pathway becomes an exon in an alternate pathway. As a result of this process, called *alternative splicing*, the same gene can code for more than one polypeptide.

The presence of introns is also thought to have had a major impact on biological evolution. When the amino acid sequence of a protein is examined, it is often found to contain sections that are homologous to parts of several other proteins (see Figures 2.36 and 7.22 for examples). Proteins of this type are encoded by genes that are almost certainly composites made up of parts of other genes. The movement of genetic "modules" among unrelated genes—a process called **exon shuffling**—is greatly facilitated by the presence of introns, which act like inert spacer elements between exons. Genetic rearrangements require breaks in DNA molecules, which can occur within introns without introducing mutations that might impair the organism. Over time, exons can be shuffled independently in various ways, allowing a nearly infinite number of combinations in search for new and useful coding sequences. As a result of exon shuffling, evolution need not occur only by the slow accumulation of point mutations but might also move ahead by "quantum leaps" with new proteins appearing in a single generation.

Creating New Ribozymes in the Laboratory

The main sticking point in the minds of many biologists concerning the feasibility of an RNA world in which RNA acted as the sole catalyst is that, to date, only a few reactions have been found to be catalyzed by *naturally occurring* RNAs. These include the cleavage and ligation of phosphodiester bonds required for RNA splicing and the formation of peptide bonds during protein synthesis. Are these the only types of reactions that RNA molecules are capable of catalyzing, or has their catalytic repertoire been sharply restricted by the evolution of more efficient protein enzymes? Several groups of researchers have explored the catalytic *potential* of RNA by creating new RNA molecules in the laboratory. Although these experiments can never prove that such RNA molecules existed in ancient organisms, they provide proof of the *principle* that such RNA molecules could have evolved through a process of natural selection.

In one approach, researchers have created catalytic RNAs from scratch without any preconceived design as to how the RNA should be constructed. The RNAs are produced by allowing automated DNA-synthesizing machines to assemble DNAs with *random* nucleotide sequences. Transcription of these DNAs produces a population of RNAs whose nucleotide sequences are also randomly determined. Once a population of RNAs is obtained, individual members can be selected from the population by virtue of particular properties they possess. This approach has been described as "test-tube evolution."

In one group of studies, researchers initially selected for RNAs that bound to specific amino acids and, subsequently, for a subpopulation that would transfer a specific amino acid

onto the 3′ end of a targeted tRNA. This is the same basic reaction carried out by aminoacyl-tRNA synthetases, which are enzymes that link amino acids to tRNAs as required for protein synthesis (page 460). It is speculated that amino acids may have been used initially as adjuncts (cofactors) to enhance catalytic reactions carried out by ribozymes. Over time, ribozymes presumably evolved that were able to string specific amino acids together to form small proteins, which were more versatile catalysts than their RNA predecessors. As we will see later in this chapter, ribosomes—the ribonucleoprotein machines responsible for protein synthesis—are essentially ribozymes at heart, which provides strong support for this evolutionary scenario.

As proteins took over a greater share of the workload in the primitive cell, the RNA world was gradually transformed into an "RNA–protein world." At a later point in time, RNA was presumably replaced by DNA as the genetic material, which propelled life forms into the present "DNA–RNA–protein world." The evolution of DNA may have required only two types of enzymes: a ribonucleotide reductase to convert ribonucleotides into deoxyribonucleotides and a reverse transcriptase to transcribe RNA into DNA. The fact that RNA catalysts do not appear to be involved in either DNA synthesis or transcription supports the idea that DNA was the last member of the DNA–RNA–protein triad to appear on the scene.

Somewhere along the line of evolutionary progress, a code had to evolve that would allow the genetic material to specify the sequence of amino acids to be incorporated into a given protein. The nature of this code is the subject of a later section of this chapter.

REVIEW ?

1. What is a split gene? How was the existence of split genes discovered?

2. What is the relationship between hnRNAs and mRNAs? How was this relationship uncovered?

3. What are the general steps in the processing of a pre-mRNA into an mRNA? What is the role of the snRNAs and the spliceosome?

4. What is meant by the term *RNA world*? What type of evidence argues for its existence?

5. What is meant by the phrase "test-tube evolution" as it applies to the catalytic activity of RNA molecules?

11.5 SMALL REGULATORY RNAs AND RNA SILENCING PATHWAYS

The idea that RNA molecules are directly involved in the regulation of gene expression began with a puzzling observation. The petals of a petunia plant are normally light purple. In 1990, two groups of investigators reported on an attempt to deepen the color of the flowers by introducing extra copies of a gene encoding a pigment-producing enzyme. To the sur-
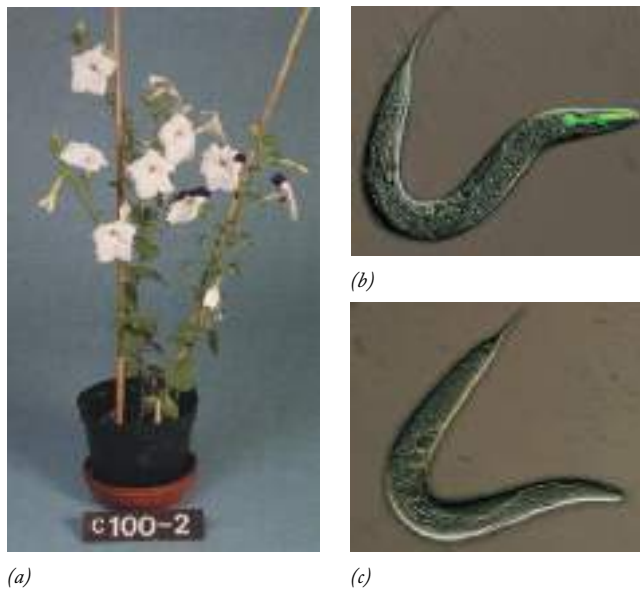
FIGURE 11.37 RNA interference. (*a*) Petunia plants normally have purple flowers. The flowers of this plant appear white because the cells contain an extra gene (a transgene) that encodes an enzyme required for pigment production. The added gene has triggered RNA interference leading to the specific destruction of mRNAs transcribed from both the transgene and the plant's own genes, causing the flowers to be largely unpigmented. (*b*) A nematode worm containing a gene encoding a GFP fusion protein (page 267) that is expressed specifically in the animal's pharynx. (*c*) This worm developed from a parent containing the same genotype as that shown in *b*, whose gonad had been injected with a solution of dsRNA that is complementary to the mRNA encoding the GFP fusion protein. The absence of visible staining reflects the destruction of the mRNA by RNA interference. (A: FROM DAVID BAULCOMBE, CURR. BIOL. 12:R82, 2002; BY PERMISSION OF CELL PRESS; B, C: COPYRIGHT © 2006, NEW ENGLAND BIOLABS. REPRINTED WITH PERMISSION.)

prise of the researchers, the presence of the extra genes caused the petals to lose their pigmentation rather than become more darkly pigmented as expected (Figure 11.37*a*). Subsequent studies indicated that, under these experimental conditions, both the added genes and their normal counterparts within the genome were being transcribed, but the resulting mRNAs were somehow degraded. The phenomenon became known as *posttranscriptional gene silencing* (*PTGS*).

It wasn't until 1998 that the molecular basis for this form of gene silencing became understood. In that year, Andrew Fire of Carnegie Institute of Washington and Craig Mello of the University of Massachusetts and their colleagues conducted an experiment on the nematode *C. elegans*. They injected these worms with several different preparations of RNA hoping to stop production of a particular muscle protein. One of the preparations contained "sense" RNA, that is, an RNA having the sequence of the mRNA that encoded the protein being targeted; another preparation contained "antisense" RNA, that is, an RNA having the complementary sequence of the mRNA in question; and a third preparation consisted of a double-stranded RNA containing both the sense and antisense sequences bound to one another. Neither of the single-stranded RNAs had much of an effect, but the

double-stranded RNA was very effective in stopping production of the encoded protein. Fire and Mello described the phenomenon as **RNA interference** (**RNAi**). They demonstrated that double-stranded RNAs (dsRNAs) were taken up by cells where they induced a response leading to the selective destruction of mRNAs having the same sequence as the added dsRNA. Let's say, for example, one sought to stop the production of the enzyme glycogen phosphorylase within the cells of a nematode worm so that the effect of this enzyme deficiency on the phenotype of the worm could be determined. Remarkably, this result can be attained by simply placing the worm in a solution of dsRNA that shares the same sequence as the target mRNA. A similar experiment is shown in Figure 11.37*b,c*. This phenomenon is similar in effect, though entirely different in mechanism, to the formation of knockout mice that are lacking a particular gene that encodes a particular protein.

The phenomenon of dsRNA-mediated RNA interference is an example of the broader phenomenon of **RNA silencing**, in which small RNAs, typically working in conjunction with protein machinery, act to inhibit gene expression in various ways. RNAi is thought to have evolved as a type of "genetic immune system" to protect organisms from the presence of foreign or unwanted genetic material. To be more specific, RNAi probably evolved as a mechanism to block the replication of viruses and/or to suppress the movements of transposons within the genome because both of these potentially dangerous processes can involve the formation of double-stranded RNAs. Cells can recognize dsRNAs as "undesirable" because such molecules are not produced by the cell's normal genetic activities.

The steps involved in the RNAi pathway are shown in Figure 11.38*a*. The double-stranded RNA that initiates the response is first cleaved into small (21–23 nucleotide), double-stranded fragments, called **small interfering RNAs** (**siRNAs**), by a particular type of ribonuclease, called Dicer. Enzymes of the class to which Dicer belongs act specifically on dsRNA substrates and generate small dsRNAs that have 3′ overhangs, as shown in Figure 11.38*a*. These small dsRNAs are then loaded into a complex (identified as a *pre-RISC* in Figure 11.38*a*) that contains a member of the Argonaute family of proteins. Argonaute proteins play a key role in all of the known RNA silencing pathways. In the RNAi pathway, one of the strands of the RNA duplex (called the passenger strand) is cleaved in two and then dissociates from the pre-RISC. The other strand of the RNA duplex (called the guide strand) is incorporated, together with its Argonaute partner, into a related protein complex named *RISC*. Several different varieties of RISCs have been identified, distinguished by the particular Argonaute protein they contain. In animal cells, RISCs that contain an siRNA typically include the Argonaute protein Ago2. The RISC provides the machinery for the tiny single-stranded siRNA to bind to an RNA having a complementary sequence. Once bound, the target RNA is cleaved at a specific site by the ribonuclease activity of the Ago2 protein. Thus, the siRNA acts like a guide to direct the complex toward a complementary target RNA, which is then destroyed by a protein partner. Each siRNA can orchestrate the destruction of numerous copies of the target RNA, which might be a viral

(a)   *(b)*
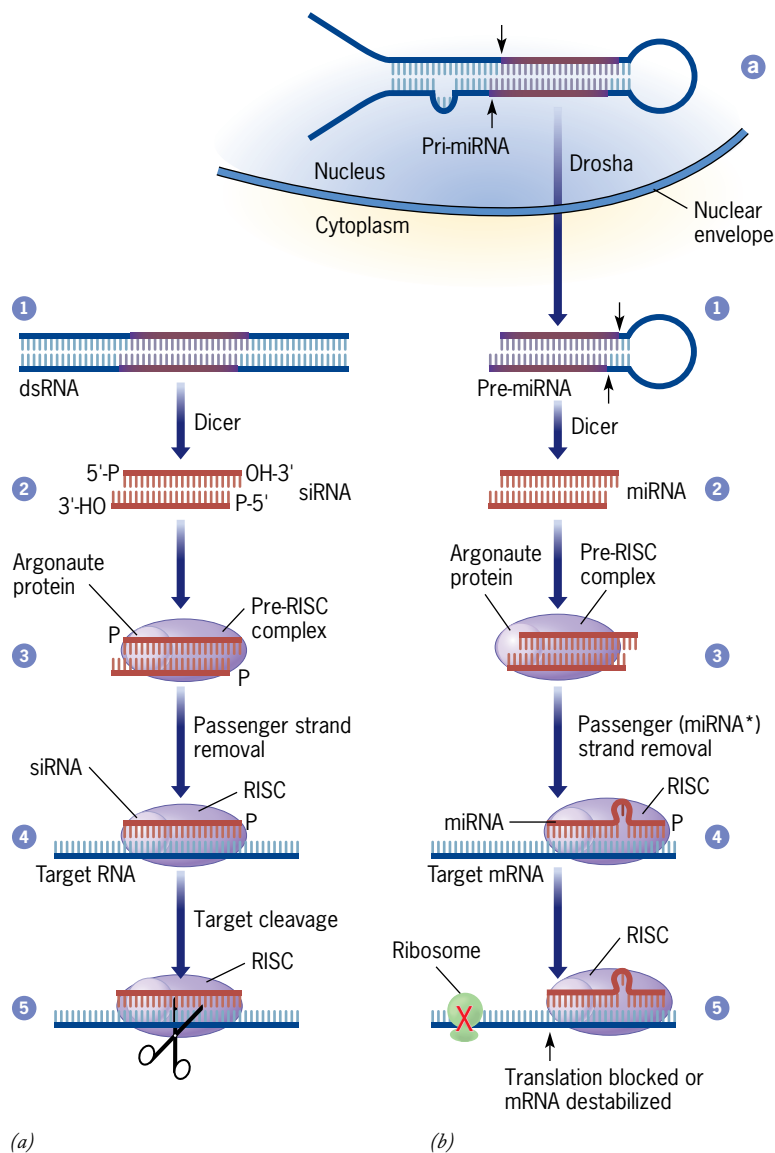
FIGURE 11.38 The formation and mechanism of action of siRNAs and miRNAs. (*a*) In step 1, both strands of a double-stranded RNA are cleaved by the endonuclease Dicer to form a small (21–23 nucleotide) siRNA, which has overhanging ends (step 2). In step 3, the siRNA becomes associated with a protein complex, a pre-RISC, that contains an Argonaute protein (typically Ago2) capable of cleaving and removing the passenger strand of the siRNA duplex. In step 4, the single-stranded guide RNA, in association with proteins of the RISC complex, binds to a target RNA that has a complementary sequence. The target RNA might be a viral RNA, a transcript from a transposon, or an mRNA, depending on circumstances. In step 5, the target RNA is cleaved at a specific site by the Argonaute protein and subsequently degraded. (*b*) MicroRNAs are derived from single-stranded precursor RNAs that contain complementary sequences that allow them to fold back on themselves to form a double-stranded RNA with a stem-loop at one end (step a). This pseudo-dsRNA (or pri-miRNA) is cleaved near its terminal loop by a protein complex containing an endonuclease named

Drosha to generate a pre-miRNA that has a 3′ overhang at one end. The pre-miRNA is exported to the cytoplasm where it is cleaved by Dicer into a small duplex miRNA (step 2) that has a 3′ overhang at both ends. The double-stranded RNA becomes associated with a protein complex containing an Argonaute protein (typically Ago1), leading to the separation of the strands and removal of the passenger strand (called miRNA*). The single-stranded guide miRNA then binds to a complementary region on an mRNA (step 4) and inhibits translation of the message as shown in step 5 (or alternatively leads to destabilization and degradation of the mRNA as discussed in Section 12.6). Unlike siRNAs, miRNAs that inhibit translation are only partially complementary to the target mRNA, hence the bulge. Most plant miRNAs and a handful of animal miRNAs are precisely complementary to the mRNA, or nearly so; in these cases, the outcome of the interaction tends to be cleavage of the mRNA in the same manner shown in *a*. (It can also be noted that a certain class of miRNAs derived from introns do not form long hairpin pri-miRNAs and do not require Drosha for processing.)

transcript, an RNA transcribed from a transposon or retrotransposon, or a host mRNA that is being targeted by a researcher as described in the beginning of this section.

RNA interference has been studied largely in plants and nematode worms. In fact, it is generally accepted that vertebrates do not utilize RNAi as a defense against viruses, relying

instead on a well-developed immune system.[5] In fact, when a dsRNA is added to mammalian cells growing in culture, or injected directly into the body of a mammal, rather than stopping the translation of a *specific* protein it usually initiates a *global* response that inhibits protein synthesis in general. This global shutdown of protein synthesis (discussed in Section 17.1) is thought to have evolved as a means to protect cells from infection by viruses. To get around this large-scale global response, researchers turned to the use of very small dsRNAs. They found that introduction into mammalian cells of synthetic dsRNAs that were 21 nucleotides long

---

[5]Mammals possess all of the components needed to generate siRNAs. Recent studies have indicated that siRNAs (or endo-siRNAs, as they are called) are produced in mammalian oocytes, in part as a mechanism of defense against the movement of transposable elements in the female germ cell line. Whether or not endo-siRNAs have a broader role in mammalian biology remains to be determined. (See *Nature Revs. Mol. Cell Biol.* 9:673, 2008 and 10:135, 2009 and *Mol. Cell* 31:309, 2008.)

(i.e., equivalent in size to the siRNAs produced as intermediates during RNA interference in other organisms) did not trigger *global* inhibition of protein synthesis. Moreover, such dsRNAs were capable of RNAi, that is, they inhibited synthesis of the specific protein encoded by an mRNA having a matching nucleotide sequence. Proteins encoded by other mRNAs were generally not affected. This technique has become a major experimental strategy to learn more about the function of specific genes. One can simply knock out (or, more accurately, knock down) the activity of the specific gene in question, using dsRNAs to destroy the transcribed mRNAs, and look for any cellular abnormalities that result from a deficiency of the encoded protein (see Figures 8.8 and 18.51). Libraries containing thousands of siRNAs have been constructed for use in experiments aimed at determining the activities of large numbers of genes in a single study. The potential clinical importance of RNAi is discussed in the accompanying Human Perspective.

THE HUMAN PERSPECTIVE

Clinical Applications of RNA Interference

Medical scientists are continually searching for "magic bullets," therapeutic compounds that fight a particular disease in a highly specific manner without toxic side effects. We can consider two major types of diseases—viral infections and cancer—that have become targets for a new type of molecular "magic bullet." Viruses are able to ravage an infected cell because they synthesize messenger RNAs that encode viral proteins that disrupt cellular activities. Most cells that become cancerous contain mutations in certain genes (called oncogenes), causing them to produce mutant mRNAs that are translated into abnormal versions of cellular proteins. Consider what might happen if a patient with one of these diseases could be treated with a drug that destroyed or inhibited the specific mRNAs transcribed from the viral genome or mutant cancer gene while, at the same time, ignoring all the other mRNAs of the cell. Several strategies have been developed in recent years that have this goal in mind; the most recent of these strategies takes advantage of the phenomenon of RNA interference.

As discussed above, mammalian cells can be subjected to RNAi—a process that leads to the degradation of a specific mRNA—by introducing the cells to a double-stranded siRNA in which one of the strands is complementary to the mRNA being targeted. When cells are incubated with a 21- to 23-nucleotide synthetic siRNA, the molecules are taken up by the cells, and incorporated into an mRNA-cleaving ribonucleoprotein complex (as in Figure 11.38a), which attacks the complementary mRNA. Alternatively, RNAi can be induced within mammalian cells that have been genetically engineered to carry a gene with inverted repeats. When the gene is transcribed, the RNA product folds back on itself to form a hairpin-shaped (i.e., double-stranded) siRNA precursor (similar to that in Figure 11.38b) that is processed into an active siRNA. The use of synthetic siRNAs is likely to have a transient effect, which may or may not be beneficial depending on the condition being treated. In contrast, the use of viral vectors has the potential to generate a long-lasting therapeutic effect with a single application. At the same time, however, the use of viruses raises other safety issues.

As with all potential drug technologies, the first stages in development of RNAi therapeutics involved preclinical studies in which siRNAs were tested in cells from diseased patients or in animal models (i.e., animals with induced diseases similar to those that afflict humans). Many of these preclinical studies suggested that RNAi held great promise in the treatment of a wide range of diseases and viral infections. We can consider just a few examples. As noted above, cancer cells typically contain one or more mutated genes that encode abnormal proteins responsible for producing the cancerous phenotype. A type of leukemia, for example, is caused by a gene, called *BCR-ABL*, that is formed by the fusion of two normal genes. An siRNA against the mRNA produced by the *BCR-ABL* fusion gene proved capable of converting the malignant cells to a normal phenotype in culture. Similarly, an siRNA against a gene carrying an expanded CAG tract of the type that causes Huntington's disease (page 396) was shown to block the production of the abnormal protein.

A large number of preclinical studies probing the therapeutic value of siRNAs have focused on viral infection. Administration of siRNAs complementary to influenza, HIV, or hepatitis viral sequences has prevented or eliminated infections by these viruses in laboratory animals. In the case of AIDS, it is hoped that stem cells can be isolated from a patient, transfected with a vector carrying an siRNA that targets an HIV-encoded mRNA, and then reinfused back into the patient's bloodstream. Such cells have the potential to produce the siRNA more or less continuously, causing the cell and its descendants to be resistant to destruction by the virus. There are roadblocks, however, for using RNAi to treat viral infections. The hallmark of RNAi is its extraordinary sequence specificity, which can be both a virtue and a handicap. Ultimately, RNAi may prove to be ineffective against viruses, because these pathogens tend to mutate rapidly. Mutations result in changes in genome sequence, leading to the production of mRNAs that are no longer fully complementary to the therapeutic siRNA.

Even though it has been barely more than a decade since RNAi first appeared as an obscure phenomenon in plants and worms

(page 449), siRNA technology is now the basis of a rapidly growing list of clinical trials. The first test of RNAi therapeutics in human patients came against a form of macular degeneration that is caused by the overgrowth of blood vessels behind the retina of elderly patients, causing progressive loss of vision. The excessive growth of these blood vessels is spurred by the production of a growth factor, VEGF, which induces its effect by binding to the cell-surface receptor VEGFR1. When siRNAs targeting the mRNAs encoding VEGF (or VEGFR1) were injected directly into the eyes of patients with this disease, the beneficial effect was substantial; the deterioration of vision was largely halted and the condition appeared to stabilize.[a] Given the success of these initial trials, this same siRNA (named Bevasiranib) is also being tested in patients experiencing retinal deterioration as a complication of diabetes. Another series of clinical tests have begun that are directed against a respiratory virus, RSV, which is a common cause of infant hospitalization. In these trials, adult subjects inhale an aerosol that contains an siRNA (named ALN-RSV01) directed against mRNAs encoding a key viral protein. This aerosol treatment had proven highly effective in combatting these respiratory infections in laboratory animals. Clinical trials of siRNAs aimed at the treatment of high blood cholesterol, certain cancers, hepatitis, pandemic flu, and other diseases are either planned or underway.

Both of the diseases discussed above, age-related macular degeneration and RSV respiratory infection, can be treated with local application of the RNAi therapeutic. As with other types of gene therapies, a major obstacle in using RNAi to treat most diseases or infections is the difficulty in delivering siRNAs (or the viral vectors that encode the hairpin siRNA precursors) to affected tissues deep within the body. In an attempt to meet this challenge, a wide variety of nanosized particles have been developed that (1) either bind to siRNAs or encapsulate them and (2) target the siRNAs to the appropriate tissue or organ following their injection into the bloodstream. We can briefly consider two representative examples of studies that utilize specific targeting strategies. In one report, liposome-like nanoparticles were constructed to contain an antibody that binds to the β_7 integrin subunit on the surface of a specific subset of white blood cells known to be involved in inflammation of the digestive tract. These nanoparticles were also constructed to contain an siRNA directed against the mRNA encoding cyclin D1, a regulatory molecule involved in the proliferation of these white blood cells.

[a]There has been a report suggesting that the beneficial effects of this agent are not the result of a direct inhibition by the siRNA of VEGF synthesis but rather an indirect effect on blood vessel growth through stimulation of a protein called TLR3, which is a receptor of the innate immune system (see Section 17.1). This issue will have to be resolved before a full interpretation of these results is possible. There is also concern, based on studies in mice, that treatment of patients with high doses of siRNAs (or vectors containing genes for siRNA precursors) could overwhelm the RNAi-producing machinery of the cell. This in turn could lead to a reduction in the synthesis of miRNAs, whose production requires this same machinery, which could have severe negative side effects.

(a) (b) (c)

FIGURE 1 **The effects of siRNA on the gut epithelium of mice with induced inflammatory disease.** (*a*) Histological appearance of a section through the intestine of a normal mouse. (*b*) Appearance of the intestine of a mouse that had developed inflammatory colitis but was not treated with a β_7-targeted siRNA. (*c*) Appearance of the intestine of a mouse that had been treated with a β_7-targeted siRNA directed against an mRNA encoding cyclin D1. The siRNA has targeted the inflammation-promoting white blood cells and led to a dramatic reduction in intestinal tissue damage. (FROM DAN PEER ET AL., COURTESY OF MOTOMU SHIMAOKA, SCIENCE 319:630, 2008; © COPYRIGHT 2008, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

These siRNA-carrying nanoparticles were then injected into the bloodstream of mice in which inflammation had been induced. The results of the experiment are shown in Figure 1; the anti-β_7 antibody has successfully targeted the therapeutic agent to the site of inflammation, and inhibition of cyclin D1 synthesis by the siRNA has led to a dramatic reversal of the disease condition. This study demonstrates the potential for siRNA therapeutics in the treatment of severe inflammatory bowel diseases (IBDs), such as colitis and Crohn's disease, which are presently treated with drugs that can have debilitating side effects. In another report, researchers have successfully targeted a virus (JEV) that infects the brains of mice, causing encephalitis. Of all the organs in the body, the brain is the most difficult to access, because even small-molecular-weight drugs cannot penetrate the blood–brain barrier (page 256). In this study, an siRNA directed against the encephalitis virus was linked to a small fragment of a protein from the envelope of the rabies virus, a pathogen that normally infects the brain. In the intact rabies virus, this envelope protein appears to promote access to the brain by binding to acetylcholine receptors on the surfaces of nerve cells and/or cells that line the blood capillaries. In the present study, the siRNA-protein complex was administered intravenously into mice that had also been infected with the encephalitis virus. The siRNA entered the brain resulting in the survival of 80 percent of the treated mice. None of the untreated animals survived the infection.

The studies described here are only the beginning steps in a long research pathway, but they strongly suggest that RNA interference will one day become a valuable therapeutic strategy against a diverse array of disorders.

MicroRNAs: Small RNAs that Regulate Gene Expression

By 1993, it was known that *C. elegans* embryos lacking the gene *lin-4* were unable to develop into normal late-stage larvae. In that year, Victor Ambros, Gary Ruvkun and their colleagues at Harvard University reported that the *lin-4* gene encodes a small RNA that is complementary to segments in the 3′ untranslated region of a specific mRNA encoding the protein LIN-14. They proposed that, during larval development, the *lin-4* RNA binds to the complementary mRNA, blocking translation of the message, which triggers a transition to the next stage of development. Mutants that are unable to produce the small *lin-4* RNA possess an abnormally high

level of the LIN-14 protein and cannot transition normally to later larval stages. This was the first clear example of RNA silencing of gene expression, but several years passed before the broader importance of these findings was appreciated. In 2000, it was shown that one of these tiny worm RNAs—a 21-nucleotide species called *let-7*—is highly conserved throughout evolution. Humans, for example, encode several RNAs that are either identical or nearly identical to *let-7*. This observation led to an explosion of interest in these RNAs.

It has become evident in recent years that both plants and animals produce a collection of tiny RNAs named **microRNAs** (**miRNAs**) that, because of their small size, had been overlooked for decades. As first discovered in nematodes, specific miRNAs such as *lin-4* and *let-7* are synthesized only at certain times during development, or in certain tissues of a plant or animal, and are thought to play a regulatory role. An example of the selective expression of specific miRNAs during the development of the zebra fish is shown in Figure 11.39. The size of miRNAs, at roughly 21–24 nucleotides in length, places them in the same size range as the siRNAs involved in RNAi. This observation is more than coincidence, since miRNAs are produced by a similar processing machinery to that responsible for the formation of siRNAs. miRNAs and siRNAs might be considered to be "cousins," as both act in posttranscriptional RNA silencing pathways. There are, however, important differences. An siRNA is derived from the double-stranded product of a virus or transposable element (or a dsRNA provided by a researcher), and it targets the same transcripts from which it arose. In contrast, an miRNA is encoded by a conventional segment of the genome and targets a specific mRNA as part of a normal cellular program. In other words, siRNAs serve primarily to maintain the integrity of the genome, whereas miRNAs serve primarily to regulate gene expression.

The pathway for synthesis of a typical miRNA is shown in Figure 11.38*b*. The miRNA is synthesized by RNA polymerase II as a primary transcript with a 5′ cap and poly(A) tail. This primary transcript folds back on itself to form a long, double-stranded, hairpin-shaped RNA called a *pri-miRNA*. The pri-miRNA is cleaved within the nucleus by an enzyme called Drosha into a shorter, double-stranded, hairpin-shaped precursor (or *pre-miRNA*) shown in step 1 of Figure 11.38*b*. The pre-miRNA is exported to the cytoplasm where it gives rise to the small double-stranded miRNA. The miRNA is carved out of the pre-miRNA by Dicer, the same ribonuclease responsible for the formation of siRNAs. As with siRNAs, the double-stranded miRNA becomes associated with an Argonaute protein in whose company the RNA duplex is disassembled, and one of the single strands is incorporated into a RISC complex as shown in Figure 11.38*b*.

A typical miRNA is partially complementary to a region in the 3′ UTR of the mRNA target (as in Figure 11.38*b*) and acts as a translational inhibitor, just as the ones originally discovered in nematodes. In these cases, base pairing involves six or seven nucleotides near the 5′ end of the miRNA (typically nucleotides 2–8 of the miRNA), which is referred to as the "seed" region, as well as a few additional base pairs elsewhere in the miRNA. The mechanism of translational inhibition by miRNAs is discussed in Section 12.6. Some miRNAs, however, particularly those found in plants, direct the cleavage of the bound RNA rather than inhibiting its translation. MicroRNAs that direct mRNA cleavage tend to be fully complementary to their mRNA target.

One of the challenges in the study of miRNAs is the difficulty in identifying genes that encode these tiny transcripts. In fact, most potential miRNA genes are initially identified from computer analysis of genomic DNA sequences, although an increasing number have been verified through direct cloning and sequencing experiments. Recent estimates suggest that humans may encode over a thousand distinct miRNA species. Roughly one-third of human messenger RNAs contain sequences that are complementary to likely miRNAs, which provides a hint of the degree to which these small regulatory RNAs may be involved in the control of gene expression in higher organisms. A single miRNA may bind to dozens or even hundreds of different mRNAs. Conversely, many mRNAs contain sequences that are complementary to several different miRNAs, which suggests that these miRNAs may act in various combinations to "fine-tune" the level of gene expression. This prediction is supported by experiments in which cells are forced to take up and express specific miRNA genes. Under these conditions, large numbers of mRNAs in the genetically engineered cells are negatively affected. When different miRNA genes are introduced into
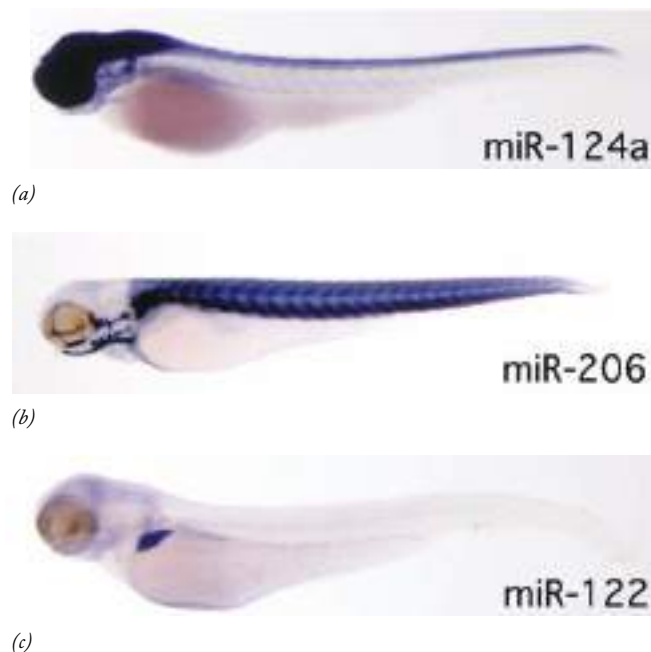
FIGURE 11.39 MicroRNAs are synthesized in specific tissues during embryonic development. These micrographs of zebra fish embryos show the specific expression of three different miRNAs whose localization is indicated by the blue stain. *miR-124a* is expressed specifically in the nervous system (*a*), *miR-206* in skeletal muscle (*b*) and *miR-122* in the liver (*c*). (From Erno Wienholds, Wigard Kloosterman, et al, Science 309:311, 2005, courtesy of Ronald H. A. Plasterk; © copyright 2005, American Association for the Advancement of Science.)

these cells, different groups of mRNAs are downregulated, indicating that the effect is sequence-specific. It can be noted that this type of fine-tuning of mRNA expression levels by miRNAs contrasts sharply with the much more dramatic effects of certain miRNAs, such as *lin-4*, which reduce the expression of a target mRNA to the point that it triggers a major change in the course of development (page 452).

MicroRNAs have been implicated in many processes, including the patterning of the nervous system, control of cell proliferation and cell death, leaf and flower development in plants, and differentiation of various cell types (Section 12.6). The roles of miRNAs in the development of cancer are explored at the end of Section 16.3. Dissecting the roles of individual miRNAs during mammalian development and tissue homeostasis promises to be a focus of research over the next decade.

piRNAs: A Class of Small RNAs that Function in Germ Cells

We have seen in Chapter 10 that the movement of transposable elements poses a threat to the genome because it can disrupt the activity of a gene into which the element happens to insert. If this type of genomic jumping were to occur during the life of an adult liver or kidney cell, the consequences would likely be minimal since the transposition event affects only that particular cell and its daughters (if the somatic cell is still capable of cell division). If, however, a transposition event were to occur in a germ cell, that is, a cell that has the capability to give rise to gametes, then the event has the potential to affect every cell in an individual of the next generation. It is not surprising, therefore, that specialized mechanisms have evolved to suppress the movement of transposable elements in germ cells.

It was noted on page 449 that one of the apparent functions of siRNAs is to prevent the movement of transposable elements. Recent studies indicate that germ cells of animals express a distinct class of small RNAs, called **piwi-interacting RNAs** (or **piRNAs**) that suppress the movement of transposable elements in the germ line. piRNAs get their name from the proteins with which they associate. These proteins are called PIWIs, and they are a subclass of the Argonaute family, the same family of proteins that associate with siRNAs and miRNAs as part of their RISC complexes. The roles of the PIWI proteins have been best studied in fruit flies, where deletion of these proteins leads to defects in the suppression of transposon movement in germ cells and ultimate failure of gamete formation. piRNAs and their associated PIWI proteins are also required for successful gamete formation in mammals, but their roles are poorly understood.

There are a number of important differences between piRNAs on one hand and si/miRNAs on the other: (1) piRNAs are longer than these other small RNAs, measuring about 24–32 nucleotides in length; (2) the majority of mammalian piRNAs can be mapped to a small number of huge genomic loci, some of which can encode thousands of different piRNAs; and (3) the formation of piRNAs does not involve the formation of dsRNA precursors or cleavage by the Dicer ribonuclease. Instead, piRNA biogenesis appears to depend on the endonuclease activity of the PIWI protein acting on a long, single-stranded primary transcript. The mechanism by which piRNAs act to silence target RNA expression remains unclear.

Other Noncoding RNAs

As the field of cell and molecular biology has matured, we have gradually learned to appreciate the remarkable structural and functional diversity of RNA molecules. New types of RNAs have been discovered frequently over the past few decades, as described in the previous pages of this text. But researchers have been unprepared for one flurry of recent findings, which indicate that *at least* two-thirds of a mouse or human genome is normally transcribed, far more than would be expected based on the number of "meaningful" DNA sequences thought to be present. In fact, most of the sequences included among this transcribed DNA are normally thought of as "junk," including large numbers of transposable elements that constitute a sizeable fraction of mammalian genomes (page 404). Some studies raise the question whether there is any basic difference between a gene and an intergenic region; they report that transcription can begin at all kinds of unexpected sites within the genome, and that transcripts overlap extensively with one another. Other studies reveal that cells often transcribe both strands of a DNA element, producing both a *sense RNA* and an *antisense RNA*. It is also reported that antisense RNAs can be synthesized by polymerases that had initially bound to the promoter of a protein-coding gene and then moved upstream from that site along the *opposite* strand of the chromosome, that is, in the opposite direction from polymerase molecules that are transcribing the gene itself.

Just as the DNA of a cell or organism constitutes its genome and the proteins it produces constitutes it proteome, the RNAs that are synthesized by a cell or organism constitutes its **transcriptome**. Why is the transcriptome of a mammalian cell so large? In other words, why does a cell transcribe all of these various types of DNA sequences? We don't know the answer to this basic question. According to one viewpoint, much of this pervasive transcription activity is simply the "background noise" that accompanies the complex process of gene expression. Proponents of this viewpoint cite the results of studies in mice in which blocks of the genome that lack protein-coding genes have been deleted. It was found in these studies that healthy mice can develop even though they cannot synthesize the noncoding RNAs (*ncRNAs*) that would normally have been produced from the deleted DNA sequences. According to a contrasting viewpoint, much of the noncoding RNA being produced is involved in diverse regulatory activities that have yet to be identified. Supporters of this position cite results suggesting that this pervasive transcription is not random; many of the noncoding transcripts display distinct and reproducible patterns of tissue-specific distribution, and their origin can be traced to specific loci in the genome. Keep in mind that it has only been in the last decade or so that the existence of siRNAs, miRNAs, and piRNAs

have come to light and it is likely that other types of small ncRNAs remain undetected. A number of long ncRNAs (e.g., *XIST* and *AIRE*, whose functions are discussed in later chapters) have also been identified, and these are probably the tip of the iceberg. It is even possible that the vast transcriptional output of the mammalian genome holds the key to explaining why we have approximately the same number of genes as organisms we believe to be much less complex (see Figure 10.27). Regardless of the actual explanation, one point is clear: there is a great deal that we do not understand about the many roles of RNAs in eukaryotic cells.

REVIEW

1. What is an siRNA, an miRNA and a piRNA? How is each of them formed in the cell? What are their proposed functions?

11.6 ENCODING GENETIC INFORMATION

Once the structure of DNA had been revealed in 1953, it became evident that the sequence of amino acids in a polypeptide was specified by the sequence of nucleotides in the DNA of a gene. It seemed very unlikely that DNA could serve as a direct, physical template for the assembly of a protein. Instead, it was presumed that the information stored in the nucleotide sequence was present in some type of **genetic code**. With the discovery of messenger RNA as an intermediate in the flow of information from DNA to protein, attention turned to the manner in which a sequence written in a ribonucleotide "alphabet" might code for a sequence in an "alphabet" consisting of amino acids.

The Properties of the Genetic Code

One of the first models of the genetic code was presented by the physicist George Gamow, who proposed that each amino acid in a polypeptide was encoded by three sequential nucleotides. In other words, the code words, or **codons**, for amino acids were nucleotide triplets. Gamow arrived at this conclusion by a bit of armchair logic. He reasoned that it would *require* at least three nucleotides for each amino acid to

have its own unique codon. Consider the number of words that can be spelled using an alphabet containing four different letters corresponding to the four possible bases that can be present at a particular site in DNA (or mRNA). There are 4 possible one-letter words, 16 (4^2) possible two-letter words, and 64 (4^3) possible three-letter words. Because there are 20 different amino acids (words) that have to be specified, codons must contain at least 3 successive nucleotides (letters). The triplet nature of the code was soon verified in a number of insightful genetic experiments conducted by Francis Crick, Sydney Brenner, and colleagues at Cambridge University.[6]

In addition to proposing that the code was triplet, Gamow also suggested that it was *overlapping*. Even though this suggestion proved to be incorrect, it raises an interesting question concerning the genetic code. Consider the following sequence of nucleotides:

—AGCAUCGCAUCGA—

If the code is overlapping, then the ribosome would move along the mRNA one nucleotide at a time, recognizing a new codon with each move. In the preceding sequence, AGC would specify an amino acid, GCA would specify the next amino acid, CAU the next, and so on. In contrast, if the code is *nonoverlapping*, each nucleotide along the mRNA would be part of one, and only one, codon. In the preceding sequence, AGC, AUC, and GCA would specify successive amino acids.

A conclusion as to whether the genetic code was overlapping or nonoverlapping could be inferred from studies of mutant proteins, such as the mutant hemoglobin responsible for sickle cell anemia. In sickle cell anemia, as in most other cases that were studied, the mutant protein was found to contain a single amino acid substitution. If the code is overlapping, a change in one of the base pairs in the DNA would be expected to affect three consecutive codons (Figure 11.40) and, therefore, three consecutive amino acids in the corresponding polypeptide. If, however, the code is nonoverlapping and each nucleotide is part of only one codon, then only one amino acid replacement would be expected. These and other data indicated that the code is nonoverlapping.

---

[6]Anyone looking to read a short research paper that conveys a feeling for the inductive power and elegance of the early groundbreaking studies in molecular genetics can find these experiments on the genetic code in *Nature* 192:1227, 1961. (See *Cell* 128:815, 2007 for discussion of this classic experiment.)

Base Sequence	Codons	
Original sequence ...AGCATCG...	Overlapping code ..., AGC GCA CAT ATC TCG ...	Nonoverlapping code ..., AGC ATC ...
Sequence after single-base substitution ...AGAATCG...	Overlapping code ..., AGA GAA AAT ATC TCG ...	Nonoverlapping code ..., AGA ATC ...

FIGURE 11.40 The distinction between an overlapping and nonoverlapping genetic code. The effect on the information content of an mRNA by a single base substitution depending on whether the code is overlapping or nonoverlapping. The affected codons are underlined in red.

Given a triplet code that can specify 64 different amino acids and the reality that there are only 20 amino acids to be specified, the question arises as to the function of the extra 44 triplets. If any or all of these 44 triplets code for amino acids, then at least some of the amino acids would be specified by more than one codon. A code of this type is said to be *degenerate*. As it turns out, the code is highly degenerate, as nearly all of the 64 possible codons specify amino acids. Those that do not (3 of the 64) have a special "punctuation" function—they are recognized by the ribosome as termination codons and cause reading of the message to stop.

The degeneracy of the code was originally predicted by Francis Crick on theoretical grounds when he considered the great range in the base composition among the DNAs of various bacteria. It had been found, for example, that the G + C content could range from 20 percent to 74 percent of the genome, whereas the amino acid composition of the proteins from these organisms showed little overall variation. This suggested that the same amino acids were being encoded by different base sequences, which would make the code degenerate.

Identifying the Codons   By 1961, the general properties of the code were known, but not one of the coding assignments of the specific triplets had been discovered. At the time, most geneticists thought it would take many years to decipher the entire code. But a breakthrough was made by Marshall Nirenberg and Heinrich Matthaei when they developed a technique to synthesize their own artificial genetic messages and then determine what kind of protein they encoded. The first message they tested was a polyribonucleotide consisting exclusively of uridine; the message was called *poly(U)*. When poly(U) was added to a test tube containing a bacterial extract with all 20 amino acids and the materials necessary for protein synthesis (ribosomes and various soluble factors), the system followed the artificial messenger's instructions and manufactured a polypeptide. The assembled polypeptide was analyzed and found to be polyphenylalanine—a polymer of the amino acid phenylalanine. Nirenberg and Matthaei had thus shown that the codon UUU specifies phenylalanine.

Over the next four years, the pursuit was joined by a number of laboratories, and synthetic mRNAs were constructed to test the amino acid specifications for all 64 possible codons. The result was the universal decoder chart for the genetic code shown in Figure 11.41. The decoder chart lists the nucleotide sequence for each of the 64 possible codons in an mRNA. Instructions for reading the chart are provided in the accompanying figure legend.

The codon assignments given in Figure 11.41 are "essentially" universal, that is, present in all living organisms. The first exceptions to the universality of the genetic code were found to occur in the codons of mitochondrial mRNAs. For example, in human mitochondria, UGA is read as tryptophan rather than stop, AUA is read as methionine rather than isoleucine, and AGA and AGG are read as stop rather than arginine. More recently, exceptions have been found, here and there, in the nuclear DNA codons of protists and fungi. Even

with these discrepancies, the similarities among genetic codes far outweigh their differences, and it is evident that minor deviations have evolved as secondary changes from the standard genetic code shown in Figure 11.41. In other words, all known organisms present on Earth today share a common evolutionary origin.

Examination of the codon chart in Figure 11.41 indicates that the amino acid assignments are distinctly nonrandom. If one looks for the codon boxes for a specific amino acid, they tend to be clustered within a particular portion of the chart. Clustering reflects the similarity in codons that specify the same amino acid. As a result of this similarity in codon sequence, spontaneous mutations causing single base changes in a gene often will not produce a change in the amino acid sequence of the corresponding protein. A change in nucleotide sequence that does not affect amino acid sequence is said to be *synonymous*, whereas a change that causes an amino acid substitution is said to be *nonsynonymous*. Because synonymous changes do not generally alter an organism's phenotype, they aren't selected for or against by natural selection. This is not the case for nonsynonymous changes, which are likely to alter an organism's phenotype and are subject to natural selection. Now that we have sequenced the genomes of related organisms, such as those of chimpanzees and humans, we can look directly at the sequences of homologous genes and see how many changes are synonymous or nonsynonymous. Genes possessing an excess of nonsynonymous substitutions in their coding regions are likely to have been influenced by natural selection (see the footnote on page 406).

The "safeguard" aspect of the code goes beyond its degeneracy. Codon assignments are such that *similar* amino acids tend to be specified by similar codons. For example, codons of the various hydrophobic amino acids (depicted as brown boxes in Figure 11.41) are clustered in the first two columns of the chart. Consequently, a mutation that results in a base substitution in one of these codons is likely to substitute one hydrophobic residue for another. In addition, the greatest similarities between amino acid–related codons occur in the first two nucleotides of the triplet, whereas the greatest variability occurs in the third nucleotide. For example, glycine is encoded by four codons, all of which begin with the nucleotides GG. An explanation for this phenomenon is revealed in the following section in which the role of the transfer RNAs is described.

REVIEW ?

1. Explain what is meant by stating that the genetic code is triplet and nonoverlapping? What did the finding that DNA base compositions varied greatly among different organisms suggest about the genetic code? How was the identity of the UUU codon established?

2. Distinguish between the effects of a base substitution in the DNA on a nonoverlapping and an overlapping code.

3. Distinguish between a synonymous and a nonsynonymous base change.
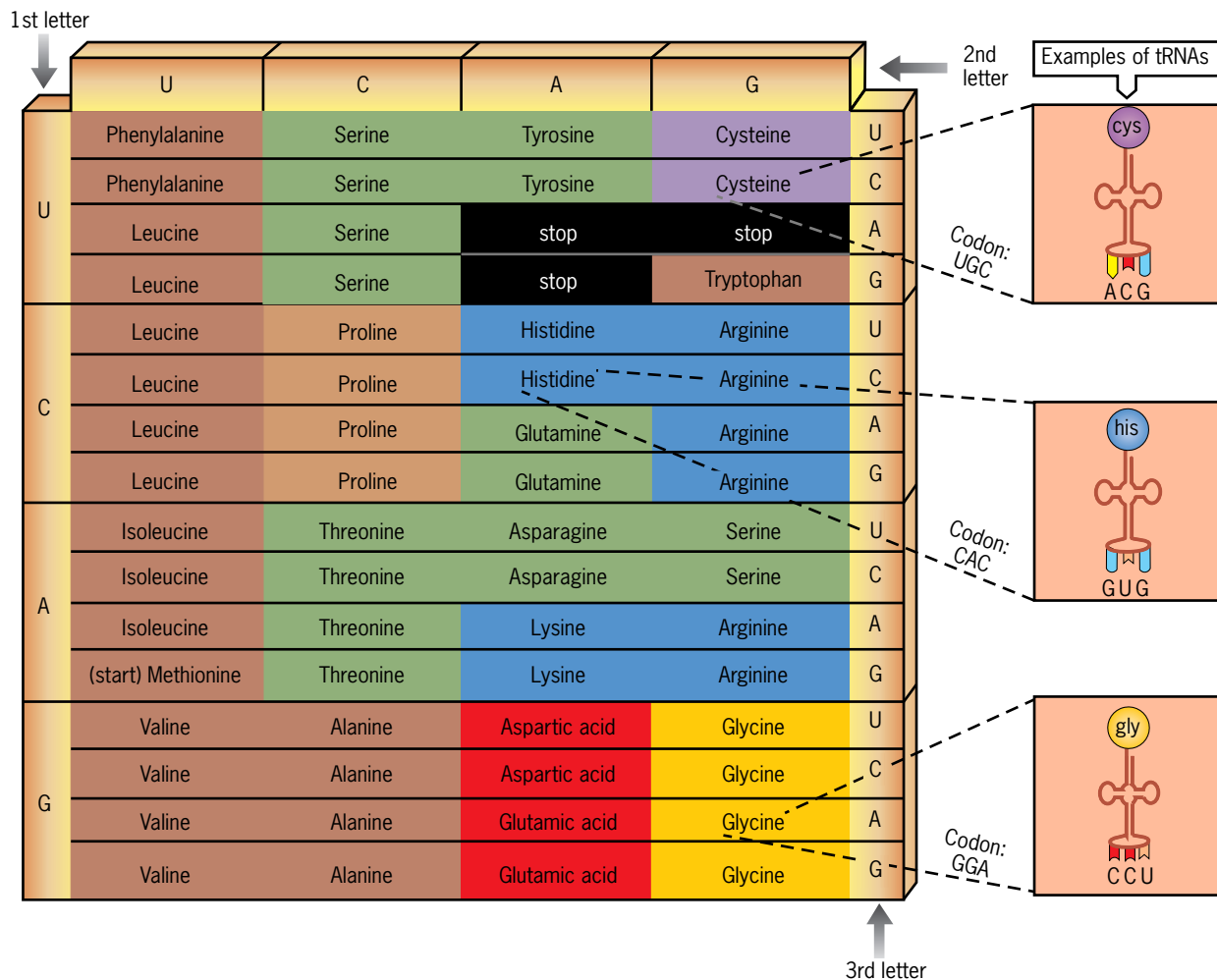
FIGURE 11.41 The genetic code. This universal decoder chart lists each of the 64 possible mRNA codons and the corresponding amino acid specified by that codon. To use the chart to translate the codon UGC, for example, find the first letter (U) in the indicated row on the left. Follow that row to the right until you reach the second letter (G) indicated at the top; then find the amino acid that matches the third letter (C) in the row on the right. UGC specifies the insertion of cysteine. Each amino acid (except two) has two or more codons that order its insertion, which makes the code degenerate. A given amino acid tends to be encoded by related codons. This feature reduces the likelihood that base substitutions will result in changes in the amino acid sequence of a protein. Amino acids with similar properties also tend to be clustered. Amino acids with acidic side chains are shown in red, those with basic side chains in blue, those with polar uncharged side chains in green, and those with hydrophobic side chains in brown. As discussed in the following section, decoding in the cell is carried out by tRNAs, a few of which are illustrated schematically in the right side of the figure.

11.7 DECODING THE CODONS: THE ROLE OF TRANSFER RNAs

Nucleic acids and proteins are like two languages written with different types of letters. This is the reason protein synthesis is referred to as *translation*. Translation requires that information encoded in the nucleotide sequence of an mRNA be decoded and used to direct the sequential assembly of amino acids into a polypeptide chain. Decoding the information in an mRNA is accomplished by transfer RNAs, which act as adaptors. On one hand, each tRNA is linked to a specific amino acid (as an aa-tRNA), while on the other hand, that same tRNA is able to recognize a particular codon in the mRNA. The interaction between successive codons in the mRNA and specific aa-tRNAs leads to the synthesis of a polypeptide with an ordered sequence of amino acids. To understand how this occurs, we must first consider the structure of tRNAs.

The Structure of tRNAs

In 1965, after seven years of work, Robert Holley of Cornell University reported the first base sequence of an RNA molecule, that of a yeast transfer RNA that carries the amino acid alanine (Figure 11.42a). This tRNA is composed of 77 nucleotides, 10 of which are modified from the standard 4 nucleotides of RNA (A, G, C, U), as indicated by the shading in the figure.
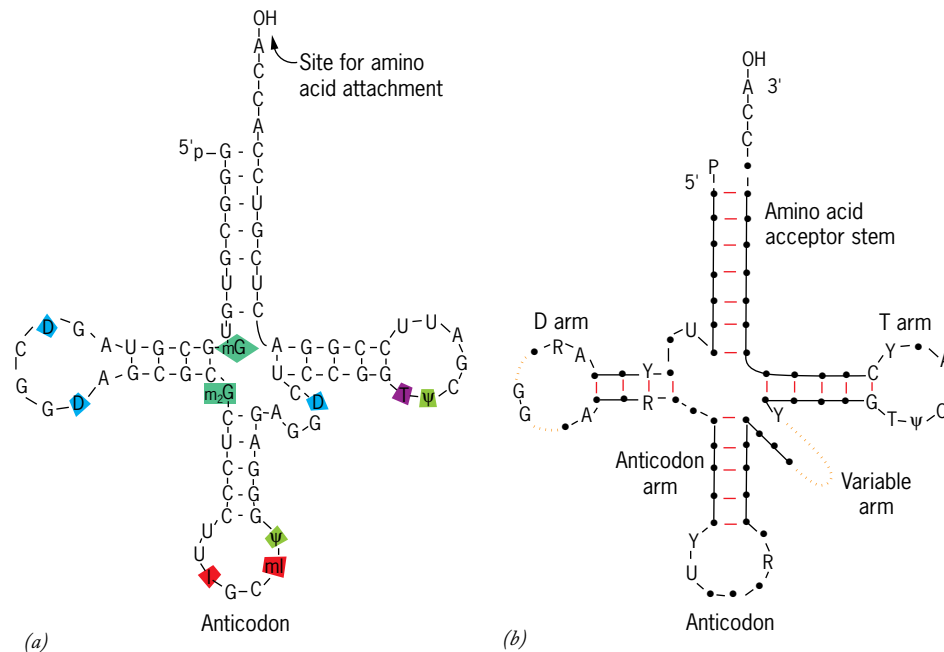
FIGURE 11.42 Two-dimensional structure of transfer RNAs. (*a*) The nucleotide sequence of the cloverleaf form of a yeast tRNA^Ala. The amino acid becomes linked to the 3′ end of the tRNA, whereas the opposite end bears the anticodon, in this case IGC. The function of the anticodon is discussed later. In addition to the four bases, A, U, G, and C, this tRNA contains ψ, pseudouridine; T, ribothymidine; mI, methyl-inosine; I, inosine; me₂G, dimethylguanosine; D, dihydrouridine; and meG, methylguanosine. The sites of the 10 modified bases in this tRNA are indicated by the color shading. (*b*) Generalized representation of tRNA in the cloverleaf form. The bases common to all tRNAs (both prokaryotic and eukaryotic) are indicated by letters; R is an invariant purine, Y is an invariant pyrimidine, ψ is an invariant pseudouridine. The greatest variability among tRNAs occurs in the V (variable) arm, which can range from 4 to 21 nucleotides. There are two sites of minor variability in the length of the D arm.

Over the following years, other tRNA species were puri-fied and sequenced, and a number of distinct similarities present in all the different tRNAs became evident (Fig-ure 11.42*b*). All tRNAs were roughly the same length—between 73 and 93 nucleotides—and all had a significant percentage of unusual bases that were found to result from enzymatic modifications of one of the four standard bases *after* it had been incorporated into the RNA chain, that is, *posttranscriptionally*. In addition, all tRNAs had sequences of nucleotides in one part of the molecule that were complemen-tary to sequences located in other parts of the molecule. Be-cause of these complementary sequences, the various tRNAs become folded in a similar way to form a structure that can be drawn in two dimensions as a cloverleaf. The base-paired stems and unpaired loops of the tRNA cloverleaf are shown in Figures 11.42 and 11.43. The unusual bases, which are con-centrated in the loops, disrupt hydrogen bond formation in these regions and serve as potential recognition sites for vari-ous proteins. All mature tRNAs have the triplet sequence CCA at their 3′ end. These three nucleotides may be encoded in the tRNA gene (in many prokaryotes) or added enzymati-cally (in eukaryotes). In the latter case, a single talented en-zyme adds all three nucleotides in the proper order without the benefit of a DNA or RNA template.

Up to this point we have considered only the secondary, or two-dimensional, structure of these adaptor molecules. Transfer RNAs fold into a unique and defined tertiary struc-ture. X-ray diffraction analysis has shown that tRNAs are constructed of two double helices arranged in the shape of an L (Figure 11.43*b*). Those bases that are found at compara-ble sites in all tRNA molecules (the *invariant* bases of Fig-ure 11.42*b*) are particularly important in generating the L-shaped tertiary structure. The common shapes of tRNAs reflects the fact that all of them take part in a similar series of reactions during protein synthesis. However, each tRNA has unique features that distinguish it from other tRNAs. As dis-cussed in the following section, it is these unique features that make it possible for an amino acid to become attached enzy-matically to the appropriate (cognate) tRNA.

Transfer RNAs translate a sequence of mRNA codons into a sequence of amino acid residues. Information contained in the mRNA is decoded through the formation of base pairs between complementary sequences in the transfer and messenger RNAs (see Figure 11.49). Thus, as with other processes involving nucleic acids, complementarity between base pairs lies at the heart of the translation process. The part of the tRNA that participates in this complementary interac-tion with the codon of the mRNA is a stretch of three sequen-tial nucleotides, called the **anticodon**, that is located in the middle loop of the tRNA molecule (Figure 11.43*a*). This loop is invariably composed of seven nucleotides, the middle three of which constitute the anticodon. The anticodon is located at one end of the L-shaped tRNA molecule opposite from the end at which the amino acid is attached (Figure 11.43*b*).

Given the fact that 61 different codons can specify an amino acid, a cell might be expected to have at least 61 differ-
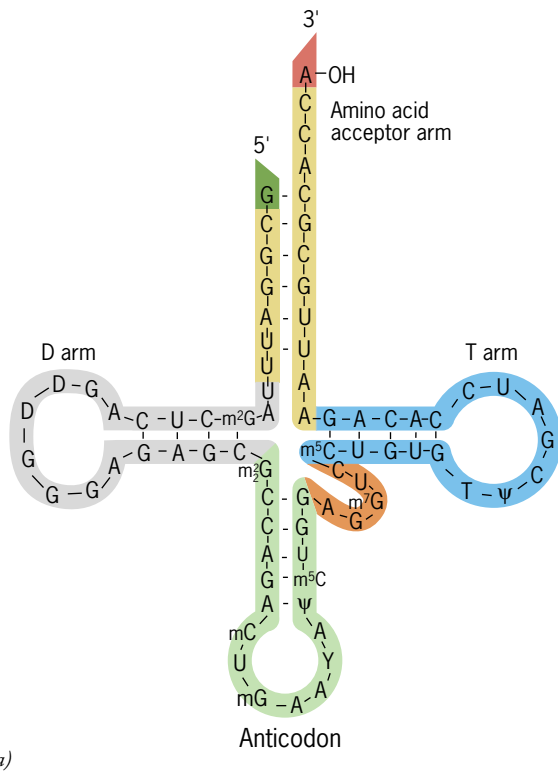
FIGURE 11.43 The structure of a tRNA. (*a*) Two-dimensional structure of a yeast phenylalanyl-tRNA with the various regions of the molecule color-coded to match part *b*. (*b*) Three-dimensional structure of tRNA^{Phe} derived from X-ray crystallography. The amino acceptor (AA) arm and the TψC (T) arm form a continuous double helix, and the anticodon (AC) arm and the D arm form a partially continuous double helix. These two helical columns meet to form an L-shaped molecule. (B: COURTESY OF MIKE CARSON.)

ent tRNAs, each with a different anticodon complementary to one of the codons of Figure 11.41. But recall that the greatest similarities among codons that specify the same amino acid occur in the first two nucleotides of the triplet, whereas the greatest variability in these same codons occurs in the third nucleotide of the triplet. Consider the 16 codons ending in U. In every case, if the U is changed to a C, the same amino acid is specified (first two lines of each box in Figure 11.41). Similarly, in most cases, a switch between an A and a G at the third site is also without effect on amino acid determination. The interchangeability of the base of the third position led Francis Crick to propose that the same transfer RNA may be able to recognize more than one codon. His proposal, termed the *wobble hypothesis*, suggested that the steric requirement between the anticodon of the tRNA and the codon of the mRNA is very strict for the first two positions but is more flexible at the third position. As a result, two codons that specify the same amino acid and differ only at the third position should use the same tRNA in protein synthesis. Once again, a Crick hypothesis proved to be correct.

The rules governing the wobble at the third position of the codon are as follows (Figure 11.44): U of the anticodon can pair with A or G of the mRNA; G of the anticodon can pair with U or C of the mRNA; and I (inosine, which is derived from guanine in the original tRNA molecule) of the anticodon can pair with U, C, or A of the mRNA. As a result of
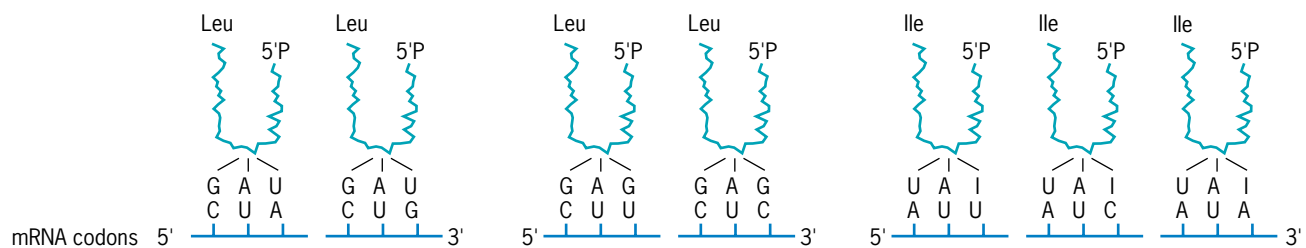
FIGURE 11.44 The wobble in the interaction between codons and anticodons. In some cases, the nucleotide at the 5′ end of the tRNA anticodon is capable of pairing with more than one nucleotide at the 3′ end (third position) of the mRNA codon. Consequently, more than one codon can use the same tRNA. The rules for pairing in the wobble scheme are indicated in the figure as well as in the text.

the wobble, the six codons for leucine, for example, require only three tRNAs.

Amino Acid Activation It is critically important during polypeptide synthesis that each transfer RNA molecule be attached to the correct (cognate) amino acid. Amino acids are covalently linked to the 3′ ends of their cognate tRNA(s) by an enzyme called an **aminoacyl-tRNA synthetase** (**aaRS**) (Figure 11.45). Although there are many exceptions, organisms typically contain 20 different aminoacyl-tRNA synthetases, one for each of the 20 amino acids incorporated into proteins. Each of the synthetases is capable of "charging" all of the tRNAs that are appropriate for that amino acid (i.e., any tRNA whose anticodon recognizes one of the various codons specifying that amino acid, as indicated in Figure 11.41). Aminoacyl-tRNA synthetases provide an excellent example of the specificity of protein–nucleic acid interactions. Certain common features must exist among all tRNA species specifying a given amino acid to allow a single aminoacyl-tRNA synthetase to recognize all of these tRNAs while, at the same time, discriminating against all of the tRNAs for other amino acids. Information concerning the structural features of tRNAs that cause them to be selected or rejected as substrates has come primarily from two sources:

1. Determination of the three-dimensional structure of these enzymes by X-ray crystallography, which allows investigators to identify which sites on the tRNA make direct contact with the protein. As illustrated in Figure 11.45, the two ends of the tRNA—the acceptor stem and anticodon—are particularly important for recognition by most of these enzymes.

2. Determination of the changes in a tRNA that cause the molecule to be aminoacylated by a noncognate synthetase. It is found, for example, that a specific base pair in a tRNA^{Ala} (the G-U base pair involving the third G from the 5′ end of the molecule in Figure 11.42a) is the primary determinant of its interaction with the alanyl-tRNA synthetase. Insertion of this specific base pair into the acceptor stem of a tRNA^{Phe} or a tRNA^{Cys} is sufficient to cause these tRNAs to be recognized by alanyl-tRNA synthetase and to be aminoacylated with alanine.

Aminoacyl-tRNA synthetases carry out the following two-step reaction:

1st step: ATP + amino acid → aminoacyl-AMP + PP$_i$

2nd step: aminoacyl-AMP + tRNA
$$\rightarrow \text{aminoacyl-tRNA} + \text{AMP}$$

In the first step, the energy of ATP activates the amino acid by formation of an adenylated amino acid, which is bound to the enzyme. This is the primary energy-requiring step in the chemical reactions leading to polypeptide synthesis. Subsequent events, such as the transfer of the amino acid to the tRNA molecule (step 2 above), and eventually to the growing polypeptide chain, are thermodynamically favorable. The PP$_i$ produced in the first reaction is subsequently hydrolyzed to P$_i$, further driving the overall reaction toward the formation of
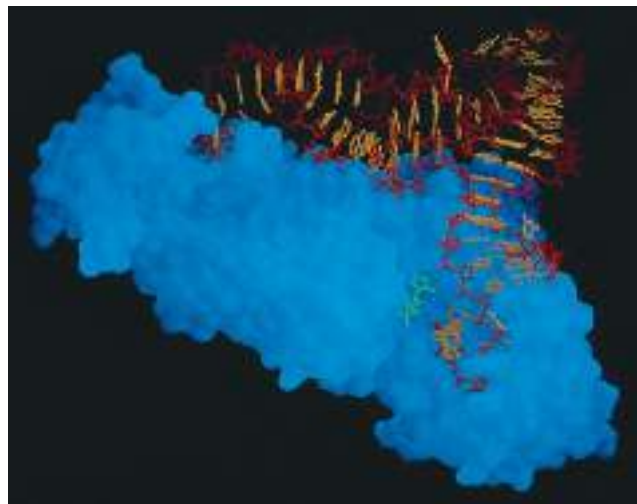
FIGURE 11.45 **Three-dimensional portrait of the interaction between a tRNA and its aminoacyl-tRNA synthetase.** The crystal structure of *E. coli* glutaminyl-tRNA synthetase (in blue) complexed with tRNA^{Gln} (in red and yellow). The enzyme recognizes this specific tRNA and discriminates against others through interactions with the acceptor stem and anticodon of the tRNA. (FROM THOMAS A. STEITZ, SCIENCE VOL. 246, COVER OF 12/1/89 ISSUE; © COPYRIGHT BY THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

products (page 423). As we will see later, energy is expended during protein synthesis, but it is not used in the formation of peptide bonds. In the second step, the enzyme transfers its bound amino acid to the 3′ end of a cognate tRNA. Should the synthetase happen to place an inappropriate amino acid on a tRNA, a proofreading mechanism of the enzyme is activated and the bond between the amino acid and tRNA is severed.[7]

Several methods have been developed that allow investigators to synthesize proteins, either in the test tube or within cells, that contain unnatural amino acids, that is, amino acids other than the 20 normally incorporated by the translation machinery. These unnatural amino acids are not generated by modification of normal amino acids after their incorporation into the polypeptide but are directly encoded within the mRNA. This "expansion" of the genetic code generally involves the use of an experimentally modified tRNA that recognizes one of the stop codons and a cognate amino acyl-tRNA synthetase that specifically recognizes the unnatural amino acid. The unnatural amino acid is then incorporated into the polypeptide chain wherever a particular stop codon is encountered within an mRNA. Using these methods, researchers can synthesize a protein containing chemical groups capable of reporting on the protein's activities, or they can design proteins with novel structures and functions for use as potential drugs or for other commercial applications.

---

[7]Not all aa-tRNA synthetases possess this type of proofreading mechanism. Some of these enzymes remove an incorrect amino acid by hydrolyzing the aminoacyl-AMP linkage following the first reaction step. The two amino acids most difficult to distinguish are valine and isoleucine, which differ by a single methylene group (Figure 2.26). The isoleucyl-tRNA synthetase employs both types of proofreading mechanisms which ensures accurate aminoacylation.

REVIEW ?

1. Why are tRNAs referred to as adaptor molecules? What aspects of their structure do tRNAs have in common?

2. Describe the nature of the interaction between tRNAs and aminoacyl-tRNA synthetases. Describe the nature of the interaction between tRNAs and mRNAs. What is the wobble hypothesis?

11.8 TRANSLATING GENETIC INFORMATION

Protein synthesis, or **translation**, may be the most complex synthetic activity in a cell. The assembly of a protein requires all the various tRNAs with their attached amino acids, ribosomes, a messenger RNA, numerous proteins having different functions, cations, and GTP. The complexity is not surprising considering that protein synthesis requires the incorporation of 20 different amino acids in the precise sequence dictated by a coded message written in a language that uses different characters. In the following discussion, we will draw most heavily on translation mechanisms as they operate in bacterial cells, which is simpler and better understood. The process is remarkably similar in eukaryotic cells.

The synthesis of a polypeptide chain can be divided into three rather distinct activities: *initiation* of the chain, *elongation* of the chain, and *termination* of the chain. We will consider each of these activities in turn.

Initiation

Once it attaches to an mRNA, a ribosome always moves along the mRNA from one codon to the next, that is, in consecutive blocks of three nucleotides. To ensure that the proper triplets are read, the ribosome attaches to the mRNA at a precise site, termed the **initiation codon**, which is specified as AUG. Binding to this codon automatically puts the ribosome in the proper **reading frame** so that it correctly reads the entire message from that point on. For example, in the following case,

—CUAGUUACAUGCUCCAGUCCGU—

the ribosome moves from the initiation codon, AUG, to the next three nucleotides, CUC, then to CAG, and so on along the entire line.

The basic steps in the initiation of translation in bacterial cells are illustrated in Figure 11.46.

Step 1: Bringing the Small Ribosomal Subunit to the Initiation Codon As seen in Figure 11.46, an mRNA does not bind to an intact ribosome, but to the small and large subunits in separate stages. The first major step of initiation is the binding of the small ribosomal subunit to the first AUG sequence (or one of the first) in the message, which serves as the initiation codon.[8] How does the small subunit select the

_____
[8]GUG is also capable of serving as an initiation codon and is found as such in a few natural messages. When GUG is used, *N*-formylmethionine is still utilized to form the initiation complex despite the fact that internal GUG codons specify valine.
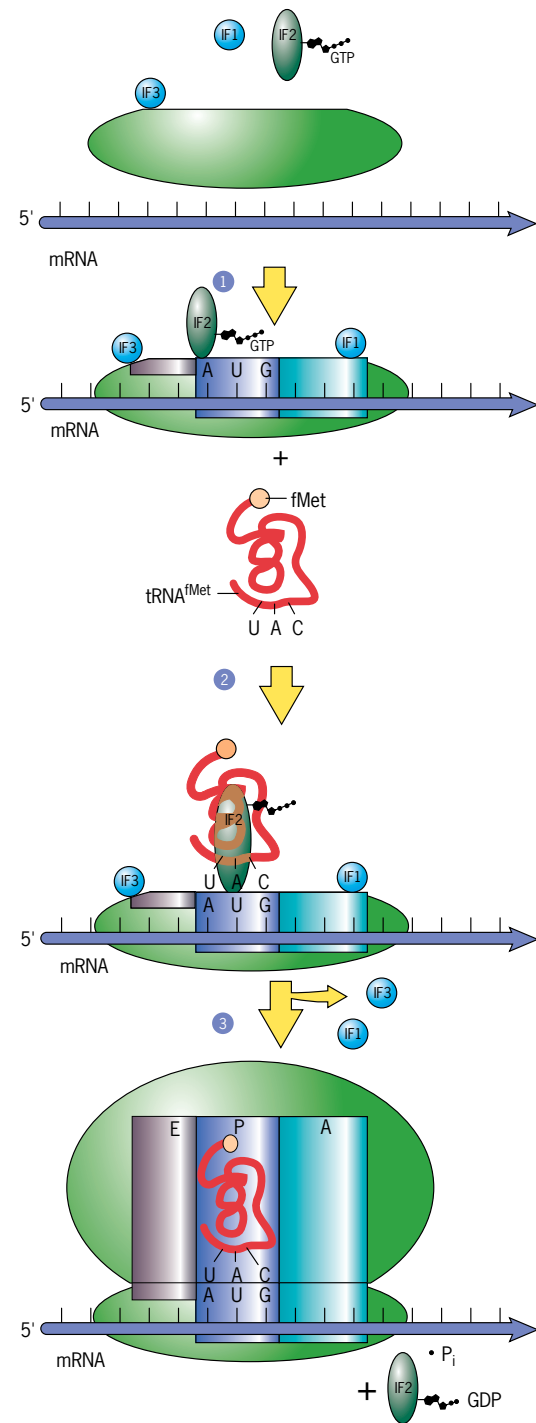
FIGURE 11.46 Initiation of protein synthesis in bacteria. In step 1, initiation of translation begins with the association of the 30S ribosomal subunit with the mRNA at the AUG initiation codon, a step that requires IF1 and IF3. The 30S ribosomal subunit binds to the mRNA at the AUG initiation codon as the result of an interaction between a complementary nucleotide sequence on the rRNA and mRNA, as discussed in the text. In step 2, the formylmethionyl-tRNA[fMet] becomes associated with the mRNA and the 30S ribosomal subunit complex by binding to IF2-GTP. In step 3, the 50S subunit joins the complex, GTP is hydrolyzed, and IF2-GDP is released. The initiator tRNA enters the P site of the ribosome, whereas all subsequent tRNAs enter the A site (see Figure 11.49).

initial AUG codon as opposed to some internal one? Bacterial mRNAs possess a specific sequence of nucleotides (called the Shine-Dalgarno sequence after its discoverers) that resides 5 to 10 nucleotides before the initiation codon. The Shine-Dalgarno sequence is complementary to a sequence of nucleotides near the 3′ end of the 16S *ribosomal RNA* of the small ribosomal subunit.

rRNA − − A C C U C C U U U A 3′
mRNA − − − G G A G G A − − − 5′

Interaction between these complementary sequences on the mRNA and rRNA leads to the attachment of a 30S subunit at the AUG initiation codon.

Initiation Requires Initiation Factors Several of the steps outlined in Figure 11.46 require the help of soluble proteins, called **initiation factors** (designated as IFs in bacteria and eIFs in eukaryotes). Bacterial cells require three initiation factors—IF1, IF2, and IF3—which attach to the 30S subunit (step 1, Figure 11.46). IF2 is a GTP-binding protein required for attachment of the first aminoacyl-tRNA. IF3 may prevent the large (50S) subunit from joining prematurely to the 30S subunit, and also facilitate entry of the initial aa-tRNA. IF1 promotes attachment of the 30S subunit to the mRNA and may prevent the aa-tRNA from entering the wrong site on the ribosome.

Step 2: Bringing the First aa-tRNA into the Ribosome If one examines the codon assignments (Figure 11.41), it can be seen that AUG is more than just an initiation codon; it is also the only codon for methionine. In fact, methionine is always the first amino acid to be incorporated at the N-terminus of a nascent polypeptide chain. (In bacteria, the initial methionine bears a formyl group, making it *N*-formylmethionine.) The methionine (or *N*-formylmethionine) is subsequently removed enzymatically from the majority of newly synthesized proteins. Cells possess two distinct methionyl-tRNAs: one used to initiate protein synthesis and a different one to incorporate methionyl residues into the interior of a polypeptide. The initiator aa-tRNA enters the P site of the ribosome (discussed below) where it binds to both the AUG codon of the mRNA and the IF2 initiation factor (step 2, Figure 11.46). IF1 and IF3 are released.

Step 3: Assembling the Complete Initiation Complex Once the initiator tRNA is bound to the AUG codon and IF3 is displaced, the large subunit joins the complex and the GTP bound to IF2 is hydrolyzed (step 3, Figure 11.46). GTP hydrolysis probably drives a conformational change in the ribosome that is required for the release of IF2-GDP.

Initiation of Translation in Eukaryotes Eukaryotic cells require at least 12 initiation factors comprising a total of more than 25 polypeptide chains. As indicated in Figure 11.47, several of these eIFs (e.g., eIF1, eIF1A, eIF5, and eIF3) bind to the 40S subunit, which prepares the subunit for binding to the
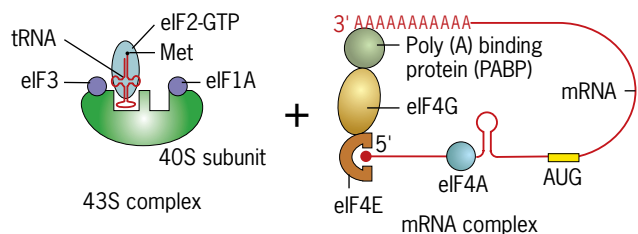
FIGURE 11.47 Initiation of protein synthesis in eukaryotes. As discussed in the text, initiation begins with the union of two complexes, one (called the 43S complex) contains the 40S ribosomal subunit bound to several initiation factors (eIFs) and the initiator tRNA, whereas the other contains the mRNA bound to a separate group of initiation factors. This union is mediated by an interaction between eIF3 on the 43S complex and eIF4G on the mRNA complex. Once the 43S complex has bound to the 5′ end of the mRNA, it scans along the message until it reaches the appropriate AUG initiation codon.

mRNA. The initiator tRNA linked to a methionine also binds to the 40S subunit prior to its interaction with the mRNA. The initiator tRNA enters the P site of the subunit in association with eIF2-GTP. Once these events have occurred, the small ribosomal subunit with its associated initiation factors and charged tRNA (which together form a 43S preinitiation complex) is ready to find the 5′ end of the mRNA, which bears the methylguanosine cap (page 442).

The 43S complex is originally recruited to the mRNA with the help of a cluster of initiation factors that are already bound to the mRNA (Figure 11.47). Among these factors: (1) eIF4E binds to the 5′ cap of the eukaryotic mRNA; (2) eIF4A moves along the 5′ end of the message removing any double-stranded regions that would interfere with the movement of the 43S complex along the mRNA; and (3) eIF4G serves as a linker between the 5′ capped end and the 3′ polyadenylated end of the mRNA (Figure 11.47). Thus EIF4G, in effect, converts a linear mRNA into a circular message.

Once the 43S complex binds to the 5′ end of the mRNA, the complex then scans along the message until it reaches a recognizable sequence of nucleotides (typically 5′—CCAC-CAUGC—3′) that contains the AUG initiation codon. Once the 43S complex reaches the appropriate AUG codon, eIF2-GTP is hydrolyzed, eIF2-GDP (and other associated eIFs) are released, and the large (60S) subunit joins the complex to complete initiation.[9]

The Role of the Ribosome Now that we have reached a point where we have assembled a complete ribosome, we can look more closely at the structure and function of this multisubunit structure. Ribosomes are molecular machines, similar in some respects to the molecular motors described in Chap-

---

[9]Not all mRNAs are translated following attachment of the small ribosomal subunit to the 5′ end of the message. Many viral mRNAs and a relatively small number of cellular mRNAs, most notably those utilized during mitosis or during periods of stress, are translated as the result of the ribosome attaching to the mRNA at an *internal ribosome entry site* (*IRES*), which may be located at some distance from the 5′ end of the message.

ter 9. During translation, a ribosome undergoes a repetitive cycle of mechanical changes that is driven with energy released by GTP hydrolysis. Unlike myosin or kinesin, which simply move along a physical track, ribosomes move along an mRNA "tape" containing encoded information. Ribosomes, in other words, are *programmable* machines: the information stored in the mRNA determines the sequence of aminoacyl-tRNAs that the ribosome accepts during translation. Another feature that distinguishes ribosomes from many other cellular machines is the importance of their component RNAs. Ribosomal RNAs play major roles in selecting tRNAs, ensuring accurate translation, binding protein factors, and polymerizing amino acids (discussed in the Experimental Pathways at the end of the chapter).

The past few years have seen great progress in our understanding of the structure of the ribosome. Earlier studies using high-resolution cryoelectron microscopic imaging techniques (Section 18.8) revealed the ribosome to be a highly irregular structure with bulges, lobes, channels, and bridges (see Figure 2.56). These studies also provided evidence of major conformational changes occurring in both the small and large subunits during translation. During the 1990s, major advances were made in the crystallization of ribosomes, and by the end of the decade the first reports of the X-ray crystallographic structure of prokaryotic ribosomes had appeared. Figures 11.48a and b show the overall structure of the two ribosomal subunits of a bacterial ribosome as revealed by X-ray crystallography.

Each ribosome has three sites for association with transfer RNA molecules. These sites, termed the **A (aminoacyl) site**, the **P (peptidyl) site**, and the **E (exit) site**, receive each tRNA in successive steps of the elongation cycle, as described in the following section. The positions of tRNAs bound to the A, P, and E sites of both the small and large ribosomal subunit are shown in Figure 11.48a,b. The tRNAs bind within these sites and span the gap between the two ribosomal subunits (Figure 11.48c). The anticodon ends of the bound tRNAs contact the small subunit, which plays a key role in decoding the information contained in the mRNA. In contrast, the amino acid–carrying ends of the bound tRNAs contact the large subunit, which plays a key role in catalyzing peptide bond formation. Other major features revealed by these high-resolution structural studies include the following:

1. The interface between the small and large subunits forms a relatively spacious cavity (Figure 11.48c) that is lined almost exclusively by RNA. The side of the small subunit that faces this cavity is lined along its length by a single continuous double-stranded RNA helix. This helix is shaded in the two-dimensional structure of the 16S rRNA in Figure 11.3. The surfaces of the two subunits that face one another contain the binding sites for the mRNA and incoming tRNAs and, thus, are of key importance for the function of the ribosome. The fact that these surfaces consist largely of RNA supports the proposal that primordial ribosomes were composed exclusively of RNA (page 448).

(a)   (b)

(a′)   (b′)

Binding sites for tRNA
A
P
E

(c)
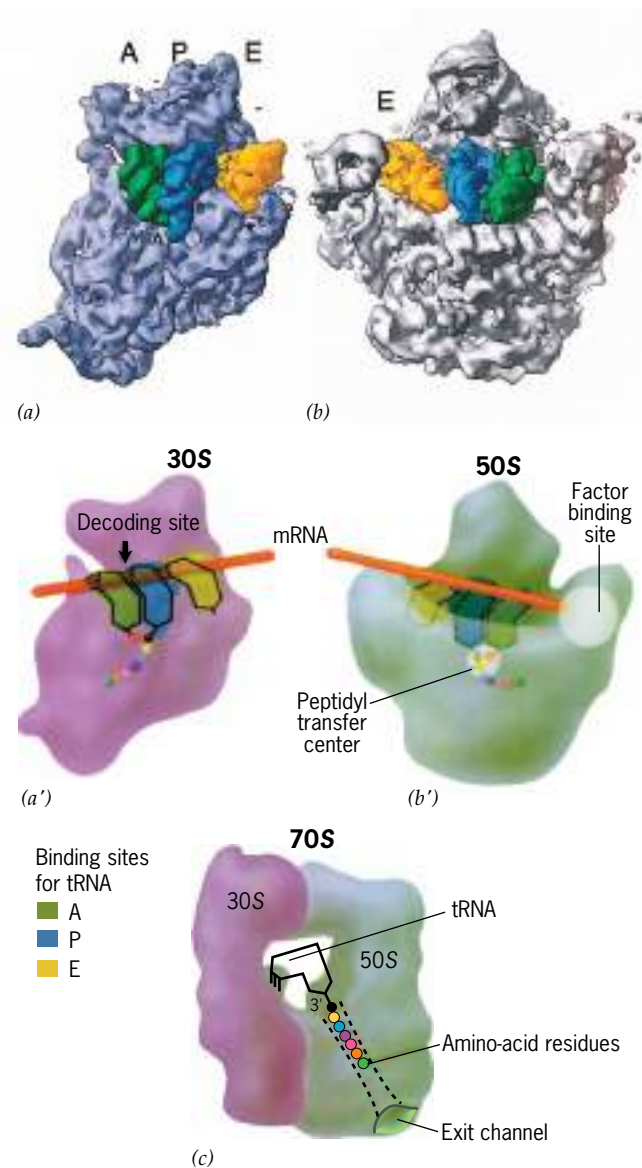
FIGURE 11.48 Model of the bacterial ribosome based on X-ray crystallographic data, showing tRNAs bound to the A, P, and E sites of the two ribosomal subunits. (*a–b*) Views of the 30S and 50S subunits, respectively, with the three bound tRNAs shown at the interface between the subunits. (*a′–b′*) Drawings corresponding to the structures shown in parts *a* and *b*. The drawing in *a′* of the 30S subunit shows the approximate locations of the anticodons of the three tRNAs and their interaction with the complementary codons of the mRNA. The drawing in *b′* of the 50S subunit shows the tRNA sites from the reverse direction. The amino acid acceptor ends of the tRNAs of the A and P sites are in close proximity within the peptidyl transfer site of the subunit, where peptide bond formation occurs. The binding sites for the elongation factors EF-Tu and EF-G are located on the protuberance at the right side of the subunit. (*c*) Drawing of the 70S bacterial ribosome showing the space between the two subunits that is spanned by each tRNA molecule and the channel within the 50S subunit through which the nascent polypeptide exits the ribosome. (A–B: FROM JAMIE H. CATE ET AL., COURTESY OF HARRY F. NOLLER, SCIENCE 285:2100, 1999; A′–B′, C: FROM A. LILJAS, SCIENCE 285:2078, 1999; © COPYRIGHT 1999, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

2. The active site, where amino acids are covalently linked to one another, also consists of RNA. This catalytic portion of the large subunit resides in a deep cleft, which protects the newly formed peptide bond from hydrolysis by the aqueous solvent.

3. The mRNA is situated in a narrow groove that winds around the neck of the small subunit, passing through the A, P, and E sites. Prior to entering the A site, the mRNA is stripped of any secondary structure it might possess by a helicase activity of the ribosome.

4. A tunnel runs completely through the core of the large subunit beginning at the active site. This tunnel provides a passageway for translocation of the elongating polypeptide through the ribosome (Figure 11.48c).

5. Most of the proteins of the ribosomal subunits have multiple RNA-binding sites and are ideally situated to stabilize the complex tertiary structure of the rRNAs.

Elongation

The basic steps in the process of elongation of translation in bacterial cells are illustrated in Figure 11.49. This series of steps is repeated over and over as amino acids are polymerized into the growing polypeptide chain.

Step 1: Aminoacyl-tRNA Selection   With the charged initiator tRNA in place within the P site, the ribosome is available for entry of the second aminoacyl-tRNA into the vacant A site, which is the first step in elongation (step 1, Figure 11.49a). Before the second aminoacyl-tRNA can effectively bind to the exposed mRNA in the A site, it must combine with a protein elongation factor bound to GTP. This particular elongation factor is called EF-Tu (or Tu) in bacteria and eEF1α in eukaryotes. EF-Tu is required to deliver aminoacyl-tRNAs to the A site of the ribosome. Although any aminoacyl-tRNA—Tu-GTP complex can enter the site, only one whose anticodon is complementary to the mRNA codon situated in the A site will trigger the necessary conformational changes within the ribosome that causes the tRNA to remain bound to the mRNA in the decoding center. Once the proper aminoacyl-tRNA—Tu-GTP is bound to the mRNA codon, the GTP is hydrolyzed and the Tu-GDP complex released, leaving the aa-tRNA bound in the ribosome's A site. Regeneration of Tu-GTP from the released Tu-GDP requires another elongation factor, EF-Ts.

Step 2: Peptide Bond Formation   At the end of the first step, the two amino acids, attached to their separate tRNAs, are juxtaposed in such a position that they can react with one another (Figure 11.48a′,b′). The second step in the elongation cycle is the formation of a peptide bond between these two amino acids (step 2, Figure 11.49a). Peptide bond formation is accomplished as the amine nitrogen of the aa-tRNA in the A site carries out a nucleophilic attack on the carbonyl carbon of the amino acid bound to the tRNA of the P site, displacing the P-site tRNA (Figure 11.49b). As a result of this reaction,

the tRNA bound to the second codon in the A site has an attached dipeptide, whereas the tRNA in the P site is deacylated. Peptide bond formation occurs spontaneously without the input of external energy. The reaction is catalyzed by **peptidyl transferase**, a component of the large subunit of the ribosome. For years it was assumed that the peptidyl transferase was one of the proteins of the ribosome. Then, as the catalytic powers of RNA became apparent, attention shifted to the ribosomal RNA as the catalyst for peptide bond formation. It has now been shown that peptidyl transferase activity does indeed reside in the large ribosomal RNA molecule of the large ribosomal subunit (see the chapter-opening photo). In other words, peptidyl transferase is a ribozyme, a subject discussed in the Experimental Pathways at the end of the chapter.

Step 3: Translocation   The formation of the first peptide bond leaves one end of the tRNA molecule of the A site still attached to its complementary codon on the mRNA and the other end of the molecule attached to a dipeptide (step 2, Figure 11.49a). The tRNA of the P site is now devoid of any linked amino acid. The following step, called **translocation**, is characterized by a ratchet-like motion of the small subunit relative to the large subunit. As a result, the ribosome moves three nucleotides (one codon) along the mRNA in the $5' \rightarrow 3'$ direction (step 3, Figure 11.49a). Translocation is accompanied by the movement of the dipeptidyl-tRNA from the A site to the P site of the ribosome, still hydrogen-bonded to the second codon of the mRNA, and the movement of the deacylated tRNA from the P site to the E site. An intermediate stage in the translocation process has been visualized by cryo-electron microscopy, which shows the tRNAs occupying partially translocated "hybrid states." In these hybrid states, the anticodon ends of the tRNAs still reside in the A and P sites of the small subunit, while the acceptor ends of the tRNAs have moved into the P and E sites of the large subunit. Translocation is driven by conformational changes in another elongation factor (EF-G in bacteria and eEF2 in eukaryotes) following hydrolysis of its bound GTP. Following this reaction, EF-G–GDP leaves the ribosome.

Step 4: Releasing the Deacylated tRNA   In the final step of elongation (step 4, Figure 11.49a), the deacylated tRNA leaves the ribosome, emptying the E site.

For each cycle of elongation, at least two molecules of GTP are hydrolyzed: one during aminoacyl-tRNA selection and one during translocation. Each cycle of elongation takes about one-twentieth of a second, most of which is probably spent sampling aa-tRNAs from the surrounding cytosol. Once the peptidyl-tRNA has moved to the P site by translocation, the A site is once again open to the entry of another aminoacyl-tRNA, in this case one whose anticodon is complementary to the third codon (Figure 11.49a). Once the third charged tRNA is associated with the mRNA in the A site, the dipeptide from the tRNA of the P site is displaced by the aa-tRNA of the A site, forming the second peptide bond and, as a result, a tripeptide attached to the tRNA of the A site. The tRNA in the P site is once again devoid of an amino acid. Pep-

(a)

(b)
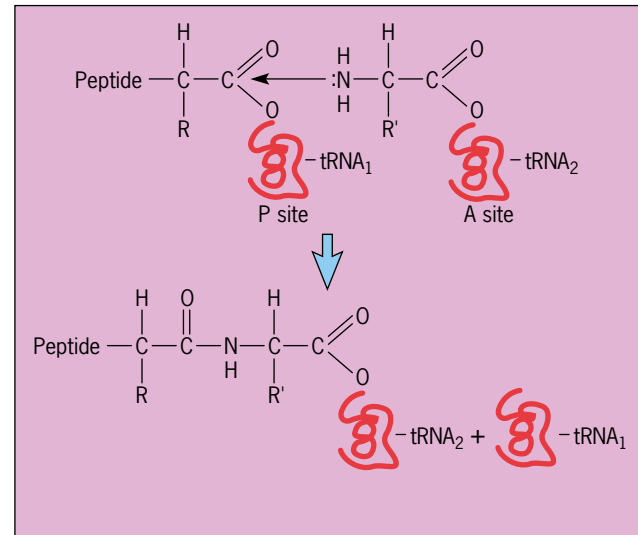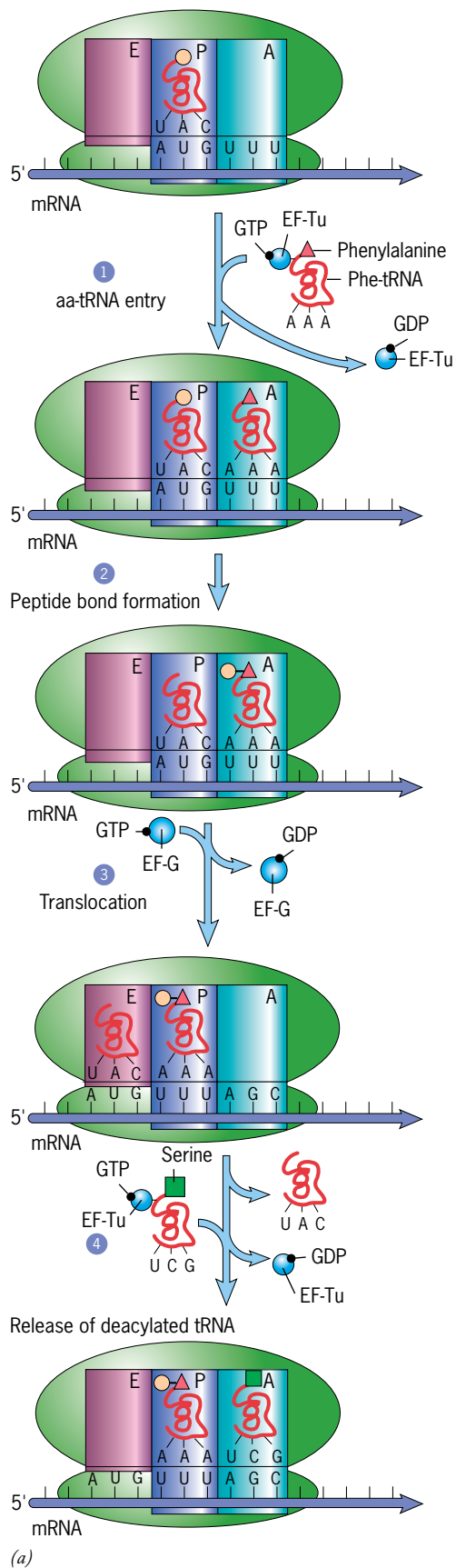
FIGURE 11.49 Steps in the elongation of the nascent polypeptide during translation in bacteria. (*a*) In step 1, an aminoacyl-tRNA whose anticodon is complementary to the second codon of the mRNA enters the empty A site of the ribosome. The binding of the tRNA is accompanied by the release of EF-Tu-GDP. In step 2, peptide bond formation is accomplished by the transfer of the nascent polypeptide chain from the tRNA in the P site to the aminoacyl-tRNA of the A site, forming a dipeptidyl-tRNA in the A site and a deacylated tRNA in the P site. The reaction is catalyzed by a part of the rRNA acting as a ribozyme. In step 3, the binding of EF-G and the hydrolysis of its associated GTP results in the translocation of the ribosome relative to the mRNA. Translocation is accompanied by the movement of the deacylated tRNA and peptidyl-tRNA into the E and P sites, respectively. In step 4, the deacylated tRNA leaves the ribosome, and a new aminoacyl-tRNA enters the A site. (*b*) Peptide bond formation and the subsequent displacement of the deacylated tRNA. A ribosome can catalyze the incorporation of approximately five amino acids into a growing polypeptide per second, which is roughly 10 million times greater than that observed in the uncatalyzed reaction using model substrates in solution.

tide bond formation is followed by translocation of the ribosome to the fourth codon and release of the deacylated tRNA, whereupon the cycle is ready to begin again.

We have seen in this section how the ribosome moves three nucleotides (one codon) at a time along the mRNA. The particular sequence of codons in the mRNA that is utilized by a ribosome (i.e., the reading frame) is fixed at the time the ribosome attaches to the initiation codon at the beginning of translation. Some of the most destructive mutations are ones in which a single base pair is either added to or deleted from the DNA. Consider the effect of an addition of a single nucleotide in the following sequence:

—AUG CUC CAG UCC GU →
—AUG CUC GCA GUC CGU—

The ribosome moves along the mRNA in the incorrect reading frame from the point of mutation through the remainder

of the coding sequence. Mutations of this type are called **frameshift mutations**. Frameshift mutations lead to the assembly of an entirely abnormal sequence of amino acids from the point of the mutation. It can be noted that, after more than two decades in which it was assumed that the ribosome always moved from one triplet to the next, several examples were discovered in which the mRNA contains a recoding signal that causes the ribosome to change its reading frame, either backing up one nucleotide (a shift to the -1 frame) or skipping a nucleotide (a shift to the $+1$ frame).

A wide variety of antibiotics have their effect by interfering with various aspects of protein synthesis in bacterial cells. Streptomycin, for example, selectively binds to the small ribosomal subunit of bacterial cells, causing certain of the codons of the mRNA to be misread, thus increasing the synthesis of aberrant proteins. Because the antibiotic does not bind to eukaryotic ribosomes, it has no effect on translation of the host cell's mRNAs. Resistance by bacteria to streptomycin can be traced to changes in ribosomal proteins, particularly S12.

Termination

As shown in Figure 11.41, 3 of the 64 trinucleotide codons function as stop codons that terminate polypeptide assembly rather than encode an amino acid. No tRNAs exist whose anticodons are complementary to a stop codon.[10] When a ribosome reaches one of these codons, UAA, UAG, or UGA, the signal is read to stop further elongation and release the polypeptide associated with the last tRNA.

Termination requires *release factors*. Release factors can be divided into two groups: class I RFs, which recognize stop codons in the A site of the ribosome, and class II RFs, which are GTP-binding proteins (G proteins) whose roles are not well understood. Bacteria have two class I RFs: RF1, which recognizes UAA and UAG stop codons, and RF2, which recognizes UAA and UGA stop codons. Eukaryotes have a single class I RF, eRF1, which recognizes all three stop codons. Class I RFs enter the A site of the ribosome, where a conserved tripeptide at one end of the release factor is thought to interact directly with the stop codon in the A site not unlike the way that an anticodon triplet of a tRNA molecule would interact with a sense codon in that site. The ester bond linking the nascent polypeptide chain to the tRNA is then hy-

---

[10]There are minor exceptions to this statement. It is stated in this chapter that codons dictate the incorporation of 20 different amino acids. In actual fact, there is a 21st amino acid, called selenocysteine, that is incorporated into a very small number of polypeptides. Selenocysteine is a rare amino acid that contains the metal selenium. In mammals, for example, it occurs in a dozen or so proteins. Selenocysteine has its own tRNA, called tRNA^{Sec}, but lacks its own aa-tRNA synthetase. This unique tRNA is recognized by the seryl-tRNA synthetase, which attaches a serine to the 3′ end of the tRNA^{Sec}. Following attachment, the serine is altered enzymatically to a selenocysteine. Selenocysteine is encoded by UGA, which is one of the three stop condons. Under most circumstances UGA is read as a termination signal. In a few cases, however, the UGA is followed by a folded region of the mRNA that binds a special elongation factor that causes the ribosome to recruit a tRNA^{Sec} into the A site rather than a termination factor. A 22nd amino acid, pyrrolysine, is encoded by another stop codon (UAG) in the genetic code of some archaea. Pyrrolysine has its own tRNA and aa-tRNA synthetase.

drolyzed, and the completed polypeptide is released. At this point, hydrolysis of the GTP bound to the class II RF (RF3 or eRF3) leads to the release of the class I RF from the ribosome's A site. The final steps in translation include the release of the deacylated tRNA from the P site, dissociation of the mRNA from the ribosome, and disassembly of the ribosome into its large and small subunits in preparation for another round of translation. These final steps require a number of protein factors. In bacterial cells, these proteins include EF-G, IF3, and RRF (ribosome recycling factor), which promotes ribosomal subunit separation.

mRNA Surveillance and Quality Control

Because the three termination codons can be readily formed by single base changes from many other codons (see Figure 11.41), one might expect mutations to arise that produce stop codons within the coding sequence of a gene. Mutations of this type, termed **nonsense mutations**, have been studied for decades and are responsible for roughly 30 percent of inherited disorders in humans. Premature termination codons, as they are also called, are also commonly introduced into mRNAs during splicing. Cells possess an mRNA surveillance mechanism capable of detecting messages with premature termination codons. In most cases, mRNAs containing such mutations are translated only once before they are selectively destroyed by a process called **nonsense-mediated decay** (**NMD**). NMD protects the cell from producing nonfunctional, shortened proteins.

How is it possible for a cell to distinguish between a legitimate termination codon that is supposed to end translation of a message and a premature termination codon? To understand this feat, we have to reconsider events that occur during pre-mRNA processing in mammalian cells. It wasn't mentioned earlier, but when an intron is removed by a spliceosome, a complex of proteins is deposited on the transcript 20–24 nucleotides upstream from the newly formed exon–exon junction. This landmark of the splicing process is called the **exon–junction complex** (**EJC**), which stays with the mRNA until it is translated. In a normal mRNA, the termination codon is typically present in the last exon and the EJC is just upstream from that site. As an mRNA undergoes its initial round of translation, the EJCs are thought to be displaced by the advancing ribosome. Consider what would happen during translation of an mRNA that contained a premature termination codon. The ribosome would stop at the site of the mutation and then dissociate, leaving any EJCs that were attached to the mRNA downstream of the site of premature termination. This sets in motion a series of events leading to the enzymatic destruction of the abnormal message.

NMD is best known for its role in eliminating mRNAs transcribed from mutant genes, such as those responsible for cystic fibrosis or muscular dystrophy. A number of biotechnology companies are developing drugs that interfere with NMD and allow RNAs with nonsense codons to be translated into proteins. Patients with both cystic fibrosis and muscular dystrophy are currently being treated with such drugs in clinical

trials. Even though the protein encoded by the mutant gene will be abnormally short, it may still possess enough residual activity to rescue the patients from an otherwise fatal disease.

NMD serves as another reminder of the opportunistic nature of biological evolution. Just as evolution has "taken advantage" of the presence of introns to facilitate exon shuffling (page 448), it has also used the process by which these gene inserts are removed to establish a mechanism of quality control, which ensures that only untainted mRNAs advance to a stage where they can be translated.

Polyribosomes

When a messenger RNA in the process of being translated is examined in the electron microscope, a number of ribosomes are invariably seen to be attached along the length of the mRNA thread. This complex of ribosomes and mRNA is called a **polyribosome**, or **polysome** (Figure 11.50*a*). Each of the ribosomes initially assembles from its subunits at the initiation codon and then moves from that point toward the 3′ end of the mRNA until it reaches a termination codon. As each ribosome moves away from the initiation codon, another ribosome attaches to the mRNA and begins its translation ac-

tivity. The rate at which translation initiation occurs varies with the mRNA being studied; some mRNAs have a much greater density of associated ribosomes than others. The simultaneous translation of the same mRNA by numerous ribosomes greatly increases the rate of protein synthesis within the cell. Recent studies utilizing cryoelectron tomography (Section 18.2) have suggested that the three-dimensional arrangement and orientation of ribosomes within a "free" (i.e., nonmembrane bound) polysome are quite highly ordered. A model of a polysome generated by this technique is shown in Figure 11.50*b*. The ribosomes that comprise this model polysome are densely packed and have adopted a "double-row" array. Moreover, each of the individual ribosomes is oriented so that its nascent polypeptide (red or green filaments) is situated at its outer surface facing the cytosol. It is suggested that this orientation maximizes the distance between nascent chains, thereby minimizing the likelihood that the nascent chains will interact with one another and, possibly, aggregate. Figure 11.50*c* shows an electron micrograph of polysomes bound to the cytosolic surface of the ER membrane. It is presumed that these polysomes had been engaged in the synthesis of membrane and/or organelle proteins at the time the cell was fixed (page 278). The ribosomes within each polysome appear to be organized at the surface of the ER membrane into a circular loop or spiral.

Now that we have described the basic events of translation, it is fitting to close the chapter with pictures of the process, one taken from a prokaryotic cell (Figure 11.51*a*) and the other from a eukaryotic cell (Figure 11.51*b*). Unlike eukaryotic cells, in which transcription occurs in the nucleus and translation occurs in the cytoplasm with many intervening steps, the corresponding activities in bacterial cells are tightly coupled. Thus protein synthesis in bacterial cells begins on mRNA templates well before the mRNA has been completely synthesized. The synthesis of an mRNA proceeds in the same direction as the movement of ribosomes translating that mes-

(a)

(b)

(c)
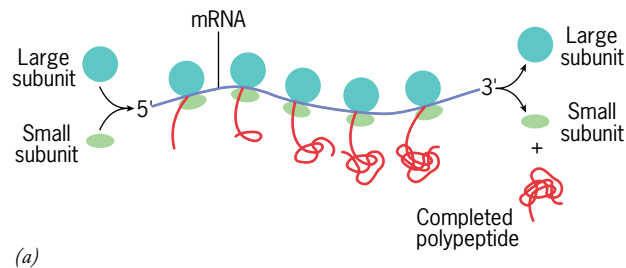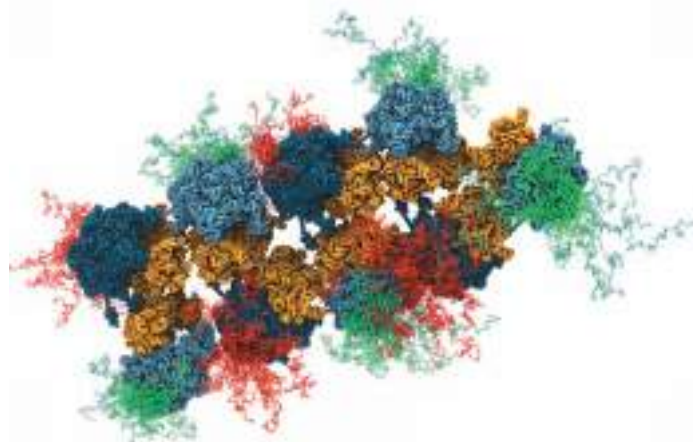
FIGURE 11.50 Polyribosomes. (*a*) Schematic drawing of a polyribosome (polysome). (*b*) This three-dimensional model was generated from cryoelectron tomograms of bacterial polysomes in the act of translation in vitro. To obtain the tomograms, preparations were vitrified (frozen into glass-like solid ice without formation of ice crystals) in liquid ethane at −196° C. Electron micrographs were then taken with the specimen positioned at various tilt angles, providing the data to generate a three-dimensional reconstruction. (*c*) Electron micrograph of a grazing section through the outer edge of a rough ER cisterna. The ribosomes are aligned in loops and spirals, indicating their attachment to mRNA molecules to form polysomes. (B: FROM FLORIAN BRANDT ET AL., COURTESY OF WOLFGANG BAUMEISTER, CELL 136:267, 2009; BY PERMISSION OF CELL PRESS; C: COURTESY OF E. YAMADA.)
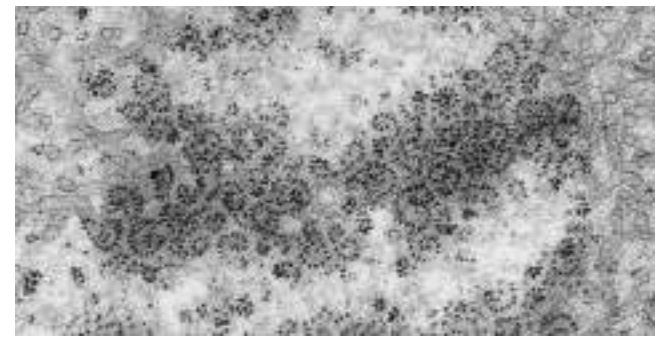
(a)

(b)
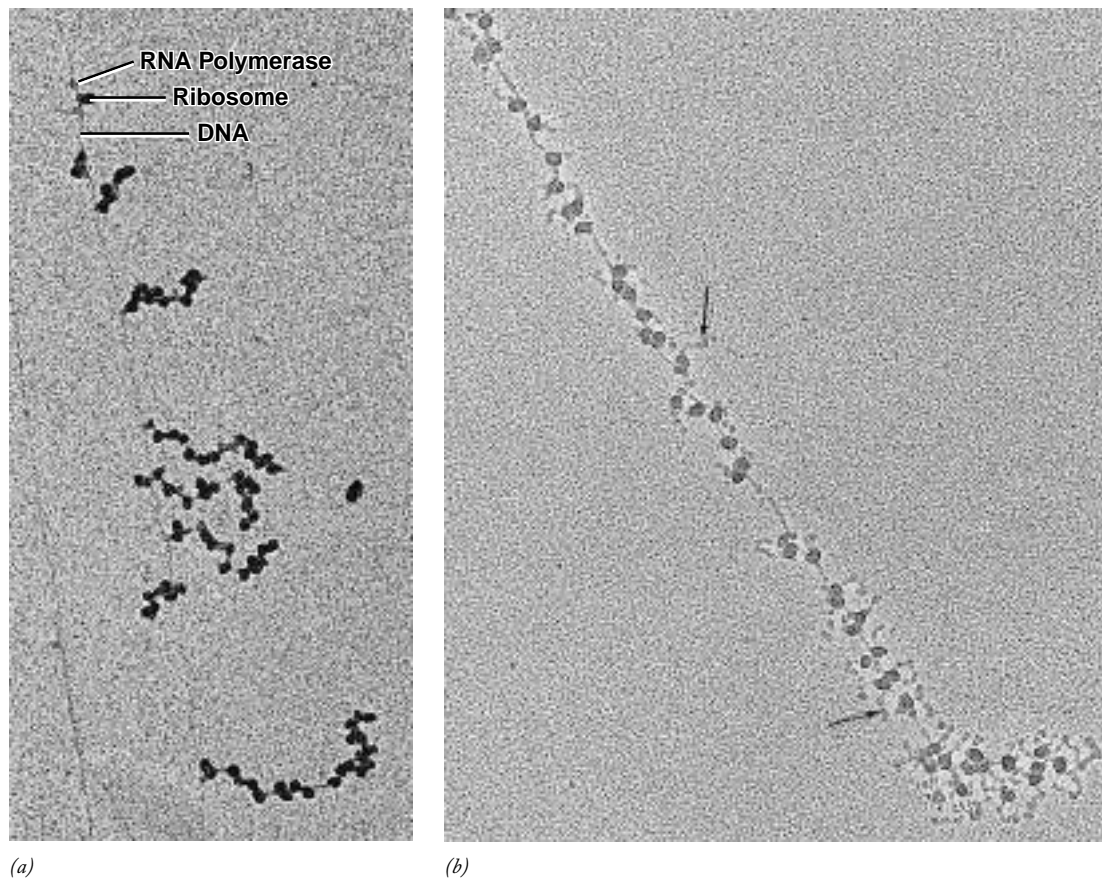
FIGURE 11.51 Visualizing transcription and translation. (*a*) Electron micrograph of portions of an *E. coli* chromosome engaged in transcription. The DNA is seen as faint lines running the length of the photo, whereas the nascent mRNA chains are seen to be attached at one of their ends, presumably by an RNA polymerase molecule. The particles associated with the nascent RNAs are ribosomes in the act of translation; in bacteria, transcription and translation occur simultaneously. The RNA molecules increase in length as the distance from the initiation site increases. (*b*) Electron micrograph of a polyribosome isolated from cells of the silk gland of the silkworm, which produces large quantities of the silk protein fibroin. This protein is large enough to be visible in the micrograph (arrows point to nascent polypeptide chains). (A: REPRINTED WITH PERMISSION FROM OSCAR L. MILLER, JR., BARBARA A. HAMKALO, AND C. A. THOMAS, SCIENCE 169:392, 1970; © COPYRIGHT 1970, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE; B: COURTESY OF STEVEN L. MCKNIGHT AND OSCAR L. MILLER, JR.)

sage, that is, from the 5′ to 3′ end. Consequently, as soon as an RNA molecule has begun to be produced, the 5′ end is available for attachment of ribosomes. The micrograph in Figure 11.51*a* shows DNA being transcribed, nascent mRNAs being synthesized, and ribosomes that are translating each of the nascent mRNAs. Nascent protein chains are not visible in the micrograph of Figure 11.51*a*, but they are visible in the micrograph of Figure 11.51*b*, which shows a single polyribosome isolated from a silk gland cell of a silkworm. The silk protein being synthesized is visible because of its large size and fibrous nature. The development of techniques to visualize transcription and translation by Oscar Miller, Jr. have provided a fitting visual demonstration of processes that had been previously described in biochemical terms.

REVIEW ?

1. Describe some of the ways that the initiation step of translation differs from the elongation steps of translation.

2. How does the effect of a nonsense mutation differ from that of a frameshift mutation? Why?

3. What is a polyribosome? How does its formation differ in prokaryotes and eukaryotes?

4. During translation elongation, it can be said that an aminocyl-tRNA enters the A site, a peptidyl-tRNA enters the P site, and a deacylated tRNA enters the E site. Explain how each of these events occurs.

EXPERIMENTAL PATHWAYS

The Role of RNA as a Catalyst

Research in biochemistry and molecular biology through the 1970s had solidified our understanding of the roles of proteins and nucleic acids. Proteins were the agents that made things happen in the cell, the enzymes that accelerated the rates of chemical reactions within organisms. Nucleic acids, on the other hand, were the informational molecules in the cell, storing genetic instructions in their nucleotide sequence. The division of labor between protein and nucleic acids seemed as well defined as any distinction that had been drawn in the biological sciences. Then, in 1981, a paper was published that began to blur this distinction.[1]

Thomas Cech and his co-workers at the University of Colorado had been studying the process by which the ribosomal RNA precursor synthesized by the ciliated protozoan *Tetrahymena thermophila* was converted to mature rRNA molecules. The *T. thermophila* pre-rRNA contained an intron of about 400 nucleotides that had to be excised from the primary transcript before the adjoining segments could be linked together (ligated).

Cech had previously found that nuclei isolated from the cells were able to synthesize the pre-rRNA precursor and carry out the entire splicing reaction. Splicing enzymes had yet to be isolated from any type of cell, and it seemed that *Tetrahymena* might be a good system in which to study these enzymes. The first step was to isolate the pre-rRNA precursor in an intact state and determine the minimum number of nuclear components that had to be added to the reaction mixture to obtain accurate splicing. It was found that when isolated nuclei were incubated in a medium containing a low concentration of monovalent cations (5 mM NH_4^+), the pre-rRNA molecule was synthesized but the intron was not excised. This allowed the researchers to purify the intact precursor, which they planned to use as a substrate to assay for splicing activity in nuclear extracts. They found, however, that when the purified precursor was incubated by itself at higher concentrations of NH_4^+ in the presence of Mg^{2+} and a guanosine phosphate (e.g., GMP or GTP), the intron was spliced from the precursor (Figure 1).[1] Nucleotide-sequence analysis confirmed that the small RNA that had been excised from the precursor was the intron with an added guanine-containing nucleotide at its 5′ end. The additional nucleotide was shown to be derived from the GTP that was added to the reaction mixture.

Splicing of an intron is a complicated reaction that requires recognition of the sequences that border the intron, cleavage of the phosphodiester bonds on both sides of the intron, and ligation of the adjoining fragments. All efforts had been made to remove any protein that might be clinging to the RNA before it was tested for its ability to undergo splicing. The RNA had been extracted with detergent–phenol, centrifuged through a gradient, and treated with a proteolytic enzyme. There were only two reasonable explanations: either splicing was accomplished by a protein that was bound very tightly to the RNA, or the pre-rRNA molecule was capable of splicing itself. The latter idea was not easy to accept.

To resolve the question of the presence of a protein contaminant, Cech and co-workers utilized an artificial system that could not possibly contain nuclear splicing proteins.[2] The DNA that encodes the rRNA precursor was synthesized in *E. coli*, purified, and used as a template for in vitro transcription by a purified bacterial RNA polymerase. The pre-rRNA synthesized in vitro was then purified and incubated by itself in the presence of monovalent and divalent ions and
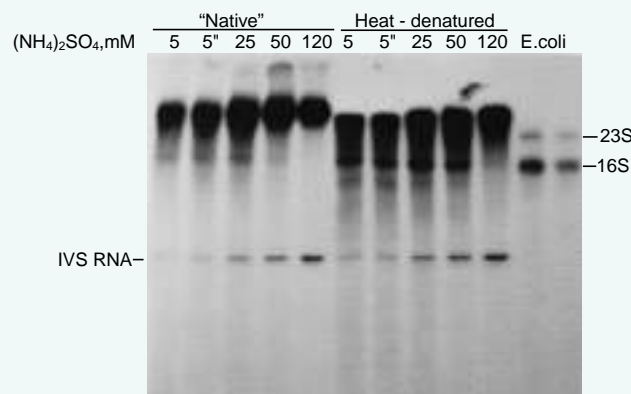
FIGURE 1 Purified ^{32}P-labeled *Tetrahymena* ribosomal RNA, transcribed at different $(NH_4)_2SO_4$ concentrations, was analyzed by polyacrylamide gel electrophoresis. The numbers at the top indicate the concentration of ammonium sulfate. Two groups of samples are shown, "native" and heat-denatured. Samples of the latter group were boiled for five minutes in buffer and cooled on ice to dissociate any molecules that might be held together by hydrogen bonding between complementary bases. The right two columns contain bacterial 16S and 23S rRNAs, which provide markers of known size with which the other bands in the gel can be compared. It can be seen from the positions of the bands that, as the ammonium sulfate concentration rises, the presence of a small RNA appears whose size is equal to that of the isolated intron (IVS, intervening sequence). These data provided the first indication that the rRNA was able to excise the intron without the aid of additional factors. (FROM T. R. CECH ET AL., CELL 27:488, 1981; BY PERMISSION OF CELL PRESS.)

a guanosine compound. Because the RNA had never been in a cell, it could not possibly be contaminated by cellular splicing enzymes. Yet the isolated pre-rRNA underwent the precise splicing reaction that would have occurred in the cell. The RNA had to have spliced itself.

As a result of these experiments, RNA was shown to be capable of catalyzing a complex, multistep reaction. Calculations indicated that the reaction had been speeded approximately 10 billion times over the rate of the noncatalyzed reaction. Thus, like a protein enzyme, the RNA was active in small concentration, was not altered by the reaction, and was able to greatly accelerate a chemical reaction. The primary distinction between this RNA and "standard proteinaceous enzymes" was that the RNA acted on itself rather than on an independent substrate. Cech called the RNA a "ribozyme."

In 1983, a second, unrelated example of RNA catalysis was discovered.[3] Sidney Altman of Yale University and Norman Pace of the National Jewish Hospital in Denver were collaborating on the study of ribonuclease P, an enzyme involved in the processing of a transfer RNA precursor in bacteria. The enzyme was unusual in being composed of both protein and RNA. When incubated in buffers containing high concentrations (60 mM) of Mg^{2+}, the purified RNA subunit was found capable of removing the 5′ end of the tRNA precursor (lane 7, Figure 2) just as the entire ribonuclease P molecule would have done inside the cell. The reaction products of the in vitro reaction included the processed, mature tRNA molecule. In contrast,
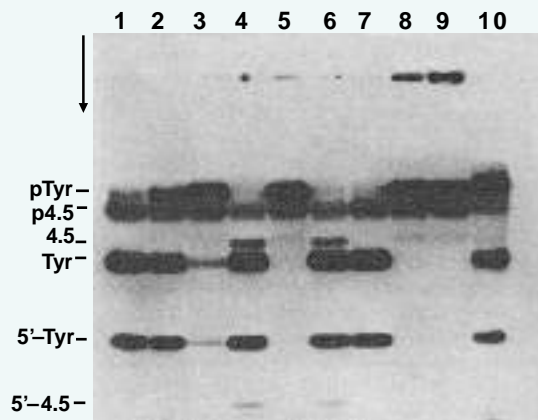
FIGURE 2 The results of polyacrylamide gel electrophoresis of reaction mixtures that had contained the precursor of tyrosinyl-tRNA (pTyr) and the precursor of another RNA called 4.5S RNA (p4.5). We will focus only on pTyr, which is normally processed by ribonuclease P into two RNA molecules, Tyr and 5′-Tyr (which is the 5′ end of the precursor). The positions at which these three RNAs (pTyr, Tyr, and 5′-Tyr) migrate during electrophoresis are indicated to the left of the gel. Lane 1 shows the RNAs that appear in the reaction mixture when pTyr (and p4.5) are incubated with the complete ribonuclease P. Very little of the pTyr remains in the mixture; it has been converted into the two products (Tyr and 5′-Tyr). Lane 5 shows the RNAs that appear in the reaction mixture when pTyr is incubated with the purified protein component of ribonuclease P. The protein shows no ability to cleave the tRNA precursor as evidenced by the absence of bands where the two products would migrate. In contrast, when pTyr is incubated with the purified RNA component of ribonuclease P (lane 7), the pTyr is processed as efficiently as if the intact ribonucleoprotein had been used. (FROM CECILIA GUERRIER-TOKADA ET AL. CELL 35:850, 1983; BY PERMISSION OF CELL PRESS.)

FIGURE 3 A molecular model of a portion of the catalytic RNA subunit of bacterial ribonuclease P (white) and its bound substrate, precursor tRNA (red). The site on the precursor tRNA where cleavage by the ribozyme occurs is indicated by a yellow sphere. (COURTESY OF MICHAEL E. HARRIS AND NORMAN R. PACE.)

the isolated protein subunit of the enzyme was devoid of catalytic activity (lane 5, Figure 2).

To eliminate the possibility that a protein contaminant was actually catalyzing the reaction, the RNA portion of ribonuclease P was synthesized in vitro from a recombinant DNA template. As with the RNA that had been extracted from bacterial cells, this artificially synthesized RNA, without any added protein, was capable of accurately cleaving the tRNA precursor.[4] Unlike the rRNA processing enzyme studied by Cech, the RNA of ribonuclease P acted on another molecule as a substrate rather than on itself. Thus, it was demonstrated that ribozymes have the same catalytic properties as proteinaceous enzymes. A model of the interaction between the catalytic RNA subunit of ribonuclease P and a precursor tRNA substrate is shown in Figure 3.

The demonstration that RNA could catalyze chemical reactions created the proper atmosphere to reinvestigate an old question: Which component of the large ribosomal subunit is the peptidyl transferase, that is, the catalyst for peptide bond formation? During the 1970s, a number of independent findings raised the possibility that ribosomal RNAs might be doing more than simply acting as a scaffold to hold ribosomal proteins in the proper position so they could catalyze translation. Included among the evidence were the following types of data:

1. Certain strains of *E. coli* carry genes that encode bacteria-killing proteins called colicins. One of these toxins, colicin E3, was known to inhibit protein synthesis in sensitive bacterial cells. Ribosomes that had been isolated from colicin E3–treated cells appeared perfectly normal by most criteria, but they were unable to support protein synthesis in vitro. Further analysis of such ribosomes revealed that the defect resides in the rRNA, not the ribosomal proteins. The 16S RNA of the small subunit is cleaved by the colicin about 50 nucleotides from its 3′ end, which renders the entire subunit unable to support protein synthesis.[5]

2. Treatment of large ribosomal subunits with ribonuclease T₁, an enzyme that cleaves bonds between accessible RNA nucleotides, destroyed the subunit's ability to carry out the peptidyl transferase reaction.[6]

3. A wide variety of studies on antibiotics that inhibited peptide bond formation, including chloramphenicol, carbomycin, and erythromycin, suggested that these drugs act on the ribosomal RNA, not the protein. For example, it was found that ribosomes became resistant to the effects of chloramphenicol as the result of substitutions in the bases in the ribosomal RNA.[7]

4. Ribosomal RNAs were shown to have a highly conserved base sequence, much more so than the amino acid sequences of ribosomal proteins. Some of the regions of the ribosomal RNAs are virtually unchanged in ribosomes isolated from prokaryotes, plants, and animals, as well as ribosomes isolated from mitochondria and chloroplasts. The fact that rRNA sequences are so highly conserved suggests that the molecules have a crucial role in the function of the ribosome.[8,9] In fact, a 1975 paper by C. R. Woese and co-workers states, "Since little or no correlation exists between these conserved regions and known ribosomal protein binding sites, the implication is strong that large areas of the RNA are directly involved in ribosomal function."[8]

Following the discovery of catalytic RNAs by Cech's and Altman's laboratories, investigation into the roles of ribosomal RNAs intensified. A number of studies by Harry Noller and his colleagues at the University of California, Santa Cruz, pinpointed the site in the ribosomal RNA that resides in or around the peptidyl transferase center.[9] In one study, it was shown that transfer RNAs

bound to the ribosome protect specific bases in the rRNA of the large subunit from attack by specific chemical agents. Protection from chemical attack is evidence that the tRNA must be situated very close to the rRNA bases that are shielded.[10] Protection is lost if the 3′ end of the tRNA (the end with the CCA bonded to the amino acid) is removed. This is the end of the tRNA involved in peptide bond formation, which would be expected to reside in close proximity to the peptidyl transferase site.

Attempts to ascribe a particular function to *isolated* ribosomal RNA had always met with failure. It was presumed that, even if ribosomal RNAs did have specific functions, the presence of ribosomal proteins were most likely required to keep the rRNAs in their proper conformation. Considering that the ribosomal proteins and rRNAs have coevolved for billions of years, the two molecules would be expected to be interdependent. Despite this expectation, the catalytic power of isolated rRNA was finally demonstrated by Noller and coworkers in 1992.[11] Working with particularly stable ribosomes from *Thermus aquaticus*, a bacterium that lives at high temperatures, Noller treated preparations of the large ribosomal subunit with a protein-extracting detergent (SDS), a protein-degrading enzyme (proteinase K), and several rounds of phenol, a protein denaturant. Together, these agents removed at least 95 percent of the protein from the ribosomal subunit, leaving the rRNA behind. Most of the 5 percent of the protein that remained associated with the rRNA was demonstrated to consist of small fragments of the ribosomal proteins. Yet, despite the removal of nearly all of the protein, the rRNA retained 80 percent of the peptidyl transferase activity of the intact subunit. The catalytic activity was blocked by chloramphenicol and by treatment with ribonuclease. When the RNA was subjected to additional treatments to remove the small amount of remaining protein, the preparation lost its catalytic activity. Because it is very unlikely that the remaining protein had any catalytic importance, it is generally agreed that these experiments demonstrated that peptidyl transferase is a ribozyme.

This conclusion gained confirmation from subsequent X-ray crystallographic studies by Thomas Steitz, Peter Moore, and their colleagues at Yale University on the large ribosomal subunit of *Haloarcula marismortui*, an archaebacterium that lives in the Dead Sea. A model of the subunit is shown in the chapter-opening image on page 419. To identify the peptidyl transferase site within the large ribosomal subunit, these researchers soaked crystals of these subunits with CCdA–phosphate–puromycin, a substance that inhibits peptide bond formation by binding tightly to the peptidyl transferase active site. Determination of the structure of these subunits at atomic resolution reveals the location of the bound inhibitor, and thus the location of the peptidyl transferase site.[12] It was found in this study that the active-site inhibitor is bound within a cleft of the subunit that is surrounded entirely by conserved nucleotide residues of the 23S rRNA. In fact, there isn't a single amino acid side chain of any of the ribosomal proteins within about 18 Å of the site where a peptide bond is synthesized. It would seem hard to argue with the conclusion that the ribosome is a ribozyme. See Ref. 13 for a review of these subjects.

References

1. CECH, T. R., ZAUG, A. J., & GRABOWSKI, P. J. 1981. In vitro splicing of the ribosomal RNA precursor of *Tetrahymena*. *Cell* 27:487–496.
2. KRUGER, K. ET AL. 1982. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31:147–157.
3. GUERRIER-TOKADA, C. ET AL. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35:849–857.
4. GUERRIER-TOKADA, C. ET AL. 1984. Catalytic activity of an RNA molecule prepared by transcription in vitro. *Science* 223:285–286.
5. BOWMAN, C. M. ET AL. 1971. Specific inactivation of 16S ribosomal RNA produced by colicin E3 in vivo. *Proc. Nat'l. Acad. Sci. USA.* 68:964–968.
6. CERNA, J., RYCHLIK, I., & JONAK, J. 1975. Peptidyl transferase activity of *Escherichia coli* ribosomes digested by ribonuclease T$_1$. *Eur. J. Biochem.* 34:551–556.
7. KEARSEY, S. & CRAIG, I. W. 1981. Altered ribosomal RNA genes in mitochondria from mammalian cells with chloramphenicol resistance. *Nature* 290:607–608.
8. WOESE, C. R. ET AL. 1975. Conservation of primary structure in 16S ribosomal RNA. *Nature* 254:83–86.
9. NOLLER, H. F. & WOESE, C. R. 1981. Secondary structure of 16S ribosomal RNA. *Science* 212:403–411.
10. MOAZED, D. & NOLLER, H. F. 1989. Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell* 57:585–597.
11. NOLLER, H. F., HOFFARTH, V., & ZIMNIAK, L. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 256:1416–1419.
12. NISSEN, P. ET AL. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science* 289:920–930.
13. CECH, T. R. 2009. Crawling out of the RNA world. *Cell* 136:599–602.

SYNOPSIS

Our understanding of the relationship between genes and proteins came as the result of a number of key observations. The first important observation was made by Garrod, who concluded that persons suffering from inherited metabolic diseases were missing specific enzymes. Later, Beadle and Tatum induced mutations in the genes of *Neurospora* and identified the specific metabolic reactions that were affected. These studies led to the concept of "one gene–one enzyme" and subsequently to the more refined version of "one gene–one polypeptide chain." The molecular consequence of a mutation was first described by Ingram, who showed that the inherited disease of sickle cell anemia results from a single amino acid substitution in a globin chain. (p. 420)

The first step in gene expression is the transcription of a DNA template strand by an RNA polymerase. Polymerase molecules are directed to the proper site on the DNA by binding to a promoter region, which in almost every case lies just upstream from the site at which transcription is initiated. The polymerase moves in the 3′ to 5′ direction along the template DNA strand, assembling a complementary, antiparallel strand of RNA that grows from its 5′ terminus in a 3′ direction. At each step along the way, the enzyme catalyzes a reaction in which ribonucleoside triphosphates (NTPs) are hydrolyzed into nucleoside monophosphates as they are incorporated. The reaction is further driven by hydrolysis of the released pyrophosphate. Prokaryotes possess a single type of RNA polymerase that can associate with a variety of different sigma factors that determine which genes are transcribed. The site at which transcription is initiated is determined by a nucleotide sequence located approximately 10 bases upstream from the initiation site. (p. 422)

Eukaryotic cells have three distinct RNA polymerases (I, II, and III), each responsible for the synthesis of a different group of

RNAs. Approximately 80 percent of a cell's RNA consists of ribosomal RNA (rRNA). Ribosomal RNAs (with the exception of the 5S species) are synthesized by RNA polymerase I, transfer RNAs and 5S rRNA by RNA polymerase III, and mRNAs by RNA polymerase II. All three types of RNA are derived from primary transcripts that are longer than the final RNA product. RNA processing requires a variety of small nuclear RNAs (snRNAs). (p. 426)

Three of the four eukaryotic rRNAs (the 5.8S, 18S, and 28S rRNAs) are synthesized from a single transcription unit, consisting of DNA (rDNA) that is localized within the nucleolus, and are processed by a series of nucleolar reactions. Nucleoli from amphibian oocytes can be dispersed to reveal actively transcribed rDNA, which takes the form of a chain of "Christmas trees." Each of the trees is a transcription unit whose smaller branches represent shorter RNAs that are at an earlier stage of transcription, that is, closer to the site where RNA synthesis was initiated. Analysis of these complexes reveals the tandem arrangement of the rRNA genes, the presence of nontranscribed spacers separating the transcription units, and the presence of associated ribonucleoprotein (RNP) particles that are involved in processing the transcripts. The steps in the processing of rRNA have been studied by exposing cultured mammalian cells to labeled precursors, such as [^{14}C]methionine, whose methyl groups are transferred to a number of nucleotides of the pre-rRNA. The presence of the methyl groups is thought to protect certain sites on the RNA from cleavage by nucleases and to aid in the folding of the RNA molecule. Approximately half of the 45S primary transcript is removed during the course of formation of the three mature rRNA products. The nucleolus is also the site of assembly of the two ribosomal subunits. (p. 428)

Kinetic studies of rapidly labeled RNAs first suggested that mRNAs were derived from much larger precursors. When eukaryotic cells are incubated for one to a few minutes in [^{3}H]uridine or other labeled RNA precursors, most of the label is incorporated into a group of RNA molecules of very large molecular weight and diverse nucleotide sequence that are restricted to the nucleus. These RNAs are referred to as heterogeneous nuclear RNAs (or hnRNAs). When cells that have been incubated briefly with [^{3}H]uridine are chased in a medium containing unlabeled precursors for an hour or more, radioactivity appears in smaller cytoplasmic mRNAs. This and other evidence, such as the presence of 5′ caps and 3′ poly(A) tails on both hnRNAs and mRNAs, led to the conclusion that hnRNAs were the precursors of mRNAs. (p. 434)

Pre-mRNAs are synthesized by RNA polymerase II in conjunction with a number of general transcription factors that allow the polymerase to recognize the proper DNA sites and to initiate transcription at the proper nucleotide. In many genes, the promoter lies between 24 and 32 bases upstream from the site of initiation in a region containing the TATA box. The TATA box is recognized by the TATA-binding protein (TBP), whose binding to the DNA initiates the assembly of a preinitiation complex. Phosphorylation of a portion of the RNA polymerase leads to the disengagement of the polymerase and the initiation of transcription. (p. 435)

One of the most important revisions in our concept of the gene came in the late 1970s with the discovery that the coding regions of a gene did not form a continuous sequence of nucleotides. The first observations in this regard were made on studies of transcription of the adenovirus genome in which it was found that the terminal portion of a number of different messenger RNAs are composed of the same sequence of nucleotides that is encoded by several discontinuous segments in the DNA. The regions between the coding segments are called intervening sequences, or introns. A similar condition was soon found to exist for cellular genes, such as those that code for β-globin and ovalbumin. In each case, the regions of DNA that encode portions of the polypeptide (the exons) are separated from one another by noncoding regions (the introns). Subsequent analysis indicated that the entire gene is transcribed into a primary transcript. The regions corresponding to the introns are subsequently excised from the pre-mRNA, and adjacent coding segments are ligated together. This process of excision and ligation is called RNA splicing. (p. 437)

The major steps in the processing of primary transcripts into mRNA include the addition of a 5′ cap, formation of a 3′ end, addition of a 3′ poly(A) tail, and removal of introns. The formation of the 5′ cap occurs by stepwise reactions in which the terminal phosphate is removed, a GMP is added in an inverted orientation, and methyl groups are transferred to both the added guanosine and the first nucleotide of the transcript itself. The final 3′ end of the mRNA is generated by cleavage of the primary transcript at a site just downstream from an AAUAAA recognition site and the addition of adenosine residues one at a time by poly(A) polymerase. Removal of the introns from the primary transcript depends on the presence of invariant residues at both the 5′ and 3′ splice sites on either side of each intron. Splicing is accomplished by a multicomponent spliceosome that contains a variety of proteins and ribonucleoprotein particles (snRNPs) that assemble in a stepwise fashion at the site of intron removal. Studies suggest that the snRNAs of the spliceosomes, possibly in conjunction with proteins, are the catalytically active components of the snRNPs. One of the apparent evolutionary benefits of split genes is the ease with which exons can be shuffled within the genome, generating new genes from portions of preexisting ones. (p. 440)

Most eukaryotic cells have a mechanism called RNA interference induced by double-stranded RNAs (siRNAs) that leads to the destruction of complementary mRNAs. RNA interference is thought to have evolved as a defense against viral infection and/or transposon mobility. Researchers have taken advantage of the phenomenon to stop the synthesis of specific proteins by targeting their mRNAs. Eukaryotic genomes encode large numbers of small microRNAs (miRNAs) (20–25 nucleotides) that regulate translation of specific mRNAs. Both siRNAs and miRNAs are generated by a common processing machinery that includes the enzyme Dicer that cleaves the precursor and an effector complex RISC, which holds the single-stranded guide RNA that either cuts the mRNA or blocks its translation. A third class of small regulatory RNAs, called piRNAs, are formed from single-stranded RNA precursors and do not require Dicer during biogenesis. piRNAs are active in suppressing transposable element mobility in germ cells. (p. 448)

Information for the incorporation of amino acids into a polypeptide is encoded in the sequence of triplet codons in the mRNA. In addition to being triplet, the genetic code is nonoverlapping and degenerate. In a nonoverlapping code, each nucleotide is part of one, and only one, codon, and the ribosome must move along the message three nucleotides at a time. The ribosome attaches to the mRNA at the initiation codon, AUG, which automatically puts the ribosome in the proper reading frame so that it correctly reads the entire message. A triplet code constructed of 4 different nucleotides can have 64 (4^3) different codons. The code is degenerate because many of the 20 different amino acids have more than 1 codon. Of the 64 possible codons, 61 specify an amino acid, while the other 3 are stop codons that cause the ribosome to terminate translation. The codon assignments are essentially universal, and their sequences are such that base substitutions in the mRNA tend to minimize the effect on the polypeptide. (p. 455)

Information in the nucleotide alphabet of DNA and RNA is decoded by transfer RNAs during the process of translation. Transfer RNAs are small RNAs (73 to 93 nucleotides long) that share a similar, L-shaped, three-dimensional structure and a number of invariant residues. One end of the tRNA bears the amino acid, and the other end contains a three-nucleotide anticodon sequence that is complementary to the triplet codon of the mRNA. The steric requirements of complementarity between codon and anticodon are relaxed at the third position of the codon to allow different codons that specify the same amino acid to use the same tRNA. It is essential that each tRNA is linked to the proper (cognate) amino acid, that is, the amino acid specified by the mRNA codon to which the tRNA anticodon binds. Linkage of tRNAs to their cognate amino acids is accomplished by a group of enzymes called aminoacyl-tRNA synthetases. Each enzyme is specific for one of the 20 amino acids and is able to recognize all of the tRNAs to which that amino acid must be linked. (p. 457)

Protein synthesis is a complex synthetic activity involving all the various tRNAs with their attached amino acids, ribosomes, messenger RNA, a number of proteins, cations, and GTP. The process is divided into three distinct activities: initiation, elongation, and termination. The primary activities of initiation include the precise attachment of the small ribosomal subunit to the initiation codon of the mRNA, which sets the reading frame for the entire translational process; the entry into the ribosome of a special initiator tRNA; and the assembly of the translational machinery. During elongation, a cycle of tRNA entry, peptide bond formation, and tRNA exit repeats itself with every amino acid incorporated. Aminoacyl-tRNAs enter the A site, where they bind to the complementary codon of the mRNA. After tRNA enters, the nascent polypeptide attached to the tRNA of the P site is transferred to the amino acid on the tRNA of the A site, forming a peptide bond. Peptide bond formation is catalyzed by a portion of the large rRNA acting as a ribozyme. In the last step of elongation, the ribosome translocates to the next codon of the mRNA, as the deacylated tRNA of the P site is transferred to the E site, and the deacylated tRNA that was in the E site is released from the ribosome. Both initiation and elongation require GTP hydrolysis. Translation is terminated when the ribosome reaches one of the three stop codons. After a ribosome assembles at the initiation codon and moves a short distance toward the 3′ end of the mRNA, another ribosome generally attaches to the initiation codon so that each mRNA is translated simultaneously by a number of ribosomes, which greatly increases the rate of protein synthesis within the cell. The complex of an mRNA and its associated ribosomes is a polyribosome. (p. 461)

ANALYTIC QUESTIONS

1. Look at the codon chart of Figure 11.41; which codons would you expect to have a unique tRNA, that is, one that is used only for that codon? Why is it that many codons do not have their own unique tRNA?

2. Proflavin is a compound that inserts itself into DNA and causes frameshift mutations (page 466). How would the effect on the amino acid sequence of a proflavin-induced mutation differ between an overlapping and a nonoverlapping code?

3. You have just isolated a new drug that has only one effect on cell metabolism; it totally inhibits the breakdown of pre-rRNA to ribosomal RNA. After treating a culture of mammalian cells with this drug, you give the cells [³H]uridine for 2 minutes, and then grow the cells in the presence of the drug in unlabeled medium for 4 hours before extracting the RNA and centrifuging it through a sucrose gradient. Draw the curves you would obtain by plotting both absorbance at 260 nm and radioactivity against the fraction number of the gradient. Label the abscissa (*X*-axis) using S values of RNA.

4. Using the same axes as in the previous question, draw the profile of radioactive RNA that you would obtain after a culture of mammalian cells had been incubated for 48 hours in [³H]uridine without any inhibiting drugs present.

5. Suppose you were to construct a synthetic RNA from a repeating dinucleotide (e.g., AGAGAGAG) and then use this RNA as a messenger to synthesize a polypeptide in an in vitro protein-synthesizing system, such as that used by Nirenberg and Matthaei to produce polyphenylalanine. What type of polypeptide would you make from this particular polynucleotide? Would you expect to have more than one type of polypeptide produced? Why or why not?

6. Suppose that you found an enzyme that incorporated nucleotides randomly into a polymer without the requirement for a template. How many different codons would you be able to find in synthetic RNAs made using two different nucleotide precursors (e.g., CTP and ATP)? (An enzyme called polynucleotide phosphorylase catalyzes this type of reaction and was used in studies that identified codons.)

7. Draw the parts of a 15S globin pre-mRNA, labeling the noncoding portions.

8. What is the minimum number of GTPs that you would need to synthesize a pentapeptide?

9. On the same set of axes as in Figure 10.17, draw two reannealing curves, one for DNA extracted from *Xenopus* brain tissue and the other for DNA extracted from *Xenopus* oocytes.

10. Would you agree with the following statement? The discovery that sickle cell anemia resulted from a single amino acid change was proof that the genetic code was nonoverlapping. Why or why not?

11. Thalassemia is a disease characterized by mutations that convert amino acid codons into stop codons. Suppose you were to compare the polypeptides synthesized in vitro from mRNAs purified from a wide variety of thalassemia patients. How would you expect these polypeptides to compare? Refer to the codon chart of Figure 11.41; how many amino acid codons can be converted into stop codons by a single base substitution?

12. Do you think it would be theoretically possible to have a genetic code with only two letters, A and T? If so, what would be the minimum number of nucleotides required to make a codon?

13. On page 448, experiments are reported that lead to the synthesis of ribozymes with unique catalytic properties. In 2001, an artificial ribozyme was isolated that was capable of incorporating up to 14 ribonucleotides onto the end of an existing RNA using an RNA strand as a template. The ribozyme could use any RNA

sequence as a template and would incorporate complementary nucleotides into the newly synthesized RNA strand with an accuracy of 98.5 percent. If you were a proponent of an ancient RNA world, how would you use this finding to argue your case? Would this prove the existence of an ancient RNA world? If not, what do you think would provide the strongest evidence for such a world?

14. How is it possible that mRNA synthesis occurs at a greater rate in bacterial cells than any other class, yet very little mRNA is present within the cell?

15. If a codon for serine is 5′-AGC-3′, the anticodon for this triplet would be 5′-— — —-3′. How would the wobble phenomenon affect this codon–anticodon interaction?

16. One of the main arguments that proteins evolved before DNA (i.e., that the RNA world evolved into an RNA–protein world rather than an RNA–DNA world) is based on the fact that the translational machinery involves a large variety of RNAs (e.g., tRNAs, rRNAs), whereas the transcriptional machinery shows no evidence of RNA involvement. Can you explain how such an argument about the stages of early evolution might be based on these observations?

17. Frameshift mutations and nonsense mutations were described on page 466. It was noted that nonsense mutations often lead to the destruction by NMD of an mRNA containing a premature termination codon. Would you expect mRNAs containing frameshift mutations to be subject to NMD?

18. The arrowheads in Figure 11.16 indicate the direction of transcription of the various tRNA genes. What does this drawing tell you about the template activity of each strand of a DNA molecule within a chromosome?

19. Genes are usually discovered by finding an abnormal phenotype resulting from a genetic mutation. Alternatively, they can often be identified by examining the DNA sequence of a genome. Why do you suppose that the genes encoding miRNAs were not discovered until very recently?

20. It was stated on page 456 that synonymous codon changes do not *generally* alter an organism's phenotype. Can you think of an occasion where this might not be true? What does this say about the coding requirements of the genetic material?

12
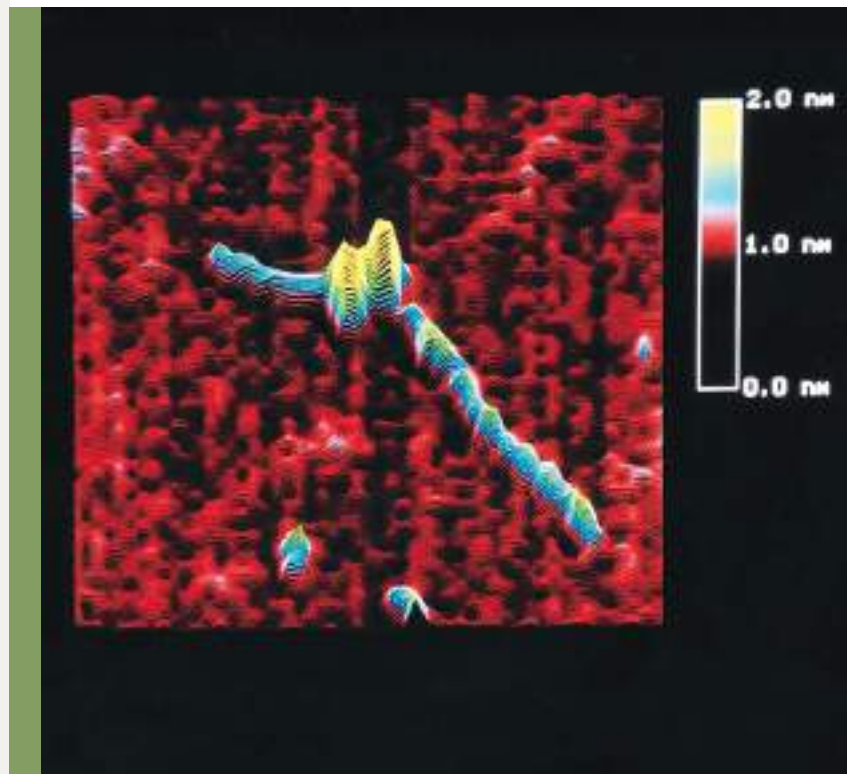
The Cell Nucleus and
the Control of Gene Expression

I n spite of their obvious differences in form and function, the cells that make up a multicellular organism contain a complete set of genes. The genetic information present in a specialized eukaryotic cell can be compared to a book of blueprints prepared for the construction of a large, multipurpose building. During the construction process, all of the blueprints will probably be needed, but only a small subset of this information has to be consulted during the work on a particular floor or room. The same is true for a fertilized egg, which contains a complete set of genetic instructions that is faithfully replicated and distributed to each cell of a developing organism. As a result, differentiated cells carry much more genetic information than they will ever have to use. Consequently, cells possess mechanisms that allow them to express their genetic information selectively, following only those instructions that are relevant for that cell at that particular time. In this chapter, we will explore some of the ways in which cells control gene expression and thereby ensure that certain proteins are synthesized, while others are not. First, however, we will begin by describing the structure and properties of the eukaryotic cell nucleus in which most of the regulatory machinery is housed. ■

Scanning force micrograph of a bacterial DNA molecule (shown in blue) with a regulatory protein (called NtrC, shown in yellowish tones) bound at a site that is upstream from a gene that encodes glutamine synthetase. Phosphorylation of NtrC leads to activation of transcription of the regulated gene. A scanning force microscope measures the height of various parts of the specimen at the atomic level, which is converted into the colors shown in the key on the right. (FROM I. ROMBEL ET AL., COURTESY OF S. KUSTU, UNIVERSITY OF CALIFORNIA, BERKELEY, COLD SPRING HARBOR SYMP. QUANT. BIOL. 63:160, 1998.)

475

12.1 THE NUCLEUS OF A EUKARYOTIC CELL

Considering its importance in the storage and utilization of genetic information, the nucleus of a eukaryotic cell has a rather undistinguished morphology (Figure 12.1). The contents of the nucleus are present as a viscous, amorphous mass of material enclosed by a complex *nuclear envelope* that forms a boundary between the nucleus and cytoplasm. Included within the nucleus of a typical interphase (i.e., nonmitotic) cell are (1) the chromosomes, which are present as highly extended nucleoprotein fibers, termed *chromatin*; (2) one or more *nucleoli*, which are irregularly shaped electron-dense structures that function in the synthesis of ribosomal RNA and the assembly of ribosomes (discussed on page 428); (3) the *nucleoplasm*, the fluid substance in which the solutes of the nucleus are dissolved; and (4) the *nuclear matrix*, which is a protein-containing fibrillar network.

The Nuclear Envelope

The separation of a cell's genetic material from the surrounding cytoplasm may be the single most important feature that distinguishes eukaryotes from prokaryotes, which makes the appearance of the nuclear envelope a landmark in biological evolution. The **nuclear envelope** consists of two cellular membranes arranged parallel to one another and separated by 10 to 50 nm (Figure 12.2*a*). The membranes of the nuclear envelope serve as a barrier that keeps ions, solutes, and macromolecules from passing freely between the nucleus and cytoplasm. The two membranes are fused at sites forming circular pores that contain complex assemblies of proteins. The average mammalian cell contains several thousand nuclear pores.
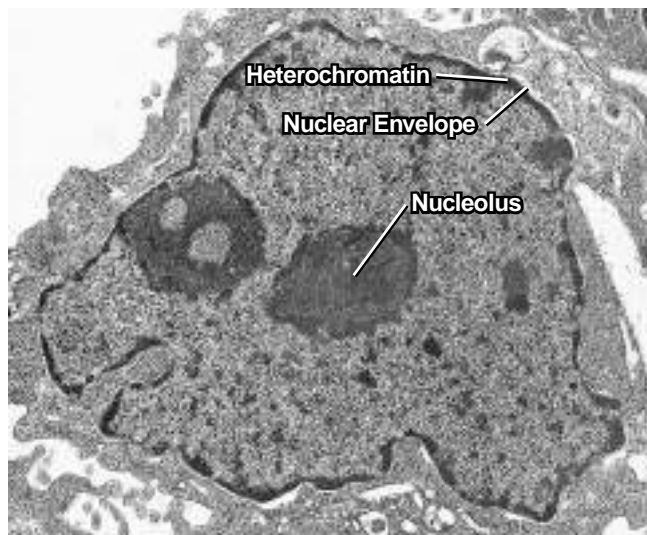
(a)   1 μm
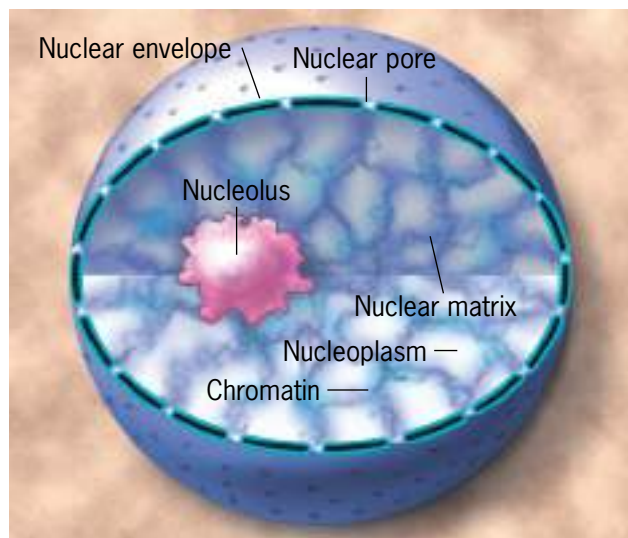
(b)

FIGURE 12.1 **The cell nucleus.** (*a*) Electron micrograph of an interphase HeLa cell nucleus. Heterochromatin (page 485) is evident around the entire inner surface of the nuclear envelope. Two prominent nucleoli are visible, and clumps of chromatin can be seen scattered throughout the nucleoplasm. (*b*) Schematic drawing showing some of the major components of the nucleus. (A: FROM WERNER W. FRANKE, INT. REV. CYTOL. (SUPPL.) 4:130, 1974.)
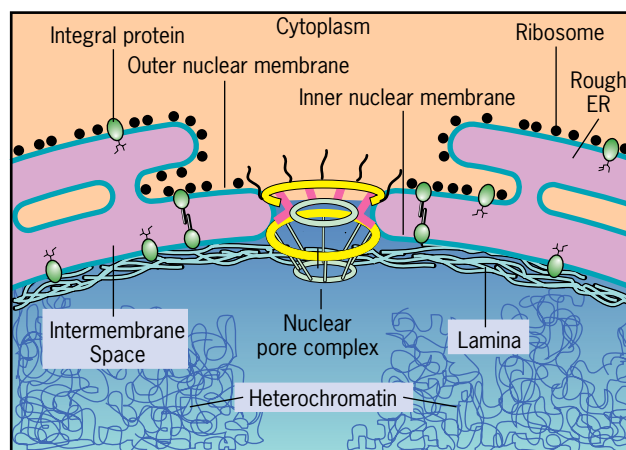
(a)
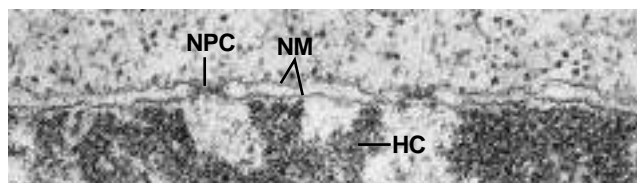
(b)   0.2 μm

FIGURE 12.2 **The nuclear envelope.** (*a*) Schematic drawing showing the double membrane, nuclear pore complex, nuclear lamina, and the continuity of the outer membrane with the rough endoplasmic reticulum (ER). Both membranes of the nuclear envelope contain their own distinct complement of proteins. (*b*) Electron micrograph of a section through a portion of the nuclear envelope of an onion root tip cell. Note the double membrane (NM) with intervening space, nuclear pore complexes (NPC), and associated heterochromatin (HC) that does not extend into the region of the nuclear pores. (B: FROM WERNER W. FRANKE ET AL. J. CELL BIOL. 91:47S, 1981; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

The outer membrane is generally studded with ribosomes and is continuous with the membrane of the rough endoplasmic reticulum. The space between the membranes is continuous with the ER lumen (Figure 12.2*a*).

The inner surface of the nuclear envelope of animal cells is bound by integral membrane proteins to a thin filamentous meshwork, called the **nuclear lamina** (Figure 12.3). The nuclear lamina provides mechanical support to the nuclear envelope, serves as a site of attachment for chromatin fibers at the nuclear periphery (Figure 12.2*b*), and has a poorly understood role in DNA replication and transcription. The filaments of the nuclear lamina are approximately 10 nm in diameter and composed of polypeptides, called *lamins*. Lamins are members of the same superfamily of polypeptides that assemble into the 10-nm intermediate filaments of the cytoplasm (see Table 9.2). As in the cytoplasm, the integrity of the intermediate filaments that make up nuclear lamina is regulated by phosphorylation and dephosphorylation. The disassembly of the nuclear lamina prior to mitosis is induced by phosphorylation of the lamins by a specific protein kinase.

Mutations in one of the lamin genes (*LMNA*) are responsible for a number of diverse human diseases, including a rare form of muscular dystrophy (called EDMD2) in which muscle cells contain exceptionally fragile nuclei. Mutations in *LMNA* have also been linked to a disease, called Hutchinson-Gilford progeria syndrome (HGPS), that is characterized by premature aging and death during teenage years from heart attack or stroke. Figure 12.3*c* shows the misshapen nuclei

from the cells of a patient with HGPS, demonstrating the importance of the nuclear lamina as a determinant of nuclear architecture. It is interesting to note that the phenotype depicted in Figure 12.3*c* has been traced to a synonymous mutation, that is, one that generated a different codon for the same amino acid. In this case, the change in DNA sequence altered the way the gene transcript was spliced, which led to production of a shortened protein, causing the altered phenotype. This example illustrates how the sequence of a gene serves as a "multiple code:" one that directs the translation machinery and others that direct the splicing machinery and protein folding.

The Structure of the Nuclear Pore Complex and Its Role in Nucleocytoplasmic Exchange   The nuclear envelope is the barrier between the nucleus and cytoplasm, and nuclear pores are the gateways across that barrier. Unlike the plasma membrane, which prevents passage of macromolecules between the cytoplasm and the extracellular space, the nuclear envelope is a hub of activity for the movement of RNAs and proteins in both directions between the nucleus and cytoplasm. The replication and transcription of genetic material within the nucleus require the participation of large numbers of proteins that are synthesized in the cytoplasm and transported across the nuclear envelope. Conversely, the mRNAs, tRNAs, and ribosomal subunits that are manufactured in the nucleus must be transported through the nuclear envelope in the opposite direction. Some components, such as the snRNAs of the

(a)   3 µm   *(b)*   3 µm   *(c)*
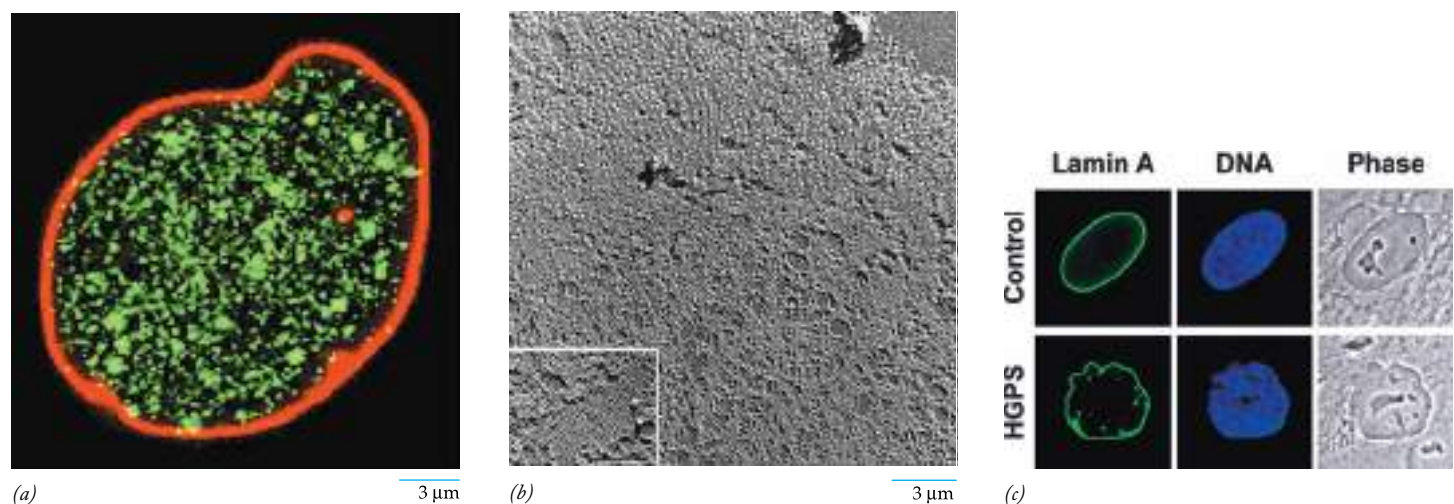
FIGURE 12.3 The nuclear lamina. (*a*) Nucleus of a cultured human cell that has been stained with fluorescently labeled antibodies to reveal the nuclear lamina (red), which lies on the inner surface of the nuclear envelope. The nuclear matrix (page 499) is stained green. (*b*) Electron micrograph of a freeze-dried, metal-shadowed nuclear envelope of a *Xenopus* oocyte that has been extracted with the nonionic detergent Triton X-100. The lamina appears as a rather continuous meshwork comprising filaments oriented roughly perpendicular to one another. Inset shows a well-preserved area from which nuclear pores have been mechanically removed. (*c*) These micrographs show the nucleus within a fibroblast that had been cultured from either a patient with HGPS (bottom row)

or a healthy subject (top row). The cells are stained for the protein lamin A (left column), for DNA (middle column), or shown in a living state under the phase-contrast light microscope (right column). The cell nucleus from the HGPS patient is misshapen due to the presence in the nuclear lamina of a truncated lamin A protein. (A: FROM H. MA, A. J. SIEGEL, AND R. BEREZNEY, J. CELL BIOL. 146:535, 1999; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS; B: REPRINTED WITH PERMISSION FROM U. AEBI, J. COHN, L. BUHLE, AND L. GERACE, NATURE 323:561, 1986; © COPYRIGHT 1986, MACMILLAN MAGAZINES LIMITED C; FROM ANNA MATTOUT ET AL., COURTESY OF ROBERT D. GOLDMAN, CURR. OPIN. CELL BIOL. 18:338, 2006.)

spliceosome (page 444), move in both directions; they are synthesized in the nucleus, assembled into RNP particles in the cytoplasm, and then shipped back to the nucleus where they function in mRNA processing. To appreciate the magnitude of the traffic between the two major cellular compartments, consider a HeLa cell, which is estimated to contain about 10,000,000 ribosomes. To support its growth, a single HeLa cell nucleus must import approximately 560,000 ribosomal proteins and export approximately 14,000 ribosomal subunits every minute.

How do all of these materials pass through the nuclear envelope? In one early approach, a suspension of tiny gold particles was injected into cells and passage of the material through the nuclear envelope was observed with the electron

(a)   *(b)*

(c)   0.1 μm

FIGURE 12.4 The movement of materials through the nuclear pore.
(*a*) Electron micrograph of the nuclear–cytoplasmic border of a frog oocyte taken minutes after injection with gold particles that had been coated with a protein normally found in the nucleus. These particles are seen to pass through the center of the nuclear pores (arrows) on their way from the cytoplasm to the nucleus. (*b*) At higher magnification, the gold particles are seen to be clustered in linear array within each pore. (*c*) Electron micrograph of a section through the nuclear envelope of an insect cell showing the movement of granular material (presumed to be a ribosomal subunit) through a nuclear pore. (A,B: COURTESY OF C. M. FELDHERR; C: FROM BARBARA J. STEVENS AND HEWSON SWIFT, J. CELL BIOL. 31:72, 1966; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

microscope. As illustrated in Figure 12.4*a,b*, these particles move from the cytoplasm into the nucleus by passing single-file through the center of the nuclear pores. Electron micrographs of cells fixed in the normal course of their activities have also shown that particulate material can pass through a nuclear pore. An example is shown in Figure 12.4*c*, in which granular material presumed to consist of a ribosomal subunit is seen squeezing through one of these pores.

Given the fact that materials as large as gold particles and ribosomal subunits can penetrate nuclear pores, one might assume that these pores are merely open channels, but just the opposite is true. Nuclear pores contain an intricate structure called the **nuclear pore complex** (**NPC**) that appears to fill the pore like a stopper, projecting into both the cytoplasm and nucleoplasm. The structure of the NPC is seen in the electron micrographs of Figure 12.5 and the model of Figure 12.6. The NPC is a huge, supramolecular complex—15 to 30 times the mass of a ribosome—that exhibits octagonal symmetry due to the eightfold repetition of a number of structures (Figure 12.6). Despite their considerable size and complexity, NPCs contain only about 30 different proteins, called *nucleoporins*, which are largely conserved between yeast and vertebrates. Each nucleoporin is present in at least eight copies, in keeping with the octagonal symmetry of the structure. The NPC is not a static structure, as evidenced by the finding that many of its component proteins are replaced with new copies over a time period of seconds to minutes.

Among the nucleoporins is a subset of proteins that possess, within their amino acid sequence, a large number of phenylalanine-glycine repeats (FG, by their single letter names). The FG repeats are clustered in a particular region of each molecule called the FG domain. Because of their unusual amino acid composition, the FG domains possess a disordered structure (page 56) that gives them an extended and flexible organization. The FG repeat-containing nucleoporins are thought to line the central channel of the NPC with their filamentous FG domains extending into the heart of the 20- to 30-nm-wide channel. The FG domains form a hydrophobic meshwork or sieve that blocks the diffusion of larger macromolecules (greater than about 40,000 Daltons) between the nucleus and cytoplasm.

In 1982, Robert Laskey and his co-workers at the Medical Research Council of England found that nucleoplasmin, one of the more abundant nuclear proteins of amphibian oocytes, contains a stretch of amino acids near its C-terminus that functions as a **nuclear localization signal** (**NLS**). This sequence enables a protein to pass through the nuclear pores and enter the nucleus. The best studied, or "classical" NLSs, consist of one or two short stretches of positively charged amino acids. The T antigen encoded by the virus SV40, for example, contains an NLS identified as -Pro-Lys-Lys-Lys-Arg-Lys-Val-. If one of the basic amino acids in this sequence is replaced by a nonpolar amino acid, the protein fails to become localized in the nucleus. Conversely, if this NLS is fused to a nonnuclear protein, such as serum albumin, and injected into the cytoplasm, the modified protein becomes concentrated in the nucleus. Thus, targeting of proteins to the nu-

(a)

0.2 μm

(b)

NEL

(c)

0.6 μm

FIGURE 12.5 **Scanning electron micrographs of the nuclear pore complex from isolated nuclear envelopes of an amphibian oocyte.** (*a*) The cytoplasmic face of the nuclear envelope showing the peripheral cytoplasmic ring of the nuclear pore complex. (*b*) The nuclear face of the nuclear envelope showing the basket-like appearance of the inner portion of the complex. (*c*) The nuclear face of the envelope showing the distribution of the NPCs and places where intact patches of the nuclear lamina (NEL) are retained. In all of these micrographs, isolated nuclear envelopes were fixed, dehydrated, dried, and metal-coated. (FROM M. W. GOLDBERG AND T. D. ALLEN, J. CELL BIOL. 119:1431, 1992; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)
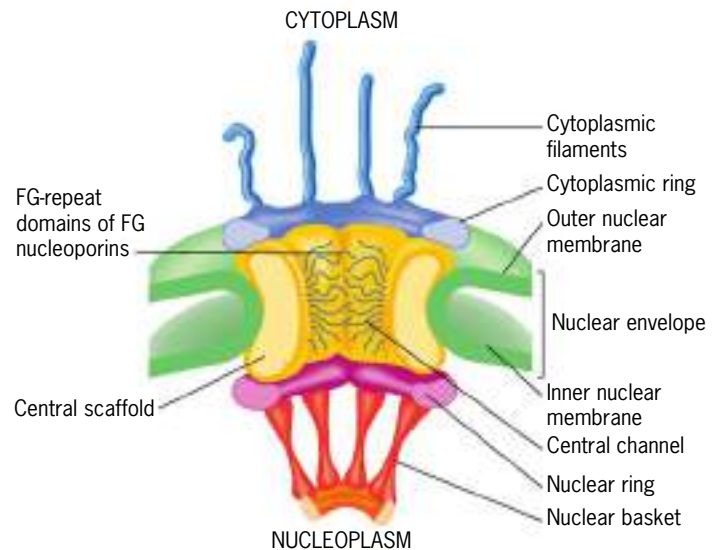
FIGURE 12.6 **A model of a vertebrate nuclear pore complex (NPC).** Three-dimensional representation of a vertebrate NPC as it is situated within the nuclear envelope. This elaborate structure consists of several parts, including a scaffold that anchors the complex to the nuclear envelope, a cytoplasmic and a nuclear ring, a nuclear basket, and eight cytoplasmic filaments. The FG-containing nucleoporins line the channel with their disordered FG-containing domains extending into the opening and forming a hydrophobic meshwork.

cleus is similar in principle to trafficking of other proteins that are destined for segregation within a particular organelle, such as a mitochondrion or a peroxisome (page 309). In all of these cases, the proteins possess a specific "address" that is recognized by a specific receptor that mediates its transport into the organelle.

The study of nuclear transport has been a very active area of research, driven by the development of in vitro systems capable of selectively importing proteins and RNPs into the nucleus. Using these systems, researchers can identify which proteins are required for the nuclear import of a particular macromolecule. These efforts have identified a family of proteins that function as mobile *transport receptors*, ferrying macromolecules across the nuclear envelope. Within this family, *importins* move macromolecules from the cytoplasm into the nucleus and *exportins* move macromolecules in the opposite direction.

Figure 12.7*a* depicts some of the major steps that occur during the nuclear import of a protein, such as nucleoplasmin, that contains a classical NLS. Import begins as the NLS-containing cargo protein binds to a heterodimeric, soluble NLS receptor, called *importin* α/β, that resides in the cytoplasm (step 1, Figure 12.7*a*). The transport receptor is thought to escort the protein cargo to the outer surface of the nucleus where it likely docks with the cytoplasmic filaments that extend from the outer ring of the NPC (step 2). Figure 12.7*b* shows a number of gold particles bound to these filaments; these particles were coated with an NLS-containing nuclear protein that was being transported through the nuclear pore

(a)

(b)

0.4 μm

FIGURE 12.7 Importing proteins from the cytoplasm into the nucleus. (*a*) Proposed steps in nuclear protein import, as described in the text. The protein bearing a nuclear localization signal (NLS) binds to the heterodimeric receptor (importin α/β) (step 1) forming a complex that associates with a cytoplasmic filament (step 2). The receptor-cargo complex moves through the nuclear pore (step 3) and into the nucleoplasm where it interacts with Ran-GTP and dissociates (step 4). The importin β subunit, in association with Ran-GTP, is transported back to the cytoplasm, where the Ran-GTP is hydrolyzed (step 5). Ran-GDP is subsequently transported back to the nucleus, where it is converted to Ran-GTP. Conversely, importin α is transported back to the cytoplasm. (*b*) Nucleoplasmin is a protein present in high concentration in the nucleoplasm of *Xenopus* oocytes. When gold particles are coated with nucleoplasmin and injected into the cytoplasm of a *Xenopus* oocyte, they are seen to bind to the cytoplasmic filaments (CF) projecting from the outer ring of the nuclear pore complex. Several particles are also seen in transit through the pore (NP) into the nucleus. (A: BASED ON A MODEL BY M. OHNO ET AL., CELL 92:327, 1998; B: FROM W. D. RICHARDSON ET AL., CELL 52:662, 1988; BY PERMISSION OF CELL PRESS, COURTESY OF A. D. MILLS.)

complex. The receptor–cargo complex then moves through the nuclear pore (step 3, Figure 12.7*a*) by engaging in a series of successive interactions with the FG domains of the FG-containing nucleoporins. These interactions are thought to "dissolve" portions of the FG-rich meshwork that fills the interior of the channel, allowing passage of the receptor-cargo complex through the NPC.

Now that we've gotten the bound cargo through the NPC and into the nuclear compartment, we need to introduce another key player, a GTP-binding protein called **Ran**. Like other GTP-binding proteins, such as Sar1 (page 290) and EF-Tu (page 464) discussed in earlier chapters, Ran can exist in an active GTP-bound form or an inactive GDP-bound form. Ran's role in regulating nucleocytoplasmic transport is based on a mechanism in which the cell maintains a high concentration of Ran-GTP in the nucleus and a very low concentration of Ran-GTP in the cytoplasm. The steep gradient

of Ran-GTP across the nuclear envelope depends on the compartmentalization of certain accessory proteins (see Figure 15.19*b* for further discussion). One of these accessory proteins (named *RCC1*) is sequestered in the nucleus where it promotes the conversion of Ran-GDP to Ran-GTP, thus maintaining the high nuclear level of Ran-GTP. Another accessory protein (named RanGAP1) resides in the cytoplasm where it promotes the hydrolysis of Ran-GTP to Ran-GDP, thus maintaining the low cytoplasmic level of Ran-GTP. Thus the energy released by GTP hydrolysis is used to maintain the Ran-GTP gradient across the nuclear envelope. As discussed below, the Ran-GTP gradient drives nuclear transport by a process that depends only on receptor-mediated diffusion; no motor proteins or ATPases have been implicated.

We can now return to our description of the classical NLS import pathway. When the importin–cargo complex arrives in the nucleus, it is met by a molecule of Ran-GTP,

which binds to the complex and causes its disassembly as indicated in step 4, Figure 12.7*a*. This is the apparent function of the high level of Ran-GTP in the nucleus: it promotes the disassembly of complexes imported from the cytoplasm. The imported cargo is released into the nucleoplasm, and one portion of the NLS receptor (the importin β subunit) is shuttled back to the cytoplasm together with the bound Ran-GTP (step 5). Once in the cytoplasm, the GTP molecule bound to Ran is hydrolyzed, releasing Ran-GDP from the importin β subunit. Ran-GDP is returned to the nucleus, where it is converted back to the GTP-bound state for additional rounds of activity. Importin α is transported back to the cytoplasm by one of the exportins.

Ran-GTP plays a key role in the escort of macromolecules from the nucleus, just as it does in their import from the cytoplasm. Recall that Ran-GTP is essentially confined to the nucleus. Whereas Ran-GTP induces the disassembly of imported complexes, as shown in step 4 of Figure 12.7*a*, Ran-GTP promotes the *assembly* of exported complexes. Proteins exported from the nucleus contain amino acid sequences (called *nuclear export signals*, or *NESs*) that are recognized by transport receptors that carry them through the nuclear envelope to the cytoplasm. Most of the traffic moving in this direction consists of various types of RNA molecules—especially mRNAs, rRNAs, and tRNAs—that are synthesized in the nucleus and function in the cytoplasm. In most cases, these RNAs move through the NPC as ribonucleoproteins (RNPs).

Transport of an mRNP from the nucleus to cytoplasm is associated with extensive remodeling; certain proteins are stripped from the mRNA, while others are added to the complex. Transport of mRNPs does not appear to require Ran but does require the activity of an RNA helicase located on the cytoplasmic filaments of the NPC. It is speculated that the helicase provides the motive force to move the mRNA into the cytoplasm. Numerous studies have demonstrated a functional link between pre-mRNA splicing and mRNA export; only mature (i.e., fully processed) mRNAs are capable of nuclear export. If an mRNA still contains an unspliced intron, that RNA is retained in the nucleus.

Chromosomes and Chromatin

Chromosomes seem to appear out of nowhere at the beginning of mitosis and disappear once again when cell division has ended. The appearance and disappearance of chromosomes provided early cytologists with a challenging question: What is the nature of the chromosome in the nonmitotic cell? We are now able to provide a fairly comprehensive answer to this question.

Packaging the Genome   An average human cell contains about 6.4 billion base pairs of DNA divided among 46 chromosomes (the value for a diploid, unreplicated number of chromosomes). Each unreplicated chromosome contains a single, continuous DNA molecule; the larger the chromosome, the longer the DNA it contains. Given that each base pair is about 0.34 nm in length, 6 billion base pairs would constitute a DNA molecule fully 2 m long. How is it possible to fit 2 meters of DNA into a nucleus only 10 μm (1×10^{-5} m) in diameter and, at the same time, maintain the DNA in a state that is accessible to enzymes and regulatory proteins? Just as important, how is the single DNA molecule of each chromosome organized so that it does not become hopelessly tangled with the molecules of other chromosomes? The answers lie in the remarkable manner in which a DNA molecule is packaged.

Nucleosomes: The Lowest Level of Chromosome Organization Chromosomes are composed of DNA and associated protein, which together is called **chromatin**. The orderly packaging of eukaryotic DNA depends on **histones**, a remarkable group of small proteins that possess an unusually high content of the basic amino acids arginine and lysine. Histones are divided into five classes, which can be distinguished by their arginine/lysine ratio (Table 12.1). The amino acid sequences of histones, particularly H3 and H4, have undergone very little change over long periods of evolutionary time. The H4 histones of both peas and cows, for example, contain 102 amino acids, and their sequences differ at only 2 amino acid residues. Why are histones so highly conserved? One reason is histones interact with the backbone of the DNA molecule, which is identical in all organisms. In addition, nearly all of the amino acids in a histone molecule are engaged in an interaction with another molecule, either DNA or another histone. As a result, very few amino acids in a histone can be replaced with other amino acids without severely affecting the function of the protein.

In the early 1970s, it was found that when chromatin was treated with nonspecific nucleases, most of the DNA was converted to fragments of approximately 200 base pairs in length. In contrast, a similar treatment of *naked* DNA (i.e., DNA devoid of proteins) produced a randomly sized population of fragments. This finding suggested that chromosomal DNA was protected from enzymatic attack, except at certain periodic sites along its length. It was presumed that the proteins associated with the DNA were providing the protection. In 1974, using the data from nuclease digestion and other types of information, Roger Kornberg, then at Harvard University, proposed an entirely new structure for chromatin. Kornberg proposed that DNA and histones are organized into repeating subunits, called **nucleosomes**. We now know that each nucleo-

TABLE 12.1 **Calf Thymus Histones**

Histone	Number of residues	Mass (kDa)	%Arg	%Lys	UEP* ($\times 10^{-6}$ year)
H1	215	23.0	1	29	8
H2A	129	14.0	9	11	60
H2B	125	13.8	6	16	60
H3	135	15.3	13	10	330
H4	102	11.3	14	11	600

*Unit evolutionary period: the time for a protein's amino acid sequence to change by 1 percent after two species have diverged.

(a)

(b)
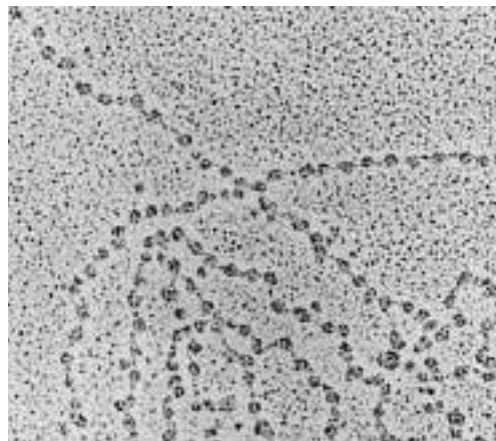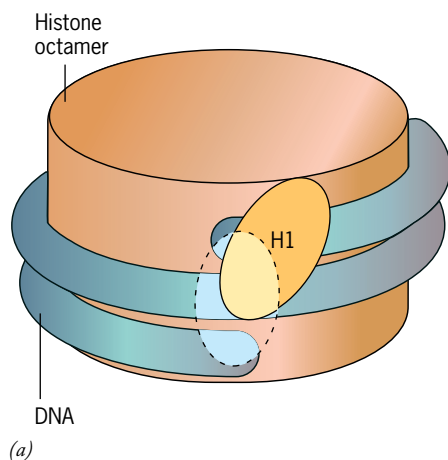
FIGURE 12.8 Nucleosomal organization of chromatin. (*a*) Schematic diagram showing the structure of a nucleosome core particle and an associated histone H1 molecule. The core particle itself consists of approximately 1.8 turns (146 base pairs) of negatively supercoiled DNA wrapped around eight core histone molecules (two each of H2A, H2B, H3, and H4). The H1 linker histone binds near the sites where DNA enters and exits the nucleosome. Two alternate positions of the H1 molecule are shown. (*b*) Electron micrograph of chromatin fibers released from the nucleus of a *Drosophila* cell in a buffer of low ionic strength. The nucleosome core particles are approximately 10 nm in diameter and are connected by short strands of naked linker DNA, which are approximately 2 nm in diameter. (B: COURTESY OF OSCAR L. MILLER, JR.)

some contains a *nucleosome core particle* consisting of 146 base pairs of supercoiled DNA (page 390) wrapped almost twice around a disk-shaped complex of eight histone molecules (Figure 12.8*a*). The histone core of each nucleosome consists of two copies each of histones H2A, H2B, H3, and H4 assembled into an octamer, as discussed below. The remaining histone—type H1—resides outside the nucleosome core particle. The H1 histone is referred to as a *linker histone* because it binds to part of the *linker DNA* that connects one nucleosome core particle to the next. Fluorescence studies indicate that H1 molecules continuously dissociate and reassociate with chromatin. Together the H1 protein and the histone octamer interact with about 168 base pairs of DNA. H1 histone molecules can be selectively removed from the chromatin fibers by subjecting the preparation to solutions of low ionic strength. When H1-depleted chromatin is observed under the electron microscope, the nucleosome core particles and naked linker DNA can be seen as separate elements, which together appear like "beads on a string" (Figure 12.8*b*).

Our understanding of DNA packaging has been greatly advanced in recent years by dramatic portraits of the nucleosome core particle obtained by X-ray crystallography (Figure 12.9). The eight histone molecules that comprise a nucleosome core particle are organized into four heterodimers: two H2A-H2B dimers and two H3-H4 dimers (Figure 12.9*a*,*c*). Dimerization of histone molecules is mediated by their C-terminal domains, which consist largely of α helices (represented by the cylinders in Figure 12.9*c*) folded into a compact mass in the core of the nucleosome. In contrast, the N-terminal segment of each core histone (and also the C-terminal segment of H2A) takes the form of a long, flexible tail (represented by the dashed lines of Figure 12.9*c*) that extends past the DNA helix and into the surroundings. These tails are targets of a variety of covalent modifications whose key functions will be explored later in the chapter.

Histone modification is not the only mechanism to alter the histone character of nucleosomes. In addition to the four "conventional" core histones discussed above, several alternate versions of the H2A and H3 histones are also synthesized in most cells. The importance of these histone variants, as they are called, remains largely unexplored, but they are thought to have specialized functions (Table 12.2). The localization and apparent function of one of these variants, CENP-A, is discussed on page 496. Another variant, H2A.X, is distributed throughout the chromatin, where it replaces conventional H2A in a fraction of the nucleosomes. H2A.X becomes phosphorylated at sites of DNA-strand breakage and may play a role in recruiting the enzymes that repair the DNA. Two other core histone variants—H2A.Z and H3.3—can be incorporated into nucleosomes of genes as they become activated and may play a role in promoting the transcription of that genetic locus.

TABLE 12.2 Histone Variants

Type	Variant	Location	Likely function
H2A			
	H2A.X	Throughout chromatin	DNA repair
	H2A.Z	Euchromatin	Transcription
	macroH2A	Inactive X chromosome	Transcriptional silencing
H3			
	CENP-A	Centromeres	Kinetochore assembly
	H3.3	Transcribed loci	Transcription

(a)   *(b)*   *(c)*
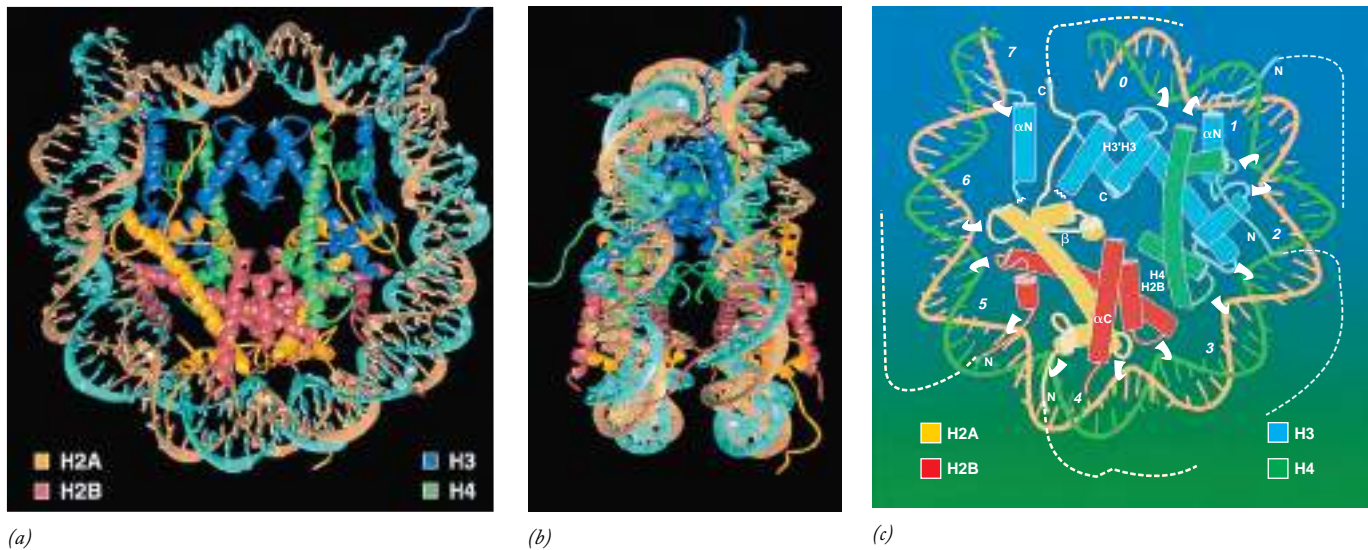
FIGURE 12.9 **The three-dimensional structure of a nucleosome as revealed by X-ray crystallography.** (*a*) A nucleosome core particle viewed down the central axis of the DNA superhelix, showing the position of each of the eight histone molecules of the core octamer. The histones are seen to be organized into four dimeric complexes. Each histone dimer binds 27 to 28 base pairs of DNA, with contacts occurring where the minor groove of the DNA faces the histone core. (*b*) The disk shape of the nucleosome core particle is evident in this view perpendicular to the central axis. The two H3-H4 dimers are associated with one another in the center of the core particle to form a tetramer, whereas the two H2A-H2B dimers are positioned on each side of the (H3-H4)$_2$ tetramer. (*c*) A simplified, schematic model of half of a nucleosome core particle, showing one turn (73 base pairs) of the DNA superhelix and four core histone molecules. The four different histones are shown in separate colors, as indicated by the key. Each core histone is seen to consist of (1) a globular region, called the "histone fold," consisting of three α helices (represented by the cylinders) and (2) a flexible, extended N-terminal tail (indicated by the letter N) that projects out of the histone disk and past the DNA double helix. The intermittent points of interaction between the histone molecules and the DNA are indicated by white hooks. The dashed lines indicate the outermost portion of the histone tails; these flexible tails lack a defined tertiary structure and therefore do not appear in the X-ray structures shown in *a* and *b*. (A,B: REPRINTED WITH PERMISSION FROM KAROLIN LUGER ET AL., NATURE 389:251, 1997; COURTESY OF TIMOTHY J. RICHMOND. C: AFTER DRAWING BY D. RHODES; © COPYRIGHT 1997, BY MACMILLAN MAGAZINES LIMITED.)

DNA and core histones are held together by several types of noncovalent bonds, including ionic bonds between negatively charged phosphates of the DNA backbone and positively charged residues of the histones. The two molecules make contact at sites where the minor groove of the DNA faces inward toward the histone core, which occurs at approximately 10 base-pair intervals (the white hooks in Figure 12.9*c*). In between these points of contact, the two molecules are seen to be separated by considerable space, which might provide access to the DNA for transcription factors and other DNA-binding proteins. For many years, histones were thought of as inert, structural molecules but, as we will see in following sections, these small proteins play critically important roles in determining the activity of the DNA with which they are associated. It has also become evident that chromatin is a dynamic cellular component in which histones, regulatory proteins, and a myriad variety of enzymes move in and out of the nucleoprotein complex to facilitate the complex tasks of DNA transcription, compaction, replication, recombination, and repair.

We began this section by wondering how a nucleus 10 μm in diameter can pack 200,000 times this length of DNA within its boundaries. The assembly of nucleosomes is the first important step in the compaction process. With a nucleotide–nucleotide spacing of 0.34 nm, the 200 base pairs of a single 10-nm nucleosome would stretch nearly 70 nm if fully extended. Consequently, it is said that the packing ratio of the DNA of nucleosomes is approximately 7:1.

Higher Levels of Chromatin Structure   A DNA molecule wrapped around nucleosome core particles of 10-nm diameter is the lowest level of chromatin organization. Chromatin does not, however, exist within the cell in this relatively extended, "beads-on-a-string" state. When chromatin is released from nuclei and prepared at physiologic ionic strength, a fiber of approximately 30-nm thickness is observed (Figure 12.10*a*). Despite more than two decades of investigation, the structure of the 30-nm fiber remains a subject of debate. Two models in which the nucleosomal filament is coiled into the higher-order, thicker fiber are shown in Figure 12.10*b,c*. The models differ in the relative positioning of nucleosomes within the fiber. Recent research favors the "zig-zag" model depicted in Figure 12.10*b*, in which successive nucleosomes along the DNA are arranged in different stacks and alternating nucleosomes become interacting neighbors. Regardless of how it is accomplished, the assembly of the 30-nm fiber increases the DNA-packing ratio an additional 6-fold, or about 40-fold altogether.
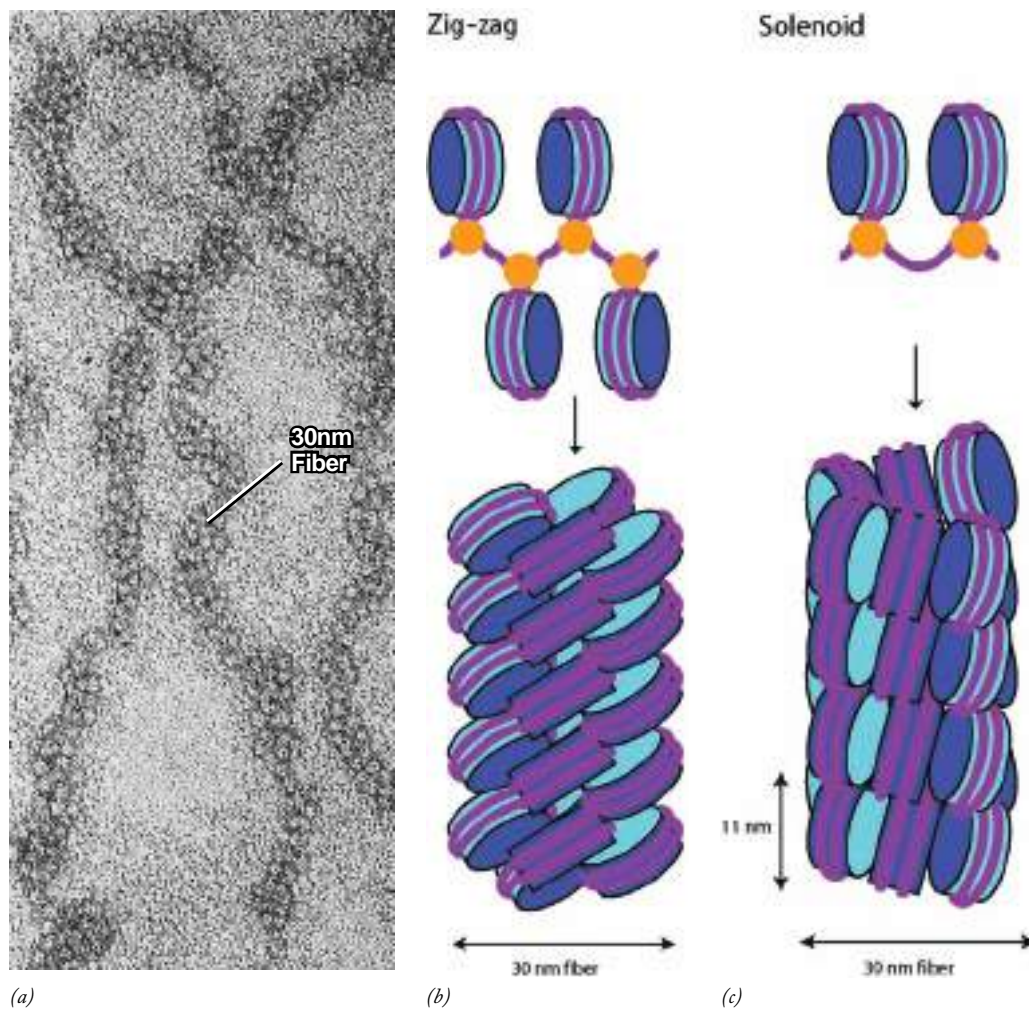
FIGURE 12.10 **The 30-nm fiber: a higher level of chromatin structure.** (*a*) Electron micrograph of a 30-nm chromatin fiber released from a nucleus following lysis of the cell in a hypotonic salt solution. (*b*) In the "zig-zag" model, the linker DNA is present in a straight, extended state that criss-crosses back and forth between consecutive core particles, which are organized into two separate stacks of nucleosomes. The lower portion of the figure shows how the two stacks of nucleosomes are coiled into a higher-order helical structure. (*c*) In the "solenoid" model, the linker DNA gently curves as it connects consecutive core particles, which are organized into a single, continuous helical array containing about 6–8 nucleosomes per turn. In these models, the histone octamer is shown in blue, the DNA in magenta, and the linker H1 histone in yellow. (A: COURTESY OF BARBARA HAMKALO AND JEROME B. RATTNER; B,C: FROM SEPIDEH KHORASANIZADEH, CELL 116:262, 2004; BY PERMISSION OF CELL PRESS.)

Maintenance of the 30-nm fiber depends on the interaction between histone molecules of neighboring nucleosomes. Linker histones and core histones have both been implicated in higher-order packaging of chromatin. If, for example, H1 linker histones are selectively extracted from compacted chromatin, the 30-nm fibers uncoil to form the thinner, more extended beaded filament shown in Figure 12.8*b*. Adding back H1 histone leads to restoration of the higher-order structure. Core histones of adjacent nucleosomes may interact with one another by means of their long, flexible tails. Structural studies indicate, for example, that the N-terminal tail of an H4 histone from one nucleosome core particle can reach out and make extensive contact with both the linker DNA between nucleosome particles and the H2A/H2B dimer of adjacent particles. These types of interactions are thought to mediate the folding of the nucleosomal filament into a thicker fiber. In fact, chromatin fibers prepared with H4 histones that lack their tails are unable to fold into higher-order fibers.

The next stage in the hierarchy of DNA packaging is thought to occur as the 30-nm chromatin fiber is gathered into a series of large, supercoiled loops, or domains, that may be compacted into even thicker (80–100 nm) fibers. The DNA loops are apparently tethered at their bases to pro-

teins that are part of an organized nuclear scaffold or matrix (discussed on page 499). Included among these proteins is a type II topoisomerase that presumably regulates the degree of DNA supercoiling. The topoisomerase would also be expected to untangle the DNA molecules of different loops should they become intertwined. Normally, loops of chromatin fibers are spread out within the nucleus and cannot be visualized, but their presence can be revealed under certain circumstances. For instance, when isolated *mitotic* chromosomes are treated with solutions that extract histones, the histone-free DNA can be seen to extend outward as loops from a protein scaffold (Figure 12.11).

The mitotic chromosome represents the ultimate in chromatin compactness; 1 μm of mitotic chromosome length typically contains approximately 1 cm of DNA, which represents a packing ratio of 10,000:1. This compaction occurs by a poorly understood process that is discussed in Section 14.2. An overview of the various levels of chromatin organization, from the nucleosomal filament to a mitotic chromosome, is depicted in Figure 12.12.

Heterochromatin and Euchromatin After mitosis has been completed, most of the chromatin in highly compacted mi-
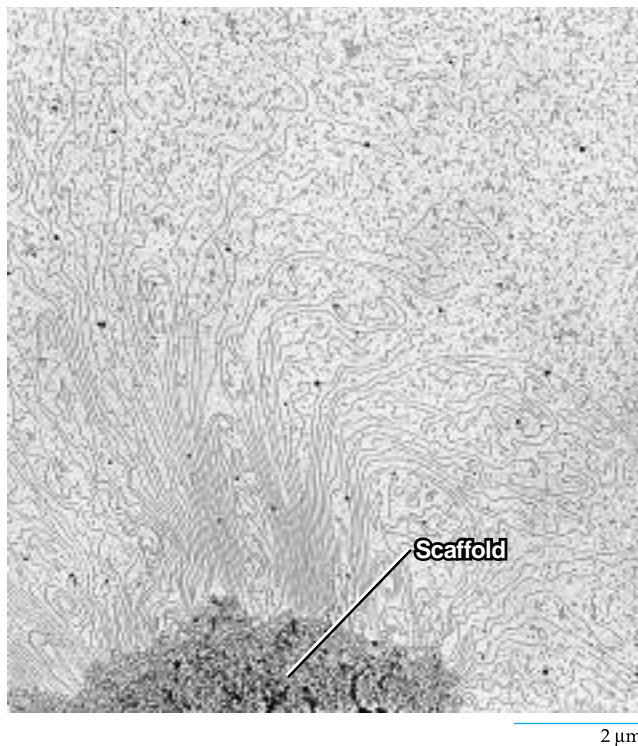
FIGURE 12.11 **Chromatin loops: a higher level of chromatin structure.** Electron micrograph of a mitotic chromosome that had been treated with a solution of dextran sulfate to remove histones. The histone-depleted chromosome displays loops of DNA that are attached at their bases to a residual protein scaffold. (FROM JAMES R. PAULSON AND U. K. LAEMMLI, CELL 12:823, 1977; BY PERMISSION OF CELL PRESS.)

FIGURE 12.12 **Levels of organization of chromatin.** Naked DNA molecules are wrapped around histones to form nucleosomes, which represent the lowest level of chromatin organization. Nucleosomes are organized into 30-nm fibers, which in turn are organized into looped domains. When cells prepare for mitosis, the loops become further compacted into mitotic chromosomes (see Figure 14.13).

totic chromosomes returns to its diffuse interphase condition. Approximately 10 percent of the chromatin, however, generally remains in a condensed, compacted form throughout interphase. This compacted, densely stained chromatin is seen at the periphery of the nucleus in Figure 12.1a. Chromatin that remains compacted during interphase is called **heterochromatin** to distinguish it from **euchromatin**, which returns to a dispersed state. When a radioactively labeled RNA precursor such as [³H]uridine is given to cells that are subsequently fixed, sectioned, and autoradiographed, the clumps of heterochromatin remain largely unlabeled, indicating that they have relatively little transcriptional activity. The state of a particular region of the genome, whether it is euchromatic or heterochromatic, is stably inherited from one cell generation to the next.

Heterochromatin is divided into two classes. **Constitutive heterochromatin** remains in the compacted state in all cells at all times and, thus, represents DNA that is permanently silenced. In mammalian cells, the bulk of the constitutive heterochromatin is found in the region that flanks the telomeres and centromere of each chromosome and in a few other sites, such as the distal arm of the Y chromosome in male mammals. The DNA of constitutive heterochromatin consists primarily of repeated sequences (page 394) and con-

tains relatively few genes. In fact, when genes that are normally active move into a position adjacent to heterochromatin (having changed position as the result of transposition or translocation), they tend to become transcriptionally silenced, a phenomenon known as *position effect*. It is thought that heterochromatin contains components whose influence can spread outward a certain distance, affecting nearby genes. The spread of heterochromatin along the chromosome is apparently blocked by specialized barrier sequences (*boundary elements*) in the genome. Constitutive heterochromatin also serves to inhibit genetic recombination between homologous repetitive sequences. This type of recombination can lead to DNA duplications and deletions (Figure 10.22).

(a)   (b)   (c)
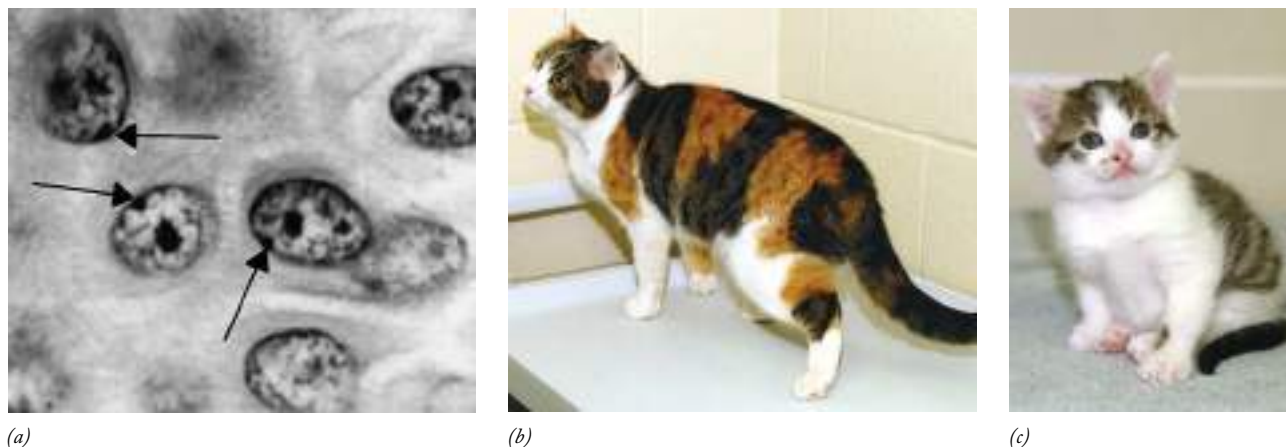
FIGURE 12.13 **The inactive X chromosome: an example of facultative heterochromatin.** (*a*) The inactivated X chromosome in the nucleus of a woman's cells appears as a darkly staining heterochromatic structure, called a Barr body (arrows). (*b*) A calico cat. Random inactivation of either X chromosome in different cells during early embryonic development creates a mosaic of tissue patches. Each patch comprises the descendants of one cell that was present in the embryo at the time of inactivation. These patches are visually evident in calico cats, which are heterozygotes with an allele for black coat color residing on one X chromosome and an allele for orange coat color on the other X. This explains why male calico cats are virtually nonexistent: because all cells in the male have either the black or orange coat color allele. (The white spots on this cat are due to a different, autosomal coat color gene.) (*c*) This kitten was cloned from the cat shown in *b*. The two animals are genetically identical but have different coat patterns, a reflection of the random nature of the X inactivation process (and likely other random developmental events). (A: COURTESY OF MURRAY L. BARR; B,C: COURTESY OF COLLEGE OF VETERINARY MEDICINE AND BIOMEDICAL SCIENCES, TEXAS A&M UNIVERSITY.)

Unlike the constitutive variety, **facultative heterochromatin** is chromatin that has been specifically inactivated during certain phases of an organism's life or in certain types of differentiated cells (as in Figure 17.9*b*). An example of facultative heterochromatin can be seen by comparing cells of a female mammal to those of a male. The cells of males have a tiny Y chromosome and a much larger X chromosome. Because the X and Y chromosomes have only a few genes in common, males have a single copy of most genes that are carried on the sex chromosomes. Although cells of females contain two X chromosomes, only one of them is transcriptionally active. The other X chromosome remains condensed as a heterochromatic clump (Figure 12.13*a*) called a *Barr body* after the researcher who discovered it in 1949. Formation of a Barr body ensures that the cells of both males and females have the same number of active X chromosomes and thus synthesize equivalent amounts of the products encoded by X-linked genes.

X Chromosome Inactivation   Based on her studies of the inheritance of coat color in mice, the British geneticist Mary Lyon proposed the following in 1961:

1. Heterochromatinization of the X chromosome in female mammals occurs during early embryonic development and leads to the inactivation of the genes on that chromosome.

2. Heterochromatinization in the embryo is a random process in the sense that the paternally derived X chromosome and the maternally derived X chromosome stand an equal chance of becoming inactivated *in any given cell*. Consequently, at the time of inactivation, the paternal X can be inactivated in one cell of the embryo, and the maternal X can be inactivated in a neighboring cell. Once an X chromosome has been inactivated, its heterochromatic state is transmitted through many cell divisions, so that the same X chromosome is inactive in all the descendants of that particular cell.

3. Reactivation of the heterochromatinized X chromosome occurs in germ cells prior to the onset of meiosis. Consequently, both X chromosomes are active during oogenesis, and all of the gametes receive a euchromatic X chromosome.

The Lyon hypothesis was soon confirmed.[1] Because maternally and paternally derived X chromosomes may contain different alleles for the same trait, adult females are in a sense *genetic mosaics*, where different alleles function in different cells. X-chromosome mosaicism is reflected in the patchwork coloration of the fur of some mammals, including calico cats (Figure 12.13*b,c*). Pigmentation genes in humans are not located on the X chromosome, hence the absence of "calico women." Mosaicism due to X inactivation can be demonstrated in women, nonetheless. For example, if a narrow beam of red or green light is shone into the eyes of a woman who is

_____
[1]The random inactivation of X chromosomes discussed here, which occurs after the embryo implants in the uterus, is actually the second wave of X chromosome inactivation to occur in the embryo. The first wave, which occurs very early in development, is not random but rather leads only to the inactivation of X chromosomes that had been donated by the father. This early inactivation of paternal X chromosomes is maintained in the cells that give rise to extra-embryonic tissues (e.g., the placenta) and is not discussed in the text. Early paternal X inactivation is erased in cells that give rise to embryonic tissue and random X inactivation subsequently occurs.

a heterozygous carrier for red-green color blindness, patches of retinal cells with defective color vision can be found interspersed among patches with normal vision.

The mechanism responsible for X inactivation has been a focus of attention since a 1992 report suggesting that inactivation is initiated by a noncoding RNA molecule—rather than a protein—that is transcribed from one of the genes (called *XIST* in humans) on the X chromosome that becomes inactivated. The *XIST* RNA is a large transcript (over 17 kb long), which distinguishes it from many other noncoding RNAs that tend to be quite small. The *XIST* RNA does not diffuse into the nucleoplasm, but accumulates along the length of the chromosome just before it is inactivated.[2] The *XIST* gene is required to initiate inactivation, but not to maintain it from one cell generation to the next. This conclusion is based on the discovery of tumor cells in certain women that contain an inactivated X chromosome whose *XIST* gene is deleted. X inactivation is thought to be maintained by DNA methylation (page 520) and repressive histone modifications, as discussed in the next section.

The Histone Code and Formation of Heterochromatin

Figure 12.9*c* shows a schematic model of the nucleosome core particle with its histone tails projecting outward. But this is only a general portrait that obscures important differences among nucleosomes. Cells contain a remarkable array of enzymes that are able to add chemical groups to or remove them from specific amino acid residues in the histone tails. Those residues that are subject to modification, most notably by methylation, acetylation, or phosphorylation, are indicated by the colored bars in Figure 12.14. The past few years has seen the emergence of a hypothesis known as the **histone code**, which postulates that the state and activity of a particular region of chromatin depend on the specific modifications, or combinations of modifications, to the histone tails in that region. In other words, the pattern of modifications adorning the tails of the core histones contains encoded information governing the properties of those nucleosomes. Studies suggest that histone tail modifications act in two ways to influence chromatin structure and function.

1. The modified residues serve as docking sites to recruit a specific array of nonhistone proteins, which then determine the properties and activities of that segment of chromatin. A sampling of some of the specific proteins that bind selectively to modified histone residues is depicted in Figure 12.15. Each of the proteins bound to the histones in Figure 12.15 is capable of modulating some aspect of chromatin activity or structure.

2. The modified residues alter the manner in which the histone tails of neighboring nucleosomes interact with one another or with the DNA to which the nucleosomes are bound. Changes in these types of interactions can lead to changes in the higher order structure of chromatin.

---

[2]Approximately 15 percent of genes on the chromosome escape inactivation by an unknown mechanism. The "escapees" include genes that are also present on the Y chromosome, which ensures that they are expressed equally in both sexes.
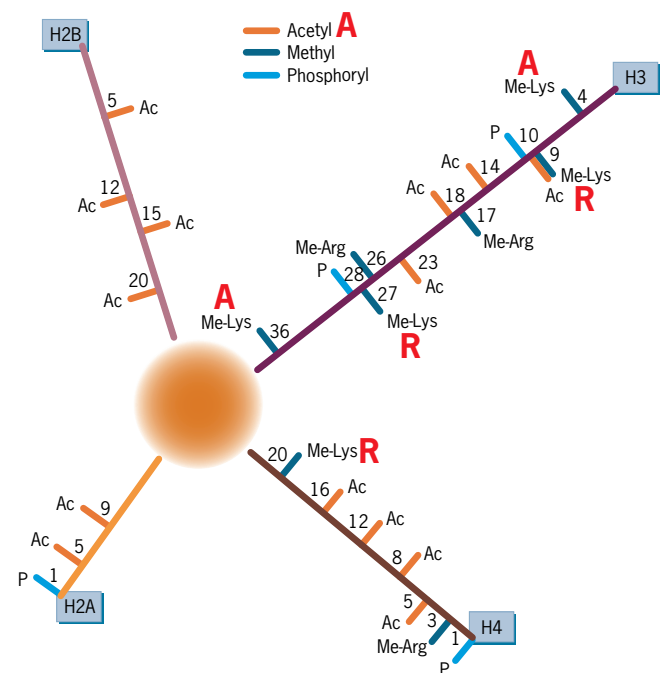
FIGURE 12.14 **Histone modifications and the "histone code."** Histones can be enzymatically modified by the covalent addition of methyl, acetyl, and phosphate groups (and others not discussed). This illustration indicates the positions in the N-terminal tails of the four core histones at which each of these three groups can be added. Methyl groups are added to either lysine or arginine residues, acetyl groups to lysine residues, and phosphate groups to serine residues. The matter is even more complex, because each lysine residue can have either one, two, or three added methyl groups, and each arginine residue can have either one or two added methyl groups. The number of added methyl groups can affect the affinity of the residue for an interacting protein. Unless otherwise noted, we will restrict the discussion to tri-methylated lysine residues (e.g., H3K9me3 or H3K36me3). Certain modifications are associated with particular chromatin activities, which has led to the concept of a histone code. The present discussion is restricted to the lysine residues, which are best understood. The red letters A and R represent transcriptional activation and repression, respectively. Acetylation of lysines on both histones H3 and H4 is closely correlated with transcriptional activation. The effects of methylation of H3 and H4 lysines depends strongly on which of these residues is modified. For example, methylation of lysine 9 of histone H3 (i.e., H3K9) is typically present in heterochromatin and associated with transcriptional repression, as discussed in the text. Methylation of H3K27 and H4K20 is also strongly associated with transcriptional repression, whereas methylation of H3K4 and H3K36 is associated with activation. Just as there are enzymes that add each of these groups, there are also enzymes (deacetylases, demethylases, and phosphatases) that specifically remove them. (C: AFTER G. FELSENFELD & M. GROUDINE, REPRINTED WITH PERMISSION FROM NATURE 421:450, 2003; © COPYRIGHT 2003, MACMILLAN MAGAZINES LIMITED.)

Acetylation of the lysine residue at position 16 on histone H4, for example, interferes with the formation of the compact 30-nm chromatin fiber.

For the moment, we will restrict the discussion to the formation of heterochromatin as it occurs, for example, during
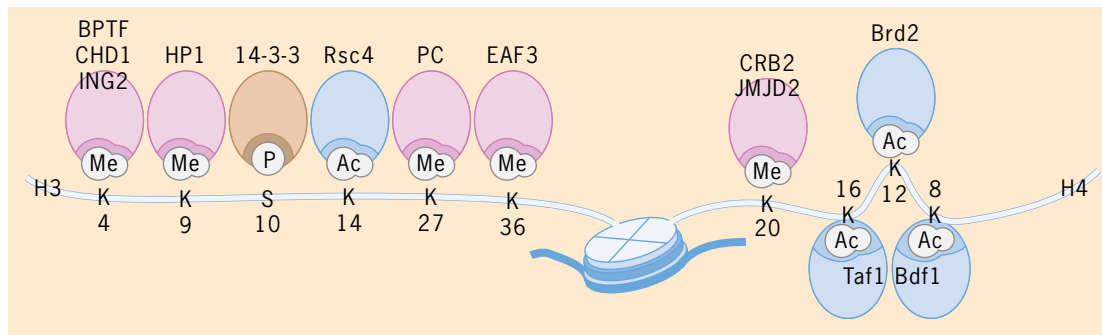
FIGURE 12.15 Examples of proteins that bind selectively to modified H3 or H4 residues. Each of the bound proteins possesses an activity that alters the structure and/or function of the chromatin. There is an added complexity that is not shown in this drawing in that modifications at one histone residue can influence events at other residues, a phenomenon known as cross-talk. For example, the binding of the heterochromatin protein HP1 to H3K9 is blocked by phosphorylation of the adjacent serine residue (H3S10), which typically occurs during mitosis. (FROM T. KOUZARIDES, CELL 128:696, 2007, BY PERMISSION OF CELL PRESS.)

X chromosome inactivation. For the sake of simplicity, we will focus on modification of a single residue—lysine 9 of H3—which will illustrate the general principles by which cells utilize the histone code. The actions of several other histone modifications are indicated in Figure 12.14 and discussed in the accompanying legend, and another example is described on page 517. As we will see throughout this chapter, techniques have been developed in recent years to analyze changes that affect genome transcription, such as histone modifications, on a genome-wide level, rather than simply looking at these changes one gene at a time. This has given us a much broader view of the general importance of each of these phenomena than was possible only a few years ago.

Comparison of the nucleosomes present within heterochromatic versus euchromatic chromatin domains revealed a striking difference. The lysine residue at the #9 position (Lys9 or K9) of the H3 histone in heterochromatic domains is largely methylated, whereas this same residue in euchromatic domains tends to be unmethylated, although it is often acetylated. Removal of the acetyl groups from H3 and H4 histones is among the initial steps in conversion of euchromatin into heterochromatin. The correlation between transcriptional repression and histone deacetylation can be seen by comparing the inactive, heterochromatic X chromosome of female cells, which contains deacetylated histones, to the active, euchromatic X chromosome, whose histones exhibit a normal level of acetylation (Figure 12.16). Histone deacetylation is accompanied by methylation of H3K9, which is catalyzed by an enzyme (a *histone methyltransferase*) that appears to be dedicated to this, and only this, particular function. This enzyme, called SUV39H1 in humans, can be found localized within heterochromatin, where it may stabilize the heterochromatic nature of the region through its methylation activity.

The formation of a methylated lysine at the #9 position endows the histone H3 tail with an important property: it becomes capable of binding with high affinity to proteins that contain a particular domain, called a *chromodomain*. The human genome contains at least 30 proteins with chromodomains, the best studied of which is *heterochromatic protein 1* (or *HP1*). HP1 has been implicated in the formation and maintenance of heterochromatin. Once bound to an H3 tail, HP1 is thought to interact with other proteins, including (1) SUV39H1, the enzyme responsible for methylating the H3K9 residue and (2) other HP1 molecules on nearby nucleosomes. These binding properties of the HP1 molecule promote the formation of an interconnecting network of methylated nucleosomes, leading to a compacted, higher-order chromatin domain. Most importantly, this state is transmitted through cell divisions from one cell generation to the next (discussed on page 497).

Studies in a number of organisms indicate that small RNAs, similar in nature to those involved in RNA interference (page 449), play an important role in targeting a particular region of the genome to undergo H3K9 methylation and subsequent heterochromatinization. If, for example, compo-

FIGURE 12.16 Experimental demonstration of a correlation between transcriptional activity and histone acetylation. This metaphase chromosome spread has been labeled with fluorescent antibodies to acetylated histone H4, which fluoresce green. It is evident that all of the chromosomes except the inactivated X stain brightly with the antibody against the acetylated histone. (FROM P. JEPPESEN AND B. M. TURNER, COVER OF CELL VOL. 74, NO. 2, 1993; BY PERMISSION OF CELL PRESS.)

FIGURE 12.17 A model showing possible events during the formation of heterochromatin. Recent studies suggest that noncoding RNAs play a role in directing heterochromatinization in at least some organisms. In this model, RNAs are being transcribed from both strands of repeated DNA sequences (step 1). The RNAs form double-stranded molecules (step 2) that are processed by the endonuclease Dicer and other components of the RNAi machinery (page 449) to form a single-stranded siRNA guide and an associated protein complex (step 3). In step 4, the siRNA-protein complex has recruited the enzyme SUV39H1 and the siRNA has bound to a complementary segment of a nascent RNA. Once there, the enzyme is able to add methyl groups to the K9 residue of H3 core histones, replacing the acetyl groups that were previously linked to the H3K9 residues. (The acetyl groups, which are characteristic of transcribed regions of euchromatin, are removed enzymatically from the lysine residues by a deacetylase, which is not shown.) In step 5, the acetyl groups have all been replaced by methyl groups, which serve as binding sites for the HP1 protein (step 6). The boundary element in the DNA prevents the spread of heterochromatinization into adjacent regions of chromatin. Once HP1 has bound to the histone tails, the chromatin can be packaged into higher-order, more compact structures by means of interaction between HP1 protein molecules (step 7). The enzyme SUV39H1 can also bind to methylated histone tails (not shown) so that additional nucleosomes can become methylated. In step 8, a region of highly compacted heterochromatin has been formed. (*Note:* HP1 can be present as several isoforms with different properties.)

nents of the RNAi machinery are deleted, the methylation of H3K9 and heterochromatinization is impaired. In addition to its role in the formation of heterochromatin, the RNAi machinery may also function in the maintenance of the heterochromatic state from one cell generation to the next. These findings point to yet another role, on a rapidly growing list of roles, performed by noncoding RNAs. A model showing the types of events thought to take place during heterochromatin assembly is depicted in Figure 12.17. It would not be surprising, given the fact that most of the genome is now thought to be transcribed (page 454), to discover that RNAs play a major role in guiding many of the changes in chromatin structure that are known to occur during embryonic development or in response to physiological stimuli.

The Structure of a Mitotic Chromosome   The relatively dispersed state of the chromatin of an interphase cell favors interphase activities, such as replication and transcription. In contrast, the chromatin of a mitotic cell exists in its most highly condensed state, which favors the delivery of an intact "package" of DNA to each daughter cell. Mitotic chromosomes have also proven useful to biologists and physicians because they contain a complete set of genetic material of a cell and are made readily visible by simple techniques.

When a chromosome undergoes compaction during mitotic prophase, it adopts a distinct and predictable shape determined primarily by the length of the DNA molecule in that chromosome and the position of the centromere (discussed later). The mitotic chromosomes of a dividing cell can be displayed by the technique depicted in Figure 12.18a. In this technique, a dividing cell is broken open, and the mitotic chromosomes from the cell's nucleus settle and attach to the
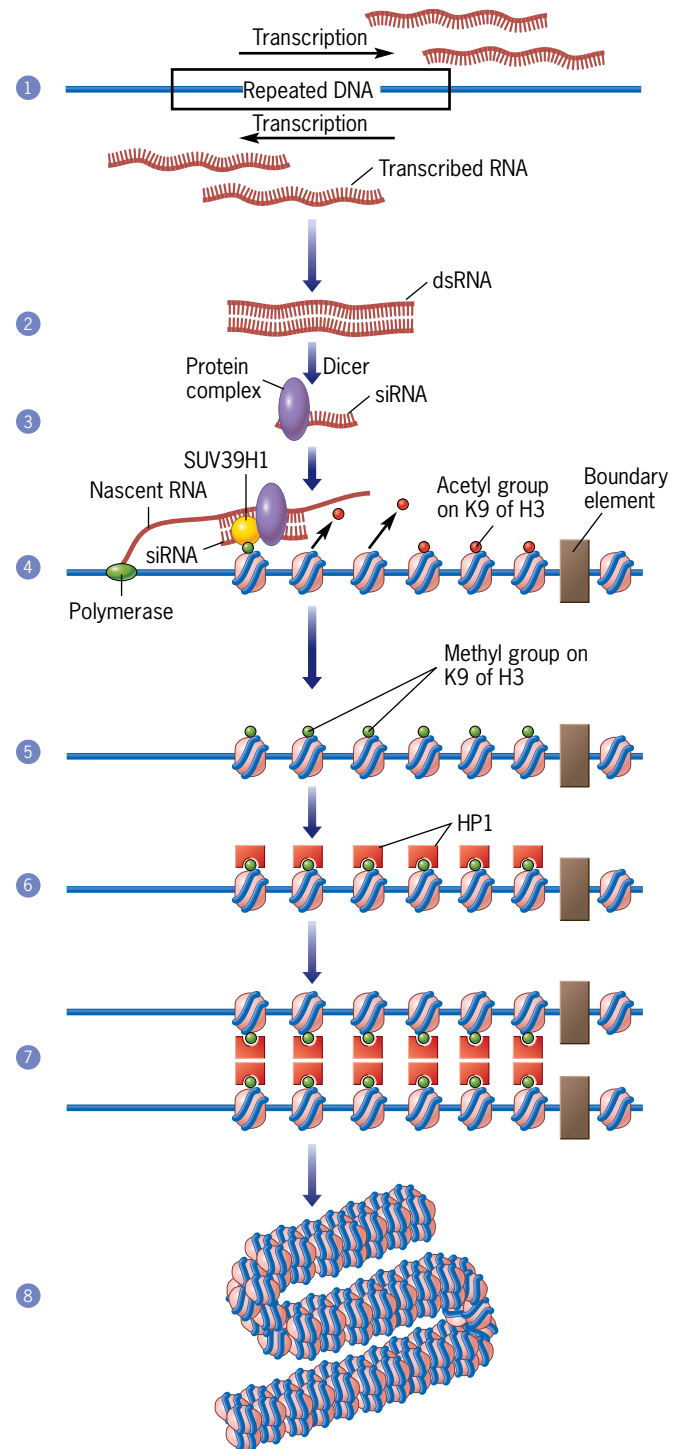
surface of the slide over a very small area (as in Figure 12.18b). The chromosomes displayed in Figure 12.18b have been prepared using a staining methodology in which chromosome preparations are incubated with multicolored, fluorescent DNA probes that bind specifically to particular chromosomes. By using various combinations of DNA probes and computer-aided visualization techniques, each chromosome can be "painted" with a different "virtual color," making it readily identifiable to the trained eye. In addition to providing a colorful image, this technique delivers superb resolution, which allows

(a)

Capillary pipette with bulb

Drop of blood

Transfer to vial for culturing.

Culture medium (includes substance which stimulates mitosis in leukocytes)

Culture approximately 72 hours, then add colchicine for 30 min to 3 hours. Collect cells by centrifugation.

Medium

Cells

Wash with fresh medium. Add hypotonic solution to cells. Let sit 10 min. Remove supernatant and add cold fixative (3:1 methanol:acetic acid). Let sit 30 min in cold, then disperse cells.

Cell suspension

Drops of fixative with cells

Wet slide

Evaporate fixative. Stain slide.

Slide

Site containing the chromosomes released from a single nucleus

(b)

(c)

FIGURE 12.18 Human mitotic chromosomes and karyotypes. (*a*) Procedure used to obtain preparations of mitotic chromosomes for microscopic observation from leukocytes in the peripheral blood. (*b*) Photograph of a cluster of mitotic chromosomes that spilled out of the nucleus of a single dividing human cell. The DNA of each chromosome has hybridized to an assortment of DNA probes that are linked covalently to two or more fluorescent dyes. Different chromosomes bind different combinations of these dyes and, consequently, emit light of different wavelengths. Emission spectra from different chromosomes are converted into distinct and recognizable display colors by computer processing. Pairs of homologous chromosomes can be identified by searching for chromosomes of the same color and size. (*c*) The stained chromosomes of a human male arranged in a karyotype. Karyotypes are prepared from a photograph of chromosomes released from a single nucleus. Each chromosome is cut out of the photograph, and homologues are paired according to size, as shown. (B: FROM E. SCHRÖCK ET AL., COURTESY OF THOMAS RIED, SCIENCE 273:495, 1996; © COPYRIGHT 1996, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE; C: FROM CNRI/SCIENCE PHOTO LIBRARY/PHOTO RESEARCHERS.)

clinical geneticists to discover chromosomal aberrations that might otherwise be missed (see Figure 2 of the accompanying Human Perspective).

If the individual chromosomes are cut out of a photograph such as that of Figure 12.18*b*, they can be matched up into homologous pairs (23 in humans) and ordered according to decreasing size, as shown in Figure 12.18*c*. A preparation of this type is called a **karyotype**. The chromosomes shown in the karyotype of Figure 12.18*c* have been prepared using a staining procedure that gives the chromosomes a cross-banded appearance. The pattern of these bands is highly characteristic for each chromosome of a species and provides a basis to identify chromosomes and compare them from one species to the next (see Figure 3 in the Human Perspective). Karyotypes are routinely prepared from cultures of blood cells and used to screen individuals for chromosomal abnormalities. As discussed in the accompanying Human Perspective, extra, missing, or grossly altered chromosomes can be detected in this manner.

THE HUMAN PERSPECTIVE

Chromosomal Aberrations and Human Disorders

In addition to mutations that alter the information content of a single gene, chromosomes may be subjected to more extensive alterations that occur most commonly during cell division. Pieces of a chromosome may be lost or segments may be exchanged between different chromosomes. Because these chromosomal aberrations follow chromosomal breakage, their incidence is increased by exposure to agents that damage DNA, such as viral infection, X-rays, or reactive chemicals. In addition, the chromosomes of some individuals contain "fragile" sites that are particularly susceptible to breakage. Persons with certain rare inherited conditions, such as Bloom syndrome, Fanconi anemia, and ataxia-telangiectasia, have unstable chromosomes that exhibit a greatly increased tendency toward breakage.

The consequences of a chromosomal aberration depend on the genes that are affected and the type of cell in which it occurs. If the aberration occurs in a somatic (nonreproductive) cell, the consequences are generally minimal because only a few cells of the body are usually affected. On rare occasions, however, a somatic cell carrying an aberration may be transformed into a malignant cell, which can grow into a cancerous tumor. Chromosomal aberrations that occur during meiosis—particularly as a result of abnormal crossing over—can be transmitted to the next generation. When an aberrant chromosome is inherited through a gamete, all cells of the embryo will have the aberration, which generally proves lethal during development. There are several types of chromosomal aberrations, including the following:

■ **Inversions.** Sometimes a chromosome breaks in two places, and the segment between the breaks becomes resealed into the chromosome in reverse orientation. This aberration is called an inversion. More than 1 percent of humans carry an inversion that can be detected during chromosome karyotyping (see Figure 10.30*b*). A chromosome bearing an inversion usually contains all the genes of a normal chromosome, and therefore, the individual is not adversely affected. However, if a cell with a chromosome inversion enters meiosis, the aberrant chromosome cannot pair properly with its normal homologue because of differences in the order of their genes. In such cases, chromosome pairing is usually accompanied by a loop (Figure 1). If crossing over occurs within the loop, as shown in the figure, the gametes generated by meiosis either possess an additional copy of certain genes (a duplication) or are missing those genes (a deletion). When a gamete containing an altered chromosome fuses with a normal gamete at fertilization, the resulting zygote has a chromosome imbalance and is usually nonviable.
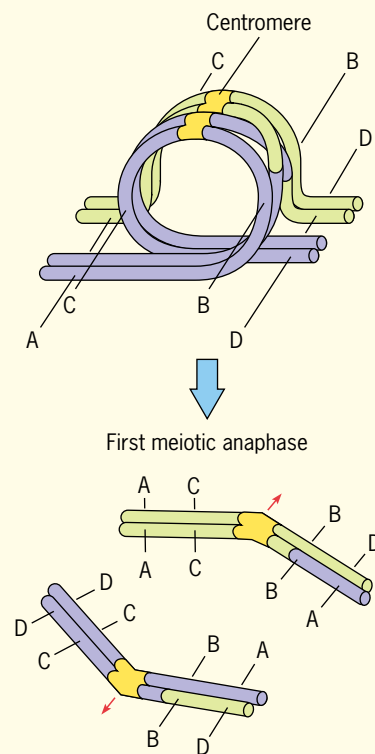
FIGURE 1 The effect of inversion. Crossing over between a normal chromosome (purple) and one containing an inversion (green) is usually accompanied by formation of a loop. The chromosomes that result from the crossover contain duplications and deficiencies, which are shown in the chromosomes at first meiotic division in the lower part of the figure.

■ **Translocations.** When all or a piece of one chromosome becomes attached to another chromosome, the aberration is called a translocation (Figure 2). Like inversions, a translocation that occurs in a somatic cell generally has little effect on the functions of that cell or its progeny. However, certain translocations increase the likelihood that the cell will become malignant. The best studied example is the Philadelphia chromosome, which is found in the malignant cells (but not the normal cells) of
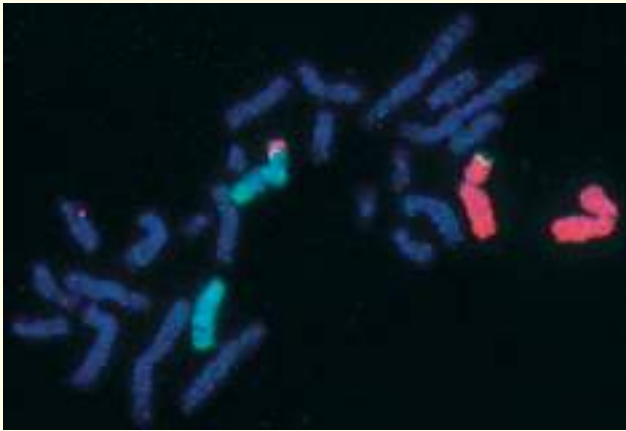
FIGURE 2 A translocation. Micrograph shows a set of human chromosomes in which chromosome 12 (bright blue) has exchanged pieces with chromosome 7 (red). The affected chromosomes have been made fluorescent by in situ hybridization with a large number of DNA fragments that are specific for each of the two chromosomes involved. The use of these "stains" makes it very evident when one chromosome has exchanged pieces with another chromosome. (COURTESY OF LAWRENCE LIVERMORE NATIONAL LABORATORY, FROM A TECHNIQUE DEVELOPED BY JOE GRAY AND DAN PINKEL.)

individuals with certain forms of leukemia. The Philadelphia chromosome, which is named for the city in which it was discovered in 1960, is a shortened version of human chromosome number 22. For years, it was thought that the missing segment represented a simple deletion, but with improved techniques for observing chromosomes, the missing genetic piece was found translocated to another chromosome (number 9). Chromosome number 9 contains a gene (*ABL*) that encodes a protein kinase that plays a role in cell proliferation. As a result of translocation, one small end of this protein is replaced by about 600 extra amino acids encoded by a gene (*BCR*) carried on the translocated piece of chromosome number 22. This novel "fusion protein" retains the catalytic activity of the original ABL but is no longer subject to the cell's normal regulatory mechanisms. As a result, the affected cell becomes malignant and causes chronic myelogenous leukemia (CML).
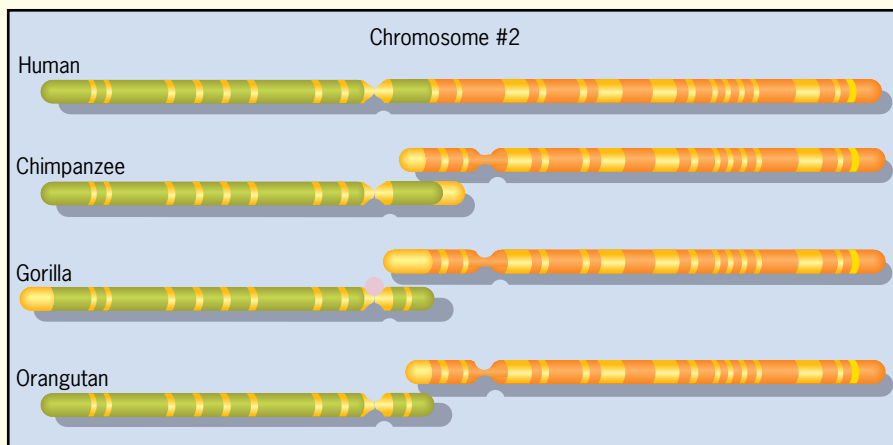
Like inversions, translocations cause problems during meiosis. A chromosome altered by translocation has a different genetic content from its homologue. As a result, the gametes formed by meiosis will either contain extra copies of genes or be missing genes. Translocations have been shown to play an important role in evolution, generating large-scale changes that may be pivotal in the branching of separate evolutionary lines from a common ancestor. Such a genetic incident probably happened during our own recent evolutionary history. A comparison of the 23 pairs of chromosomes in human cells with the 24 pairs of chromosomes in the cells of chimpanzees, gorillas, and orangutans reveals a striking similarity. Close examination of the two ape chromosomes that have no counterpart in humans reveals that together they are equivalent, band for band, to human chromosome number 2 (Figure 3). At some point during the evolution of humans, an entire chromosome was apparently translocated to another, creating a single fused chromosome and reducing the haploid number from 24 to 23.

■ **Deletions.** A deletion occurs when a portion of a chromosome is missing. As noted above, zygotes containing a chromosomal deletion are produced when one of the gametes is the product of an abnormal meiosis. Forfeiting a portion of a chromosome often results in a loss of critical genes, producing severe consequences, even if the individual's homologous chromosome is normal. Most human embryos that carry a significant deletion do not develop to term, and those that do exhibit a variety of malformations. The first correlation between a human disorder and a chromosomal deletion was made in 1963 by Jerome Lejeune, a French geneticist who had earlier discovered the chromosomal basis of Down syndrome. Lejeune discovered that a baby born with a variety of facial malformations was missing a portion of chromosome 5. A defect in the larynx (voice box) caused the infant's cry to resemble the sound of a suffering cat. Consequently, the scientists named the disorder cri-du-chat syndrome, meaning cry-of-the-cat syndrome.

■ **Duplications.** A duplication occurs when a portion of a chromosome is repeated. The role of duplications in the formation of multigene families was discussed on page 399. More substantial chromosome duplications create a condition in which a number of genes are present in three copies rather than the two copies normally present (a condition called *partial trisomy*). Cellular activities are very sensitive to the number of copies of genes, and thus extra copies of genes can have serious deleterious effects.

FIGURE 3 Translocation and evolution. If the only two ape chromosomes that have no counterpart in humans are hypothetically fused, they match human chromosome number 2, band for band.

Telomeres   Each chromosome contains a single, continuous, double-stranded DNA molecule. The tips of each DNA molecule are composed of an unusual stretch of repeated sequences that, together with a group of specialized proteins, forms a cap at each end of the chromosome called a **telomere**. Human telomeres contain the sequence $\frac{\text{TTAGGG}}{\text{AATCCC}}$ repeated from about 500 to 5000 times (Figure 12.19a). Unlike most repeated sequences that vary considerably from species to species, the same telomere sequence is found throughout the vertebrates, and similar sequences are found in most other organisms. This similarity in sequence suggests that telomeres have a conserved function in diverse organisms. A number of DNA-binding proteins have been identified that bind specifically to the telomere sequence and are essential for telomere function. The protein that is bound to the chromosomes in Figure 12.19b plays a role in regulating telomere length in yeast. As recently discovered, the short repetitive DNA sequence of the telomeres also serves as a template for the synthesis of noncoding RNAs whose functions have become a focus of current interest.

As discussed in Chapter 13, the DNA polymerases that replicate DNA do not initiate the synthesis of a strand of DNA but can only add nucleotides to the 3′ end of an existing strand. Replication is initiated at the 5′ end of each newly synthesized strand by synthesis of a short RNA primer (step 1, Figure 12.20a) that is subsequently removed (step 2). Because of this mechanism, and the additional processing steps shown in step 3, the 5′ end of each newly synthesized strand is missing a short segment of DNA that is present at the 3′ end of the complementary template strand. As a result, the strand with the 3′ end overhangs the strand with the 5′ end. Rather than existing as an unprotected single-stranded terminus, the overhanging strand is "tucked back" into the double-stranded portion of the telomere to form a loop as illustrated in Figure 12.20b. This conformation is thought to protect the telomeric end of the DNA.

If cells were not able to replicate the ends of their DNA, the chromosomes would be expected to become shorter and shorter with each round of cell division (Figure 12.20a). This predicament has been called "the end-replication problem." The primary mechanism by which organisms solve the "end-replication problem," came to light in 1984 when Elizabeth Blackburn and Carol Greider of the University of California, Berkeley, discovered a novel enzyme, called **telomerase**, that can add new repeat units to the 3′ end of the overhanging strand (Figure 12.20c). Once the 3′ end of the strand has been lengthened, a conventional DNA polymerase can use the newly synthesized 3′ segment as a template to return the 5′ end of the complementary strand to its previous length.

(a)
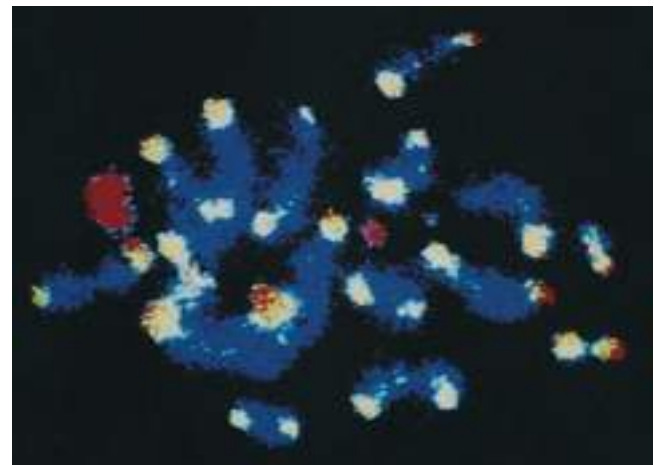
(b)   2 μm

FIGURE 12.19 Telomeres. (*a*) In situ hybridization of a DNA probe containing the sequence TTAGGG, which localizes to the telomeres of human chromosomes. (*b*) Demonstration that certain proteins bind specifically to telomeric DNA. These chromosomes were prepared from a meiotic nucleus of a yeast cell and incubated with the protein RAP1, which was subsequently localized at the telomeres by a fluorescent anti-RAP1 antibody. Blue areas indicate DNA staining, yellow areas repre-sent anti-RAP1 antibody labeling, and red shows RNA stained with propidium iodide. Humans have a homologous telomere protein. (A: FROM J. MEYNE, IN R. P. WAGNER, CHROMOSOMES: A SYNTHESIS, COPYRIGHT © 1993. REPRINTED BY PERMISSION OF WILEY-LISS, INC., A SUBSIDIARY OF JOHN WILEY & SONS, INC.; B: FROM FRANZ KLEIN ET AL., J. CELL BIOL. 117:940, 1992, COURTESY OF SUSAN M. GASSER; BY COPY-RIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

(a)

(b)

(c)
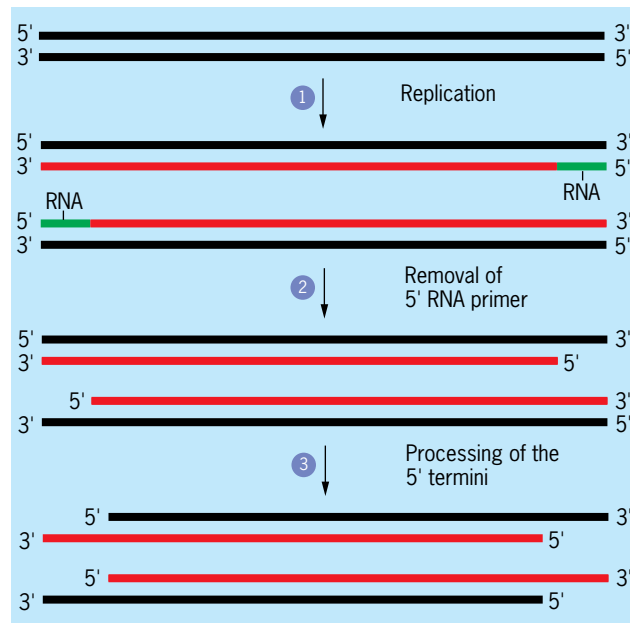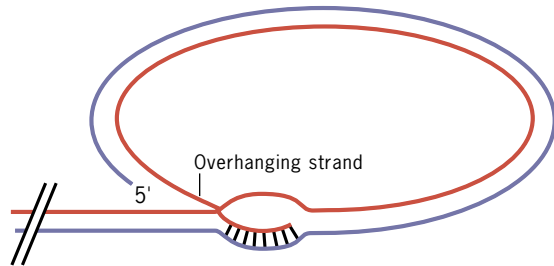
FIGURE 12.20 **The end-replication problem and the role of telomerase.** (*a*) When the DNA of a chromosome is replicated (step 1), the 5′ ends of the newly synthesized strands (red) contain a short segment of RNA (green), which had functioned as a primer for synthesis of the adjoining DNA. Once this RNA is removed (step 2), the 5′ end of the DNA becomes shorter relative to that of the previous generation. The 5′ ends at each end of the chromosome are further processed by nuclease activities (step 3), which increases the lengths of the single-stranded overhangs. (*b*) The single-stranded overhang does not remain as a free extension, but invades the duplex as shown here, displacing one of the strands, which forms a loop. The loop is a binding site for a team of specific telomere-capping proteins that protect the ends of the chromosomes and regulate telomere length. (*c*) The mechanism of action of telomerase. The enzyme contains an RNA molecule that is complementary to the end of the G-rich strand, which extends past the C-rich strand. The telomerase RNA binds to the protruding end of the G-rich strand (step 1) and then serves as a template for the addition of nucleotides onto the 3′ terminus of the strand (step 2). After a segment of DNA is synthesized, the telomerase RNA slides to the new end of the strand being elongated (step 3) and serves as the template for the incorporation of additional nucleotides (step 4). The gap in the complementary strand is filled by the replication enzymes polymerase α-primase (see Figure 13.21). (The TTGGGG sequence depicted in this drawing is that of the ciliated protist *Tetrahymena*, in which telomerase was discovered.) (c: AFTER C. W. GREIDER AND E. H. BLACKBURN, REPRINTED WITH PERMISSION FROM NATURE 337: 336, 1989; © COPYRIGHT 1989, MACMILLAN MAGAZINES LIMITED.)
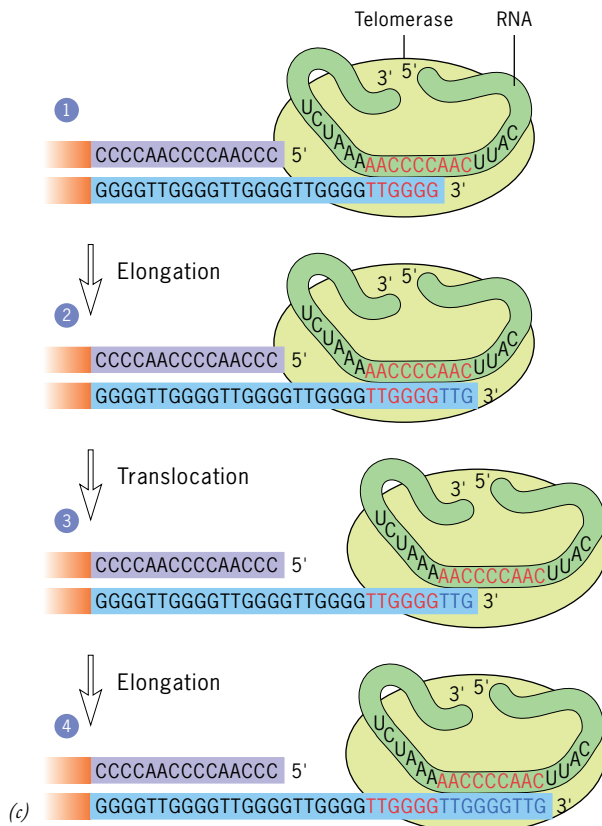
Telomerase is a reverse transcriptase that synthesizes DNA using an RNA template. Unlike most reverse transcriptases, the enzyme itself contains the RNA that serves as its template (Figure 12.20*c*).

Telomeres are very important parts of a chromosome: they are required for the complete replication of the chromosome; they form caps that protect the chromosomes from nucleases and other destabilizing influences; and they prevent the ends of chromosomes from fusing with one another. Figure 12.21 shows mitotic chromosomes from a mouse that has been genetically engineered to lack telomerase. Many of the chromosomes have undergone end-to-end fusion, which leads to catastrophic consequences as the chromosomes are torn apart in subsequent cell divisions. Recent experiments suggest additional roles for telomeres, which have made them a focus of current research.

Suppose a researcher were to take a small biopsy of your skin, isolate a population of fibroblasts from the dermis, and allow these cells to grow in an enriched culture medium. The fibroblasts would divide every day or so in culture and eventu-
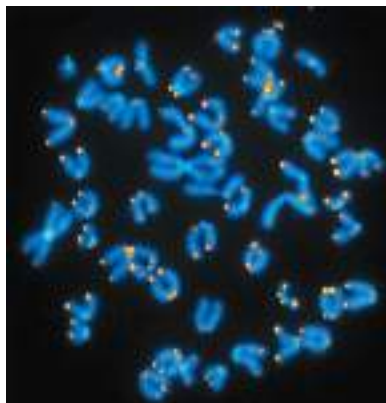
FIGURE 12.21 **The importance of telomerase in maintaining chromosome integrity.** The chromosomes in this micrograph are from the cell of a knockout mouse that lacks a functional gene for the enzyme telomerase. The telomeres appear as yellow spots following in situ hybridization with a fluorescently labeled telomere probe. It can be seen that some chromosomes lack telomeres entirely and that a number of chromosomes have fused to one another at their ends. Chromosome fusion produces chromosomes with more than one centromere, which in turn leads to chromosome breakage during cell division. The genetic instability resulting from loss of a telomere may be a major cause of cells becoming cancerous. (FROM MARIA A. BLASCO ET AL., COURTESY OF CAROL W. GREIDER, CELL, VOL. 91, COVER #1, 1997; BY PERMISSION OF CELL PRESS.)

ally cover the dish. If a fraction of these cells were removed from the first dish and replated on a second dish, they would once again proliferate and cover the second dish. You might think you could subculture these cells indefinitely—as was thought in the first half of the last century—but you would be wrong. After about 50 to 80 population doublings, the cells stop dividing entirely and enter a stage referred to as *replicative senescence*. If the average length of the telomeres in the fibroblasts at the beginning and end of the experiment are compared, one finds a dramatic decrease in the length of the telomeres over time in culture. Telomeres shrink because most cells lack telomerase and are unable to prevent the loss of their chromosome ends. With each cell division, the telomeres of the chromosomes become shorter and shorter. Telomere shortening continues to a critical point at which a physiological response is triggered within each cell that causes that cell to cease continued growth and division. In contrast, cells that are forced to express telomerase continue to proliferate for hundreds of extra divisions. Not only do the telomerase-expressing cells continue to divide, they do so without showing signs of the normal aging processes seen in control cultures.

Although telomerase is absent from most of the body's cells, there are notable exceptions. The germ cells of the gonads retain telomerase activity, and the telomeres of their chromosomes do not shrink as the result of cell division. Consequently, each offspring begins life as a zygote that contains telomeres of maximal length. Similarly, the stem cells located in the lining of the skin and intestine and the hematopoietic stem cells of the blood-forming tissues also express this en-

zyme, which allows these cells to continue to proliferate and generate the huge numbers of differentiated cells required in these organs. The importance of telomerase is revealed by a rare condition in which individuals have greatly reduced telomerase levels, because they are heterozygous for the gene that encodes either the telomerase RNA or protein subunit. These individuals suffer from bone marrow failure due to the inability of this blood-forming tissue to produce a sufficient number of blood cells over a normal human lifetime.

If telomeres are such an important factor in limiting the number of times that a cell can divide, one might expect telomeres to be a major factor in normal human aging. Although the subject is controversial, there is scientific support for this idea. For example, studies have reported that older persons whose white blood cells have shorter telomeres are more likely to have cardiovascular disease or to contract serious infections than persons of comparable age whose blood cells have longer telomeres. Another series of investigations have found that patients with Werner's syndrome, an inherited disease that causes patients to age much more rapidly than normal, is characterized by abnormal telomere maintenance. It has even been reported that women who are under chronic stress as a consequence of caring for seriously ill children have shorter telomeres and reduced telomerase activity. Before you conclude that having an overactive telomerase is the key to extending the human lifespan, you might consider the following information.

According to current consensus, telomere shortening plays a key role in protecting humans from cancer by limiting the number of divisions of a potential tumor cell. Malignant cells, by definition, are cells that have escaped the body's normal growth control and continue to divide indefinitely. How is it that malignant tumor cells can divide repeatedly without running out of telomeres and bringing on their own death? Unlike normal cells that lack detectable telomerase activity, approximately 90 percent of human tumors consist of cells that contain an active telomerase enzyme.[3] It is speculated that the growth of tumors is accompanied by intense selection for cells in which the expression of telomerase has been reactivated. The vast majority of tumor cells fail to express telomerase and die out, whereas the rare cells that express the enzyme are "immortalized." This does not mean that activation of telomerase, by itself, causes cells to become malignant. As discussed in Chapter 16, the development of cancer is a multistep process in which cells typically develop abnormal chromosomes, changes in cell adhesion, and the ability to invade normal tissues. Unlimited cell division is thus only one property of cancer cells. It is interesting to note that the initial discoveries of telomeric DNA sequences and telomerase were carried out on *Tetrahymena*, the same single-celled pond creature exploited in the discovery of ribozymes (page 469). This point serves as another reminder that one never knows which experimental pathways will lead to discoveries that have great medical significance.

---

[3]The other 10 percent or so have an alternate mechanism based on genetic recombination that maintains telomere length in the absence of telomerase.
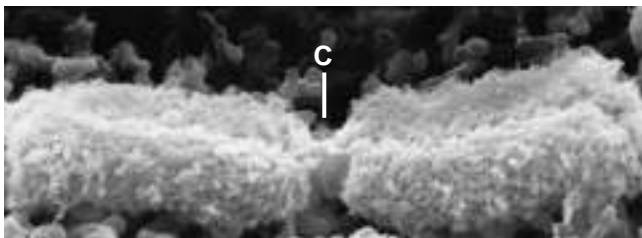
FIGURE 12.22　Each mitotic chromosome has a centromere whose site is marked by a distinct indentation. Scanning electron micrograph of a mitotic chromosome. The centromere (C) contains highly repeated DNA sequences (satellite DNA) and a protein-containing structure called the kinetochore that serves as a site for the attachment of spindle microtubules during mitosis and meiosis (discussed in Chapter 14). (FROM JEROME B. RATTNER, BIOESS. 13:51, 1991.)

Centromeres　Each chromosome depicted in Figure 12.18 contains a site where the outer surfaces are markedly indented. The site of the constriction marks the **centromere** of the chromosome (Figure 12.22). In humans the centromere contains a tandemly repeated, 171-base-pair DNA sequence (called α-*satellite DNA*) that extends for at least 500 kilobases. This stretch of DNA associates with specific proteins that distinguish it from other parts of the chromosome. For instance, centromeric chromatin contains a unique H3 histone variant, called CENP-A, which replaces conventional H3 in a certain fraction of the centromeric nucleosomes. During the formation of mitotic chromosomes, the CENP-A-containing nucleosomes become situated on the outer surface of the centromere where they serve as the platform for the assembly of the kinetochore. The kinetochore, in turn, serves as the attachment site for the microtubules that separate the chromosomes during cell division (see Figure 14.16). Chromosomes lacking CENP-A fail to assemble a kinetochore and are lost during cell division.

It has been suggested in previous chapters that DNA sequences responsible for essential cellular functions tend to be conserved. It came as a surprise, therefore, to discover that centromeric DNA exhibited marked differences in nucleotide sequence, even among closely related species. This finding suggests that the DNA sequence itself is not an important determinant of centromere structure and function, a conclusion that is strongly supported by the following studies on humans. Approximately 1 in every 2000 humans is born with cells that have an excess piece of chromosomal DNA that forms an additional diminutive chromosome, called a *marker chromosome*. In some cases, marker chromosomes are devoid of α-satellite DNA, yet they still contain a primary constriction and a fully functional centromere that allows the duplicated marker chromosomes to be separated normally into daughter cells at each division. Clearly, some other DNA sequence in these marker chromosomes becomes "selected" as the site to contain CENP-A and other centromeric proteins. The centromere appears at the same site on a marker chromosome in all of the person's cells, indicating that the property is transmitted to the daughter chromosomes during cell division. In one study, marker chromosomes were found to be transmitted stably through three generations of family members.

Epigenetics: There's More to Inheritance than DNA

As described in the previous paragraph, α-satellite DNA is not required for the development of a centromere. In fact, dozens of unrelated DNA sequences have been found at the centromeres of marker chromosomes. It is not the DNA that indelibly marks the site as a centromere but the CENP-A-containing chromatin that it contains. These findings raise a larger issue. Not all inherited traits are dependent on DNA sequences. Inheritance that is not encoded in DNA sequence is referred to as **epigenetic** as opposed to *genetic*. The inactivation of the X chromosome discussed on page 486 is another example of an epigenetic phenomenon: the two X chromosomes can have identical DNA sequences, but one is inactivated and the other is not. Furthermore, the state of inactivation is transmitted from each cell to its daughters throughout the life of the person. However, unlike genetic inheritance, an epigenetic state can usually be reversed; X chromosomes, for example, are reactivated prior to formation of gametes. Inappropriate changes in epigenetic state are associated with numerous diseases. There is also evidence to suggest that differences in disease susceptibility and longevity between genetically identical twins may be due, in part, to epigenetic differences that appear between the twins as they age.

Biologists have discussed epigenetic phenomena for decades, but have struggled to understand (1) the mechanisms by which epigenetic information is stored and (2) the mechanisms by which an epigenetic state can be transmitted from cell to cell and from parent to offspring. In this discussion, we will focus primarily on one type of epigenetic phenomenon: the state of a cell's gene activity. Consider a stem cell residing at the base of the epidermis (as in Figure 7.1). Certain genes in these cells are transcriptionally active and others are repressed, and it is important that this characteristic pattern of gene activity is transmitted from one cell to its daughters. However, not all of the daughter cells continue life as stem cells; some of them take on a new commitment and begin the process of differentiation into mature epidermal cells. This step requires a change in the transcriptional state of that cell. Recent attention has focused on the histone code (page 487) as a critical factor in both the determination of the transcriptional state of a particular region of chromatin and its transmission to subsequent cellular generations.

When the DNA of a cell is replicated, histones associated with the DNA as part of the nucleosomes are distributed randomly to the daughter cells along with the DNA molecules. As a result, each daughter DNA strand receives roughly half of the core histones that were associated with the parental strand (see Figure 13.24). The other half of the core histones that become associated with the daughter DNA strands are recruited from a pool of newly synthesized histone molecules.

The modifications present on the histone tails in the parental chromatin are thought to determine the modifications that will be introduced in the newly synthesized histones in the daughter chromatin. We saw on page 488, for example, that heterochromatic regions of chromatin have methylated lysine residues at position 9 of histone H3. The enzyme responsible for this methylation reaction is present as one of the components of heterochromatin. As heterochromatin is replicated, this histone methyltransferase is thought to methylate the newly synthesized H3 molecules that become incorporated into the daughter nucleosomes. In this way, the methylation pattern of the chromatin, and thus its condensed heterochromatic state, is transmitted from the parent cell to its offspring. In contrast, euchromatic regions tend to contain acetylated H3 tails, and this modification is also transmitted from parental chromatin to progeny chromatin, which may serve as the epigenetic mechanism by which active euchromatic regions are perpetuated in daughter cells. Histone modifications represent one carrier of epigenetic information, covalent modifications to the DNA are another type. This latter subject is taken up on page 520.

The Nucleus as an Organized Organelle

Examination of the cytoplasm of a eukaryotic cell under the electron microscope reveals the presence of a diverse array of membranous organelles and cytoskeletal elements. Examination of the nucleus, on the other hand, typically reveals little more than scattered clumps of chromatin and one or more irregular nucleoli. As a result, researchers were left with the impression that the nucleus is largely a "sack" of randomly positioned components. With the development of new micro-
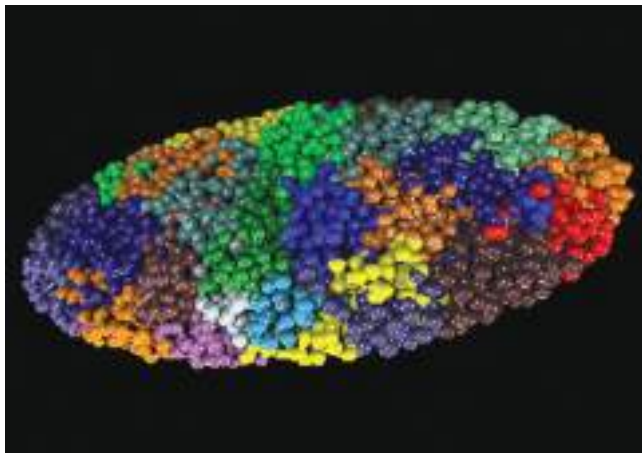
FIGURE 12.23 Three-dimensional map of all of the chromosomes present in a human fibroblast nucleus. This computer-generated image is based on fluorescence in situ hybridization analysis similar to that described for Figure 12.18*b* that allows each human chromosome to be distinguished from others and represented by an identifiable color. Each chromosome is found to occupy a distinct territory within the nucleus. (FROM ANDREAS BOLZER ET AL, PLoS BIOL. 3:E157, 2005, COURTESY OF THOMAS CREMER.)

scopic techniques, including fluorescence in situ hybridization (FISH, page 397) and imaging of live GFP-labeled cells (page 267), it became possible to localize specific gene loci within the interphase nucleus. It became evident from these studies that the nucleus maintains considerable order. For example, the chromatin fibers of a given interphase chromosome are not strewn through the nucleus like a bowl of spaghetti, but are concentrated into a distinct territory that does not overlap extensively with the territories of other chromosomes (Figure 12.23).

Even though genes are typically transcribed while they reside within their territories, individual chromatin fibers can extend away from these territories for considerable distances. Furthermore, DNA sequences that participate in a common biological response but reside on different chromosomes can apparently come together within the nucleus where they can influence gene transcription. Interactions between different loci have been revealed by the invention of a variety of techniques that involve "chromosome conformation capture (3C)." In this approach, cells are killed (i.e., fixed) by treatment with formaldehyde, which causes DNA sequences residing in close proximity to become covalently cross-linked to one another (they are said to be "captured"). After the fixation process, the DNA is isolated and subjected to digestion by restriction enzymes (Section 18.13), and the digestion products are analyzed to determine which DNA sequences in the genome were interacting with a given DNA sequence (the "bait" sequence) at the time of fixation. For example, a researcher might want to know which DNA sequences throughout the genome interact with the β-globin locus during the differentiation of erythrocytes in the bone marrow. DNA sequences that are found to interact using this technique are typically sequences that are present on the same chromosome. This is what you would expect, for example, between an enhancer and a promoter for the same gene, which come into close proximity as depicted in Figure 12.44. But numerous examples have also been found where the interacting DNA sequences are present on different chromosomes. A good example of such interchromosomal interactions comes from a study in which human cultured breast cells (both normal and malignant versions) were treated with the hormone estrogen (Figure 12.24). Estrogen induces the transcription of a large number of target genes through its binding to an estrogen receptor (ERα). Two of estrogen's target genes in humans are *GREB1*, located on chromosome 2, and *TRFF1*, located on chromosome 21. Prior to treating the cells with estrogen, the chromosome 2 and 21 territories are at distant locations from one another, and the *GREB1* and *TRFF1* gene loci are tucked away within the interior of their respective chromosome territories (Figure 12.24*a*.). However, within minutes after the cells are exposed to estrogen, these two chromosome territories move into close physical proximity to one another, and the two gene loci become localized together (colocalized) on the periphery of their territories (Figure 12.24*b*). These studies support the idea that genes are physically moved to sites within the nucleus called transcription factories, where the transcription machinery is concentrated, and that genes

(a)

(b)

(c)
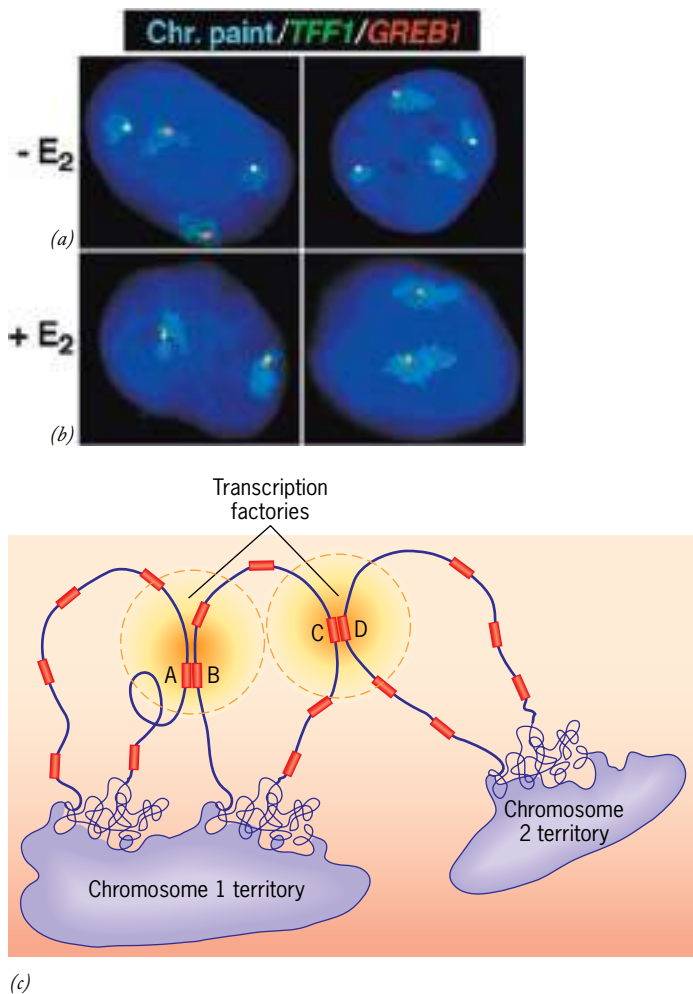
FIGURE 12.24 Interactions can occur between distantly located genes in response to physiological stimuli. (*a*) Two cells derived from a breast cancer line that have not been treated with hormone. The chromosome 2 and 21 territories, which have been fluorescently labeled in each cell, are positioned independently of one another within the nucleus. (*b*) Two of the same types of breast cancer cells visualized 60 minutes after estrogen treatment. Interactions between the chromosome 2 and 21 territories are now evident. Close examination shows the colocalization of the *TFF1* locus (in green) on chromosome 21 and the *GREB1* locus on chromosome 2 (in red). In this study, approximately half of the hormone-treated cells exhibited biallelic interactions (i.e., both *TFF1* alleles interacting with *GREB1* alleles) as depicted here. (*c*) A drawing illustrating how genes on different regions of the same chromosome (A and B), or on different chromosomes (C and D), can come together within the nucleus. In some cases, DNA sequences from distant loci may influence one another's transcription activity, and in other cases these distant genes may simply share a common pool of proteins involved in the transcription process. (FROM QIDONG HU ET AL., COURTESY OF MICHAEL G. ROSENFELD, PROC. NAT'L ACAD. SCI. U.S.A. 105: 19200, 2008.)

(a)   2 μm

(b)   1.5 μm

FIGURE 12.25 Nuclear compartmentalization of the cell's mRNA processing machinery. (*a*) The nucleus of a cell stained with fluorescent antibodies against one of the factors involved in processing pre-mRNAs. The mRNA processing machinery is localized to approximately 30 to 50 discrete sites, or "speckles." The cell shown in this micrograph had been infected with cytomegalovirus, whose genes (shown as a green dot) are being transcribed near one of these domains. (*b*) Cultured cells were transfected with a virus, and transcription was activated by addition of cyclic AMP. Images are seen at various times after transcriptional activation. The site of transcription of the viral genome in this cell is indicated by the arrows. This site was revealed at the end of the experiment by hybridizing the viral RNA to a fluorescently labeled probe (indicated by the white arrow in the fourth frame). Pre-mRNA splicing factors (orange) form a trail from existing speckles in the direction of the genes being transcribed. (A: FROM TOM MISTELI AND DAVID L. SPECTOR, CURR. OPIN. CELL BIOL. 10:324, 1998; B: REPRINTED WITH PERMISSION FROM TOM MISTELI, JAVIER F. CÁCERES, AND DAVID L. SPECTOR, NATURE 387:525, 1997; COPYRIGHT 1997, MACMILLAN MAGAZINES LIMITED.)

involved in the same response tend to become colocalized in the same factory (Figure 12.24*c*).

Another example of the interrelationship between nuclear organization and gene expression is illustrated in Figure 12.25*a*. This micrograph shows a cell that has been stained with a fluorescent antibody against one of the protein factors

involved in pre-mRNA splicing. Rather than being spread uniformly throughout the nucleus, the processing machinery is concentrated within 20 to 50 irregular domains, referred to as "speckles." According to current opinion, these speckles function as dynamic storage depots that supply splicing factors for use at nearby sites of transcription. The green dot in the nucleus in Figure 12.25*a* is a viral gene that is being transcribed near one of the speckles. The micrographs of Figure 12.25*b* show a trail of splicing factors extending from a speckle domain toward a nearby site where pre-mRNA synthesis has recently been activated. The various structures of the nucleus, including nucleoli and speckles, are dynamic, steady-state compartments whose existence depends on their

continued activity. If that activity is blocked, the compartment simply disappears as its materials are dispersed into the nucleoplasm.

In addition to nucleoli and speckles, several other types of nuclear bodies (e.g., Cajal bodies, GEMs, and PML bodies) are often seen under the microscope. Each of these nuclear bodies contains large numbers of proteins that move in and out of the structure in a dynamic manner. Because none of these nuclear bodies are bounded by a membrane, no special transport mechanisms are required for these large-scale movements. Various functions have been attributed to these nuclear structures, but they remain poorly defined. Moreover, they appear not to be essential for cell viability and will not be discussed further.
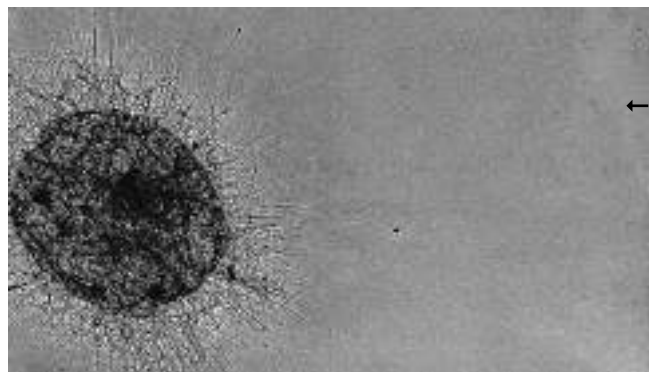
(a)   5 μm
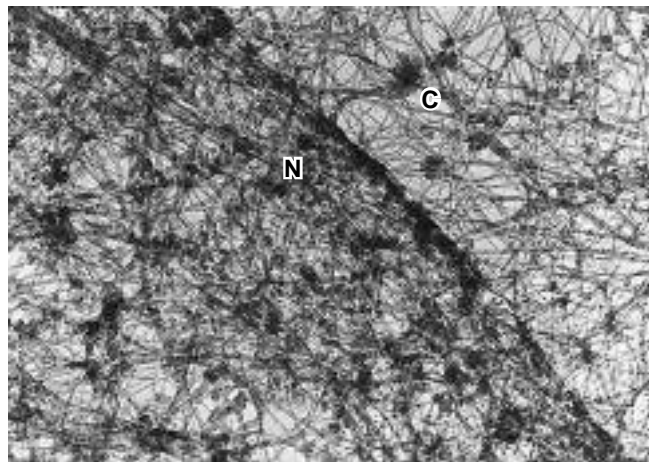
(b)   0.2 μm

FIGURE 12.26 The nuclear matrix. (*a*) Electron micrograph of a nucleus isolated in the presence of detergent and 2 M salt, which leaves the nuclear matrix surrounded by a halo of DNA loops. Arrow denotes the outer limit of the DNA loops. (*b*) Electron micrograph of a portion of a mouse fibroblast that has been extracted with detergent and depleted of its chromatin and DNA by treatment with nucleases and high salt. The nucleus (N) is seen to consist of a residual filamentous matrix whose elements terminate at the site of the nuclear envelope. The cytoplasm (C) contains a different cytoskeletal matrix whose structure is discussed in Chapter 9. (A: REPRINTED WITH PERMISSION FROM D. A. JACKSON, S. J. MCCREADY, AND P. R. COOK, NATURE 292:553, 1981; © COPYRIGHT 1981, MACMILLAN MAGAZINES LIMITED; B: FROM DAVID G. CAPCO, KATHERINE M. WAN, AND SHELDON PENMAN, CELL 29:851, 1982; BY PERMISSION OF CELL PRESS.)

The Nuclear Matrix   When isolated nuclei are treated with nonionic detergents and high salt (e.g., 2 M NaCl), which remove lipids and nearly all of the histone and nonhistone proteins of the chromatin, the DNA is seen as a halo surrounding a residual nuclear core (Figure 12.26*a*). If the DNA fibers are subsequently digested with DNase, the structure that remains possesses the same shape as the original nucleus but is composed of a network of thin protein-containing fibrils crisscrossing through the nuclear space (Figure 12.26*b*). This insoluble fibrillar network has been named the **nuclear matrix**. The nuclear matrix has been a subject of major contention in cell biology, with a number of researchers arguing that the protein network seen in Figure 12.26 is an artifact of preparation.

According to its many proponents, the nuclear matrix serves as more than a skeleton to maintain the shape of the nucleus or a scaffold on which to organize loops of chromatin (page 484); it also serves to anchor much of the machinery that is involved in the various activities of the nucleus, including transcription, RNA processing, and replication. For instance, if cells are incubated with fluorescently or radioactively labeled RNA or DNA precursors for a brief period, nearly all of the newly synthesized nucleic acid is found to be associated with fibrils of the type shown in Figure 12.26.

REVIEW ?

1. Describe the components that make up the nuclear envelope. What is the relationship between the nuclear membranes and the nuclear pore complex? How does the nuclear pore complex regulate bidirectional movement of materials between the nucleus and cytoplasm?

2. What is the relationship between the histones and DNA of a nucleosome core particle? How was the existence of nucleosomes first revealed? How are nucleosomes organized into higher levels of chromatin?

3. What is the difference in structure and function between heterochromatin and euchromatin? Between constitutive and facultative heterochromatin? Between an active and inactivated X chromosome in a female mammalian cell? How is the histone code a determinant of the state of a chromatin region?

4. What is the difference in structure and function between the centromeres and telomeres of a chromosome?

5. Describe some of the observations that suggest that the nucleus is an ordered compartment.

12.2 CONTROL OF GENE EXPRESSION IN BACTERIA

A bacterial cell lives in direct contact with its environment, which may change dramatically in chemical composition from one moment to the next. At certain times, a particular compound may be present, while at other times that compound
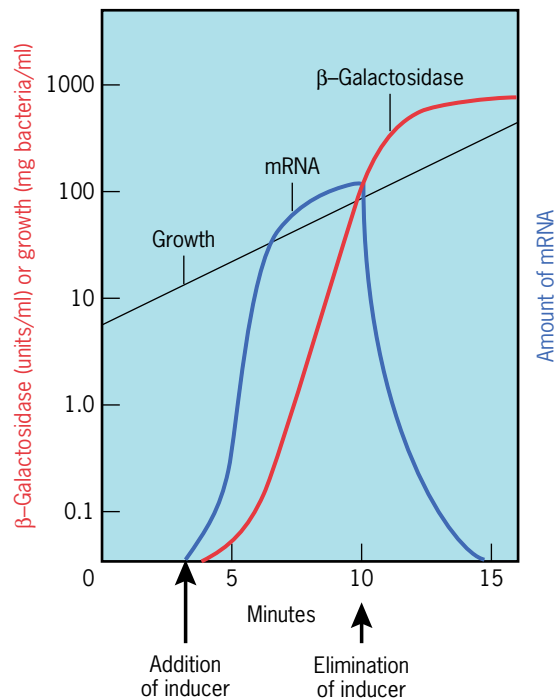
FIGURE 12.27 The kinetics of β-galactosidase induction in *E. coli*. When an appropriate inducer (a β-galactoside) is added, the production of mRNA for the enzyme β-galactosidase begins very rapidly, followed within a minute or so by the appearance of the enzyme, whose concentration increases rapidly. Removal of the inducer leads to a precipitous drop in the level of the mRNA, which reflects its rapid degradation. The amount of enzyme then levels off because new molecules are no longer synthesized.

is absent. Consider the consequences of transferring a culture of bacteria from a minimal medium to one containing either (1) lactose or (2) tryptophan.

1. Lactose is a disaccharide (see Figure 2.16) composed of glucose and galactose whose oxidation can provide the cell with metabolic intermediates and energy. The first

step in the catabolism (i.e., degradation) of lactose is the hydrolysis of the bond (a β-galactoside linkage) that joins the two sugars, a reaction catalyzed by the enzyme β-galactosidase. When bacterial cells are growing under minimal conditions, the cells have no need for β-galactosidase. Under minimal conditions, an average cell contains fewer than five copies of β-galactosidase and a single copy of the corresponding mRNA. Within a few minutes after adding lactose to the culture medium, cells contain approximately 1000 times the number of β-galactosidase molecules. The presence of lactose has induced the synthesis of this enzyme (Figure 12.27).

2. Tryptophan is an amino acid required for protein synthesis. In the absence of this compound in the medium, a bacterium must expend energy synthesizing this amino acid. Cells growing in the absence of tryptophan contain the enzymes, and their corresponding mRNAs, needed for tryptophan manufacture. If, however, this amino acid should become available in the medium, the cells no longer have to synthesize their own tryptophan, and within a few minutes, the production of the enzymes of the tryptophan pathway stops. In the presence of tryptophan, the genes that encode these enzymes are repressed.

The Bacterial Operon

In bacteria, the genes that encode the enzymes of a metabolic pathway are usually clustered together on the chromosome in a functional complex called an **operon**. All of the genes of an operon are coordinately controlled by a mechanism first described in 1961 by Francois Jacob and Jacques Monod of the Pasteur Institute in Paris. A typical bacterial operon consists of structural genes, a promoter region, an operator region, and a regulatory gene (Figure 12.28).

- **Structural genes** code for the enzymes themselves. The structural genes of an operon usually lie adjacent to one another, and the RNA polymerase moves from one structural gene to the next, transcribing all of the genes into a
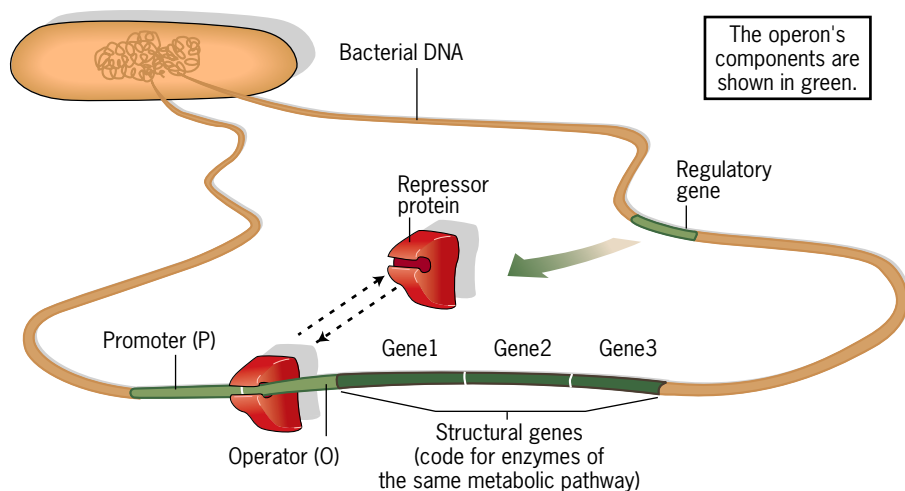
FIGURE 12.28 Organization of a bacterial operon. The enzymes that make up a metabolic pathway are encoded by a series of structural genes that reside in a contiguous array within the bacterial chromosome. All of the structural genes of an operon are transcribed into a continuous mRNA, which is translated into separate polypeptides. Transcription of the structural genes is controlled by a repressor protein that is synthesized by a regulatory gene. When bound to the operator site of the DNA, the repressor protein blocks movement of the RNA polymerase from the promoter to the structural genes.

single mRNA. This extended mRNA is then translated into the various individual enzymes of the metabolic pathway. Consequently, turning on one gene turns on all the enzyme-producing genes of an operon.

- The **promoter** is the site where the RNA polymerase binds to the DNA prior to beginning transcription (discussed on page 426).

- The **operator**, which typically resides adjacent to or overlaps with the promoter (see Figure 12.30), serves as the binding site for a protein, called the **repressor**. The repressor is an example of a **gene regulatory protein**—a protein that recognizes a specific sequence of base pairs within the DNA and binds to that sequence with high affinity. As will be evident in the remaining sections of this chapter, DNA-binding proteins, such as bacterial repressors, play a predominant role in determining whether or not a particular segment of the genome is transcribed.

- The **regulatory gene** encodes the repressor protein.

The key to operon expression lies in the repressor. When the repressor binds the operator (Figure 12.29), the promoter is shielded from the polymerase, and transcription of the structural genes is prevented. The capability of the repressor to bind the operator and inhibit transcription depends on the conformation of the repressor, which is regulated allosterically by a key compound in the metabolic pathway, such as lactose or tryptophan, as described shortly. It is the concentration of this key metabolic substance that determines whether the operon is active or inactive at any given time.

The *lac* Operon   The interplay among these various elements is illustrated by the *lac operon*—the cluster of genes that regulates production of the enzymes needed to degrade lactose in bacterial cells. The *lac* operon is an example of an **inducible operon**, that is, one in which the presence of a key metabolic substance (in this case, lactose) induces transcription of the structural genes (Figure 12.29*a*). The *lac* operon contains three tandem structural genes: the *z* gene, which encodes β-galactosidase; the *y* gene, which encodes galactoside permease, a protein that promotes the entry of lactose into the cell; and the *a* gene, which encodes thiogalactoside transacetylase, an enzyme whose physiologic role is unclear. If lactose is present in the medium, the disaccharide enters the cell where it binds to the *lac* repressor, changing the conformation of the repressor and making it unable to attach to the DNA of the operator. In this state, the structural genes are transcribed, the enzymes are synthesized, and the lactose molecules are catabolized. Thus, in an inducible operon, such as the *lac* operon, the repressor protein is able to bind to the DNA only in the absence of lactose, which functions as the **inducer**.[4] As the concentration of lactose in the medium decreases, the disaccharide dissociates from its binding site on the repressor molecule. Release of lactose enables the repressor to bind to the operator, which physically blocks the polymerase from reaching the structural genes, turning off transcription of the operon.

Positive Control by Cyclic AMP   Repressors, such as those of the *lac* and *trp* operons, exert their influence by *negative control*, as the interaction of the DNA with this protein inhibits gene expression. The *lac* operon is also under *positive control*, as was discovered during an early investigation of a phenomenon called the *glucose effect*. If bacterial cells are supplied with glucose as well as a variety of other substrates, such as lactose or galactose, the cells catabolize the glucose and ignore the other compounds. The glucose in the medium acts to suppress the production of various catabolic enzymes, such as β-galactosidase, needed to degrade these other substrates. In 1965, a surprising finding was made: cyclic AMP (cAMP), previously thought to be involved only in eukaryotic metabolism, was detected in cells of *E. coli*. It was found that the concentration of cAMP in the cells was related to the presence of glucose in the medium; the higher the glucose concentration, the lower the cAMP concentration. Furthermore, when cAMP was added to the medium in the presence of glucose, the catabolic enzymes that were normally absent were suddenly synthesized by the cells.

Although the exact means by which glucose lowers the concentration of cAMP has still not been elucidated, the mechanism by which cAMP overcomes the effect of glucose is well understood. As might be expected, a small molecule such as cAMP (see Figure 15.11) could not serve on its own to stimulate the expression of a specific battery of genes. As in eukaryotic cells, cAMP acts in prokaryotic cells by binding to a protein, in this case the *cAMP receptor protein* (*CRP*). By itself, CRP is unable to bind to DNA. However, the cAMP–CRP complex recognizes and binds to a specific site in the *lac* control region (Figure 12.30). The presence of the bound CRP causes a change in the conformation of the DNA, which makes it possible for RNA polymerase to transcribe the *lac* operon. Thus the presence of the bound cAMP–CRP complex is necessary for transcription of the operon, even when lactose is present and the repressor is inactivated. As long as glucose is abundant, cAMP concentrations remain below that required to promote transcription of the operon.

The *trp* Operon   In a *repressible* operon, such as the tryptophan (or *trp*) operon, the repressor is unable to bind to the operator DNA by itself. Instead, the repressor is active as a DNA-binding protein only when complexed with a specific factor, such as tryptophan (Figure 12.29*b*), which functions as a *corepressor*. In the absence of tryptophan, the operator site is open to binding by RNA polymerase, which transcribes the structural genes of the *trp* operon and leads to the production of the enzymes that synthesize tryptophan. Once tryptophan becomes available, the enzymes of the tryptophan synthetic pathway are no longer required. Under these conditions, the increased concentration of tryptophan leads to the formation of the tryptophan–repressor complex, which blocks transcription.

---

[4]The actual inducer is allolactose, which is derived from and differs from lactose by the type of linkage joining the two sugars. This feature is ignored in the discussion.
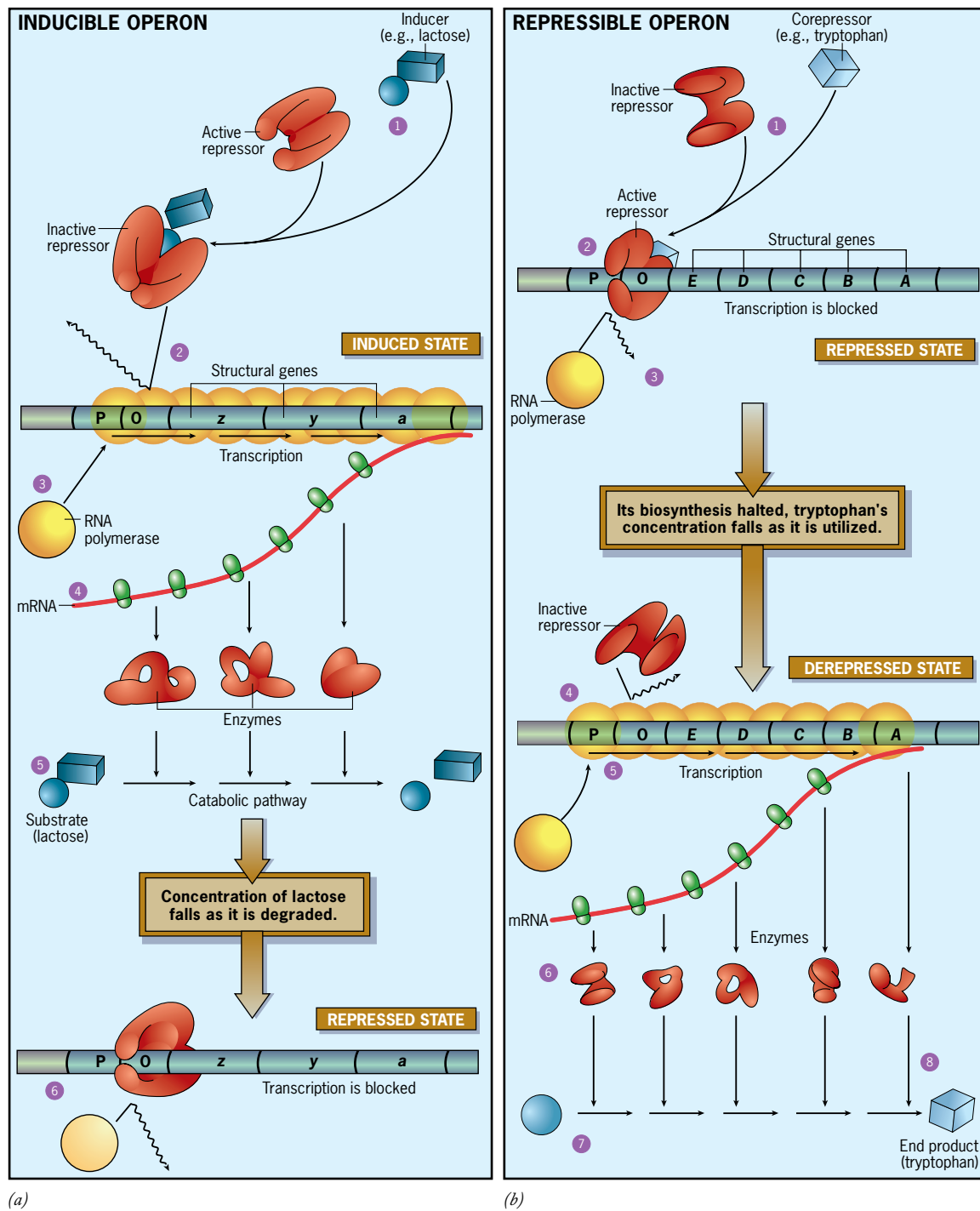
(a)   *(b)*

FIGURE 12.29 Gene regulation by operons. Inducible and repressible operons work on a similar principle: if the repressor is able to bind to the operator, genes are turned off; if the repressor is inactivated and unable to bind to the operator, genes are expressed. (*a*) In an inducible operon, (1) the inducer (in this case, the disaccharide lactose) binds to the repressor protein and (2) prevents its attachment to the operator (O). (3) Without the repressor in the way, RNA polymerase attaches to the promoter (P) and (4) transcribes the structural genes. Thus, when the lactose concentration is high, the operon is induced, and the needed sugar-digesting enzymes are manufactured. (5) As the sugar is metabolized, its concentration dwindles (6), causing bound inducer molecules to dissociate from the repressor, which then regains its ability to attach to the operator and prevent transcription. Thus, when the inducer concentration is low, the operon is repressed, preventing synthesis of unneeded enzymes. (*b*) In a repressible operon, such as the *trp* operon, the repressor, by itself, is *unable* to bind to the operator, and the structural genes coding for the enzymes are actively transcribed. The enzymes of the *trp* operon catalyze the reactions in which this essential enzyme is synthesized. (1) When available, tryptophan molecules act as corepressors by binding to the inactive repressor and (2) changing its shape, allowing it to attach to the operator, (3) preventing transcription of the structural genes. Thus, when tryptophan concentration is high, the operon is repressed, preventing overproduction of tryptophan. (4) When the tryptophan concentration is low, most repressor molecules lack a corepressor and therefore fail to attach to the operator. (5) Genes are transcribed, (6) enzymes are synthesized, and (7) the needed end-product (tryptophan) is manufactured (8).
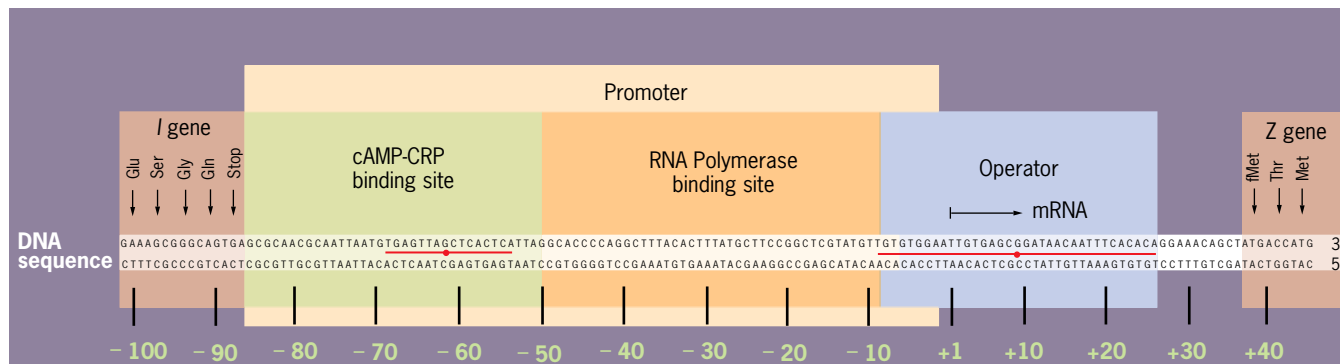
FIGURE 12.30 The nucleotide sequence of binding sites in the control region of the *lac* operon. The promoter region contains the binding site for the CRP protein as well as the RNA polymerase. The site of initiation of transcription is denoted as +1, which is approximately 40 nucleotides upstream from the site at which translation is initiated. Regions of sequence symmetry in the CRP site and operator are indicated by the red horizontal line. (REPRINTED WITH PERMISSION FROM R. C. DICKSON ET AL., SCIENCE 187:32, 1975; © COPYRIGHT 1975, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

Riboswitches

Over the past few years a different type of mechanism has captured the interest of researchers studying bacterial gene regulation. It has been found that proteins, such as the *lac* and *trp* repressors, are not the only gene regulatory molecules that are influenced by interaction with small metabolites. A number of bacterial messenger RNAs have been identified that can bind a small metabolite, such as glucosamine or adenine, with remarkable specificity. The metabolite binds to a highly structured 5′ noncoding region of the mRNA. Once bound to the metabolite, these mRNAs, or **riboswitches**, as they are called, undergo a change in their folded conformation that allows them to alter the expression of a gene involved in production of that metabolite. Most riboswitches suppress gene expression by blocking either termination of transcription or initiation of translation. Like the repressors that function in conjunction with operons, riboswitches allow cells to adjust their level of gene expression in response to changes in the available levels of certain metabolites. Given that they act without the participation of protein cofactors, riboswitches are likely another legacy from an ancestral RNA world (page 448). One type of riboswitch has been discovered in plants and fungi, and a search for this new type of gene regulation in animal cells is underway.

REVIEW ❓

1. Describe the cascade of events responsible for the sudden changes in gene expression in a bacterial cell following the addition of lactose. How does this compare with the events that occur in response to the addition of tryptophan?

2. What is the role of cyclic AMP in the synthesis of β-galactosidase?

3. What is a riboswitch?

12.3 CONTROL OF GENE EXPRESSION IN EUKARYOTES

In addition to possessing a genome containing tens of thousands of genes, complex plants and animals are composed of many different types of cells. Vertebrates, for example, are composed of hundreds of different cell types, each far more complex than a bacterial cell and each requiring a distinct battery of proteins that enable it to carry out specialized activities.

There was a time when biologists thought that cells achieved a particular differentiated state by retaining those parts of the chromosomes required for the functions of that cell type while eliminating those parts of the chromosomes that would not be needed. This idea that differentiation was accompanied by the loss of genetic information was finally put to rest in the 1950s and 1960s by a series of key experiments in both plants and animals demonstrating that differentiated cells retain the genes required to become any other cell in that organism. As one example, Frederick Steward and his colleagues at Cornell University demonstrated that a root cell isolated from a mature plant could be induced to grow into a fully developed plant that contained all the various cell types normally present.

Although single cells from an adult animal are not able to give rise to new individuals, the nuclei from such cells have been shown to contain all the information necessary for the development of a new organism. This was most dramatically demonstrated in 1997 when a team of Scottish researchers reported the first cloning of a mammal—a lamb they named Dolly. To achieve this controversial feat, the researchers prepared two types of cells: (1) unfertilized sheep's eggs whose chromosomes had been removed and (2) cultured cells derived from the mammary gland (udder) of an adult sheep. Each of the enucleated eggs was then fused with one of the cultured cells (Figure 12.31). Cell fusion was accomplished by bringing the two types of cells into contact and subjecting them to
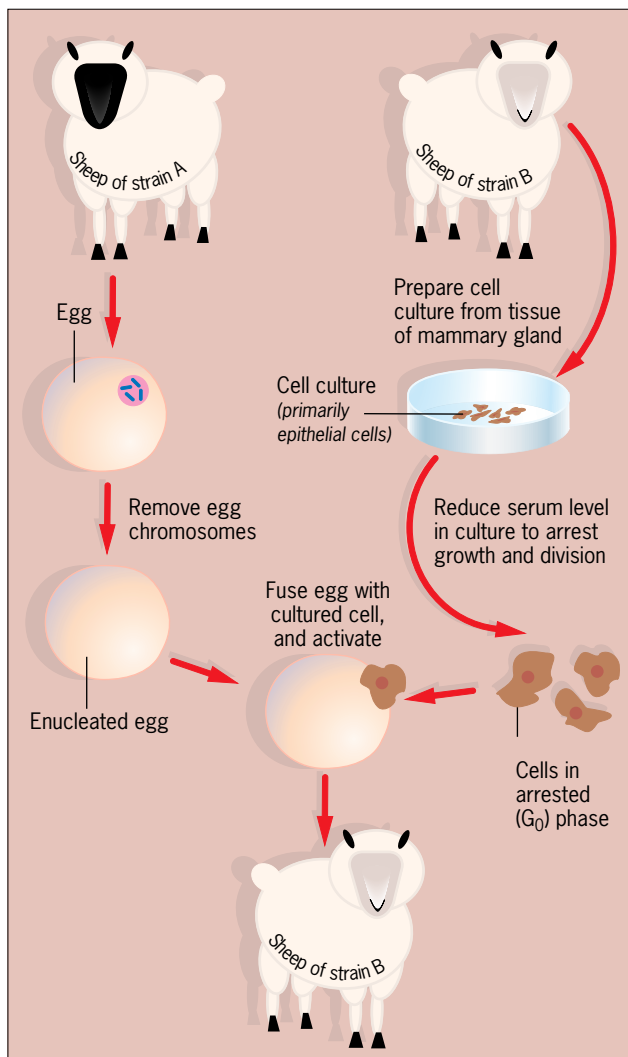
FIGURE 12.31 **The cloning of animals demonstrates that nuclei retain a complete complement of genetic information.** In this experiment, an enucleated egg from a sheep of one breed was fused with a mammary gland cell from a female of another breed. The activated egg developed into a healthy lamb. Because all of the genes in the newborn lamb had to have been derived from the transplanted nucleus (which was demonstrated by use of genetic markers), this experiment confirms the widely held belief that differentiated cells retain all of the genetic information originally present in the zygote. [The primary difficulty that is generally encountered in nuclear transplantation experiments occurs when the nucleus from an active somatic (nongerm) cell is suddenly plunged into the cytoplasm of a relatively inactive egg. To avoid damaging the donor nucleus, the cultured cells were forced into a quiescent state (called G_0) by drastically reducing the content of serum in the culture medium.]

a brief electric pulse, which also served to stimulate the egg to begin embryonic development. This procedure accomplishes, in essence, the transplantation of a nucleus from an adult cell into an egg lacking its own genetic material. Using the genetic instructions supplied by its new nucleus, one egg developed into a lamb containing all of the fully differentiated cells normally found in this animal. Since the birth of Dolly, researchers have successfully cloned over a dozen other

mammals, including mice, cattle, goats, pigs, rabbits, and cats (see Figure 12.13c).

It is evident from these experiments that nuclei of specialized cells, such as those from a plant root or an animal gland, contain the genetic information required for differentiation of other types of cells. In fact, every cell in your body contains a complete set of genes. Consequently, it is not the presence or absence of genes in a cell that determines the properties of that cell, but how those genes are utilized. Those cells that become liver cells, for example, do so because they express a specific set of "liver genes," while at the same time repressing those genes whose products are not involved in liver function. The success of the cloning experiments described above led to the conclusion that the transcriptional state of a differentiated cell, which in turn is dependent on the state of its chromatin, is not irreversible. During a cloning experiment, the nucleus of a differentiated cell is able to stop expressing the genes of the adult tissue from which it is taken and begin to selectively express the genes that are appropriate for the activated egg in which it suddenly finds itself. We can conclude from these cloning experiments that a nucleus from a differentiated cell can be *reprogrammed* by factors that reside in the cytoplasm of its new environment. The subject of selective gene expression, which resides at the heart of molecular biology, will occupy us for the remainder of this chapter.

The average bacterial cell contains enough DNA to encode approximately 3000 polypeptides, of which about one-third are typically expressed at any particular time. Compare this to a human cell that contains enough DNA (6 billion base pairs) to encode several million different polypeptides. Even though the vast majority of this DNA does not actually contain protein-coding information, it is estimated that a typical mammalian cell manufactures at least 5000 different polypeptides at any given time. Many of these polypeptides, such as the enzymes of glycolysis and the electron carriers of the respiratory chain, are synthesized by virtually all the cells of the body. At the same time, each cell type synthesizes proteins that are unique to that differentiated state. It is these proteins, more than any other components, that give each cell type its unique characteristics. Because of the tremendous amount of DNA found in a eukaryotic cell and the large number of different proteins being assembled, regulating eukaryotic gene expression is an extraordinarily complex process and is only beginning to be understood.

Consider the situation that faces a cell developing into a red blood cell inside the marrow of a human bone. Hemoglobin accounts for more than 95 percent of the cell's protein, yet the genes that code for the hemoglobin polypeptides represent less than one-millionth of its total DNA. Not only does the cell have to find this genetic needle in the chromosomal haystack, it has to regulate its expression to such a high degree that production of these few polypeptides becomes the dominant synthetic activity of the cell. Because the chain of events leading to the synthesis of a particular protein includes a number of discrete steps, there are several levels at which control might be exercised. Regulation of gene expression in eukaryotic cells occurs primarily at three distinct levels, as illustrated in the overview of Figure 12.32:
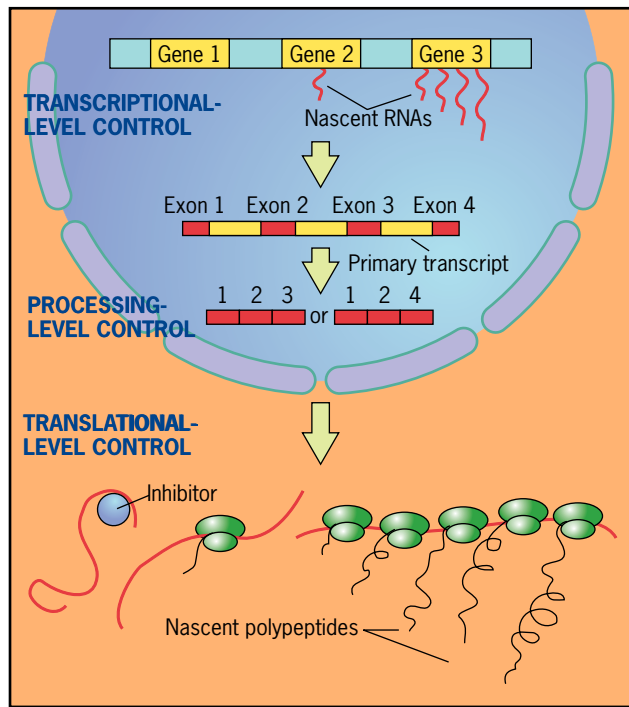
FIGURE 12.32 Overview of the levels of control of gene expression. Transcriptional-level controls operate by determining which genes are transcribed and how often; processing-level controls operate by determining which parts of the primary transcripts become part of the pool of cellular mRNAs; and translational-level controls regulate whether or not a particular mRNA is translated and, if so, how often and for how long.

1. **Transcriptional-level control** mechanisms determine whether a particular gene can be transcribed and, if so, how often.

2. **Processing-level control** mechanisms determine the path by which the primary mRNA transcript (pre-mRNA) is processed into a messenger RNA that can be translated into a polypeptide.

3. **Translational-level control** mechanisms determine whether a particular mRNA is actually translated and, if so, how often and for how long a period.

In the following sections of this chapter, we will consider each of these regulatory strategies in turn.

12.4 TRANSCRIPTIONAL-LEVEL CONTROL

As in prokaryotic cells, differential gene transcription is the single most important mechanism by which eukaryotic cells synthesize selected proteins at any given time. A large body of evidence indicates that different genes are expressed by cells at different stages of embryonic development, by cells in different tissues, and by cells that are exposed to different types of stimuli. An example of tissue-specific gene expression is shown in Figure 12.33. In this case, a gene encoding a muscle-specific protein is being transcribed in the cells of a mouse embryo that will give rise to the animal's muscle tissue.

FIGURE 12.33 Experimental demonstration of the tissue-specific expression of a gene involved in muscle cell differentiation. Transcription of the myogenin gene is activated specifically in those parts of this 11.5-day-old mouse embryo (the myotome of the somites) that will give rise to muscle tissue. The photograph shows a transgenic mouse embryo that contains the regulatory region of the myogenin gene placed upstream from a bacterial β-galactosidase gene, which acts as a *reporter*. The β-galactosidase gene is commonly employed to monitor the tissue-specific expression of a gene because the presence of the enzyme β-galactosidase is easily revealed by the blue color produced in a simple histochemical test. The activation of transcription, which results from transcription factors binding to the regulatory region of the myogenin gene, is indicated by the blue-stained cells. (FROM T. C. CHENG ET AL., COURTESY OF ERIC N. OLSON, J. CELL BIOL. 119:1652, 1992; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

Every so often a new technology is invented that fundamentally changes the ways in which certain types of problems in biology are addressed. One such methodology is that of **DNA microarrays** (or "DNA chips"), which can provide a very different type of visual portrait of transcriptional-level control. Using this technology, researchers can monitor the expression of thousands of genes expressed in a particular cell population in a single experiment.[5] Several different types of DNA microarrays have been developed. The use of one type to compare the populations of mRNAs present in yeast cells under two different growth conditions is illustrated in Figure 12.34 on page 506. Part *a* of this figure provides an outline

---

[5]As with many other important technologies in molecular biology, the use of microarrays is being largely replaced in many laboratories. As the cost of DNA sequencing has dropped dramatically in recent years, researchers can directly sequence large populations of DNAs or RNAs, rather than analyzing them indirectly by hybridization to immobilized DNAs on microarrays. This shift in technology can be seen in Figure 12.41, which shows two approaches to analyzing DNA sequences, one that employs microarrays (ChIP-chip) and the other that simply sequences the DNA fragments being analyzed. A few years ago, the ChIP-seq option was not available.
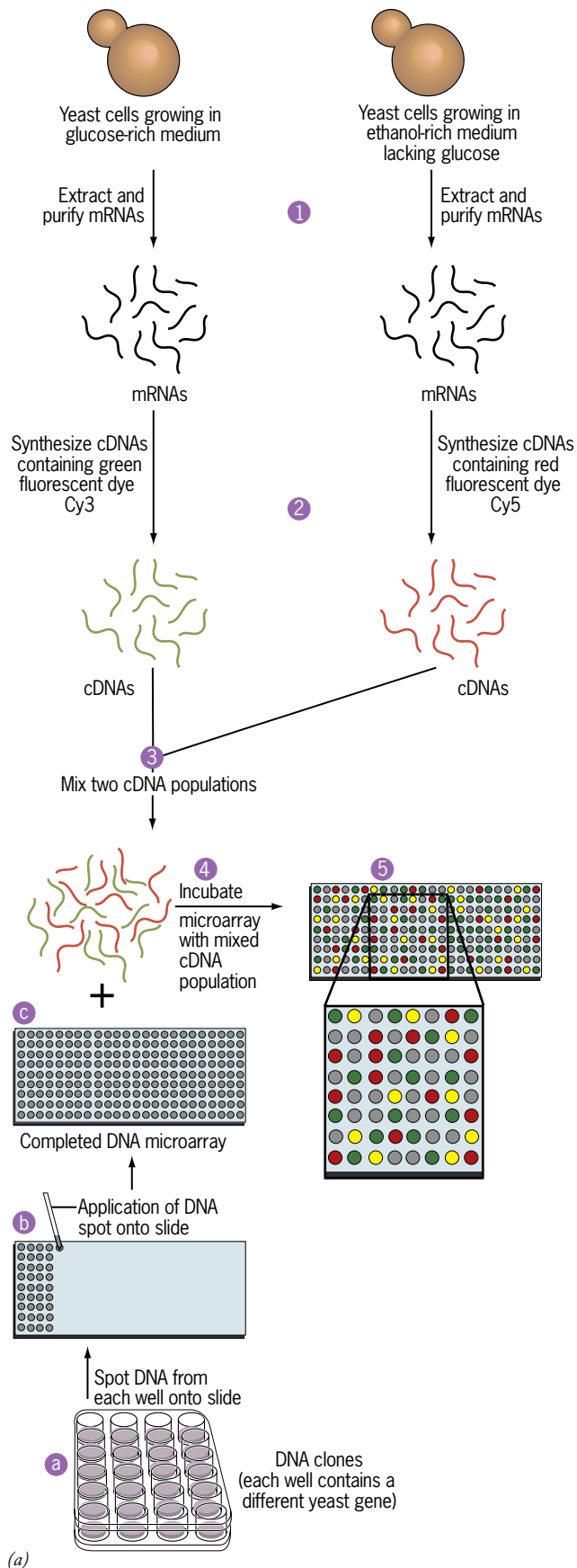
(a)
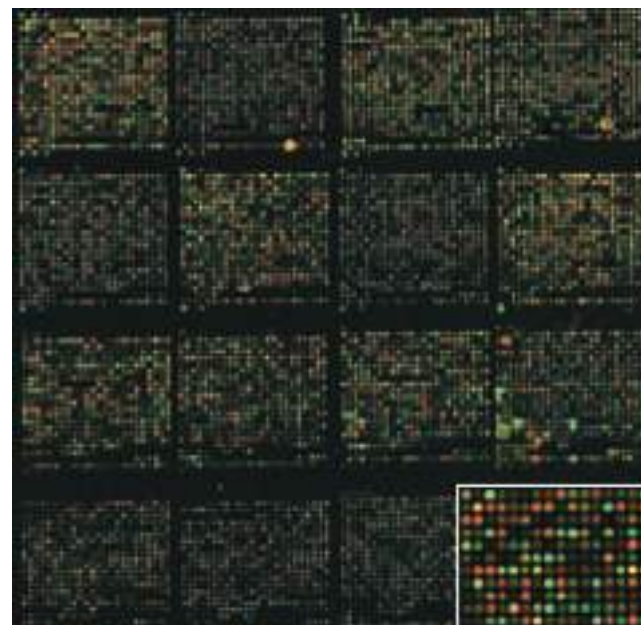
(b)
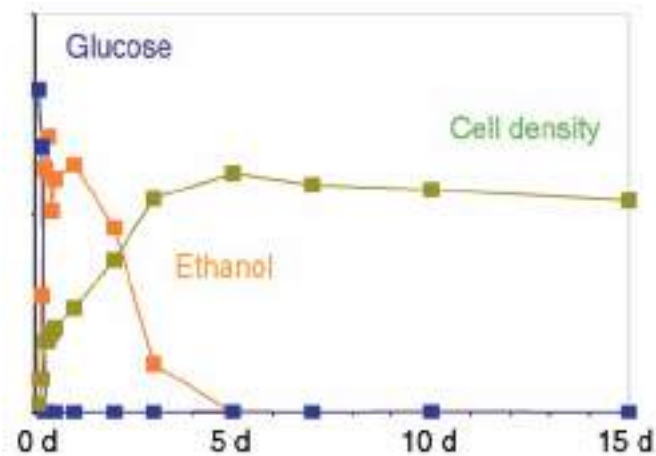
(c)
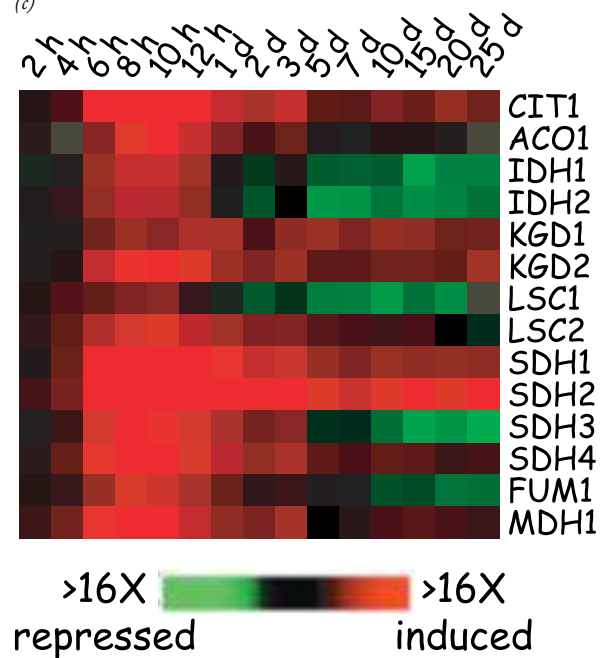
(d)

FIGURE 12.34 The construction of DNA microarrays and their use in monitoring gene transcription. (*a*) Steps in the construction of a DNA microarray. Preparation of the cDNAs (i.e., DNAs that represent the mRNAs present in a cell) used in the experiment are shown in steps 1–3. Preparation of the DNA microarray is shown in steps a–c. The cDNA mixture is incubated with the microarray in step 4 and a hypothetical result is shown in step 5. The intensity of the color of the spot is proportional to the number of cDNAs bound there. (*b*) Results of an experiment using a mixture of cDNAs representing mRNAs transcribed from yeast cells (1) in the presence of glucose (green-labeled cDNAs) and (2) in the presence of ethanol after the glucose has been depleted (red-labeled cDNAs). Those spots displaying yellow fluorescence correspond to genes that are expressed under both growth conditions. The inset at the lower right shows a close-up of a small portion of the microarray. The details of this experiment are discussed in the text. (*c*) Plot showing the changes in glucose and ethanol concentrations in the media and in cell density during the experiment. Initially, the yeast cells consume glucose, then the ethanol they had produced by fermentation, and finally they cease growth after exhausting both sources of chemical energy. (*d*) Changes in expression of genes encoding TCA-cycle enzymes during the course of the experiment. Each horizontal row depicts the level of expression of a particular gene at different time periods. The names of the genes are provided on the right (see Figure 5.7 for the enzymes). Bright red squares indicate the highest level of gene expression, bright green squares indicate the lowest level of expression. Expression of genes encoding TCA-cycle enzymes is seen to be induced as the cells begin metabolizing ethanol and repressed when the ethanol has been exhausted. (B–D: COURTESY OF PATRICK O. BROWN. SEE JOSEPH DERISI ET AL., SCIENCE 278:680, 1997 AND TRACY L. FEREA AND PATRICK O. BROWN, CURR. OPIN. GEN. DEVELOP. 9:715, 1999.)

of the basic steps taken in this type of experiment; these steps can be described briefly as follows:

1. DNA fragments representing individual genes to be studied are generated using techniques discussed in Chapter 18 (PCR, Figure 18.43; DNA cloning, Figure 18.42). In the case depicted in step a at the bottom of Figure 12.34*a*, each well on the plate holds a small volume containing a specific, cloned yeast gene. Different wells contain different genes. The cloned DNAs are then spotted, one at a time, in an ordered array on a glass slide by an automated instrument that delivers a few nanoliters of a concentrated DNA solution onto a specific spot on the slide (step b). A completed DNA microarray is shown in step c. Using this technique, DNA fragments from thousands of different genes can be spotted at known locations on a glass surface.

2. Meanwhile, the mRNAs present in the cells being studied are purified (step 1, Figure 12.34) and converted to a population of fluorescently labeled, complementary DNAs (cDNAs) (step 2). The method for preparation of cDNAs is also described in Chapter 18 (see Figure 18.45). In the example illustrated in Figure 12.34, and discussed below, cDNAs are prepared from two different cell populations, one labeled with a fluorescent green dye and the other with a fluorescent red dye. The two preparations of labeled cDNAs are mixed (step 3) and then incubated with the slide containing the immobilized DNA (step 4).

3. Those DNAs in the microarray that have hybridized to a labeled cDNA are identified by examination of the slide under the microscope (step 5). Any spot in the microarray that exhibits fluorescence represents a gene that has been transcribed in the cells being studied.

Figure 12.34*b* shows the actual results of this experiment. Each spot on the microarray of Figure 12.34*b* contains an immobilized DNA fragment from a different gene in the yeast genome. Taken together, the spots contain DNA from all of the roughly 6200 protein-coding genes present in a yeast cell. As discussed above, this particular DNA microarray was hybridized with a mixture of two different cDNA populations. One population of cDNAs, which was labeled with a green fluorescent dye, was prepared from the mRNAs of yeast cells grown in the presence of a high concentration of glucose. Unlike most cells, when baker's yeast cells are grown in the presence of glucose and oxygen, they obtain their energy by glycolysis and fermentation, rapidly converting the glucose into ethanol (Section 3.3). The other population of cDNAs, which was labeled with a red fluorescent dye, was prepared from the mRNAs of yeast cells that were growing aerobically in a medium rich in ethanol but lacking glucose. Cells growing under these conditions obtain their energy by oxidative phosphorylation, which requires the enzymes of the TCA cycle (page 180). The two cDNA populations were mixed, hybridized to the DNA in the microarray, and the slide examined in a fluorescence microscope. Genes that are active in one or the other growth media appear as either green or red spots in the microarray (step 5, Figure 12.34*a,b*). Those spots that remain devoid of color in Figure 12.34 represent genes that are not transcribed in these cells in either growth medium, whereas those spots that have a yellow fluorescence represent genes that are transcribed in cells in both types of media. A close-up of a small portion of the microarray is shown in the inset.

The experimental results shown in Figure 12.34*b* identify those genes transcribed in yeast cells under two different growth conditions. But just as importantly, they also provide information about the abundance of individual mRNAs in the cells, which is proportional to the intensity of a spot's fluorescence. A spot that displays a very strong green fluorescence, for example, represents a gene whose transcripts are abundant in yeast cells grown in glucose but is repressed in yeast cells grown in ethanol. Concentrations of individual mRNAs can vary by more than a 100-fold in yeast cells. The technique is so sensitive that an mRNA can be detected at a level of less than one copy per cell.[6]

Figure 12.34*c* shows the changes in concentration of glucose and ethanol over the course of an experiment. The glucose is rapidly metabolized by the yeast cells, leading to the disappearance of the sugar within a few hours. The ethanol produced by glucose fermentation is then gradually metabolized by the yeast cells over the next five days until it eventually disappears from the medium. Figure 12.34*d* shows the

---

[6]Differences in mRNA abundance likely reflect differences in mRNA stability (page 526) as well as differences in rates of transcription. Consequently, results obtained from DNA microarrays cannot be interpreted *solely* on the basis of transcriptional-level control.

changes in the level of expression of the genes encoding the enzymes of the TCA cycle during the course of this experiment. The levels of expression of each gene (labeled on the right) are shown in intervals of time (labeled at the top), using shades of red to represent increases in expression, and shades of green to represent decreases in expression. It is evident that transcription of genes encoding TCA enzymes is stimulated when the cells adapt to growth on a carbon source (ethanol) that is metabolized by aerobic respiration, and then repressed when the ethanol is exhausted.

DNA microarrays are currently being employed to study changes in gene expression that occur during a wide variety of biological events, such as cell division and the transformation of a normal cell into a malignant cell. It is even possible to study the diversity of RNAs being produced by a single tumor cell, once the cDNAs are amplified by PCR. Now that a number of eukaryotic genomes have been sequenced, researchers have an unlimited variety of genes whose expression can be monitored under different conditions.

DNA microarrays have many potential uses in addition to providing a visual portrait of gene expression. For example, DNA microarrays can be used to determine the degree of genetic variation in human populations or to identify the alleles for particular genes that a person carries in their chromosomes. It is hoped, one day, that this latter information may advise a person of the diseases to which they may be susceptible during their life, giving them an opportunity to take early preventive measures.

The Role of Transcription Factors in Regulating Gene Expression

Great progress has been made in elucidating how certain genes can be transcribed in a particular cell, while other genes remain inactive. Transcriptional control is orchestrated by a large number of proteins, called **transcription factors**. As discussed in Chapter 11, these proteins can be divided into two functional classes: general transcription factors that bind at core promoter sites in association with RNA polymerase (page 435), and sequence-specific transcription factors that bind to various regulatory sites of particular genes. This latter group of transcription factors can act either as *transcriptional activators* that stimulate transcription of the adjacent gene or as *transcriptional repressors* that inhibit its transcription.[7] We will focus in this section on transcriptional activators, whose job is to bind to specific regulatory sites within the DNA and initiate the recruitment of a large number of protein complexes that bring about the actual transcription of the gene itself. The steps involved in this process will occupy us for the next several pages.

Even though we know the detailed structure of many transcription factors and how they interact with their target

DNA sequences, a unified picture of their mechanism of operation remains a distant goal. It is apparent, for example, that a single gene may be controlled by many different DNA regulatory sites that bind a variety of different transcriptional factors. Conversely, a single transcription factor may become attached to numerous sites around the genome, thereby controlling the expression of a host of different genes. Each type of cell has a characteristic pattern of gene transcription, which is determined by the particular complement of transcription factors contained in that cell.

The control of gene transcription is complex and is influenced by various circumstances, including the affinity of transcription factors for particular DNA sequences and the ability of transcription factors bound at nearby sites on the DNA to interact directly with one another. An example of this type of cooperative interaction between neighboring transcription factors is illustrated in Figure 12.35 and discussed further on page 510. In a sense, the regulatory region of a gene can be thought of as a type of integration center for that gene's expression. Cells exposed to different stimuli respond by synthesizing different transcription factors, which bind to different sites in the DNA. The extent to which a given gene is transcribed presumably depends upon the particular *combination* of transcription factors bound to its upstream regulatory elements. Given that roughly 5 to 10 percent of genes encode transcription factors, it becomes apparent that a virtually unlimited number of possible combinations of interactions
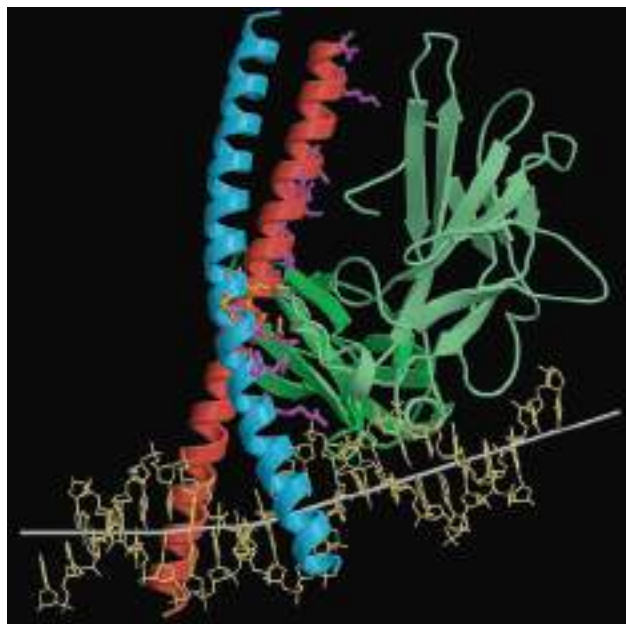
FIGURE 12.35 Interactions between transcription factors bound to different sites in the regulatory region of a gene. This image depicts the tertiary structure of two separate transcription factors, NFAT-1 (green) and AP-1 (whose two subunits are shown in red and blue) bound to the DNA upstream from a cytokine gene involved in the immune response. Cooperative interaction between these two proteins alters the expression of the cytokine gene. The gray bar traces the path of the DNA helix, which is seen to be bent as the result of this protein–protein interaction. (FROM TOM K. KERPPOLA, STRUCTURE 6:550, 1998.)

---

[7]Reports are beginning to appear in the literature in which noncoding regulatory RNAs rather than proteins are functioning to activate or repress the transcription of specific genes. Examples of such cases can be found in *Nature* 445:666, 2007; *Mol. Cell* 29:415, 2008; and *Nature Med.* 14:723, 2008. It may be evident in the future that noncoding RNAs play a major role in transcriptional-level control, just as they do at the posttranscriptional level (page 527).

among these proteins is possible. The complexity of these interactions between DNA and regulatory proteins is revealed in the marked variation in gene expression patterns between cells of different type, different tissue, different stage of development, and different physiologic state.

The Role of Transcription Factors in Determining a Cell's Phenotype

As discussed in the Human Perspective in Chapter 1, embryonic stem (ES) cells appear very early in the development of a mammalian embryo and exhibit two key properties: they are (1) capable of indefinite self-renewal and (2) pluripotent; that is, they are capable of differentiating into all of the different types of cells in the body. It had been known for several years that these pluripotent stem cells had a particular set of transcription factors that played an important role in maintaining their self-renewing, pluripotent state. Just how important these transcription factors were in the biology of ES cells was demonstrated by Shinya Yamanaka and Kazutoshi Takahashi in 2006 when they introduced the genes for 24 of these various factors into adult mouse fibroblasts. The genes were introduced (transduced) into the fibroblasts as part of viral vectors so that once the viral genome had become integrated into that of the host cell, the transcription factors were expressed. These researchers found that introducing a combination of genes encoding only four specific transcription factors—Oct4, Sox2, Myc, and Klf4—was sufficient to reprogram the fibroblasts and convert them into undifferentiated cells that behaved like pluripotent ES cells. The cellular reprogramming process does not occur within a single cell generation but occurs gradually as the cells divide in culture. As the cells become reprogrammed, the four introduced genes are silenced as the cells' endogenous genes governing pluripotency are activated. Reprogramming is accompanied by a reorganization of the chromatin in such a way that the epigenetic marks (histone modifications and DNA methylation patterns) of the differentiated cell are "erased" and the epigenetic marks characteristic of an ES cell are installed. As discussed on page 20, the induced pluripotent cells, or *iPS* cells as they are called, that were generated in these early experiments were capable of dividing indefinitely in culture and of differentiating into all of the various types of the body's cells.

The Structure of Transcription Factors

The three-dimensional structure of numerous DNA–protein complexes has been determined by X-ray crystallography and NMR spectroscopy, providing a basic portrait of the way that these two giant macromolecules interact with one another. Like most proteins, transcription factors contain different domains that mediate different aspects of the protein's function. Transcription factors typically contain at least two domains: a *DNA-binding domain* that binds to a specific sequence of base pairs in the DNA, and an *activation domain* that regulates transcription by interacting with other proteins. In addition, many transcription factors contain a surface that promotes the binding of the protein with another protein of identical or similar structure to form a dimer (Figure 12.36). The formation of dimers has proved to be a common feature of many different types of transcription factors and plays an important role in regulating gene expression.

Transcription Factor Motifs   The DNA-binding domains of most transcription factors can be grouped into several broad classes whose members possess related structures (*motifs*) that interact with DNA sequences. The existence of several families of DNA-binding proteins indicates that evolution has found a number of different solutions to the problem of constructing polypeptides that can bind to the DNA double helix. As we will see shortly, most of these motifs contain a segment (often an α helix, as in Figure 12.36) that is inserted into the major groove of the DNA, where it recognizes the sequence of base pairs that line the groove. Binding of the protein to the DNA is accomplished by a specific combination of
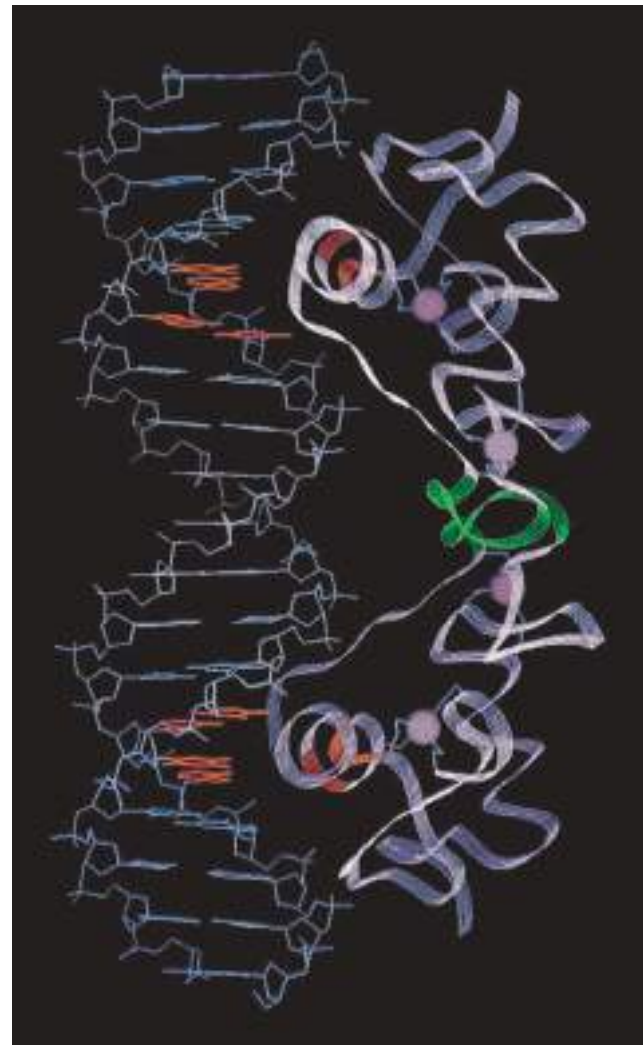
FIGURE 12.36 Interaction between a transcription factor and its DNA target sequence. A model of the interaction between the two DNA-binding domains of the dimeric glucocorticoid receptor (GR) and the target DNA. Two α helices, one from each subunit of the dimer, are symmetrically extended into two adjacent major grooves of the target DNA. Residues in the dimeric GR that are important for interactions between the two monomers are colored green. The four zinc ions (two per monomer) are shown as purple spheres. (REPRINTED WITH PERMISSION FROM T. HÄRD ET AL., COURTESY OF ROBERT KAPTEIN, SCIENCE 249:159, 1990; © COPYRIGHT 1990, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

van der Waals forces, ionic bonds, and hydrogen bonds be-tween amino acid residues and various parts of the DNA, including the backbone.

Among the most common motifs that occur in eukaryotic DNA-binding proteins are the zinc finger, the helix–loop–helix, and the leucine zipper. Each provides a structurally stable framework on which the specific DNA-recognition surfaces of the protein can be properly positioned to interact with the double helix.

1. ***The zinc-finger motif.*** The largest class of mammalian transcription factors contains a motif called the *zinc finger*. In most cases, the zinc ion of each finger is coordi-nated to two cysteines and two histidines. The two cys-teine residues are part of a two-stranded β sheet on one side of the finger, and the two histidine residues are part of a short α-helix on the opposite side of the finger (inset, Figure 12.37*a*). These proteins typically have a number of such fingers that act independently of one another and are spaced apart so as to project into successive major grooves in the target DNA, as illustrated in Figure 12.37*a*. The first zinc-finger protein to be discov-ered, TFIIIA, has nine zinc fingers (Figure 12.37*b*). Other zinc-finger transcription factors include Egr, which is involved in activating genes required for cell division, and GATA, which is involved in cardiac muscle develop-ment. Comparison of a number of zinc-finger proteins indicates that the motif provides the structural framework for a wide variety of amino acid sequences that recognize a diverse set of DNA sequences. In fact, researchers are attempting to design new species of zinc-finger proteins that are capable of targeting DNA sequences that govern the expression of particular genes of interest. It is hoped that such proteins might have the potential to act as therapeutic drugs by turning on or off the expression of disease-related genes.

2. ***The helix–loop–helix (HLH) motif.*** As the name implies, this motif is characterized by two α-helical segments sep-arated by an intervening loop. The HLH domain is often preceded by a stretch of highly basic amino acids whose positively charged side chains contact the DNA and determine the sequence specificity of the transcription factor. Proteins with this basic-HLH (or bHLH) motif always occur as dimers, as illustrated in the example of the transcription factor MyoD in Figure 12.38. The two subunits of the dimer are usually encoded by different genes, making the protein a heterodimer.

Heterodimerization greatly expands the diversity of regulatory factors that can be generated from a limited number of polypeptides (Figure 12.39). Suppose, for ex-ample, that a cell were to synthesize five different bHLH-containing polypeptides that were capable of forming heterodimers with one another in any combination; then 32 (2^5) different transcription factors that recognize 32 different DNA sequences could conceivably be formed. In actuality, combinations between polypeptides are re-stricted, not unlike the formation of heterodimeric inte-grin molecules (page 239).

(a)

(b)
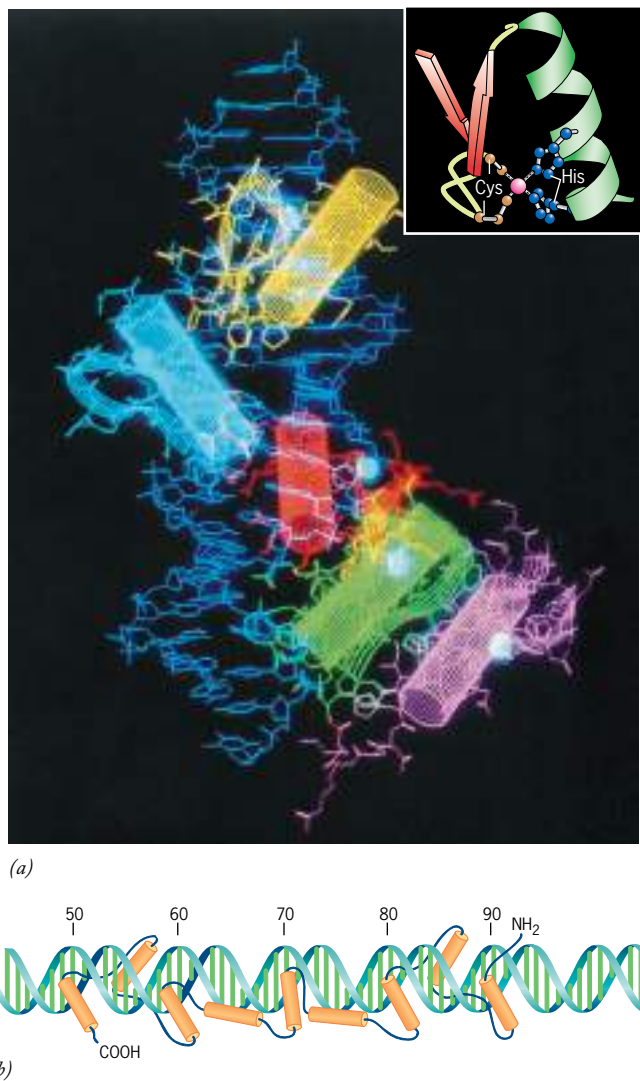
FIGURE 12.37 Transcription factors that bind by means of a zinc-finger motif. (*a*) A model of the complex between a protein with five zinc fin-gers (called GLI) and DNA. Each of the zinc fingers is colored differ-ently; the DNA is a darker blue. Cylinders and ribbons highlight the α helices and β sheets, respectively. The inset shows the structure of a single zinc finger. (*b*) A model of TFIIIA bound to the DNA of the 5S RNA gene. TFIIIA is required for the transcription of the 5S rRNA gene by RNA polymerase III. (A: Reprinted with permission from Nikola Pavletich and Carl O. Pabo, Science 261:1702, 1993; © copy-right 1993, American Association for the Advancement of Sci-ence; b: after K. R. Clemens et al., Proc. Nat'l. Acad. Sci. USA 89:10825, 1992.)

HLH-containing transcription factors play a key role in the differentiation of certain types of tissues, including skeletal muscle (as illustrated in Figure 12.33). HLH-containing transcription factors also participate in the control of cell proliferation and have been implicated in the formation of certain cancers. It was noted on page 491 that chromosome translocations can generate abnormal genes whose expression causes the cell to be-come cancerous. Genes encoding at least four different bHLH proteins (MYC, SCL, LYL-1, and E2A) have
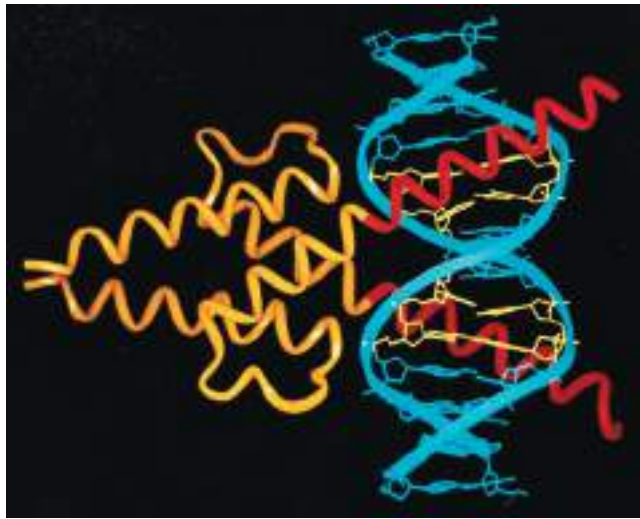
(a)
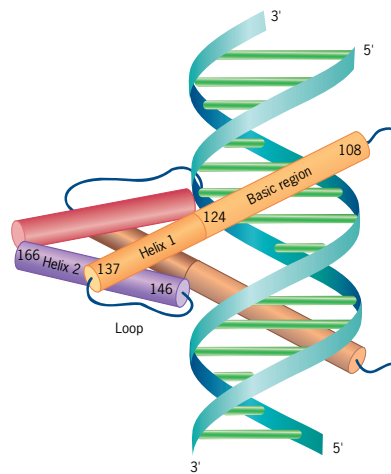
(b)

FIGURE 12.38 A transcription factor with a basic helix–loop–helix (bHLH) motif. (*a*) MyoD, a dimeric transcription factor involved in triggering muscle cell differentiation, is a bHLH protein that binds to the DNA by an associated basic region. The 14-base-pair binding site in the DNA is shown in light blue. The basic region of each MyoD monomer is red, whereas the helix–loop–helix region of each MyoD monomer is shown in brown. The DNA bases bound by the transcription factor are indicated in yellow. (*b*) A sketch of the dimeric MyoD complex in the same orientation as in part *a*. The α helices are represented as cylinders. (FROM P. C. M. MA ET AL., COURTESY OF CARL O. PABO, CELL 77:453, 1994; BY PERMISSION OF CELL PRESS.)

been found in chromosome translocations that lead to the development of specific cancers. The most prevalent of these cancers is Burkitt's lymphoma in which the *MYC* gene on chromosome 8 is translocated to a locus on chromosome 14 that contains a regulatory site for a gene encoding part of an antibody molecule. The overexpression of the *MYC* gene in its new location is thought to be a significant factor in development of this lymphoma.

3. ***The leucine zipper motif.*** This motif gets its name from the fact that leucines occur every seventh amino acid along a stretch of α helix. Because an α helix repeats every 3.5 residues, all of the leucines along this stretch of polypeptide face the same direction. Two α helices of this type are capable of zipping together to form a *coiled coil*. Thus, like most other transcription factors, proteins with a leucine zipper motif exist as dimers. The leucine zipper motif can bind DNA because it contains a stretch of basic amino acids on one side of the leucine-containing α helix. Together the basic segment and leucine zipper are referred to as a *bZIP* motif. Thus, like bHLH proteins, the α-helical portions of bZIP proteins are important in dimerization, while the stretch of basic amino acids allows the protein to recognize a specific nucleotide sequence in the DNA. AP-1, whose structure and interaction with DNA are shown in Figure 12.35, is an example of a bZIP transcription factor. AP-1 is a heterodimer whose two subunits (Fos and Jun, shown in red and blue in Figure 12.35, respectively) are encoded by the genes *FOS* and *JUN*. Both of these genes play an important role in cell proliferation and, when mutated, can cause a cell to become malignant. Mutations in either of these genes that prevent the proteins from forming heterodimers also prevent the proteins from binding to DNA, indicating the importance of dimer formation in regulating their activity as transcription factors.

DNA Sites Involved in Regulating Transcription

The complexity inherent in the control of gene transcription can be illustrated by examining the DNA in and around a single gene. In the following discussion, we will focus on the gene that codes for phosphoenolpyruvate carboxykinase (PEPCK). PEPCK is one of the key enzymes of gluconeogenesis, the metabolic pathway that converts pyruvate to glucose (see Figure 3.31). The enzyme is synthesized in the liver

A-A homodimer   B-B homodimer   A-B heterodimer
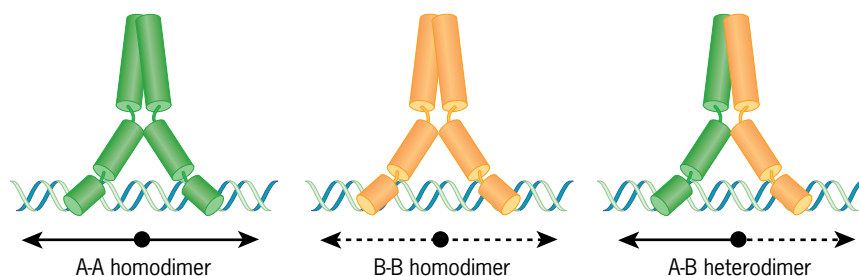
FIGURE 12.39 Increasing the DNA-binding specificities of transcription factors through dimerization. In this model of a bHLH protein, three different dimeric transcription factors that recognize different DNA-binding sites can be formed when the two subunits associate in different combinations. The human genome encodes approximately 118 different bHLH monomers, which could potentially generate thousands of different dimeric transcription factors.

when glucose levels are low, as might occur, for example, when a considerable period of time has passed since one's last meal. Conversely, synthesis of the enzyme drops sharply after ingestion of a carbohydrate-rich meal, such as pasta. The level of synthesis of *PEPCK* mRNA is controlled by a variety of different transcription factors, including a number of receptors for hormones that are involved in regulating carbohydrate metabolism. The keys to understanding the regulation of *PEPCK* gene expression lie in (1) unraveling the functions of the numerous DNA regulatory sequences that reside upstream from the gene itself, (2) identifying the transcription factors that bind these sequences, and (3) elucidating the signaling pathways that activate the machinery responsible for selective gene expression (discussed in Chapter 15).

The closest (most proximal) regulatory sequence upstream of the *PEPCK* gene is the TATA box, which is a major element of the gene's promoter (Figure 12.40). As discussed on page 435, a **promoter** is a region upstream from a gene that regulates the initiation of transcription. For our purposes, we will divide the eukaryotic promoter into separate regions, which are not that well delineated. The region that stretches from roughly the TATA box to the transcription start site is designated the *core promoter*. The core promoter is the site of assembly of a preinitiation complex consisting of RNA polymerase II and a number of general transcription factors that are required before a eukaryotic gene can be transcribed.

The TATA box is not the only short sequence that is shared by large numbers of genes. Two other promoter sequences, called the CAAT box and GC box, which are located farther upstream from the gene, are often required for a polymerase to initiate transcription of that gene. The CAAT and GC boxes bind transcription factors (e.g., NF1 and SP1)

that are found in many tissues and widely employed in gene expression. Whereas the TATA box determines the site of initiation of transcription, the CAAT and GC boxes regulate the frequency with which the polymerase transcribes the gene. The TATA, CAAT, and GC boxes, when present, are typically located within 100–150 base pairs upstream from the transcription start site. Because of their proximity to the start of the gene, these shared sequences can be referred to as *proximal promoter elements*, and are indicated in line 1 of Figure 12.40.

The more genes that are examined, the more variability that is found in the nature and locations of the promoter elements that regulate gene expression. In addition, a significant number of mammalian genes (possibly as many as 50 percent of them) have more than one promoter (i.e., *alternative promoters*) allowing initiation of transcription to occur at more than one site upstream from a gene. Alternative promoters are typically separated from one another by several hundred bases so that the primary transcripts produced by their differential usage will differ considerably at their 5′ ends. In some cases, alternative promoters are utilized in different tissues, but they lead to synthesis of the same polypeptide. In such cases, the mRNAs that encode the proteins are likely to have different 5′ UTRs, which makes them subject to distinct types of translational-level control. In other cases, alternative promoters promote the synthesis of related proteins that differ in their N-terminal amino acid sequence. In a number of cases, alternative promoters govern the transcription of mRNAs that are translated in different reading frames (page 461) to produce entirely different polypeptides.

How do researchers learn which sites in the genome interact with a particular transcription factor? The following general strategies are often employed to make this determination.
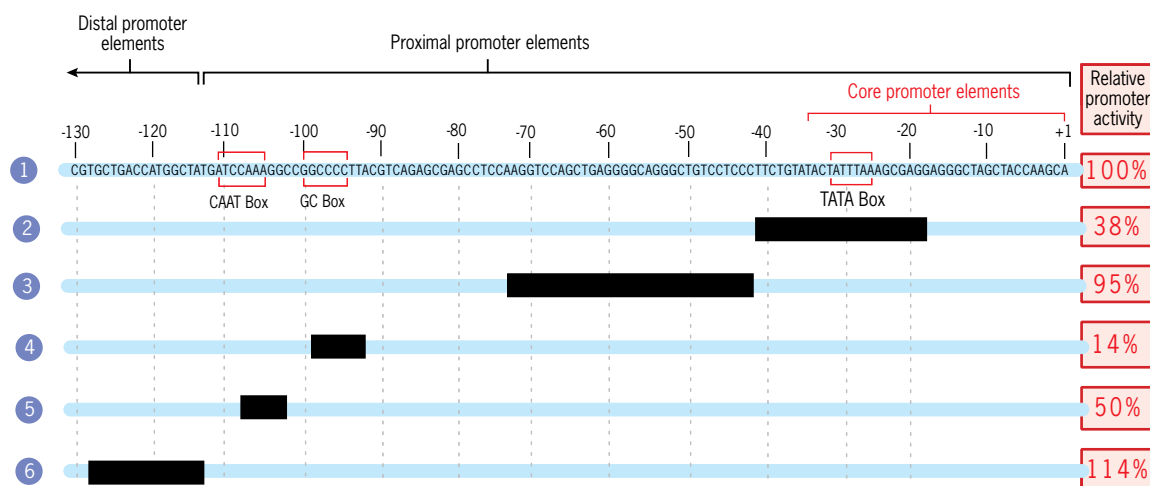
FIGURE 12.40 Identifying promoter sequences required for transcription. Line 1 shows the nucleotide sequence of one strand of the *PEPCK* gene promoter. The TATA box, CAAT box, and GC box are indicated. The other five lines show the results of experiments in which particular regions of the promoter have been deleted (indicated by black boxes) prior to transfecting cells with the DNA. The observed level of transcription of the *PEPCK* gene in each of these cases is indicated at the right. Deletions that remove all or part of the three boxes cause a marked decrease in the level of transcription, whereas deletions that affect other regions have either a lesser or no effect. (As noted in Chapter 11, many mammalian promoters lack one or more of these elements, including the TATA box. Promoters lacking the TATA box often have a conserved element at a position downstream of the start site, called the *downstream promoter element*, or *DPE*.) (DATA TAKEN FROM STUDIES BY RICHARD W. HANSON AND DARYL K. GRANNER AND THEIR COLLEAGUES.)

■ **Deletion mapping.** In this procedure, DNA molecules are prepared that contain deletions of various parts of the gene's promoter (Figure 12.40). Cells are then transfected with the altered DNA molecules; that is, they are caused to take up the DNA from the medium. Finally, the ability of the cells to transcribe the transfected DNA is determined. In many cases, the deletion of a few nucleotides has little or no effect on the transcription of an adjacent gene. However, if the deletion falls within any of the three boxes described above, the level of transcription tends to be sharply decreased (as in lines 2, 4, and 5 of Figure 12.40). Deletions in other parts of the promoter have less of an effect on transcription (lines 3 and 6 of Figure 12.40).

■ **DNA footprinting.** When a transcription factor binds to a DNA sequence, it protects that sequence from digestion by nucleases. Researchers take advantage of this property by isolating chromatin from cells and treating it with DNA-digesting enzymes, such as DNase I, that destroy those sections of the DNA that are not protected by bound transcription factors. Once the chromatin has been digested, the bound protein is removed, and the DNA sequences that had been protected are identified.

■ **Genome-wide location analysis.** As the name implies, this strategy allows researchers to simultaneously monitor all of the sites within the genome that carry out a particular activity. In this case, the goal is to identify all of the sites that are bound by a given transcription factor under a given set of physiologic conditions. An outline of the approach is shown in Figure 12.41. To carry out this analysis, cells are cultured under the desired conditions or isolated from a particular tissue or developmental stage, and then treated with an agent, such as formaldehyde, that will kill the cells and cross-link transcription factors to the DNA sites at which they were bound in the living cell (step 1, Figure 12.41). Following this cross-linking step, the cells' chromatin is isolated, broken mechanically into small fragments, and treated with an antibody that binds to the transcription factor of interest (step 2). This antibody treatment induces the precipitation of chromatin fragments containing the bound transcription factor (step 3a), while leaving all of the unbound chromatin fragments in solution (step 3b). Once this process of *chromatin immuno-precipitation* (or *ChIP*) has been carried out, the cross-links between the protein and DNA in the precipitate can be reversed and the segments of DNA can be purified (step 4). The next step is to identify where in the genome these transcription-factor binding sites are located. Two different approaches can be taken to make this determination. In most studies, the DNA fragments are identified using a DNA microarray similar to that depicted in Figure 12.34. However, unlike the microarray shown in Figure 12.34, which contains DNA representing genes (i.e., protein-coding regions), the microarrays used in these ChIP experiments contain DNA prepared from the intergenic regions of the genome. (Remember, it is the intergenic regions that contain the regulatory sites.) Each spot on the mi-
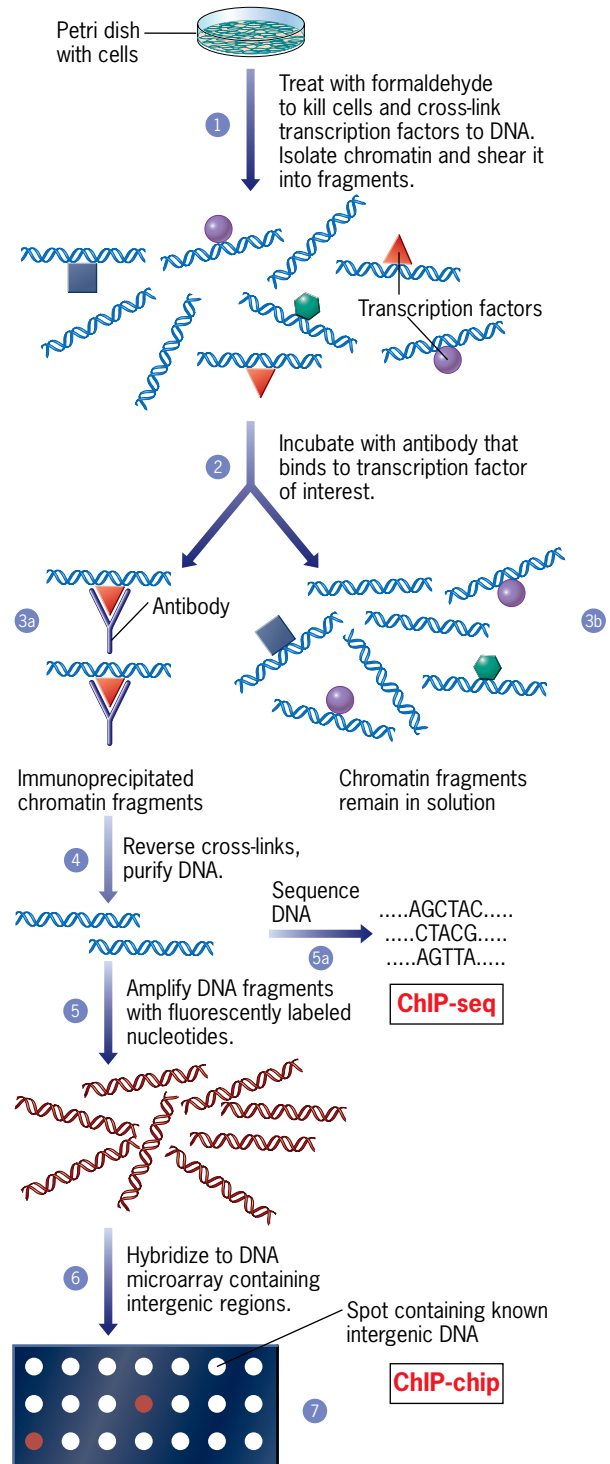
FIGURE 12.41 **Use of chromatin immunoprecipitation (ChIP) to identify transcription-factor binding sites on a global scale.** The steps are described in the text.

croarray contains DNA from a specific, identified intergenic region. The intergenic-DNA microarray is then incubated with the fluorescently labeled DNA fragments obtained from the ChIP experiment (steps 5-6) and the sites in the microarray containing bound fluorescent DNA are determined (step 7). When microarrays are used to

identify the sites in the genome that are bound by the proteins, the technique is referred to as ChIP-chip. In recent years, an increasing number of laboratories have been turning to an alternate technique, referred to as ChIP-seq, in which the bound genomic sites are identified by directly sequencing the DNA fragments precipitated by the antibody (step 5a, Figure 12.41). This approach has become possible with the development of rapid, less expensive DNA-sequencing technologies.

Interestingly, when these types of experiments are performed with mammalian transcription factors, a significant percentage of the DNA binding sites are located at considerable distances from known gene promoters. It is speculated that some of these sites are involved in regulating transcription of noncoding RNAs, such as the microRNAs discussed on page 527. Plans are underway to map the binding sites of all known mammalian transcription factors, which should provide a body of knowledge about the regulatory networks that operate to coordinately control the transcription of thousands of genes. In addition, the ChIP technique can be modified to locate the positions within the genome of any other type of DNA-bound protein, such as a particular type of modified histone or even the locations of bound RNA polymerase molecules.

The Glucocorticoid Receptor: An Example of Transcriptional Activation

Among the various hormones that affect the expression of the *PEPCK* gene are insulin, thyroid hormone, glucagon, and glucocorticoids, all of which act by means of specific transcription factors that bind to the DNA. The sites at which these transcription factors bind to the region flanking the *PEPCK* gene are called **response elements** and are shown in Figure 12.42. In the following discussion, we will focus on the glucocorticoids, a group of steroid hormones (e.g., cortisol) synthesized by the adrenal gland. Analogues of these hormones, such as prednisolone, are prescribed as potent anti-inflammatory agents.

Secretion of glucocorticoids is highest during periods of stress, as occurs during starvation or following a severe physical injury. For a cell to respond to glucocorticoids, it must possess a specific receptor capable of binding the hormone. The *glucocorticoid receptor* (*GR*) is a member of a superfamily of nuclear receptors (including receptors for thyroid hormone, retinoic acid, and estrogen) that are thought to have evolved from a common ancestral protein. Members of this superfamily are more than just hormone receptors, they are also DNA-binding transcription factors. When a glucocorticoid hormone enters a target cell, it binds to a glucocorticoid receptor protein in the cytosol, changing the conformation of the protein. This change exposes a nuclear localization signal (page 478), which facilitates translocation of the receptor into the nucleus (Figure 12.43). The ligand-bound receptor binds to a specific DNA sequence, called a *glucocorticoid response element* (*GRE*), located upstream from the *PEPCK* gene, activating transcription of the gene. Increased PEPCK levels promote the conversion of amino acids to glucose. The same GRE sequence is located upstream from different genes on different chromosomes. Consequently, a single stimulus (ele-
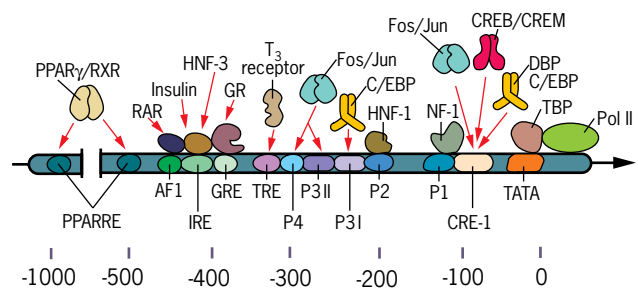
FIGURE 12.42 Regulating transcription from the rat *PEPCK* gene. Transcription of this gene, like others, is controlled by a variety of transcription factors that interact with specific DNA sequences located in a regulatory region upstream from the gene's coding region. Included within this region is a glucocorticoid response element (GRE) that, when bound by a glucocorticoid receptor, stimulates transcription from the promoter. Also included within the regulatory region are binding sites for a thyroid hormone receptor (labeled TRE); a protein that binds cyclic AMP, which is produced in response to the hormone glucagon (labeled CRE-1); and the hormone insulin (labeled IRE). A number of other transcription factors are also seen to bind to regulatory sites in this region upstream from the *PEPCK* gene. (FROM S. E. NIZIELSKI ET AL., J. NUTRITION 126:2699, 1996; © AMERICAN SOCIETY FOR NUTRITIONAL SCIENCES.)

vated glucocorticoid concentration) can simultaneously activate a number of genes whose function is necessary for a comprehensive response to stress.

Glucocorticoid response elements consist of the following sequence

$$5'\text{-AGAACAnnnTGTTCT-}3'$$
$$3'\text{-TCTTGTnnnACAAGA-}5'$$

where n can be any nucleotide. (A symmetrical sequence of this type is called a *palindrome* because the two strands have the same 5′ to 3′ sequence.) The GRE is seen to consist of two defined stretches of nucleotides separated by three undefined nucleotides. The twofold nature of a GRE is important because pairs of GR polypeptides bind to the DNA as dimers in which each subunit of the dimer binds to one-half of the DNA sequence indicated above (see Figure 12.36). The importance of the GRE in mediating a hormone response is most clearly demonstrated by introducing one of these sequences into the upstream region of a gene that normally does not respond to glucocorticoids. When cells containing DNA that has been engineered in this way are exposed to glucocorticoids, transcription of the gene downstream from the transplanted GRE is initiated. Visual evidence that gene transcription can be stimulated by foreign regulatory regions is presented in Figure 18.47.

Transcriptional Activation: The Role of Enhancers, Promoters, and Coactivators

The GRE situated upstream from the *PEPCK* gene, and the other response elements illustrated in Figure 12.42, are generally considered to be part of the promoter; they are referred to as *distal promoter elements* to distinguish them from the proxi-
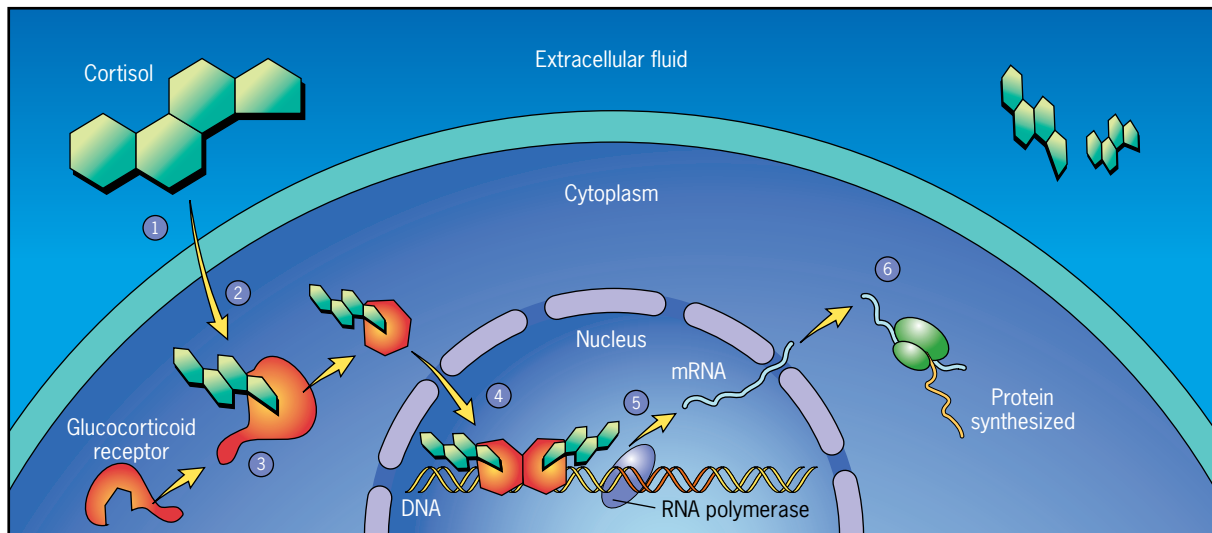
FIGURE 12.43 Activation of a gene by a steroid hormone, such as the glucocorticoid cortisol. The hormone enters the cell from the extracellular fluid (step 1), diffusing through the lipid bilayer (step 2) and into the cytoplasm, where it binds to a glucocorticoid receptor (step 3), changing its conformation and causing it to translocate into the nucleus, where it acts as a transcription factor and binds to a glucocorticoid response element (GRE) of the DNA (step 4). The glucocorticoid receptor binds to the GRE as a dimer, which activates transcription of the DNA (step 5), leading to the synthesis of specific proteins in the cytoplasm (step 6).

mal elements situated closer to the gene (Figure 12.40). The expression of most genes is also regulated by even more distant DNA elements called **enhancers**. An enhancer typically extends about 200 base pairs in length and contains multiple binding sites for sequence-specific transcriptional activators. Enhancers are often distinguished from promoter elements by a unique property: they can be moved experimentally from one place to another within a DNA molecule, or even inverted (rotated 180°), without affecting the ability of a bound transcription factor to stimulate transcription. Deletion of an enhancer can decrease the level of transcription by 100-fold or more. A typical mammalian gene may have a number of enhancers scattered within the DNA in the vicinity of the gene. Different enhancers typically bind different sets of transcription factors and respond independently to different stimuli. Some enhancers are located thousands or even tens of thousands of base pairs upstream or downstream from the gene whose transcription they stimulate.[8]

Even though enhancers and promoters may be separated by large numbers of nucleotides, enhancers are thought to stimulate transcription by influencing events that occur at the core promoter. Enhancers and core promoters can be brought into proximity because the intervening DNA can form a loop. DNA can be held in a loop through the interactions of bound proteins (Figure 12.44). If enhancers can interact with promoters over such long distances, what is to prevent an enhancer from binding to an inappropriate promoter located even farther downstream on the DNA molecule? A promoter

and its enhancers are, in essence, "cordoned off" from other promoter/enhancer elements by specialized boundary sequences called **insulators**. According to one model, insulator sequences bind to proteins of the nuclear matrix, and the DNA segments between insulators correspond to the looped domains depicted in Figure 12.12.

One of the most active areas of molecular biology in the past decade or so has focused on the mechanism by which a
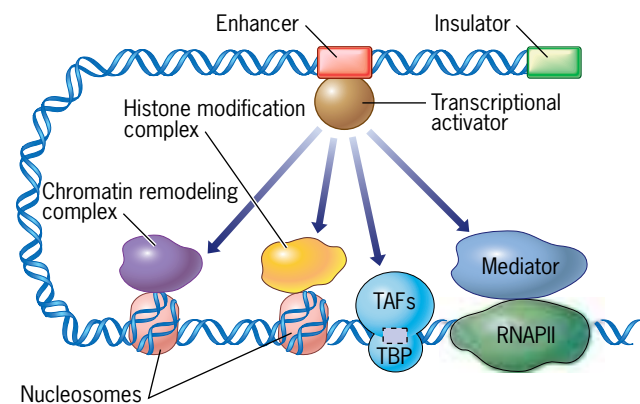
FIGURE 12.44 A survey of the means by which transcriptional activators bound at distant sites can influence gene expression. Transcriptional activators bound at upstream enhancers influence gene expression through interaction with coactivators. Four different types of coactivators are depicted here, two of them labeled "Histone modification complex" and "Chromatin remodeling complex" act by altering the structure of chromatin. Two others, labeled "TAFs" and "Mediator" act on components of the basal transcription machinery that assembles at the core promoter. These various types of coactivators are discussed in the following sections.

[8]There is wide variation in the terminology used to describe the various types of regulatory elements that control gene transcription. The terms employed here—core promoter, proximal promoter elements, distal promoter elements, and enhancers—are not universally adopted, but they describe elements that are usually present and often (but not always) capable of being distinguished.

transcriptional activator bound to an enhancer is able to stimulate the initiation of transcription at the core promoter. Transcription factors accomplish this feat through the action of intermediaries known as **coactivators**. Coactivators are large complexes that consist of numerous subunits. Coactivators can be broadly divided into two functional groups: (1) those that interact with components of the basal transcription machinery (the general transcription factors and RNA polymerase II) and (2) those that act on chromatin, converting it from a state that is relatively inaccessible to the transcription machinery to a state that is much more transcription "friendly." Figure 12.44 shows a schematic portrait of four types of coactivators, two of each of the major groups. These various types of coactivators work together in an orderly manner to activate the transcription of particular genes in response to specific intracellular signals. Given the large number of transcription factors encoded in the genome, and the limited diversity of coactivators, each coactivator complex operates in conjunction with a wide variety of different transcription factors. The coactivator CBP, for example, which is discussed below, participates in the activities of hundreds of different transcription factors.

Coactivators that Interact with the Basal Transcription Machinery   Transcription is accomplished through the recruitment and subsequent collaboration of large protein complexes. TFIID, one of the GTFs required for initiation of transcription (page 435), consists of a dozen or more subunits denoted as TAFs. Some transcription factors are thought to influence events at the core promoter by interacting with one or more of these TFIID subunits. Another coactivator that communicates directly between enhancer-bound transcription factors and the basal transcription machinery is called Mediator. Mediator is a huge, multisubunit complex that interacts directly with RNA polymerase II. Mediator is required by a wide variety of transcriptional activators and may be an essential element in the transcription of most, if not all, protein-coding genes. The mechanism of action of Mediator remains poorly defined.

Coactivators that Alter Chromatin Structure   As discussed on page 481, the DNA in a eukaryotic nucleus is not present in a naked state but is wrapped around histone octamers to form nucleosomes. The discovery of nucleosomes in the 1970s raised an important question that still has not been satisfactorily answered: how are nonhistone proteins (such as transcription factors and RNA polymerases) able to interact with DNA that is tightly associated with core histones? A large body of evidence suggests that incorporation of DNA into nucleosomes does, in fact, impede access to DNA and markedly inhibits both the initiation and elongation stages of transcription. How do cells overcome this inhibitory effect that results from chromatin structure?

As discussed on page 482, each of the histone molecules of the nucleosome core has a flexible N-terminal tail that extends outside the core particle and past the DNA helix. Covalent modifications of these tails have a profound impact on chromatin structure and function. We have already seen

how the addition of methyl groups on H3K9 residues can promote chromatin compaction and transcriptional silencing (page 488). The addition of acetyl groups to specific lysine residues in core histones tends to have the opposite effect. On a larger scale, acetylation of histone residues is thought to prevent chromatin fibers from folding into compact structures, which helps to maintain active, euchromatic regions. On a finer scale, histone acetylation increases access of specific regions of the DNA template to interacting proteins, which promotes transcriptional activation.

Techniques have been developed in recent years to determine the nature of the modifications in the histones of nucleosomes on a genome-wide scale. These techniques involve a ChIP technique similar to that illustrated in Figure 12.41, but rather than determining the genome-wide location of particular transcription factors, the goal is to pinpoint the precise locations of particular histone modifications within the genome. Rather than using antibodies against transcription factors to precipitate a particular fraction of protein-bound DNA segments, the researchers use antibodies that specifically recognize particular histone modifications, such as acetylated H3K9, acetylated H4K16, or methylated H3K36. All three of these histone marks had been known to be associated with transcriptionally active genes (Figure 12.14), but it wasn't known how these marks might be distributed within such genes. Are specific histone modifications spread uniformly throughout each gene, or are there differences from one end to the other? Results from analysis of active genes in the yeast genome are shown in Figure 12.45 and reveal marked differences within various parts of these genes. Acetylated histones are clustered in the promoter regions and largely absent from the main body of active genes, suggesting that this modification is primarily important in the activation of a gene or the initiation of its transcription. In contrast, methylated H3K36 residues are largely absent from the promoters and instead are concentrated in the transcribed regions of active genes. Sev-
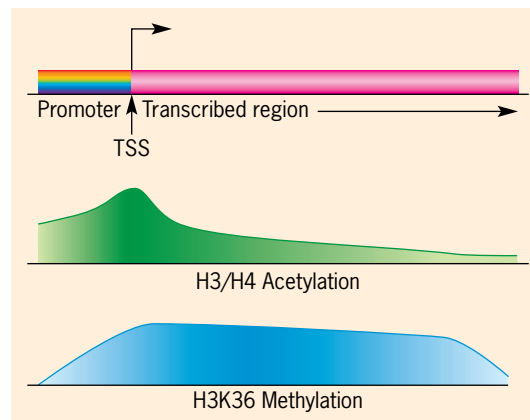
FIGURE 12.45 Selective localization of histone modifications within the chromatin of transcribed genes based on genome-wide analyses in yeast. Histone acetylation is localized primarily in the promoter region of active genes and decreases markedly in the transcribed portion of the gene, whereas H3K36 methylation displays the reverse localization pattern. TSS, transcription start site.

eral studies provide a rationale for these differences in location and give an example of the complex relationships between histone modifications. Methylation of H3K36 is catalyzed by an enzyme (Set2) that travels with the elongating polymerase. Once this residue is methylated, it serves as a recruitment site for another enzyme complex (Rpd3S) that catalyzes the removal of acetyl groups from lysine residues in the transcribed portion of the gene. Evidence suggests that removal of acetyl groups from nucleosomes in the wake of an elongating RNA polymerase prevents the inappropriate initiation of transcription within the internal coding region of a gene.

Let's look more closely at events that occur at the promoter regions of genes during the activation of transcription. Acetyl groups are added to specific lysine residues on the core histones by a family of enzymes called **histone acetyltransferases** (**HATs** or **KATs**). In the late 1990s, it was discovered that a number of coactivators possessed HAT activity. If the HAT activity of these coactivators was eliminated by mutation, so too was their ability to stimulate transcription. The discovery that coactivators contain HAT activity provided a crucial link between histone acetylation, chromatin structure, and gene activation. Figure 12.46 shows an ordered series of reactions that has been proposed to occur following the binding of a transcriptional activator, such as the glucocorticoid receptor, to its response element on the DNA. Once bound to the DNA, the activator recruits a coactivator (e.g., CBP) to a region of the chromatin that is targeted for transcription. Once positioned at the target region, the coactivator acetylates the core histones of nearby nucleosomes, which creates a binding site for a chromatin remodeling complex (discussed below). The combined actions of these various complexes increase the accessibility of the promoter to the components of the transcription machinery, which assembles at the site where transcription will be initiated.

Figure 12.46 depicts the activity of two coactivators that affect the state of chromatin. We have already seen how the HATs act to modify the histone tails; let us look more closely at the other type of coactivator, the **chromatin remodeling complexes**. Chromatin remodeling complexes use the energy released by ATP hydrolysis to alter nucleosome structure and location along the DNA. This in turn may allow the binding

FIGURE 12.46 A model for the events that occur at a promoter following the binding of a transcriptional activator. Transcription factors, such as glucocorticoid receptor (GR), bind to the DNA and recruit coactivators, which facilitate the assembly of the transcription preinitiation complex. Step 1 of this drawing depicts a region of a chromosome that is in a repressed state because of the association of its DNA with deacetylated histones. In step 2, the GR is bound to the GRE, and the coactivator CBP has been recruited. CBP contains a subunit that has histone acetyltransferase (HAT) activity. These enzymes transfer acetyl groups from an acetyl CoA donor to the amino groups of specific lysine residues. As a result, histones of the nucleosome core particles in the regions both upstream and downstream from the TATA box are being acetylated. In step 3, the acetylated histones have recruited SWI/SNF, which is a chromatin remodeling complex. Together, the two coactivators CBP and SWI/SNF have changed the structure of the chromatin to a more open, accessible state. In step 4, TFIID binds to the open region of the DNA. Although it wasn't mentioned in the text, one of the subunits of TFIID (called TAF$_{II}$250 or TAF1) also possesses histone acetyltransferase activity as indicated by the red arrow. Together, CBP and TAF$_{II}$250 disrupt additional nucleosomes to allow transcription to be initiated. In step 5, the remaining nucleosomes of the promoter have been acetylated, the RNA polymerase is bound to the promoter, and transcription is about to begin.

of various proteins to regulatory sites in the DNA. The best studied chromatin remodeling "machines" are members of the SWI/SNF family, which is shown in Figure 12.46. SWI/SNF complexes consist of 9–12 subunits, including actin or actin-related proteins. It is speculated that actin may tether the remodeling complexes to the nuclear matrix. SWI/SNF complexes are thought to be recruited to specific promoters in various ways. In Figure 12.46, the coactivator CBP has acetylated the core histones, providing a high-affinity binding site for the remodeling complex. Once recruited to a promoter, chromatin remodeling complexes are thought to disrupt histone–DNA interactions, which can

1. promote the mobility of the histone octamer so that it slides along the DNA to new positions (Figure 12.47, path 1). In the best studied case, the binding of transcriptional activators to an enhancer upstream from the *IFN*-β gene leads to the sliding of a key nucleosome approximately 35 base pairs along the DNA, exposing the

TATA box that had been previously covered by histones. Sliding occurs as the remodeling complex translocates along the DNA.

2. change the conformation of the nucleosome. In the example depicted in Figure 12.47, path 2, the DNA has formed a transient loop or bulge on the surface of the histone octamer, making that site more accessible for interaction with DNA-binding regulatory proteins.

3. facilitate the replacement within the histone octamer of a standard core histone by a histone variant (page 482) that is correlated with active transcription. For example, the SWR1 complex exchanges H2A/H2B dimers with H2A.Z/H2B dimers (Figure 12.47, path 3).

4. displace the histone octamer from the DNA entirely (Figure 12.47, path 4). For example, nucleosomes are thought to be temporarily disassembled as an elongating RNA polymerase complex moves along the DNA within a gene.

A number of attempts have been made to determine whether nucleosomes are preferentially present or absent from particular sites within the genome in a given type of cell. To carry out these studies, the chromatin is typically treated with a nuclease that digests those portions of the DNA that are not protected by their association with histone octamers. The DNA that has been protected from nuclease digestion is then dissociated from its associated histones and sequenced. These techniques have allowed researchers to prepare genome-wide maps of nucleosome positions along the DNA. The most complete analyses of nucleosome positioning have been carried out on yeast cells, and they provide a somewhat different picture from the classical view that is based on years of biochemical studies on the transcription of individual genes in mammalian cells that was presented in Figure 12.46. Studies on yeast cells suggest that the majority of genes share a common chromatin architecture, which is depicted in Figure 12.48. As indicated in this figure, nucleosomes are not randomly distributed along the DNA but instead are surprisingly well positioned. Most notably, the promoter sequences tend to reside within nucleosome-free regions (NFRs of Figure 12.48), which are flanked on either side by two very well-positioned nucleosomes identified as +1 and −1 in Figure 12.48. The nucleosome-free region surrounding the promoter is likely an important factor in allowing access by regulatory factors to these target sites in the DNA. Of all the nucleosomes in a yeast gene, the −1 nucleosome undergoes the most extensive modifications upon transcriptional activation. This nucleosome may be moved to a new site, evicted from the DNA, or undergo extensive histone modifications or histone exchange. The downstream nucleosomes (+1, +2, +3, etc.) are also subject to many of these same alterations during transcriptional activation, but the degree to which a nucleosome is affected diminishes with the distance it is located from the transcriptional start site. It is not clear yet whether the promoter regions of most nontranscribed (i.e., repressed) genes in higher eukaryotes tend to be relatively covered in nucleosomes as depicted in Figure 12.46 or rela-

FIGURE 12.47 Several alternative actions of chromatin remodeling complexes. In pathway 1, a key nucleosome has been caused to slide along the DNA, thereby exposing the TATA binding site, where the preinitiation complex can be assembled. In pathway 2, the histone octamer of a nucleosome has been reorganized. Although the TATA box is not completely free of histone association, it is now able to bind the proteins of the preinitiation complex. In pathway 3, the standard H2A/H2B dimers of a nucleosome have been exchanged with H2A.Z/H2B dimers. In pathway 4, the histone octamer has been disassembled and is lost from the DNA entirely.

FIGURE 12.48 The nucleosomal landscape of yeast genes. The top portion of the illustration shows a typical region of DNA in the vicinity of a gene that is being actively transcribed by an RNA polymerase II complex. The consensus positions of nucleosomes are illustrated by the gray ovals. The lower portion of the illustration shows the distribution of nucleosomes along the DNA with the height of the line reflecting the likelihood that a nucleosome would be found at that site. The 5′ region of the gene has the most highly defined chromatin architecture. There is a very high likelihood that the promoter region is bare of nucleosomes (forming a nucleosome-free region or NFR) that is bounded by two very well-positioned nucleosomes that lie on either side of the transcription start site (green light); these two nucleosomes are identified as +1 and −1 at the top of the figure. The subsequent nucleosomes near the 5′ end of the transcribed region also tend to be well positioned, as indicated by the distinct locations of these nucleosomes at the top of the drawing. A nucleosome is also seen to be displaced by the polymerase as it transcribes the gene. The green shading corresponds to the region of chromatin with high levels of the H2A.Z histone variant, high histone acetylation and H3K4 methylation, and a high likelihood of nucleosomal positioning. (FROM CIZHONG JIANG AND B. FRANKLIN PUGH, NATURE REVS. GEN. 10:164, 2009; © COPYRIGHT 2009, MACMILLAN MAGAZINES LIMITED, ADAPTED FROM T. N. MAVRICH, ET AL., GENOME RES. 18:1074, 2008.)

tively free of nucleosomes as suggested in Figure 12.48. It may well vary from one gene to another. Regardless, in either model, the promoter regions of the chromatin are the targets of a wide variety of both histone modifying enzymes (e.g., histone acetyltransferases, deacetylases, methyltransferases, and demethylases) and chromatin remodeling complexes (e.g., SWI/SNF).

Transcriptional Activation from Poised Polymerases
Throughout this discussion of transcriptional activation, we have described how transcription factors bind to promoters and induce the recruitment of general transcription factors (GTFs), chromatin modification complexes, and RNA polymerase II, which leads to the initiation of transcription of the adjacent gene. It came as a surprise to learn that RNA polymerases are also bound to the promoters of many genes that show no evidence they are being transcribed. In some cases, an RNA polymerase situated at one of these "transcriptionally silent" genes actually inititates the synthesis of RNA molecules, but the polymerase fails to transition to the elongation stage of transcription, so that a full-length primary transcript is never generated. The polymerases associated with these transcriptionally silent genes can be thought of as being poised and ready for action but requiring some additional regulatory event to push them into the productive phase in which the gene is fully transcribed. According to one model, RNA polymerase molecules situated downstream of promoters are held in a nonproductive (or paused) state by bound inhibitory factors (e.g., DSIF and NELF); the inhibition is then relieved as these factors are phosphorylated by kinases (e.g., P-TEFb, Figure 11.20). These studies have suggested that the regulation of gene expression at the level of transcription elongation

(rather than just transcription initiation) may play an important role in facilitating the rapid activation of genes in response to developmental or environmental signals, but this remains to be confirmed by more extensive experimentation.

Transcriptional Repression

As is evident from Figures 12.28 and 12.29, control of transcription in prokaryotic cells relies heavily on repressors, which are proteins that bind to the DNA and block transcription of a nearby gene. Although research in eukaryotes has focused primarily on factors that activate or enhance transcription of specific genes, it is evident that eukaryotic cells also possess negative regulatory mechanisms.

We've seen that transcriptional activation is associated with changes in the state and/or position of nucleosomes in a particular region of chromatin. The state of acetylation of chromatin is a dynamic property; just as there are enzymes (HATs) to add acetyl groups, there are also enzymes to remove them. Removal of acetyl groups is accomplished by **histone deacetylases** (**HDACs**). Whereas HATs are associated with transcriptional activation, HDACs are associated with transcriptional repression. HDACs are present as subunits of larger complexes described as *corepressors*. Corepressors are similar to coactivators, except that they are recruited to specific genetic loci by transcriptional factors (repressors) that cause the targeted gene to be silenced rather than activated (Figure 12.49). The progression of several types of cancer may depend on the tumor cells being able to repress the activity of certain genes. A number of anti-cancer drugs (e.g., Zolinza) are currently being tested that act by inhibiting HDAC enzymes.

ACTIVE CHROMATIN

Binding site for
repressor (silencer)

Acetylated
histones

HDAC (e.g. Sin3 complex)

Deacetylation

Transcriptional
repressor

Corepressor
e.g. SMRT/
N-CoR or
CoREST

Deacetylated
histones

INACTIVE CHROMATIN

Histone methyltransferase
(e.g. SUV39H1)

H3-K9 methylation

Methyl
group

Transcriptional
repressor

Corepressor
e.g. SMRT/
N-Cor or
CoREST

Deacetylated
histones

FIGURE 12.49 A model for transcriptional repression. The histone tails in the promoter regions of active chromatin are acetylated. When a transcriptional repressor binds to its DNA binding site, it recruits a corepressor complex (e.g., SMRT/N-CoR or CoREST) and an associated HDAC activity. The HDAC removes acetyl groups from the histone tails. A separate protein containing histone methyltransferase activity adds methyl groups to the K9 residue of H3 histone tail. Together, the loss of acetyl groups and addition of methyl groups lead to chromatin inactivation and gene silencing.

Recent studies indicate that the removal of acetyl groups from histone tails is accompanied by another histone modification: methylation of the lysine residue at position #9 of histone H3 molecules. You might recall that this modification, H3K9me, was discussed at some length on page 488 as a key event in the formation of heterochromatin. It now appears that this same modification is also involved in the more dynamic types of transcriptional repression that occur within euchromatic regions of the genome. Figure 12.49 suggests one of numerous possible models of transcriptional repression that incorporates several of these various aspects of chromatin modification.

Transcriptional repression is not very well understood, but one of the key factors in silencing a region of the genome involves a phenomenon known as DNA methylation.

DNA Methylation   Examination of the DNA of mammals and other vertebrates indicates that as many as 1 out of 100 nucleotides bears an added methyl group, which is always attached to carbon 5 of a cytosine. Methyl groups are added to the DNA by a small family of enzymes called *DNA methyl–transferases* encoded in humans by *DNMT* genes. This simple chemical modification is thought to serve as an epigenetic mark or "tag" that allows certain regions of the DNA to be identified and utilized differently from other regions. In mammals, the methylcytosine residues are part of a 5′-CpG-3′ dinucleotide within a symmetrical sequence, such as

C C G G
G G C C

G C G C
C G C G

A C G T
T G C A
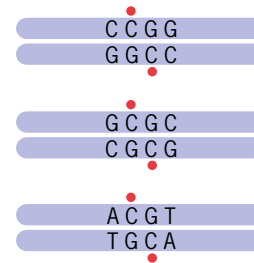
in which the red dots indicate the position of the methyl groups. The majority of methylcytosine residues in DNA are located within noncoding, repeated sequences, primarily transposable elements (page 402). Methylation is thought to maintain these elements in an inactive state. Organisms that harbor mutant DNMTs tend to exhibit a marked increase in transposition activity, which can be detrimental to the health of the organism. At least one recent study suggests that piRNAs, which were discussed on page 454 as mediators of TE suppression in germ cells, act by guiding the DNA-methylation machinery to sites where TEs reside within the genome.

DNA Methylation and Transcriptional Repression   DNA methylation of promoter regions is generally correlated with gene repression. Most evidence suggests that DNA methylation serves more to maintain a gene in an inactive state than as a mechanism for initial inactivation. As an example, inactivation of the genes on one of the X chromosomes of female mammals (page 486) occurs prior to a wave of methylation of gene promoters that is thought to convert the DNA into a more permanently repressed condition.[9] Maintenance of the repressed state is mediated by a class of proteins, including MeCP2, that bind to methylated CpG dinucleotides in key regulatory regions upstream from genes. According to one model, these proteins bind to sites of DNA methylation and then recruit corepressor complexes with associated HDAC and SUV39H1 activity. As discussed earlier, these enzymes act on histones to remove acetyl groups and methylate H3K9 residues, respectively, which together lead to chromatin compaction and gene repression (as in Figure 12.49). According to

---

[9]Recent genome-wide analyses have indicated that, whereas transcriptionally repressed genes typically have their promoter regions methylated, the same is not true of the body of these genes (i.e., the region downstream from the transcription start site). It now appears that the bodies of actively transcribed genes are much more heavily methylated than the corresponding regions of repressed genes. The reason for this unexpected methylation pattern is unclear. This subject is discussed at length in *Nature Revs. Gen.* 9:465, 2008.

this scheme, one type of epigenetic alteration, DNA methylation, serves as a guide to orchestrate a second type of epigenetic alteration, histone modification. Abnormal DNA methylation patterns are often associated with disease. For example, the development of tumors often depends on aberrant methylation and subsequent silencing of genes whose expression would normally suppress tumor growth.

Although DNA methylation is a relatively stable epigenetic mark, the transmission of these marks from a parental cell to its daughters is subject to regulation. The remarkable shifts in DNA-methylation levels that occur during the life of a mammal are depicted in Figure 12.50. The first major change in the level of methylation occurs between fertilization and the first few divisions of the zygote, when the DNA loses the methylation "tags" that were inherited from the previous generation. Then, at about the time the embryo implants into the uterus, a wave of new (or *de novo*) methylation spreads through the cells, establishing a new pattern of methylation throughout the DNA. We don't know the signals that determine whether a given gene in a particular cell is targeted for methylation or spared at this time. Regardless, once the pattern of DNA methylation is established within a cell, it is thought to be maintained through cell divisions by an enzyme (likely Dnmt1) that methylates the daughter DNA strands according to the methylation pattern of the parental strands.

DNA methylation is not a universal mechanism for inactivating eukaryotic genes. DNA methylation has not, for example, been found in yeast or nematodes. Plant DNA, in contrast, is often heavily methylated, and studies on cultured plant cells indicate that, as in animals, DNA methylation is associated with gene inactivation. In one experiment, plants treated with compounds that interfere with DNA methylation produced a greatly increased number of leaves and flower stalks. Moreover, flowers that developed on these stalks had a markedly altered morphology.

One of the most dramatic examples of the role of DNA methylation in silencing gene expression occurs as part of an epigenetic phenomenon known as genomic imprinting, which is unique to mammals.

Genomic Imprinting It had been assumed until the mid-1980s that the set of chromosomes inherited from a male parent was functionally equivalent to the corresponding set of chromosomes inherited from the female parent. But, as with many other long-standing assumptions, this proved not to be the case. Instead, certain genes are either active or inactive during early mammalian development depending solely on whether they were brought into the zygote by the sperm or the egg. For example, the gene that encodes a fetal growth factor called IGF2 is only active on the chromosome transmitted from the male parent. In contrast, the gene that encodes a specific potassium channel (KVLQT1) is only active on the chromosome transmitted from the female parent. Genes of this type are said to be **imprinted** according to their parental origin. Imprinting can be considered an epigenetic phenomenon (page 496), because the differences between alleles are inherited from one's parents but are not based on differences in DNA sequence. It is estimated that the mammalian genome contains at least 80 imprinted genes located primarily in several distinct chromosomal clusters.

Genes are thought to become imprinted as the result of selective DNA methylation of one of the two alleles. It is found, for example, that (1) the maternal and paternal versions of imprinted genes differ consistently in their degree of methylation, and (2) mice that lack a key DNA methyltransferase (Dnmt1) are unable to maintain the imprinted state of the genes they inherit. The methylation state of imprinted genes is not affected by the waves of demethylation and remethylation that sweep through the early embryo (Figure 12.50). Consequently, the same alleles that are inactive due to imprinting in the fertilized egg will be inactive in the cells of the fetus and most adult tissues. The major exception occurs in the germ cells, where the imprints inherited from the parents are erased during early development and then reestablished when that individual produces his or her own gametes. Some mechanism must exist whereby specific genes (e.g., *KVLQT1*) are selected for inactivation during sperm formation, whereas other genes (e.g., *IGF2*) are selected for inactivation during formation of the egg. Each of the imprinted gene clusters produces at least one noncoding RNA. These RNAs play a key role in directing the silencing of the nearby protein-coding genes in several of the clusters.

Disturbances in imprinting patterns have been implicated in a number of rare human genetic disorders, particularly those involving a cluster of imprinted genes residing on chromosome 15. Prader-Willi syndrome (PWS) is an inherited neurological disorder characterized by mental retardation, obesity, and underdevelopment of the gonads. The disorder
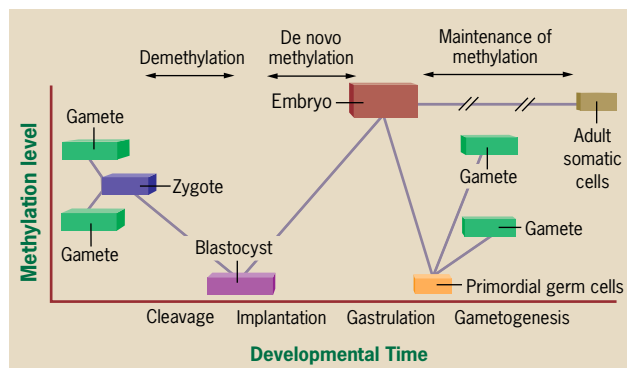
FIGURE 12.50 Changes in DNA methylation levels during mammalian development. The DNA of the fertilized egg (zygote) is substantially methylated. During cleavage, the genome undergoes global demethylation. Interestingly, DNA inherited from one's father undergoes demethylation at an earlier stage and by a different mechanism than DNA inherited from one's mother. After implantation, the DNA is subjected to new (de novo) methylation, which is maintained in the somatic (nongerm) cells at a high level throughout the remainder of development and adulthood. In contrast, the DNA of the primordial germ cells, which give rise to the gametes in the adult, is subsequently demethylated. The DNA of the germ cells then becomes remethylated at later stages of gamete formation. (BASED ON A FIGURE FROM R. JAENISCH, TRENDS GENET. 13:325, 1997; COPYRIGHT 1997, WITH PERMISSION OF ELSEVIER SCIENCE.)

often occurs when chromosome 15 inherited from the father carries a deletion in a small region containing the imprinted genes. Because the paternal chromosome carries a deletion of one or more genes and the maternal chromosome carries the inactive, imprinted version of the homologous region, the individual lacks a functional copy of the gene(s). Although genes are usually imprinted for the life of an individual, cases are known where imprinting can be lost. In fact, loss of imprinting of the *IGF2* gene occurs in about 10 percent of the population, leading to increased production of the encoded growth factor. Individuals with this epigenetic alteration are at greatly increased risk of developing colorectal cancer. One case has been uncovered of a woman who, because of a presumed deficiency in a DNA methylating enzyme, produced oocytes that are totally lacking imprinted genes. When fertilized, these oocytes fail to develop past implantation, demonstrating the essential nature of this epigenetic contribution.

What possible role could genomic imprinting play in the development of an embryo? Although there are several different thoughts on this question, there is no definitive answer. According to one researcher, genomic imprinting is a "phenomenon in search of a reason," which is where we will leave the matter.

REVIEW

1. How are the functions of a bacterial *lac* repressor and a mammalian glucocorticoid receptor similar? How are they different?

2. What types of regulatory sequences are found in the regulatory regions of the DNA upstream from a gene, such as that which codes for PEPCK? What is the role of these various sequences in controlling the expression of the nearby gene(s)?

3. What is meant by the term *epigenetic*? How is it that such diverse phenomena as histone methylation, DNA methylation, and centromere determination can all be described as epigenetic?

4. What are some of the properties that tend to be found in several groups of transcription factors?

5. What is the difference between a transcriptional activator and a transcriptional repressor? between a coactivator and a corepressor? between an HAT and an HDAC?

6. How does methylation of DNA affect gene expression? How is it related to histone acetylation or histone methylation? What is meant by genomic imprinting?

12.5 PROCESSING-LEVEL CONTROL

Alternative splicing regulates gene expression at the level of RNA processing and provides a mechanism by which a single gene can encode two or more related proteins. The genes of complex plants and animals contain numerous introns and exons (page 438). We saw in Chapter 11 how introns are re-

moved from a primary transcript and exons are retained, but that is only the beginning of the story. In many cases, there is more than one pathway by which a particular primary transcript can be processed. The particular processing pathway that is followed may depend on the particular stage of development, or the particular cell type or tissue being considered. In the simplest case, which is the only one we will consider, a specific segment can either be spliced out of the transcript or retained as part of the final mRNA, depending on the particular regulatory system operating in the cell. An example of this type of alternative splicing occurs during synthesis of fibronectin, a protein found in both blood plasma and the extracellular matrix (page 236). Fibronectin produced by fibroblasts and retained in the matrix contains two extra peptides compared to the version of the protein produced by liver cells and secreted into the blood (Figure 12.51). The extra peptides are encoded by portions of the pre-mRNA that are retained during processing in the fibroblast but are removed during processing in the liver cell.

In most cases, the proteins produced from a given gene by alternative splicing are identical along most of their length but differ in key regions that may affect such important properties as their cellular location, the types of ligands they can bind, or the kinetics of their catalytic activity. A number of transcription factors are produced from genes that can be alternately spliced, producing variants that can determine the pathway of differentiation that is taken by a cell. In the fruit fly, for example, the developmental pathway that leads an embryo into becoming a male or female is determined by alternative splicing of the transcripts from certain genes.

Alternative splicing can become very complex, allowing a wide variety of different combinations of possible exons in the final mRNA product. The mechanism by which a particular exon is included or excluded depends primarily on whether or not specific 3′ and 5′ splice sites are selected by the splicing machinery as sites to be cleaved (page 443). There are many factors that can influence splice site selection. Some splice sites
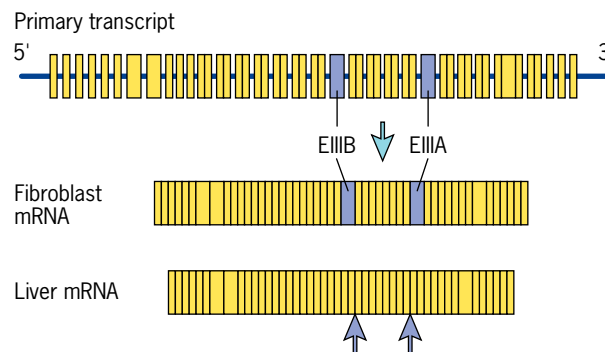
FIGURE 12.51 Alternative splicing of the fibronectin gene. The fibronectin gene consists of a number of exons shown in the top drawing (the introns shown in black are not drawn to scale). Two of these exons encode portions of the polypeptide called EIIIA and EIIIB, which are included in the protein produced by fibroblasts, but which are excluded from the protein produced in the liver. The difference is due to alternative splicing; those portions of the pre-mRNA that encode these two exons are excised from the transcript in liver cells. The sites of the missing exons are indicated by the arrows in the liver mRNA.

are described as "weak," which indicates that they can be bypassed by the splicing machinery under certain conditions. The recognition and use of weak splice sites is governed by sequences in the RNA, including exonic splicing enhancers (page 443) that are located within the exons whose inclusion is regulated. Exonic splicing enhancers serve as binding sites for specific regulatory proteins. If a particular regulatory protein is present and active in a given cell, that protein can bind to the splicing enhancer and recruit the necessary splicing factors to a nearby weak 3′ or 5′ splice site. Use of these splice sites results in the inclusion of the exon into the mRNA. A model of how this may work is depicted in Figure 12.52. If the regulatory protein is not present and active in the cell, the neighboring splice sites are not recognized and the exon is excised along with the flanking introns. Numerous other mechanisms that regulate alternative splicing have also been discovered.

Alternative processing is difficult to study, because it requires the preparation of full-length mRNAs from many different tissues and stages of development. It is estimated that more than 70 percent of human genes are subject to alternative splicing. We're also unsure as to how many different mRNAs are produced from most transcripts. In one recent study based on RNA sequencing data, approximately 100,000 different alternative splicing events were detected in a handful of major tissues, but it is uncertain how many of these transcripts have a biological function. Even if alternative splicing does not affect the amino acid sequence of a particular protein, it can generate mature mRNAs with variable 5′ or 3′ UTRs, which in turn can affect that mRNA's translation efficiency or stability (discussed below). Making it even more difficult to study, homologous transcripts are often spliced differently in different species. It is likely, for example, that a major reason that humans and mice are so different, despite the fact that they have
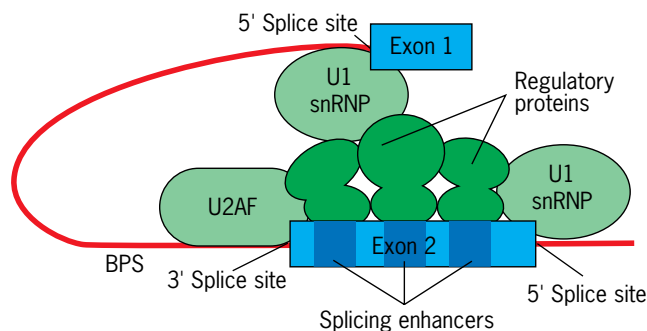
FIGURE 12.52 **A model for the role of exonic splicing enhancers in regulating alternative splicing.** In this case, exon 2 contains several splicing enhancers that bind specific regulatory proteins (typically SR proteins, page 446). These bound regulatory proteins recruit key splicing factors (U2AF) and U1 snRNP to nearby 3′ and 5′ splice sites, respectively. [U2AF plays a direct role in recruiting U2 snRNP to the branch-point site (BPS), as is required for formation of the lariat]. If U2AF and U1 snRNP were not recruited to the splice sites on each side of exon 2, this exon would not be recognized and, instead, would be excised as part of the intron. Exons may also contain sequences called exonic splicing suppressors (ESSs) that bind proteins that block spliceosome assembly and cause the exon to be left out of the mature mRNA. (AFTER K. J. HERTEL ET AL., CURR. OPIN. CELL BIOL. 9:351, 1997.)

similar genes, is the difference in the ways that many of their homologous gene transcripts are spliced.

Alternative splicing has the *potential* to generate vast numbers of related polypeptides from a single gene. Consider a gene that contained ten alternatively spliced exons, that is, exons that could either be included or excluded in the mature mRNA. A gene of this type could potentially generate tens of thousands of different polypeptide isoforms, more in fact than the number of predicted genes in the entire genome. This type of protein diversity has been proposed to play a role in directing the formation of specific synapses. In fact, recent studies suggest that certain genes involved in neural function may be subject to extensive alternative splicing. Whether or not alternative splicing occurs on such a grand scale, it is evident that the number of polypeptides encoded by the genome is at least several times the number identified by DNA sequencing alone.

RNA Editing Another way in which gene expression can be regulated at the posttranscriptional level is by **mRNA editing**, in which specific nucleotides are converted to other nucleotides after the RNA has been transcribed. RNA editing can create new splice sites, generate stop codons, or lead to amino acid substitutions. Although not nearly as widespread as alternative splicing, RNA editing is particularly important in the nervous system, where a significant number of messages appear to have one or more adenines (A) converted to inosines (I). This modification involves the enzymatic removal of an amino group from the nucleotide. I is subsequently read as a G by the translational machinery. The glutamate receptor, which mediates excitatory synaptic transmission in the brain (page 165), is a product of RNA editing. In this case, an A-to-I modification generates a glutamate receptor whose internal channel is impermeable to Ca^{2+} ions. Genetically engineered mice that are unable to carry out this specific RNA-editing step develop severe epileptic seizures and die within weeks after birth. Another important example of RNA editing affects the cholesterol-carrying protein apolipoprotein B. The LDL complexes discussed on page 306 are produced in the liver and contain the protein apolipoprotein B-100, which is translated from a full-length mRNA approximately 14,000 nucleotides long. In the intestine, the cytidine at nucleotide residue 6666 in the RNA is converted enzymatically to a uridine, which generates a stop codon (UAA) that terminates translation. The shortened version of the protein, apolipoprotein B-48, is produced only in cells of the small intestine where it plays an essential role in the absorption of fats.

REVIEW ?

1. How is it that alternative splicing can effectively increase the number of genes in the genome?

2. Describe an example of alternative splicing? What is the value to a cell of this type of control? How might a cell regulate the sites on the pre-mRNA that are chosen for splicing?

3. What is RNA editing, and how can it increase the number of proteins that can be formed from a single pre-mRNA transcript?

12.6 TRANSLATIONAL-LEVEL CONTROL

Translational-level control encompasses a wide variety of regulatory mechanisms affecting the translation of mRNAs previously transported from the nucleus to the cytoplasm. Subjects considered under this general regulatory umbrella include (1) localization of mRNAs to certain sites within a cell; (2) whether or not an mRNA is translated and, if so, how often; and (3) the half-life of the mRNA, a property that determines how long the message is translated.

Translational-level control mechanisms generally operate through interactions between specific mRNAs and various proteins and microRNAs present within the cytoplasm. It was noted on page 437 that mRNAs contain noncoding segments, called **untranslated regions** (**UTRs**), at both their 5′ and 3′ ends. The 5′ UTR extends from the methylguanosine cap at the start of the message to the AUG initiation codon, whereas the 3′ UTR extends from the termination codon at the end of the coding region to the end of the poly(A) tail that is attached to nearly all eukaryotic mRNAs (see Figure 11.21). For many years the untranslated sections of the message were largely ignored, but it has become evident that the UTRs contain nucleotide sequences used by the cell to mediate translational-level control. In following sections, we will consider three distinct aspects of translational-level control: mRNA localization, mRNA translation, and mRNA stability.

Cytoplasmic Localization of mRNAs

The localization of specific mRNAs to specific regions of a cell's cytoplasm is a major mechanism by which cells establish functionally distinct cytoplasmic domains. We will briefly consider the fruit fly, whose egg, larval, and adult stages are illustrated in Figure 12.53a. The development of the anterior–posterior (head–abdomen) axis of a larval fly and subsequent adult is foreshadowed by localization of specific mRNAs along this same axis in the oocyte. For example, mRNAs transcribed during oogenesis from the *bicoid* gene become preferentially localized at the anterior end of the oocyte, while mRNAs transcribed from the *oskar* gene become localized at the opposite end (Figure 12.53b,c). The mRNAs are subsequently translated at the site of localization where the newly synthesized protein accumulates. The protein encoded by *bicoid* mRNA plays a critical role in the development of the head and thorax, whereas the protein encoded by *oskar* mRNA is required for the formation of germ cells, which develop at the posterior end of the larva.

The information that governs the cytoplasmic localization of an mRNA resides in the 3′ UTR. This can be demonstrated using fruit flies that carry a foreign gene whose coding region is coupled to a DNA sequence that encodes the 3′ UTR of ei-

(a)

(b)

(c)   50 μm
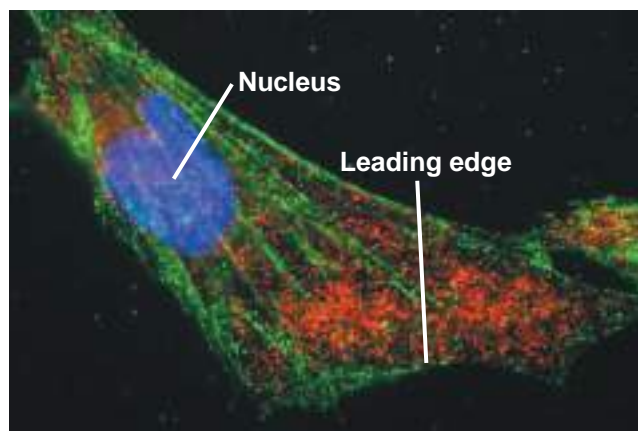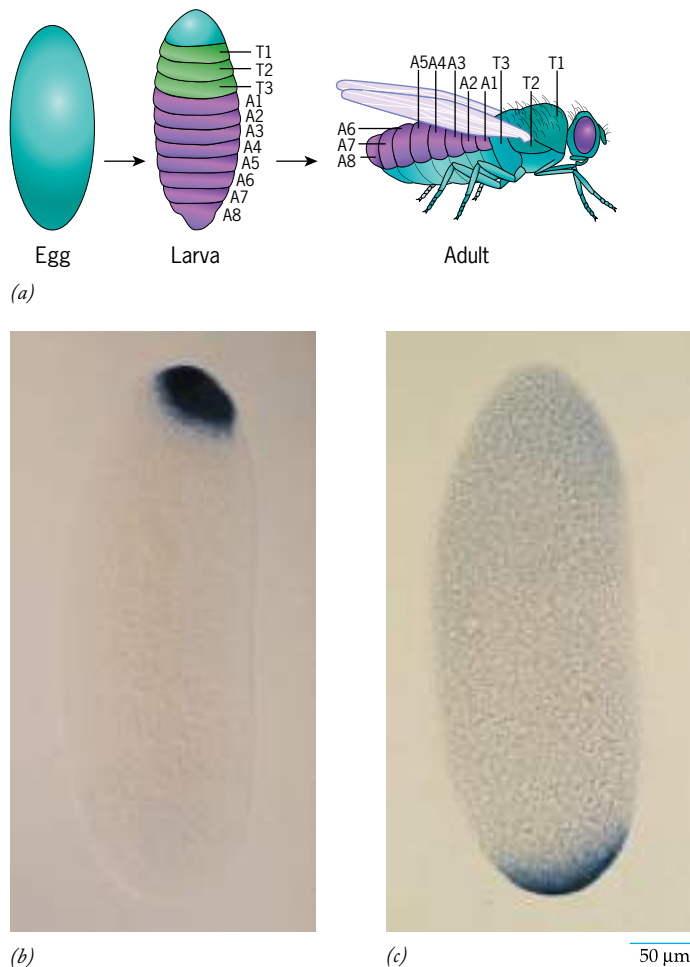
(d)

FIGURE 12.53 Cytoplasmic localization of mRNAs. (*a*) Schematic drawings showing three stages in the life of a fruit fly: the egg, larva, and adult. The segments of the thorax and abdomen are indicated. (*b*) Localization of *bicoid* mRNA at the anterior pole of an early cleavage stage of a fly embryo by in situ hybridization. (*c*) Localization of *oskar* mRNA at the posterior pole of a comparable stage to that shown in *b*. Both of these localized RNAs play an important role in the development of the anterior–posterior axis of the fruit fly. (*d*) Localization of β-actin mRNA (red) near the leading edge of a migrating fibroblast. This is the region of the cell where actin is utilized during locomotion (see Figure 9.71). (B: COURTESY OF DANIEL ST JOHNSTON; C: COURTESY OF ANTOINE GUICHET AND ANNE EPHRUSSI; D: FROM V. M. LATHAM ET AL., COURTESY ROBERT H. SINGER, CURR. BIOL. 11:1010, 2001.)

ther the *bicoid* or *oskar* mRNA. When the foreign gene is transcribed during oogenesis, the mRNA becomes localized in the site determined by the 3′ UTR. Localization of mRNAs is mediated by RNA-binding proteins that recognize localization sequences (called *zip codes*) in this region of the mRNA.

Microtubules, and the motor proteins that use them as tracks, play a key role in transporting mRNA-containing particles to particular locations. The localization of *oskar* mRNAs in a fruit fly oocyte, for example, is disrupted by agents such as colchicine that depolymerize microtubules and by mutations that alter the activity of the kinesin I motor protein. Microfilaments, on the other hand, are thought to anchor mRNAs after they have arrived at their destination. The localization of mRNAs is not restricted to eggs and oocytes but occurs in all types of polarized cells. For instance, actin mRNAs are localized near the leading edge of a migrating fibroblast, which is the site where actin molecules are needed for locomotion (Figure 12.53*d*). During the process of localization, translation of the mRNAs is specifically inhibited by associated proteins, which brings us to the subject of the control of translation.

The Control of mRNA Translation

A number of important biological processes depend on mRNAs that were synthesized at a previous time and stored in the cytoplasm in an inactive state. We will look briefly at one of these processes—the early development of an animal embryo. The mRNAs stored in an unfertilized egg, such as those transcribed from the *bicoid* or *oskar* genes in a fruit fly, are templates for proteins synthesized during early stages of development; they are not utilized for protein synthesis in the egg itself. The inactive mRNAs that are stored in an egg typically have shortened poly(A) tails. The initiation of translation of these mRNAs during early development involves at least two distinct events: removal of bound inhibitory proteins and increase in the length of the poly(A) tails by action of an enzyme residing in the egg cytoplasm. These events are illustrated in the model of translational activation in *Xenopus* embryos in Figure 12.54.

A number of mechanisms have been discovered that regulate the rate of translation of mRNAs in response to changing cellular requirements. Some of these mechanisms can be considered to act *globally* because they affect translation of all messages. When a human cell is subjected to certain stressful stimuli, a protein kinase is activated that phosphorylates the initiation factor eIF2, which blocks further protein synthesis. As discussed on page 462, eIF2-GTP delivers the initiator tRNA to the small ribosomal subunit, after which it is converted to eIF2-GDP and released. The phosphorylated version of eIF2 cannot exchange its GDP for GTP, which is required for eIF2 to become engaged in another round of initiation of translation. It is interesting to note that four different protein kinases have been identified that phosphorylate the same Ser residue of the eIF2α subunit to trigger translational inhibition. Each of these kinases becomes activated after a different type of cellular stress, including heat shock, viral infection, the presence of unfolded proteins, or amino acid starvation. Thus at least four different stress pathways converge to induce the same response.

Other mechanisms influence the rate of translation of *specific* mRNAs through the action of proteins that recognize specific elements in the UTRs of those mRNAs. One of the best studied examples involves the mRNA that codes for the protein ferritin. Ferritin sequesters iron atoms in the cytoplasm of cells, thereby protecting the cells from the toxic effects of the free metal. The translation of ferritin mRNAs is regulated by a specific repressor, called the *iron regulatory protein* (*IRP*), whose activity depends on the concentration of unbound iron in the cell. At low iron concentrations, IRP binds to a specific sequence in the 5′ UTR of the message called the *iron-response element* (*IRE*) (Figure 12.55). The bound IRP interferes physically with the binding of a ribosome to the 5′ end of the message, thereby inhibiting the initiation of translation. At high iron concentrations, the IRP is modified such that it loses its affinity for the IRE. Dissociation of the

FIGURE 12.54 **A model for the mechanism of translational activation of mRNAs following fertilization of a *Xenopus* egg.** According to this model, messenger RNAs are maintained in the cytoplasm in an inactive state by a combination of their short poly(A) tails and a bound inhibitory protein called Maskin. Maskin is tethered at one surface to CPEB, a protein that binds to sequences in the 3′ UTR of specific mRNAs, and at another surface to the cap-binding protein eIF4E. Following fertilization, CPEB is phosphorylated, which displaces Maskin, leading to changes in the mRNP. The phosphorylated version of CPEB recruits another protein CPSF, which recruits poly(A) polymerase (PAP), an enzyme that adds adenosine residues to the poly(A) tail. The elongated poly(A) tail serves as a binding site for PABP molecules, which help to recruit eIF4G, an initiation factor required for translation. As a result of these changes, the mRNA can be translated. (REPRINTED WITH PERMISSION FROM R. D. MENDEZ AND J. D. RICHTER, NATURE REVIEWS MOL. CELL BIOL. 2:524, 2001, © COPYRIGHT 2001, BY MACMILLAN MAGAZINES LIMITED.)
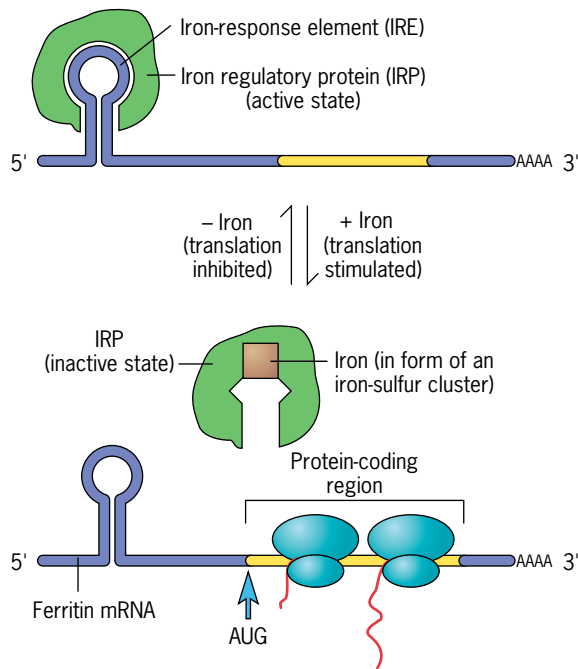
FIGURE 12.55 **The control of ferritin mRNA translation.** When iron concentrations are low, an iron-binding repressor protein, called the iron regulatory protein (IRP), binds to a specific sequence in the 5′ UTR of the ferritin mRNA, called the iron-response element (IRE), which is folded into a hairpin loop. When iron becomes available, it binds to the IRP, changing its conformation and causing it to dissociate from the IRE, allowing the translation of the mRNA to form ferritin.

IRP from the ferritin mRNA gives the translational machinery access to the mRNA, and the encoded protein is synthesized.

The Control of mRNA Stability

The longer an mRNA is present in a cell, the more times it can serve as a template for synthesis of a polypeptide. If a cell is to control gene expression, it is just as important to regulate the survival of an mRNA as it is to regulate the synthesis of that mRNA in the first place. Unlike prokaryotic mRNAs, which begin to be degraded at their 5′ end even before their 3′ end has been completed, most eukaryotic mRNAs are relatively long-lived. Even so, the half-life of eukaryotic mRNAs is quite variable. The *FOS* mRNA, for example, which is involved in the control of cell division, is degraded rapidly in the cell (half-life of 10 to 30 minutes). As a result, Fos is produced for only a short period. In contrast, mRNAs that encode the production of the dominant proteins of a particular cell, such as hemoglobin in an erythrocyte precursor or ovalbumin in a cell of a hen's oviduct, typically have half-lives of more than 24 hours. Thus, as with mRNA localization or the rate of initiation of mRNA translation, specific mRNAs can be recognized by the cell's regulatory machinery and given differential treatment.

It was shown in an early experiment that mRNAs lacking poly(A) tails were rapidly degraded following injection

(a)

(b)
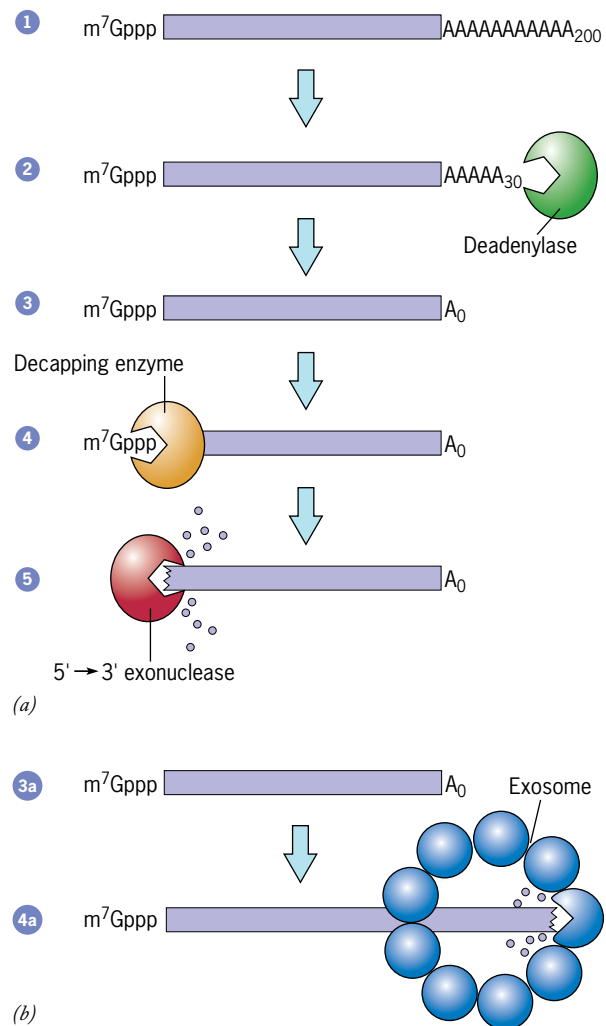
FIGURE 12.56 **mRNA degradation in mammalian cells.** The steps depicted in the drawing are described in the text.

into cells, whereas the same mRNAs possessing such tails were relatively stable. This was the first piece of evidence that suggested that the longevity of an mRNA is related to the length of its poly(A) tail. When a typical mRNA leaves the nucleus, it contains a tail of approximately 200 adenosine residues (Figure 12.56a, step 1). As an mRNA remains in the cytoplasm, its poly(A) tail tends to be gradually reduced in length as it is nibbled away by a deadenylase. No effect on the stability of the mRNA is observed until the tail becomes reduced to approximately 30 residues (step 2). Once the tail is shortened to this length, the mRNA is usually degraded rapidly by either of two pathways. In one of these pathways (shown in Figure 12.56a), degradation of the mRNA begins at its 5′ end following removal of the remaining poly(A) at the 3′ end of the message. The fact that the poly(A) tail at the 3′ end of the message protects the cap at the 5′ end of the molecule suggests that the two ends of the mRNA are held in close proximity (see Figure 11.47). Once the 3′ tail is removed (step 3, Figure 12.56a), the message is decapped

(step 4) and degraded from the 5′ end toward the 3′ end (step 5). Deadenylation, decapping, and 5′ → 3′ degradation occur within small transient cytoplasmic granules called **P-bodies**. In addition to destroying "unwanted" mRNAs, P-bodies can also act as sites where mRNAs that are no longer being translated are stored temporarily. In the alternate mRNA degradation pathway shown in Figure 12.56b, removal of the poly(A) tail (step 3a) is followed by continued digestion of the mRNA from its 3′ end (step 4a). The digestion of mRNAs in the 3′ → 5′ direction is carried out by an exonuclease that is part of a complex of 3′ → 5′ exonucleases called the *exosome*.

There must be more to mRNA longevity than simply the length of the poly(A) tail, as mRNAs having very different half-lives begin with a similar-sized tail. Once again, differences in the nucleotide sequence of the 3′ UTR have been shown to play a role in the rate at which the poly(A) tail becomes shortened. The 3′ UTR of a globin mRNA, for example, contains a number of CCUCC repeats that serve as binding sites for specific proteins that stabilize the message. If these sequences are mutated, the mRNA is destabilized. In contrast, short-lived mRNAs often contain AU-rich elements (e.g., AUUUA repeats) in their 3′ UTR that destabilize the message. If one of these destabilizing sequences is introduced into the 3′ UTR of a globin gene, the stability of the mRNA transcribed from the modified gene is reduced from a half-life of 10 hours to a half-life of 90 minutes. The importance of these destabilizing sequences (and the overall mRNA instability they produce) can be appreciated by considering the normally short-lived *FOS* mRNA mentioned above. If the destabilizing sequence from the *FOS* gene is lost through a deletion, the half-life of *FOS* mRNA increases, and the cells often become malignant. Destabilizing sequences in the 3′ UTR are thought to serve as binding sites both for proteins (e.g., AUF1) and microRNAs (e.g., miR-430) that bring about the deadenylation and subsequent destruction of the mRNA, as discussed below.

The Role of MicroRNAs in Translational-Level Control

Proteins are not the only molecules that can act as regulators of mRNA translation and stability. The formation and mechanism of action of microRNAs was discussed in Chapter 11 (page 452). Like most of the proteins we have been discussing that carry out translational-level control, miRNAs act primarily by binding to sites in the 3′ UTR of their target mRNAs. It is becoming increasingly apparent that miRNAs are important translational-level regulators of virtually every biological process. Even the earliest stages of embryonic development require the involvement of miRNAs, as evidenced by the fact that animals lacking the miRNA-producing enzyme Dicer fail to develop beyond gastrulation. Similarly, when Dicer is absent only from a particular tissue, the development of the cells of that tissue displays obvious abnormalities. It is also becoming apparent that abnormalities in

miRNA levels play a major role in the development of many common diseases.

The role of individual miRNAs in a biological process is best studied by interfering with the production of that particular species, which can be accomplished through mutation or deletion of the gene that encodes the miRNA. Studies of this type suggest that miRNAs govern the expression of batteries of genes that are involved in various pathways of cellular differentiation. We can illustrate the importance of miRNAs both in embryonic development and in adult tissues by considering examples from studies on the heart. Two members of the *miR-1* class of miRNAs play important roles in the development of muscle tissue. The deletion of one of these miRNAs (*miR-1-2*) in mouse embryos is accompanied by dramatic defects in development of the heart. Many of these animals die before birth as the result of holes in the partition between the heart's ventricles, while others die after birth from disturbances in the rhythm of their heartbeat. These deleterious effects may result from the overexpression of certain key transcription factors whose translation is normally inhibited by *miR-1* miRNAs. These same miRNAs have also been implicated in heart disease in adult humans. Patients who have had previous heart attacks and develop subsequent electrophysiological abnormalities tend to have elevated *miR-1* levels in their heart tissue. Two of the predicted targets of *miR-1* are the mRNAs that encode (1) the protein Cx43, which forms the gap junctions that conduct ionic current between the muscle cells of the heart's ventricles (page 257), and (2) the protein Kir2.1, which is a potassium channel involved in repolarization of the membrane potential (page 161) in heart cells. The expression of both of these proteins would be expected to be reduced in the presence of higher-than-normal levels of *miR-1*. The importance of these observations is supported by studies in rats that have been induced to overexpress *miR-1*; these animals exhibit a phenotype that is similar to that seen in patients who have suffered heart attacks. Most importantly, blocking the activity of *miR-1* in these rats leads to a return of Cx43 and Kir2.1 to their normal levels and a normalization of the electrophysiological state of their hearts. These findings offer hope that miRNAs may become a valid target of therapeutic drugs in the treatment of heart disease.

The manner in which miRNAs carry out translational-level control has been a matter of considerable controversy. Taken as a whole, the studies in this field suggest that miRNAs likely exert their regulatory activities through several distinct mechanisms, affecting both mRNA translation and mRNA stability, as shown in Figure 12.57.

1. Early studies on worms and recent studies on cultured mammalian cells indicate that miRNAs can repress the synthesis of proteins at some point after the initiation of translation has occurred (Figure 12.57b). This view was based initially on the observation that miRNAs bound to mRNAs are part of actively translating polysomes. One line of evidence suggests that the bound RISC (the miRNA-Argonaute protein complex of Figure 11.38b)

(a) Active translation

(b) Post-initiation inhibition

(c) Inhibition of initiation

(d) mRNA degradation
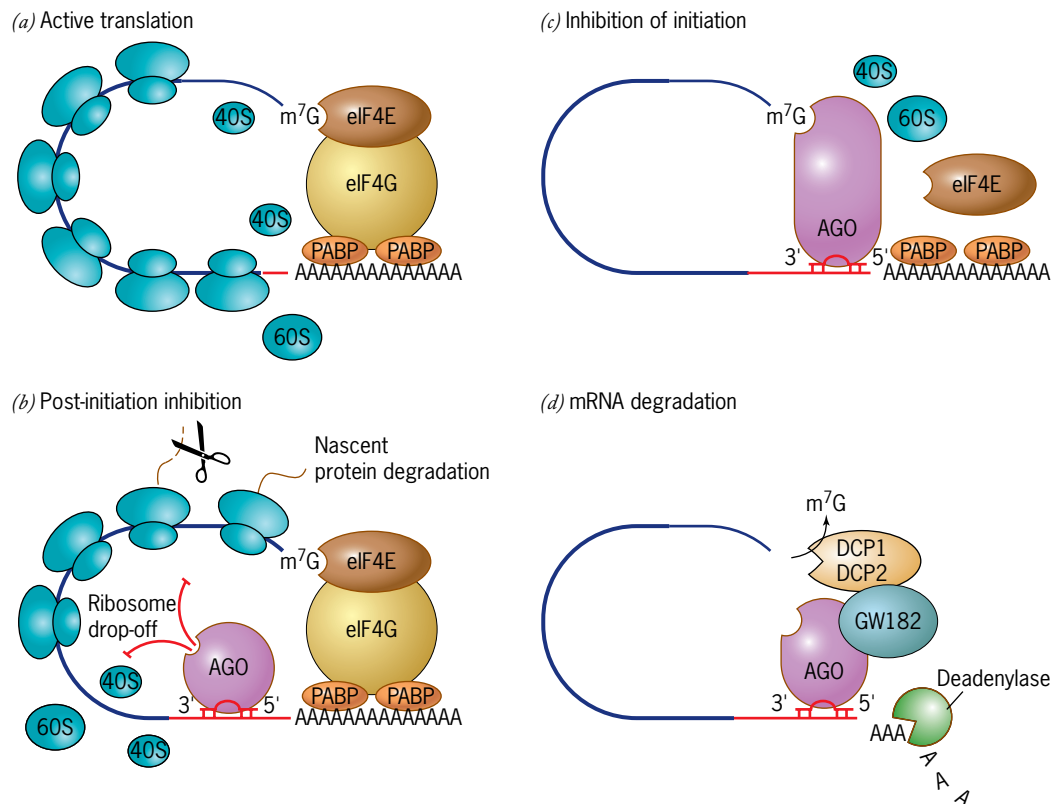
FIGURE 12.57 Potential mechanisms by which miRNAs might decrease gene expression at the translational level. (*a*) Active translation in the absence of miRNA regulation. (*b*) Inhibition of gene expression at some point after translation has been initiated. In this model, the binding of the RISC complex containing the Argonaute protein (AGO) leads either to degradation of the nascent protein or to the release of the ribosomes from the mRNA. (*c*) Inhibition of gene expression at the point where the translation process is initiated. In this model, the RISC complex either inhibits the interaction between eIF4E and the 5′ end of the mRNA or prevents the assembly of the complete ribosome from its subunits. (*d*) Inhibition of gene expression by promoting the degradation of the mRNA. In this model, the RISC complex has recruited proteins that either promote the decapping of the mRNA (by proteins DCP1/2) and/or degrade the poly(A) tail. GW182 is a component of P-bodies. (FROM G. STEFANI AND F. J. SLACK, NAT. REVS. MOL. CELL BIOL. 9:220, 2008; © COPYRIGHT 2008, MACMILLAN MAGAZINES LIMITED.)

causes premature drop-off of ribosomes as they are engaged in translational elongation, whereas another line implicates the RISC in the degradation of the nascent polypeptide.

2. Other studies suggest that the bound RISC interferes with translation initiation (Figure 12.57*c*). This view is based on numerous observations, including data suggesting that the Argonaute protein within a RISC is capable of binding to the methyl-guanine cap at the 5′ end of the mRNA and blocking the interaction of the cap with the required translation initiation factor eIF4E. Other data suggest that the RISC prevents the interaction between the large and small ribosomal subunits, which is also required to initiate translation.

3. Still other studies have suggested that some miRNAs carry out translational-level regulation by inducing the degradation of the target mRNA (Figure 12.57*d*). This is illustrated by a study on zebra fish embryos that reported a dramatic increase in the expression of one particular miRNA (miR-430) at the time when the embryo began to synthesize its own mRNAs. The appearance of this miRNA was accompanied by the degradation of large numbers of maternal mRNAs that are stored in the egg at the time of fertilization and contain binding sites for the miR-430 miRNA in their 3′ UTR.

REVIEW

1. Describe three different ways in which gene expression might be controlled at the translational level. Cite an example of each of these control mechanisms.

2. What is the role of poly(A) in mRNA stability? How might the cell regulate the stability of different mRNAs?

3. Describe the different levels at which gene expression is regulated to allow a β-globin gene with the following structure to direct the formation of a protein that accounts for over 95 percent of the protein of the cell.

    exon 1—intron—exon 2—intron—exon 3

4. Describe some of the ways in which miRNAs might regulate gene expression.

12.7 POSTTRANSLATIONAL CONTROL: DETERMINING PROTEIN STABILITY

We have seen how cells possess elaborate mechanisms to control the rates at which proteins are synthesized. It is not unexpected that cells would also possess mechanisms to control the length of time that specific proteins survive once they are fully functional. Although the subject of protein stability does not fall technically under the heading of control of gene expression, it is a logical extension of that topic and thus appears in this place in the text. Pioneering studies in the area of selective protein degradation were carried out by Avram Hershko and Aaron Ciechanover in Israel and Irwin Rose and Alexander Varshavsky in the United States.

Degradation of cellular proteins is carried out within hollow, cylindrical, protein-degrading machines called **proteasomes** that are found in both the nucleus and cytosol of cells. Proteasomes consist of four rings of polypeptide subunits stacked one on top of the other with a cap attached at either end of the stack (Figure 12.58a,b). The two central rings consist of polypeptides (β subunits) that function as proteolytic enzymes. The active sites of these subunits face the enclosed central chamber, where proteolytic digestion can occur in a protected environment.

Proteasomes digest proteins that have been specially selected and marked for destruction as described below. Some proteins are selected because they are recognized as abnormal—either misfolded or incorrectly associated with other proteins. Included in this group are abnormal proteins that had been produced on membrane-bound ribosomes of the rough ER (page 282). The selection of "normal" proteins for proteasome destruction is based on the protein's biological stability. Every protein is thought to have a characteristic longevity. Some protein molecules, such as the enzymes of glycolysis or the globin molecules of an erythrocyte, are present for days to weeks. Other proteins that are required for a specific, fleeting activity, such as regulatory proteins that initiate DNA replication or trigger cell division, may survive only a few minutes. The destruction of such key regulatory proteins by proteasomes plays a crucial role in the progression of cellular processes (as illustrated in Figure 14.26). The importance of proteasomes in life-and-death cellular processes can be demonstrated using drugs that specifically inhibit proteasomal digestion.

The factors that control a protein's lifetime are not well understood. One of the determinants is the specific amino acid that resides at the N-terminus of a polypeptide chain. Polypeptides that terminate in arginine or lysine, for exam-

FIGURE 12.58 Proteasome structure and function. (a) High-resolution electron micrograph of an isolated *Drosophila* proteasome. (b) Model of a proteasome based on high-resolution electron microscopy and X-ray crystallography. Each proteasome consists of two multisubunit caps (or regulatory particles) on either end of a tunnel-shaped cylinder (or core particle) that is formed by a stack of four rings. Each ring consists of seven subunits that are divided into two classes, α-type and β-type. The two inner rings are composed of β subunits, which surround a central chamber. The subunits are drawn in different shades of color because they are similar, but not identical, polypeptides. Three of the seven β subunits in each ring possess proteolytic activity; the other four are inactive in eukaryotic cells. (Prokaryotic cells also possess proteasomes, but they have a simpler construction, and all β subunits are active.) The two outer rings are composed of enzymatically inactive α subunits that form a narrow opening (approximately 13 Å) through which unfolded polypeptide substrates are threaded to reach the central chamber, where they are degraded. (c) Steps in the degradation of proteins by a proteasome. In step 1, the protein to be degraded is linked covalently to a string of ubiquitin molecules. Ubiquitin attachment requires the participation of three distinct enzymes (E1, E2, and E3) in a process that is not discussed in the text. In step 2, the polyubiquitinated target protein binds to the cap of the proteasome. The ubiquitin chain is then removed, and the unfolded polypeptide is threaded into the central chamber of the proteasome (step 3), where it is degraded by the catalytic activity of the β subunits (steps 4 and 5). (A: COURTESY H. HÖLZL AND WOLFGANG BAUMEISTER, J. CELL BIOL. 150:126, 2000; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

ple, are typically short-lived. A number of proteins that act at specific times within the cell cycle are marked for destruction when certain residues become phosphorylated. Still other proteins carry a specific internal sequence of amino acids called a *degron* that ensures they will not survive long within the cell.

Ubiquitin is a small, highly conserved protein with a number of functions in diverse cellular processes. We saw on page 305, for example, that membrane proteins bearing a single attached ubiquitin molecule are selectively incorporated into endocytic vesicles. Thus a single attached ubiquitin functions primarily as a sorting signal. In contrast, proteins are targeted for destruction by attachment of a number of ubiquitin molecules, one-to-another, to form a polyubiquitin chain (step 1, Figure 12.58c). In the first stage of this process, ubiquitin is transferred enzymatically from a carrier protein to a lysine residue on the condemned protein. Enzymes that transfer ubiquitin to target proteins comprise a huge family of *ubiquitin ligases* in which different members recognize proteins bearing different degradation signals. These enzymes play a crucial role in determining the life or death of key proteins and are a focus of current research.

Once it is polyubiquitinated, a protein is recognized by the cap of the proteasome (step 2, Figure 12.58c), which removes the ubiquitin chain and unfolds the target protein using energy provided by ATP hydrolysis. The unfolded, linear polypeptide is then threaded through the narrow opening in the ring of α subunits and passed into the central chamber of the proteasome (step 3), where in most cells it is digested into small peptides (steps 4 and 5). The peptide products are released back into the cytosol where they are degraded into their component amino acids.

SYNOPSIS

The nucleus of a eukaryotic cell is a complex structure bounded by the nuclear envelope, which controls the exchange of materials between the nucleus and cytoplasm, maintaining the unique composition of the cell's two major compartments. The nuclear envelope consists of several distinct components, including an inner and outer nuclear membrane separated by a perinuclear space and a variable number of nuclear pores. Nuclear pores are sites where the inner and outer nuclear membranes are fused to form a circular opening that is filled with a complex structure called the nuclear pore complex (NPC). The NPC has a basket-like structure with octagonal symmetry composed of rings, spokes, and filaments. The nuclear pores are the sites through which materials pass between the nucleus and cytoplasm. Proteins that normally reside within the nucleus contain a stretch of amino acids called the nuclear localization signal (NLS) that allows them to bind to a receptor (an importin) that transports them through the NPC. Nuclear transport is a process of diffusion facilitated by a gradient of the protein Ran, with Ran-GTP in the nucleus and Ran-GDP in the cytoplasm. The inner surface of the nuclear envelope is lined by a fibrillar meshwork called the nuclear lamina, which consists of proteins (lamins) that are members of the family of proteins that make up intermediate filaments. The fluid of the nucleus is called the nucleoplasm. (p. 476)

The chromosomes of the nucleus contain a defined complex of DNA and histone proteins that form characteristic nucleoprotein filaments, representing the first step in packaging the genetic material. Histones are small, basic proteins that are divided into five distinct classes. Histones and DNA are organized into nucleosome core complexes, which consist of two molecules each of histones H2A, H2B, H3, and H4 encircled by almost two turns of DNA. Nucleosome core particles are connected to one another by stretches of linker DNA. Together, the core particles and DNA linkers generate a nucleosome filament that resembles a chain of beads on a string. Covalent modifications to specific residues in the N-terminal tails of the core histones—including methylation, acetylation, and phosporylation—play a key role in determining the state of compaction and transcriptional activity of chromatin. (p. 481)

Chromatin is not present in cells in the highly extended nucleosome filament but is compacted into higher levels of organization. Each nucleosome core particle contains a molecule of H1 histone bound to the DNA. Both core and H1 histones mediate an interaction between neighboring nucleosomes that generates a 30-nm fiber, which represents a higher level of chromatin organization. The 30-nm fibers are in turn organized into looped domains, which are most readily visualized when mitotic chromosomes are subjected to procedures that remove histones. Mitotic chromosomes represent the most compact state of chromatin. A certain fraction of the chromatin, called heterochromatin, remains in a highly compact state throughout interphase. Heterochromatin is categorized as constitutive heterochromatin, which remains condensed in all cells at all times, and facultative heterochromatin, which is specifically inactivated during certain phases of an organism's life. One X chromosome in each cell of a female mammal is subjected to an inactivation process during embryonic development that converts it into transcriptionally inactive, facultative heterochromatin. Because X inactivation occurs randomly, the paternally derived X chromosome is inactivated in half the cells of the embryo, while the maternally derived X chromosome is inactivated in the other half. As a result, adult females are genetic mosaics with respect to genes present on the X chromosome. X-chromosome inactivation is an example of an epigenetic modification because it is a transmissable alteration in chromatin structure and function that does not involve a change in DNA sequence. (p. 483)

Mitotic chromosomes possess several clearly recognizable features. Mitotic chromosomes can be visualized by lysing cells that have been arrested in mitosis and then staining them to generate identifiable chromosomes with predictable banding patterns. Each mitotic chromosome contains a marked indentation, called the centromere, which contains highly repeated DNA sequences and serves as a site for the attachment of microtubules during mitosis. The ends of the chromosome are the telomeres, which are maintained from one cell generation to the next by a special enzyme, called telomerase, that contains RNA as an integral component.

Telomeres play an important role in maintaining chromosome integrity. (p. 489)

The nucleus is an ordered cellular compartment. Observations that support this include the following: specific chromosomes are confined to particular regions of the nucleus; the telomeres may be associated with the nuclear envelope; RNPs involved in pre-mRNA splicing are restricted to particular sites. Evidence indicates that the network of protein filaments that makes up the nuclear matrix plays a key role in maintaining the ordered structure of the nucleus. (p. 497)

In bacteria, genes are organized into regulatory units called operons. Operons are clusters of structural genes that typically encode different enzymes in the same metabolic pathway. Because all of the structural genes are transcribed into a single mRNA, their expression can be regulated in a coordinated manner. The level of gene expression is controlled by a key metabolic compound, such as the inducer lactose, which attaches to a protein repressor and changes its shape. This event alters the ability of the repressor to bind to the operator site on the DNA, thereby blocking transcription. (p. 499)

The rate of synthesis of a particular polypeptide in eukaryotic cells is determined by a complex series of regulatory events that operate primarily at three distinct levels. (1) Transcriptional control mechanisms determine whether or not a gene is transcribed and, if so, how often; (2) processing-level control mechanisms determine the path by which the primary RNA transcript is processed; and (3) translational control mechanisms determine the localization of an mRNA, whether a particular mRNA is actually translated and, if so, how often and for how long a period. (p. 503)

All of the differentiated cells of a eukaryotic organism retain all of the genetic information. Different genes are expressed by cells at different stages of development, by cells in different tissues, and by cells exposed to different stimuli. The transcription of a particular gene is controlled by transcription factors, which are proteins that bind to specific sequences located at sites outside of a gene's coding region. The closest upstream regulatory sequence is the TATA box, which is a major component of the core promoter of many genes and the site of assembly of the preinitiation complex. The activity of proteins at the TATA box is thought to depend on interactions with other proteins bound to other sites, including various types of response elements and enhancers. Enhancers are distinguished by the fact that they can be moved from place to place within the DNA or even reversed in orientation. Some enhancers may be located tens of thousands of base pairs upstream from the gene whose transcription they stimulate. Proteins bound to an enhancer and a promoter are thought to be brought into contact by loops in the DNA. (p. 505)

Determination of the three-dimensional structure of a number of complexes between transcription factors and DNA indicate that these proteins bind to the DNA through a limited number of structural motifs. Transcription factors typically contain at least two domains, one whose function is to recognize and bind to a specific sequence of base pairs in the DNA, and another whose function is to activate transcription by interacting with other proteins. Most transcription factors bind to DNA as dimers, either heterodimers or homodimers, that recognize sequences in the DNA that possess twofold symmetry. Most of the DNA-binding motifs contain a segment, often an α helix, that is inserted into the major groove of the DNA, where it recognizes the sequence of base pairs that line the groove. Among the most common motifs that occur in DNA-binding proteins are the zinc finger, the helix–loop–helix, and the leucine zipper. Each of these motifs provides a structurally stable framework on which the specific DNA-recognizing surfaces of the protein can interact with the DNA double helix. Though most transcription factors are probably stimulatory, some act by inhibiting transcription. (p. 508)

The activation and repression of transcription is mediated by a number of large complexes that function as coactivators or corepressors. Coactivators include complexes that serve as bridges between transcriptional activators bound at upstream regulatory sites and the basal transcription machinery bound at the core promoter. Other types of coactivators act by modifying core histones or by remodeling chromatin. Acetylation of core histones by histone acetyltransferases (HATs) is associated with transcriptional activation. Chromatin remodeling complexes (e.g., SWI/SNF) can cause nucleosomes to slide along the DNA or modify the nucleosome so as to increase its ability to bind regulatory proteins. (p. 514).

Eukaryotic genes are silenced when the cytosine bases of certain nucleotides residing in GC-rich regions are methylated. Methylation is a dynamic, epigenetic modification. Addition of methyl groups to DNA in key regulatory regions upstream from genes is correlated with a decrease in transcription of the gene. Methylation is particularly evident in chromatin that has been rendered transcriptionally inactive by heterochromatization, such as the inactive X chromosome in the cells of female mammals. Another phenomenon associated with DNA methylation is genomic imprinting, which affects a small number of genes that are either transcriptionally active or inactive in the embryo depending on their parental source. (p. 520)

Gene expression is regulated at the processing level by alternative splicing in which a single gene can encode two or more related proteins. Many primary transcripts can be processed by more than one pathway, producing mRNAs containing different combinations of exons. In the simplest case, a specific intron can be either spliced out of the transcript or retained as part of the final mRNA. The specific pathway taken during pre-mRNA processing is thought to depend on the presence of proteins that control the splice sites that are recognized for cleavage. (p. 522)

Gene expression is regulated at the translational level by various processes, including mRNA localization, control of translation of existing mRNAs, and mRNA longevity. Most of these regulatory activities are mediated by interactions with the 5′ and 3′ untranslated regions (UTRs) of the mRNA. For example, cytoplasmic localization of mRNAs in particular regions of the eggs of fruit flies is thought to be mediated by proteins that recognize localization sequences in the 3′ UTR. The presence of an mRNA in the cytoplasm does not ensure its translation. The protein synthetic machinery of a cell may be controlled globally so that translation of all mRNAs is affected, or the translation of specific mRNAs can be targeted, as occurs in the control of ferritin synthesis by iron levels acting on proteins that bind to the 5′ UTR of the ferritin mRNA. One of the primary factors controlling the longevity (stability) of mRNAs is the length of the poly(A) tail. Specific nucleases in the cell gradually shorten the poly(A) tail until a point is reached where a protective protein can no longer stay bound to the tail. Once the protein is lost, the mRNA is degraded in either a 5′ → 3′ or 3′ → 5′ direction. Translation-level regulation may be mediated by microRNAs that can inhibit either the initiation or progression of translation or promote mRNA degradation. An example of the importance of miRNAs can be seen in the development of the mammalian heart. (p. 524)

ANALYTIC QUESTIONS

1. Methylation of K9 of histone H3 (by an enzyme SUV39H1) is associated with heterochromatinization and gene silencing. It has been reported that methylation of H3 by other enzymes can lead to transcriptional activation. How can methylation lead to opposite effects?

2. How many copies of each type of core histone would it take to wrap the entire human genome into nucleosomes? How has evolution solved the problem of producing such a large number of proteins in a relatively short period of time?

3. Suppose you discovered a temperature-sensitive mutant whose nucleus failed to accumulate certain nuclear proteins at an elevated (restrictive) temperature but continued to accumulate other nuclear proteins. What conclusions might you draw about nuclear localization and the nature of this mutation?

4. Humans born with three X chromosomes but no Y chromosomes often develop into females of normal appearance. How many Barr bodies would you expect the cells of these women to have? Why?

5. Suppose that X-inactivation were not a random process, but always led to the inactivation of the X chromosome derived from the father. What effect would you expect this to have on the phenotype of females?

6. The chromosomes shown in Figure 12.18*b* were labeled by incubating the preparation with DNA fragments known to be specific for each of the chromosomes. Suppose one of the chromosomes in the field contained regions having two different colors. What might you conclude about this chromosome?

7. What advantage might be gained by having transcripts synthesized and processed in certain regions of the nucleus rather than randomly throughout the nucleoplasm?

8. Compare and contrast the effect of a deletion in the operator of the lactose operon with one in the operator of the tryptophan operon.

9. If you were to find a mutant of *E. coli* that produced continuous polypeptide chains containing both β-galactosidase and galactoside permease (encoded by the *y* gene), how might you explain how this happened?

10. You suspect that a new hormone you are testing functions to stimulate myosin synthesis by acting at the transcriptional level. What type of experimental evidence would support this contention?

11. Suppose you had conducted a series of experiments in which you had transplanted nuclei from several different adult tissues into an activated, enucleated mouse egg and found that the egg did not develop past the blastocyst stage. Could you conclude that the transplanted nucleus had lost genes that were required for postblastula development? Why or why not? What does this type of experiment tell you more generally about interpreting negative results?

12. It was noted on page 513 that DNA footprinting allows isolation of DNA sequences that bind specific transcription factors. Describe an experimental protocol to identify transcription factors that bind to an isolated DNA sequence. (You might consider the techniques discussed in Section 18.7.)

13. How do you explain why enhancers can be moved around within the DNA without affecting their activity, whereas the TATA box can only operate at one specific site?

14. Suppose that you are working with a cell that exhibits a very low level of protein synthesis, and you suspect the cells are subject to a global translational-control inhibitor. What experiment might you perform to determine whether this is the case?

15. The signal sequences that direct the translocation of proteins into the endoplasmic reticulum are cleaved by a signal peptidase, whereas the NLSs and NESs required for movement of a protein into or out of the nucleus remain as part of that protein. Consider a protein such as hnRNPA1, which is involved in the export of mRNA to the cytoplasm. Why is it important that the transport signal sequences for this protein remain as part of the protein, whereas the signal sequence for ER proteins can be cleaved?

16. When methylated DNA is introduced into cultured mammalian cells, it is generally transcribed for a period before it becomes repressed. Why would you expect this type of delay before inhibition of transcription would occur?

17. Suppose you had isolated a new transcription factor and wanted to know which genes this protein might regulate. Is there any way that you could use a cDNA microarray of the type shown in Figure 12.34 to approach this question? (*Note:* the microarray of Figure 12.34 contains the DNA of protein-coding regions, unlike that of Figure 12.41).

18. Although several different mammalian species have been cloned, the efficiency of this process is extremely low. Often tens or even hundreds of oocytes must be implanted with donor nuclei to obtain one healthy live birth. Many researchers believe the difficulties with cloning reside in the epigenetic modifications, such as DNA and histone methylation, that occur within various cells during an individual's life. How do you suspect such modifications might affect the success of an experiment such as that depicted in Figure 12.31?

19. A recent study in a British medical journal found there was a correlation in telomere length between fathers and their daughters and between mothers and both their sons and daughters but not between fathers and their sons. How can you explain this finding?

20. Some scientific reports are best described as *correlations,* in that they report on two events or conditions that tend to accompany one another. Correlations are often interpreted as evidence of a causative relationship between the two events or conditions. Other scientific reports involve experimental intervention and generally make a stronger case for a causal relationship. Pick one scientific conclusion that is put forth in this chapter that is based on both types of evidence. Which type of report do you find more convincing?

13
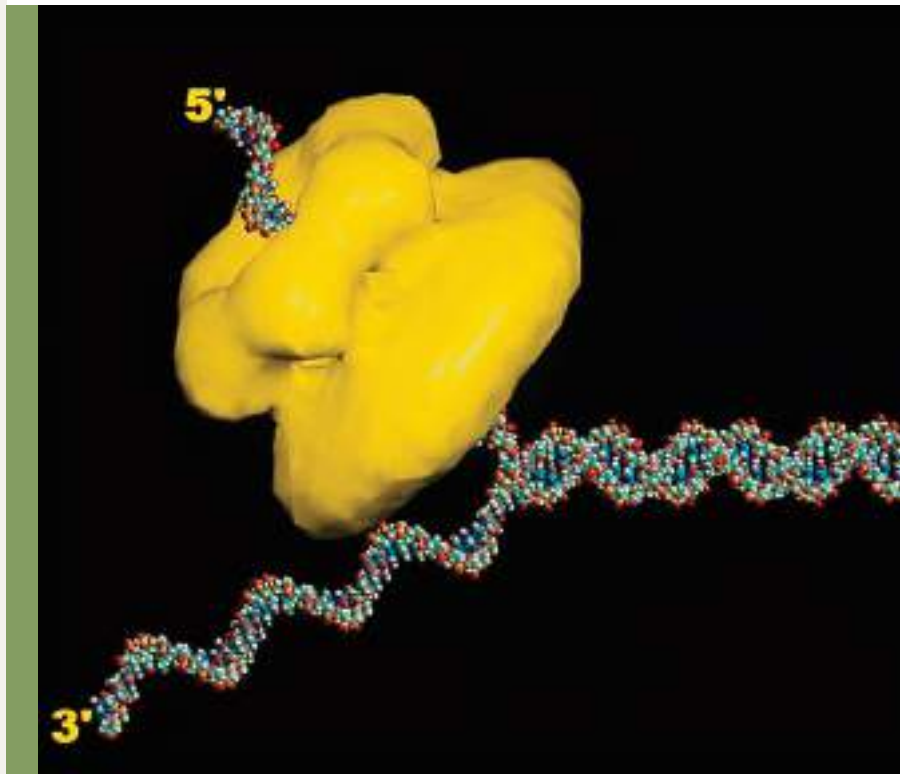
DNA Replication and Repair

Reproduction is a fundamental property of all living systems. The process of reproduction can be observed at several levels: organisms duplicate by asexual or sexual reproduction; cells duplicate by cellular division; and the genetic material duplicates by **DNA replication**. The machinery that replicates DNA is also called into action in another capacity: to repair the genetic material after it has sustained damage. These two processes—DNA replication and DNA repair—are the subjects addressed in this chapter.

The capacity for self-duplication is presumed to have been one of the first critical properties to have appeared in the evolution of the earliest primitive life forms. Without the ability to propagate, any primitive assemblage of biological molecules would be destined for oblivion. The early carriers of genetic information were probably RNA molecules that were able to self-replicate. As evolution progressed and RNA molecules were replaced by DNA molecules as the genetic material, the process of replication became more complex, requiring a large number of auxiliary components. Thus, although a DNA molecule contains the information for its own duplication, it lacks the ability to perform the activity itself. As Richard Lewontin expressed it, "the common image of DNA as a self-replicating molecule is about as true as describing a letter as a self-replicating document. The letter needs a photocopier; the DNA needs a cell." Let us see then how the cell carries out this activity. ■

Three-dimensional model of a DNA helicase encoded by the bacteriophage T7. The protein consists of a ring of six subunits. Each subunit contains two domains. In this model, the central hole encircles only one of the two DNA strands. Driven by ATP hydrolysis, the protein moves in a 5′ → 3′ direction along the strand to which it is bound, displacing the complementary strand and unwinding the duplex. DNA helicase activity is required for DNA replication. (COURTESY OF EDWARD H. EGELMAN, UNIVERSITY OF VIRGINIA.)

13.1 DNA REPLICATION

The proposal for the structure of DNA by Watson and Crick in 1953 was accompanied by a suggested mechanism for its "self-duplication." The two strands of the double helix are held together by hydrogen bonds between the bases. Individually, these hydrogen bonds are weak and readily broken. Watson and Crick envisioned that replication occurred by gradual separation of the strands of the double helix (Figure 13.1), much like the separation of two halves of a zipper. Because the two strands are complementary to each other, each strand contains the information required for construction of the other strand. Thus once the strands are separated, each can act as a template to direct the synthesis of the complementary strand and restore the double-stranded state.

Semiconservative Replication

The Watson-Crick proposal shown in Figure 13.1 made certain predictions concerning the behavior of DNA during replication. According to the proposal, each of the daughter duplexes should consist of one complete strand inherited from the parental duplex and one complete strand that has been newly synthesized. Replication of this type (Figure 13.2, scheme 1) is said to be **semiconservative** because each daughter duplex contains one strand from the parent structure. In the absence of information on the mechanism responsible for replication, two other types of replication had to be consid-
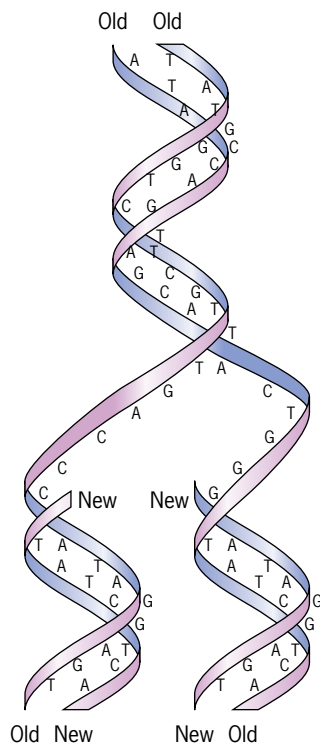
FIGURE 13.1 The original Watson-Crick proposal for the replication of a double-helical molecule of DNA. During replication, the double helix unwinds, and each of the parental strands serves as a template for the synthesis of a new complementary strand. As discussed in this chapter, these basic tenets have been borne out.
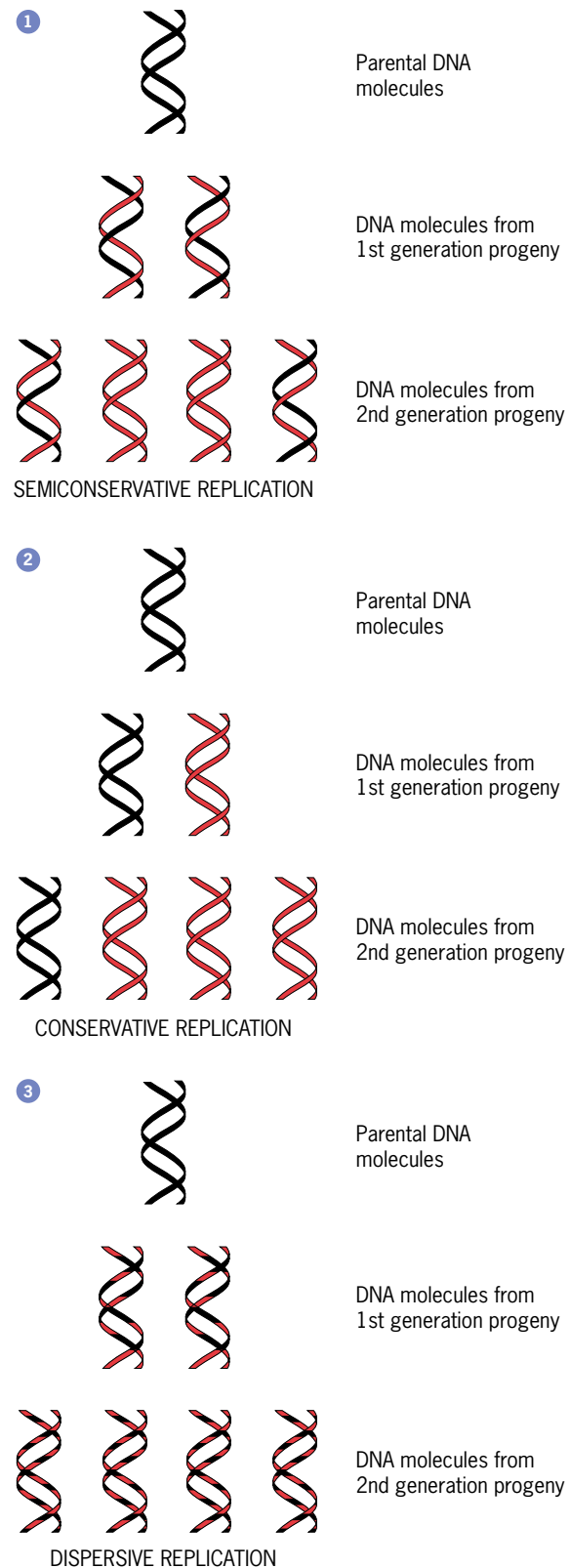
FIGURE 13.2 Three alternate schemes of replication. Semiconservative replication is depicted in scheme 1, conservative replication in scheme 2, and dispersive replication in scheme 3. A description of the three alternate modes of replication is given in the text.

ered. In *conservative* replication (Figure 13.2, scheme 2), the two original strands would remain together (after serving as templates), as would the two newly synthesized strands. As a result, one of the daughter duplexes would contain only parental DNA, while the other daughter duplex would contain only newly synthesized DNA. In *dispersive* replication (Figure 13.2, scheme 3), the parental strands would be broken into fragments, and the new strands would be synthesized in short segments. Then the old fragments and new segments would be joined together to form a complete strand. As a result, the daughter duplexes would contain strands that were composites of old and new DNA. At first glance, dispersive replication might seem like an unlikely solution, but it appeared to Max Delbrück at the time as the only way to avoid the seemingly impossible task of unwinding two intertwined strands of a DNA duplex as it replicated (discussed on page 538).

To decide among these three possibilities, it was necessary to distinguish newly synthesized DNA strands from the original DNA strands that served as templates. This was accomplished in studies on bacteria in 1957 by Matthew Meselson and Franklin Stahl of the California Institute of Technology who used heavy (^{15}N) and light (^{14}N) isotopes of nitrogen to distinguish between parental and newly synthesized DNA strands (Figure 13.3). These researchers grew bacteria in medium containing ^{15}N-ammonium chloride as the sole nitrogen source. Consequently, the nitrogen-containing bases of the DNA of these cells contained only the heavy nitrogen isotope. Cultures of "heavy" bacteria were washed free of the old medium and incubated in fresh medium with light, ^{14}N-containing compounds, and samples were removed at increasing intervals over a period of several generations. DNA was extracted from the samples of bacteria and subjected to equilibrium density-gradient centrifugation (see Figure 18.35). In this procedure, the DNA is mixed with a concentrated solution of cesium chloride and centrifuged until the double-stranded DNA molecules reach equilibrium according to their density.

(a)

(b)

FIGURE 13.3 Experiment demonstrating that DNA replication in bacteria is semiconservative. DNA was extracted from bacteria at different stages in the experiment, mixed with a concentrated solution of the salt cesium chloride (CsCl), placed into a centrifuge tube, and centrifuged to equilibrium at high speed in an ultracentrifuge. Cesium ions have sufficient atomic mass to be affected by the centrifugal force, and they form a density gradient during the centrifugation period with the lowest concentration (lowest density) of Cs at the top of the tube and the greatest concentration (highest density) at the bottom of the tube. During centrifugation, DNA fragments within the tube become localized at a position having a density equal to their own density, which in turn depends

on the ratio of ^{15}N/^{14}N that is present in their nucleotides. The greater the ^{14}N content, the higher in the tube the DNA fragment is found at equilibrium. (*a*) The results expected in this type of experiment for each of the three possible schemes of replication. The single tube on the left indicates the position of the parental DNA and the positions at which totally light or hybrid DNA fragments would band. (*b*) Experimental results obtained by Meselson and Stahl. The appearance of a hybrid band and the disappearance of the heavy band after one generation eliminates conservative replication. The subsequent appearance of two bands, one light and one hybrid, eliminates the dispersive scheme. (B: FROM M. MESELSON AND F. STAHL, PROC. NAT'L. ACAD. SCI. U.S.A. 44:671, 1958.)

In the Meselson-Stahl experiment, the density of a DNA molecule is directly proportional to the percentage of ^{15}N or ^{14}N atoms it contains. If replication is semiconservative, one would expect that the density of DNA molecules would decrease during culture in the ^{14}N-containing medium in the manner shown in the upper set of centrifuge tubes of Figure 13.3a. After one generation, all DNA molecules would be ^{15}N-^{14}N hybrids, and their buoyant density would be halfway between that expected for totally heavy and totally light DNA (Figure 13.3a). As replication continued beyond the first generation, the newly synthesized strands would continue to contain only light isotopes, and two types of duplexes would appear in the gradients: those containing ^{15}N–^{14}N hybrids and those containing only ^{14}N. As the time of growth in the light medium continued, a greater and greater percentage of the DNA molecules present would be light. However, as long as replication continued semiconservatively, the original heavy parental strands would remain intact and present in hybrid DNA molecules that occupied a smaller and smaller percentage of the total DNA (Figure 13.3a). The results of the density-gradient experiments obtained by Meselson and Stahl are shown in Figure 13.13b, and they demonstrate unequivocally that replication occurs semiconservatively. The results that would have been obtained if replication occurred by conservative or dispersive mechanisms are indicated in the two lower sets of centrifuge tubes of Figure 13.3a.[1]

By 1960, replication had been demonstrated to occur semiconservatively in eukaryotes as well. The original experiments were carried out by J. Herbert Taylor of Columbia University. The drawing and photograph of Figure 13.4 show the results of a more recent experiment in which cultured mammalian cells were allowed to undergo replication in bromodeoxyuridine (BrdU), a compound that is incorporated into DNA in place of thymidine. Following replication, a chromosome is made up of two chromatids. After one round of replication in BrdU, both chromatids of each chromosome contained BrdU (Figure 13.4a). After two rounds of replication in BrdU, one chromatid

---

[1]Anyone looking to explore the circumstances leading up to this heralded research effort and examine the experimental twists and turns as they unfolded might want to read the book *Meselson, Stahl, and the Replication of DNA* by Frederick Lawrence Holmes, 2001. A discussion of the experiment can also be found in *PNAS* 101:17889, 2004, which is on the Web.
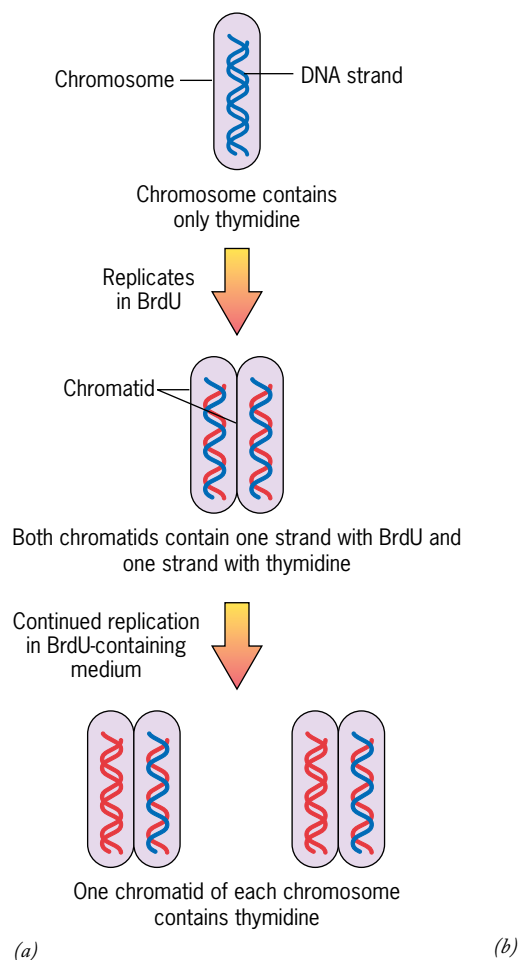
FIGURE 13.4 **Experimental demonstration that DNA replication occurs semiconservatively in eukaryotic cells.** (*a*) Schematic diagram of the results of an experiment in which cells were transferred from a medium containing thymidine to one containing bromodeoxyuridine (BrdU) and allowed to complete two successive rounds of replication. DNA strands containing BrdU are shown in red. (*b*) The results of an experiment similar to that shown in *a*. In this experiment, cultured mammalian cells were grown in BrdU for *two* rounds of replication before mitotic chromosomes were prepared and stained by a procedure using fluorescent dyes and Giemsa stain. Using this procedure, chromatids containing thymidine within one or both strands stain darkly, whereas chromatids containing only BrdU stain lightly. The photograph indicates that, after two rounds of replication in BrdU, one chromatid of each duplicated chromosome contains only BrdU, while the other chromatid contains a strand of thymidine-labeled DNA. (Some of the chromosomes are seen to have exchanged homologous portions between sister chromatids. This process of sister chromatid exchange is common during mitosis but is not discussed in the text.) (B: COURTESY OF SHELDON WOLFF.)

of each chromosome was composed of two BrdU-containing strands, whereas the other chromatid was a hybrid consisting of a BrdU-containing strand and a thymidine-containing strand (Figure 13.4a,b). The thymidine-containing strand had been part of the original parental DNA molecule prior to addition of BrdU to the culture.

Replication in Bacterial Cells

We will focus in this section of the chapter on replication in bacterial cells, which is better understood than the corresponding process in eukaryotes. The early progress in bacterial research was driven by genetic and biochemical approaches including:

- The availability of mutants that cannot synthesize one or another protein required for the replication process. The isolation of mutants unable to replicate their chromosome may seem paradoxical: how can cells with a defect in this vital process be cultured? This paradox was solved by the isolation of **temperature-sensitive** (**ts**) mutants, in which the deficiency only reveals itself at an elevated temperature, termed the *nonpermissive* (or *restrictive*) temperature. When grown at the lower (*permissive*) temperature, the mutant protein can function sufficiently well to carry out its required activity, and the cells can continue to grow and divide. Temperature-sensitive mutants have been isolated that affect virtually every type of physiologic activity (see also page 269), and they have been particularly important in the study of DNA synthesis as it occurs in replication, DNA repair, and genetic recombination.

- The development of in vitro systems in which replication can be studied using purified cellular components. In some studies, the DNA molecule to be replicated is incubated with cellular extracts from which specific proteins suspected of being essential have been removed. In other studies, the DNA is incubated with a variety of purified proteins whose activity is to be tested.

Taken together, these approaches have revealed the activity of more than 30 different proteins that are required to replicate the chromosome of *E. coli*. In the following pages, we will discuss the activities of several of these proteins whose functions have been clearly defined. Replication in bacteria and eukaryotes occurs by very similar mechanisms, and thus most of the information presented in the discussion of bacterial replication applies to eukaryotic cells as well.

Replication Forks and Bidirectional Replication
Replication begins at a specific site on the bacterial chromosome called the **origin**. The origin of replication on the *E. coli* chromosome is a specific sequence called *oriC* where a number of proteins bind to **initiate** the process of replication.[2] Once initiated, replication proceeds outward from the origin in both directions, that is, *bidirectionally* (Figure 13.5). The sites in Figure 13.5 where the pair of replicated segments come to-

[2]The subject of initiation of replication is discussed in detail on page 547 as it occurs in eukaryotes.
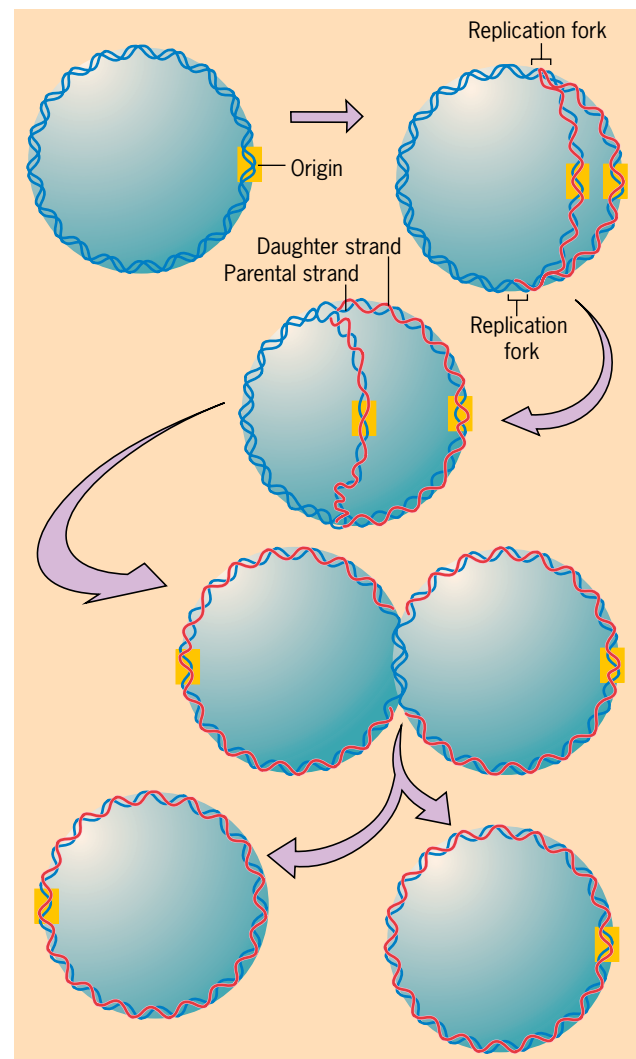
FIGURE 13.5 **Model of a circular bacterial chromosome undergoing bidirectional, semiconservative replication.** Two replication forks move in opposite directions from a single origin. When the replication forks meet at the opposite point on the circle, replication is terminated, and the two replicated duplexes detach from one another. New DNA strands are shown in red.

gether and join the nonreplicated DNA are termed **replication forks**. Each replication fork corresponds to a site where (1) the parental double helix is undergoing strand separation, and (2) nucleotides are being incorporated into the newly synthesized complementary strands. The two replication forks move in opposite directions until they meet at a point across the circle from the origin, where replication is terminated. The two newly replicated duplexes detach from one another and are ultimately directed into two different cells.

Unwinding the Duplex and Separating the Strands
Separation of the strands of a circular, helical DNA duplex poses major topological problems. To visualize the difficulties, we can briefly consider an analogy between a DNA duplex and a two-stranded helical rope. Consider what would happen if you placed a linear piece of this rope on the ground, took hold

of the two strands at one end, and began to pull the strands apart just as DNA is pulled apart during replication. It is apparent that separation of the strands of a double helix is also a process of *unwinding* the structure. In the case of a rope, which is free to rotate around its axis, separation of the strands at one end would be accompanied by rotation of the entire fiber as it resisted the development of tension. Now, consider what would happen if the other end of the rope were attached to a hook on a wall (Figure 13.6*a*). Under these circumstances, separation of the two strands at the free end would generate increasing torsional stress in the rope and cause the unseparated portion to become more tightly wound. Separation of the two strands of a circular DNA molecule (or a linear molecule that is not free to rotate, as is the case in a large eukaryotic chromosome) is analogous to having one end of a linear molecule attached to a wall; in all of these cases, tension that develops in the molecule cannot be relieved by rotation of the entire molecule. Unlike a rope, which can become tightly overwound (as in Figure 13.6*a*), an overwound DNA molecule becomes positively supercoiled (page 391). Consequently, movement of the replication fork generates positive supercoils in the unreplicated portion of the DNA ahead of the fork (Figure 13.6*b*). When one considers that a complete circular chromosome of *E. coli* contains approximately 400,000 turns and is replicated by two forks within 40 minutes, the magnitude of the problem becomes apparent.

(a)

Point of attachment of DNA

Replication machinery

(b)
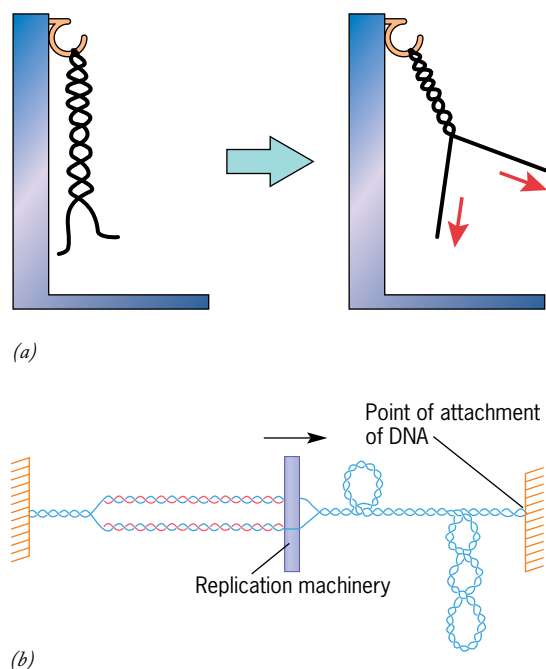
FIGURE 13.6 The unwinding problem. (*a*) The effect of unwinding a two-stranded rope that has one end attached to a hook. The unseparated portion becomes more tightly wound. (*b*) When a circular or attached DNA molecule is replicated, the DNA ahead of the replication machinery becomes overwound and accumulates positive supercoils. Cells possess topoisomerases, such as the *E. coli* DNA gyrase, that remove positive supercoils. (b: Reprinted with permission from J. C. Wang, Nature Reviews Mol. Cell Biol. 3:434, 2002; © copyright 2002, by Macmillan Magazines Limited.)

It was noted on page 391 that cells contain enzymes, called topoisomerases, that can change the state of supercoiling in a DNA molecule. One enzyme of this type, called **DNA gyrase**, a type II topoisomerase, relieves the mechanical strain that builds up during replication in *E. coli*. DNA gyrase molecules travel along the DNA ahead of the replication fork, removing positive supercoils. DNA gyrase accomplishes this feat by cleaving both strands of the DNA duplex, passing a segment of DNA through the double-stranded break to the other side, and then sealing the cuts, a process that is driven by the energy released during ATP hydrolysis (shown in detail in Figure 10.14*b*). Eukaryotic cells possess similar enzymes that carry out this required function.

The Properties of DNA Polymerases   We begin our discussion of the mechanism of DNA replication by describing some of the properties of **DNA polymerases**, the enzymes that synthesize new DNA strands. Study of these enzymes was begun in the 1950s by Arthur Kornberg at Washington University. In their initial experiments, Kornberg and his colleagues purified an enzyme from bacterial extracts that incorporated radioactively labeled DNA precursors into an acid-insoluble polymer identified as DNA. The enzyme was named *DNA polymerase* (and later, after the discovery of additional DNA-polymerizing enzymes, it was named *DNA polymerase I*). For the reaction to proceed, the enzyme required the presence of DNA and all four deoxyribonucleoside triphosphates (dTTP, dATP, dCTP, and dGTP). The newly synthesized, radioactively labeled DNA had the same base composition as the original unlabeled DNA, which strongly suggested that the original DNA strands had served as **templates** for the polymerization reaction.

As additional properties of the DNA polymerase were uncovered, it became apparent that replication was more complex than previously thought. When various types of template DNAs were tested, it was found that the template DNA had to meet certain structural requirements if it was to promote the incorporation of labeled precursors (Figure 13.7). An intact, double-stranded DNA molecule, for example, did not stimulate incorporation. This was not surprising considering the requirement that the strands of the helix must be separated for replication to occur. It was less obvious why a single-stranded, circular molecule was also devoid of activity; one might expect this structure to be an ideal template to direct the manufacture of a complementary strand. In contrast, addition of a partially double-stranded DNA molecule to the reaction mixture produced an immediate incorporation of nucleotides.

It was soon discovered that a single-stranded DNA circle cannot serve as a template for DNA polymerase because the enzyme cannot *initiate* the formation of a DNA strand. Rather, it can only add nucleotides to the 3′ hydroxyl terminus of an existing strand. The strand that provides the necessary 3′ OH terminus is called a **primer**. All DNA polymerases—both prokaryotic and eukaryotic—have these same two basic requirements (Figure 13.8*a*): a template DNA strand to copy and a primer strand to which nucleotides can be added. These requirements explain why certain DNA structures fail to promote DNA synthesis (Figure 13.7*a*). An intact, linear double helix provides the 3′ hydroxyl terminus but lacks a template. A
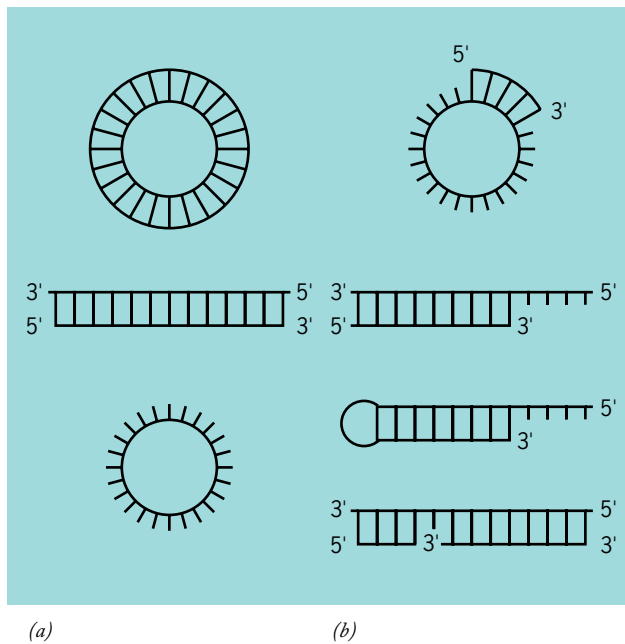
FIGURE 13.7 Templates and nontemplates for DNA polymerase activity. (*a*) Examples of DNA structures that do not stimulate the synthesis of DNA in vitro by DNA polymerase isolated from *E. coli.* (*b*) Examples of DNA structures that stimulate the synthesis of DNA in vitro. In all cases, the molecules in *b* contain a template strand to copy and a primer strand with a 3′ OH on which to add nucleotides.

circular single strand, on the other hand, provides a template but lacks a primer. The partially double-stranded molecule (Figure 13.7*b*) satisfies both requirements and thus promotes nucleotide incorporation. The finding that DNA polymerase cannot initiate the synthesis of a DNA strand raises a critical question: how is the synthesis of a new strand initiated in the cell? We will return to this question shortly.

The DNA polymerase purified by Kornberg had another property that was difficult to understand in terms of its presumed role as a replicating enzyme: it only synthesized DNA in a 5′-to-3′ (written 5′ → 3′) direction. The diagram first presented by Watson and Crick (see Figure 13.1) depicted events as they would be *expected* to occur at the replication fork. The diagram suggested that one of the newly synthesized strands is polymerized in a 5′ → 3′ direction, while the other strand is polymerized in a 3′ → 5′ direction. Is there some other enzyme responsible for the construction of the 3′ → 5′ strand? Does the enzyme work differently in the cell than under in vitro conditions? We will return to this question as well.

During the 1960s, there were hints that the "Kornberg enzyme" was not the only DNA polymerase in a bacterial cell. Then in 1969, a mutant strain of *E. coli* was isolated that had less than 1 percent of the normal activity of the enzyme, yet was able to multiply at the normal rate. Further studies revealed that the Kornberg enzyme, or *DNA polymerase I*, was only one of several distinct DNA polymerases present in bacterial cells. The major enzyme responsible for DNA replication (i.e., the replicative polymerase) is *DNA polymerase III*. A typical bacterial cell contains 300 to 400 molecules of DNA polymerase I but
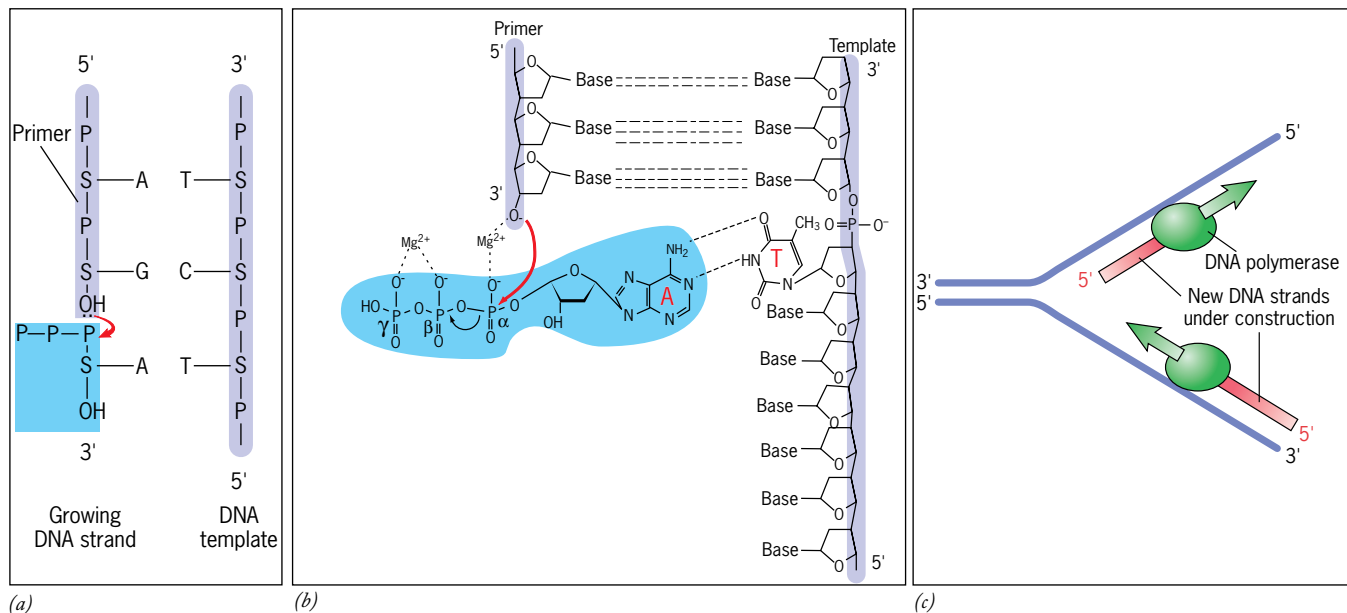
FIGURE 13.8 The activity of a DNA polymerase. (*a*) The polymerization of a nucleotide onto the 3′ end of the primer strand. The enzyme selects nucleotides for incorporation based on their ability to pair with the nucleotide of the template strand. (*b*) A simplified model of the two-metal ion mechanism for the reaction in which nucleotides are incorporated into a growing DNA strand by a DNA polymerase. In this model, one of the magnesium ions draws the proton away from the 3′ hydroxyl group of the terminal nucleotide of the primer, facilitating the nucleophilic attack of the negatively charged 3′ oxygen atom on the α phosphate of the incoming nucleoside triphosphate. The second magnesium ion promotes the release of the pyrophosphate. The two metal ions are bound to the enzyme by highly conserved aspartic acid residues of the active site. (*c*) Schematic diagram showing the direction of movement of each polymerase along the two template strands.

only about 10 copies of DNA polymerase III. The presence of DNA polymerase III had been masked by the much greater amounts of DNA polymerase I in the cell. But the discovery of other DNA polymerases did not answer the two basic questions posed above; none of the enzymes can initiate DNA chains, nor can any of them construct strands in a $3' \rightarrow 5'$ direction.

Semidiscontinuous Replication The lack of polymerization activity in the $3' \rightarrow 5'$ direction has a straightforward explanation: DNA strands cannot be synthesized in that direction. Rather, both newly synthesized strands are assembled in a $5' \rightarrow 3'$ direction. During the polymerization reaction, the —OH group at the 3' end of the primer carries out a nucleophilic attack on the 5' α-phosphate of the incoming nucleoside triphosphate, as shown in Figure 13.8b. The polymerase molecules responsible for construction of the two new strands of DNA both move in a 3'-to-5' direction *along the template*, and both construct a chain that grows from its 5'-P terminus (Figure 13.8c). Consequently, one of the newly synthesized strands grows toward the replication fork where the parental DNA strands are being separated, while the other strand grows away from the fork.

Although this solves the problem concerning an enzyme that synthesizes a strand in only one direction, it creates an even more complicated dilemma. It is apparent that the strand that grows toward the fork in Figure 13.8c can be constructed by the continuous addition of nucleotides to its 3' end. But how is the other strand synthesized? Evidence was soon gathered to indicate that the strand that grows away from the replication fork is synthesized *discontinuously*, that is, as fragments (Figure 13.9). Before the synthesis of a fragment can be initiated, a suitable stretch of template must be exposed by movement of the replication fork. Once initiated, each fragment grows away from the replication fork toward the 5' end of a previously synthesized fragment to which it is subsequently linked. Thus, the two newly synthesized strands of the daughter duplexes are synthesized by very different processes. The strand that is synthesized continuously is called the **leading strand** because its synthesis continues as the replication fork advances. The strand that is synthesized discontinuously is called the **lagging strand** because initiation of each fragment must wait for the parental strands to separate and expose additional template (Figure 13.9). As discussed on page 542, both strands are probably synthesized simultaneously, so that the terms *leading* and *lagging* may not be as appropriate as thought when they were first coined. Because one strand is synthesized continuously and the other discontinuously, replication is said to be *semidiscontinuous*.

The discovery that one strand was synthesized as small fragments was made by Reiji Okazaki of Nagoya University, Japan, following various types of labeling experiments. Okazaki found

(a)

(b)
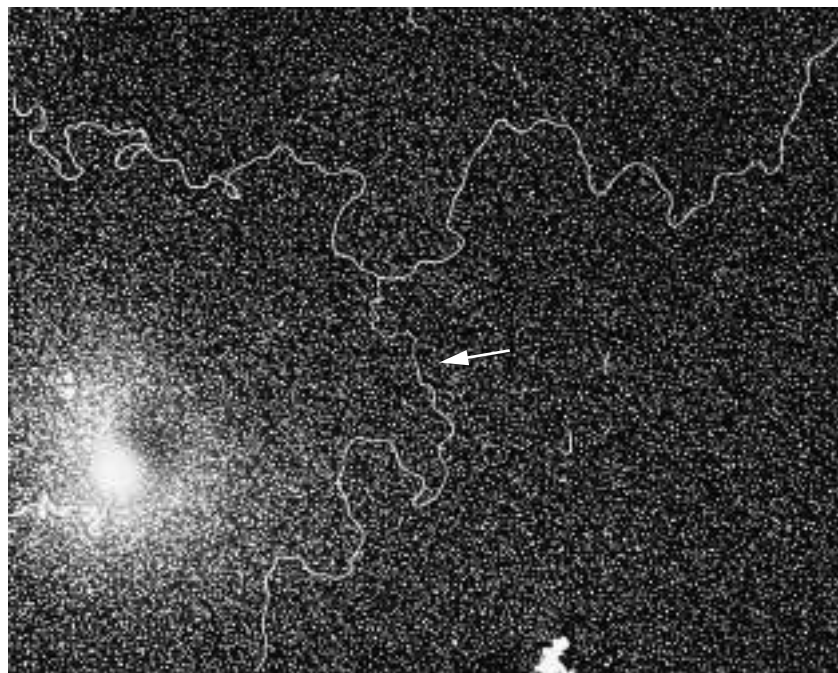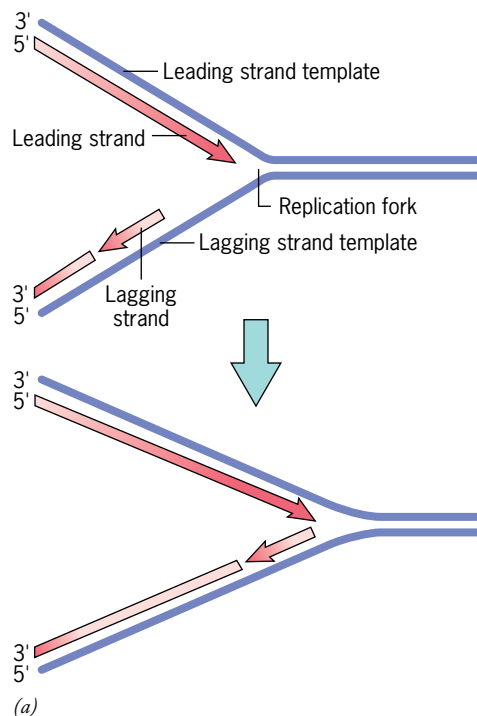
FIGURE 13.9 The two strands of a double helix are synthesized by a different sequence of events. DNA polymerase molecules move along a template only in a $3' \rightarrow 5'$ direction. As a result, the two newly assembled strands grow in opposite directions, one growing toward the replication fork and the other growing away from it. One strand is assembled in continuous fashion, the other as fragments that are joined together enzymatically. (a) Schematic diagram depicting the differences in synthesis of the two strands. (b) Electron micrograph of a replicating bacteriophage DNA molecule. The left two limbs are the replicated duplexes, and the right end is the unreplicated duplex. The lagging strand of the newly replicated DNA is seen to contain an exposed, single-stranded (thinner) portion, which runs from the replication fork to the arrow. (B: FROM J. WOLFSON AND DAVID DRESSLER, REPRINTED WITH PERMISSION FROM THE ANNUAL REVIEW OF MICROBIOLOGY, VOLUME 29; © 1975, BY ANNUAL REVIEWS, INC.)

that if bacteria were incubated in [³H]thymidine for a few seconds and immediately killed, most of the radioactivity could be found as part of small DNA fragments 1000 to 2000 nucleotides in length. In contrast, if cells were incubated in the labeled DNA precursor for a minute or two, most of the incorporated radioactivity became part of much larger DNA molecules (Figure 13.10). These results indicated that a portion of the DNA was constructed in small segments (later called **Okazaki fragments**) that were rapidly linked to longer pieces that had been synthesized previously. The enzyme that joins the Okazaki fragments into a continuous strand is called **DNA ligase**.

The discovery that the lagging strand is synthesized in pieces raised a new set of perplexing questions about the initiation of DNA synthesis. How does the synthesis of each of these fragments begin when none of the DNA polymerases are capable of strand initiation? Further studies revealed that initiation is not accomplished by a DNA polymerase but, rather, by a distinct type of RNA polymerase, called **primase**, that con-

structs a short primer composed of RNA, not DNA. The leading strand, whose synthesis begins at the origin of replication, is also initiated by a primase molecule. The short RNAs synthesized by the primase at the 5′ end of the leading strand and the 5′ end of each Okazaki fragment serve as the required primer for the synthesis of DNA by a DNA polymerase. The RNA primers are subsequently removed, and the resulting gaps in the strand are filled with DNA and then sealed by DNA ligase. These events are illustrated schematically in Figure 13.11. The formation of transient RNA primers during the process of DNA replication is a curious activity. It is thought that the likelihood of mistakes is greater during initiation than during elongation, and the use of a short removable segment of RNA avoids the inclusion of mismatched bases.

The Machinery Operating at the Replication Fork Replication involves more than incorporating nucleotides. Unwinding the duplex and separating the strands require the aid of two types of proteins that bind to the DNA, a **helicase** (or DNA unwinding enzyme) and **single-stranded DNA-binding** (**SSB**) **proteins**. DNA helicases unwind a DNA duplex in a reaction that uses energy released by ATP hydrolysis to move along one of the DNA strands, breaking the hydrogen bonds that hold the two strands together and exposing the single-stranded DNA templates. *E. coli* has at least 12 different helicases for use in various aspects of DNA (and RNA)

FIGURE 13.10 Results of an experiment showing that part of the DNA is synthesized as small fragments. Sucrose density gradient profiles of DNA from a culture of phage-infected *E. coli* cells. The cells were labeled for increasing amounts of time, and the sedimentation velocity of the labeled DNA was determined. When DNA was prepared after very short pulses, a significant percentage of the radioactivity appeared in very short pieces of DNA (represented by the peak near the top of the tube on the left). After periods of 60–120 seconds, the relative height of this peak falls as labeled DNA fragments become joined to the ends of high-molecular-weight molecules. (FROM R. OKAZAKI ET AL., COLD SPRING HARBOR SYMP. QUANT. BIOL. 33:130, 1968.)

FIGURE 13.11 The use of short RNA fragments as removable primers in initiating synthesis of each Okazaki fragment of the lagging strand. The major steps are indicated in the drawing and discussed in the text. The role of various accessory proteins in these activities is indicated in the following figures.
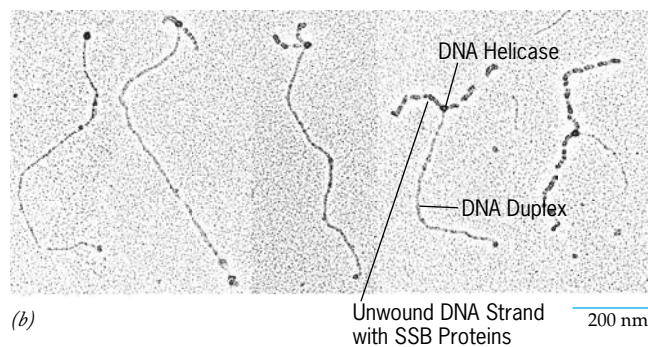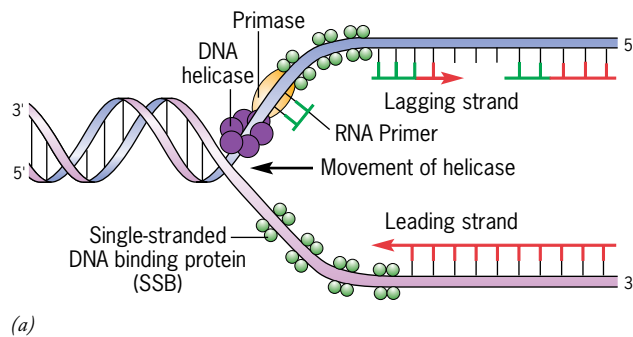
FIGURE 13.12 **The role of the DNA helicase, single-stranded DNA-binding proteins, and primase at the replication fork.** (*a*) The helicase moves along the DNA, catalyzing the ATP-driven unwinding of the duplex. As the DNA is unwound, the strands are prevented from reforming the duplex by single-stranded DNA-binding proteins (SSBs). The primase associated with the helicase synthesizes the RNA primers that begin each Okazaki fragment. The RNA primers, which are about 10 nucleotides long, are subsequently removed. (*b*) A series of five electron micrographs showing DNA molecules incubated with a viral DNA helicase (T antigen, page 550) and *E. coli* SSB proteins. The DNA molecules are progressively unwound from left to right. The helicase appears as the round particle at the fork, and the SSB proteins are bound to the single-stranded ends, giving them a thickened appearance. (B: FROM RAINER WESSEL, JOHANNES SCHWEIZER, AND HANS STAHL, J. VIROL. 66:807, 1992; COPYRIGHT © 1992, AMERICAN SOCIETY FOR MICROBIOLOGY.)

metabolism. One of these helicases—the product of the *dnaB* gene—serves as the major unwinding machine during replication. The DnaB helicase consists of six subunits arranged to form a ring-shaped protein that encircles a single DNA strand (Figure 13.12*a*). The DnaB helicase is first loaded onto the DNA at the origin of replication (with the help of the protein DnaC) and translocates in a $5' \rightarrow 3'$ direction along the lagging-strand template, unwinding the helix as it proceeds (Figure 13.12). A three-dimensional model of a similar shaped bacteriophage helicase engaged in strand separation during replication is depicted on page 533. DNA unwinding by the helicase is aided by the attachment of SSB proteins to the separated DNA strands (Figure 13.12). These proteins bind selectively to single-stranded DNA, keeping it in an extended state and preventing it from becoming rewound or damaged. A visual portrait of the combined action of a DNA helicase and SSB proteins on the structure of the DNA double helix is illustrated in the electron micrographs of Figure 13.12*b*.

Recall that an enzyme called primase initiates the synthesis of each Okazaki fragment. In bacteria, the primase and the helicase associate transiently to form what is called a "primosome." Of the two members of the primosome, the helicase moves along the lagging-strand template processively (i.e., without being released from the template strand during the lifetime of the replication fork). As the helicase "motors" along the lagging-strand template, opening the strands of the duplex, the primase periodically binds to the helicase and synthesizes the short RNA primers that begin the formation of each Okazaki fragment. As noted above, the RNA primers are subsequently extended as DNA by a DNA polymerase, specifically DNA polymerase III.

A body of evidence suggests that the same DNA polymerase III molecule synthesizes successive fragments of the lagging strand. To accomplish this, the polymerase III molecule is recycled from the site where it has just completed one Okazaki fragment to the next site along the lagging-strand template closer to the site of DNA unwinding. Once at the new site, the polymerase attaches to the 3′ OH of the RNA primer that has just been laid down by a primase and begins to incorporate deoxyribonucleotides onto the end of the short RNA.

How does a polymerase III molecule move from one site on the lagging-strand template to another site that is closer to the replication fork? The enzyme does this by "hitching a ride" with the DNA polymerase that is moving in that direction along the leading-strand template. Thus even though the two polymerases are moving in opposite directions with respect to the linear axis of the DNA molecule, they are, in fact, part of a single protein complex (Figure 13.13). The two tethered polymerases can replicate both strands by looping the DNA of the lagging-strand template back on itself, causing this template to have the same orientation as the leading-strand template. Both polymerases then can move together as part of a single replicative complex without violating the "$5' \rightarrow 3'$ rule" for synthesis of a DNA strand (Figure 13.13). Once the polymerase assembling the lagging strand reaches the 5′ end of the Okazaki fragment synthesized during the previous round, the lagging-strand template is released and the polymerase begins work at the 3′ end of the next RNA primer toward the fork. The model depicted in Figure 13.13 is often referred to as the "trombone model" because the looping DNA repeatedly grows and shortens during the replication of the lagging strand, reminiscent of the movement of the brass "loop" of a trombone as it is played.

The Structure and Functions of DNA Polymerases

DNA polymerase III, the enzyme that synthesizes DNA strands during replication in *E. coli*, is part of a large "replication machine" called the *DNA polymerase III holoenzyme* (Figure 13.14). One of the noncatalytic components of the holoenzyme, called the **β clamp**, keeps the polymerase associated with the DNA template. DNA polymerases (like RNA polymerases) possess two somewhat contrasting properties: (1) they must remain associated with the template over long
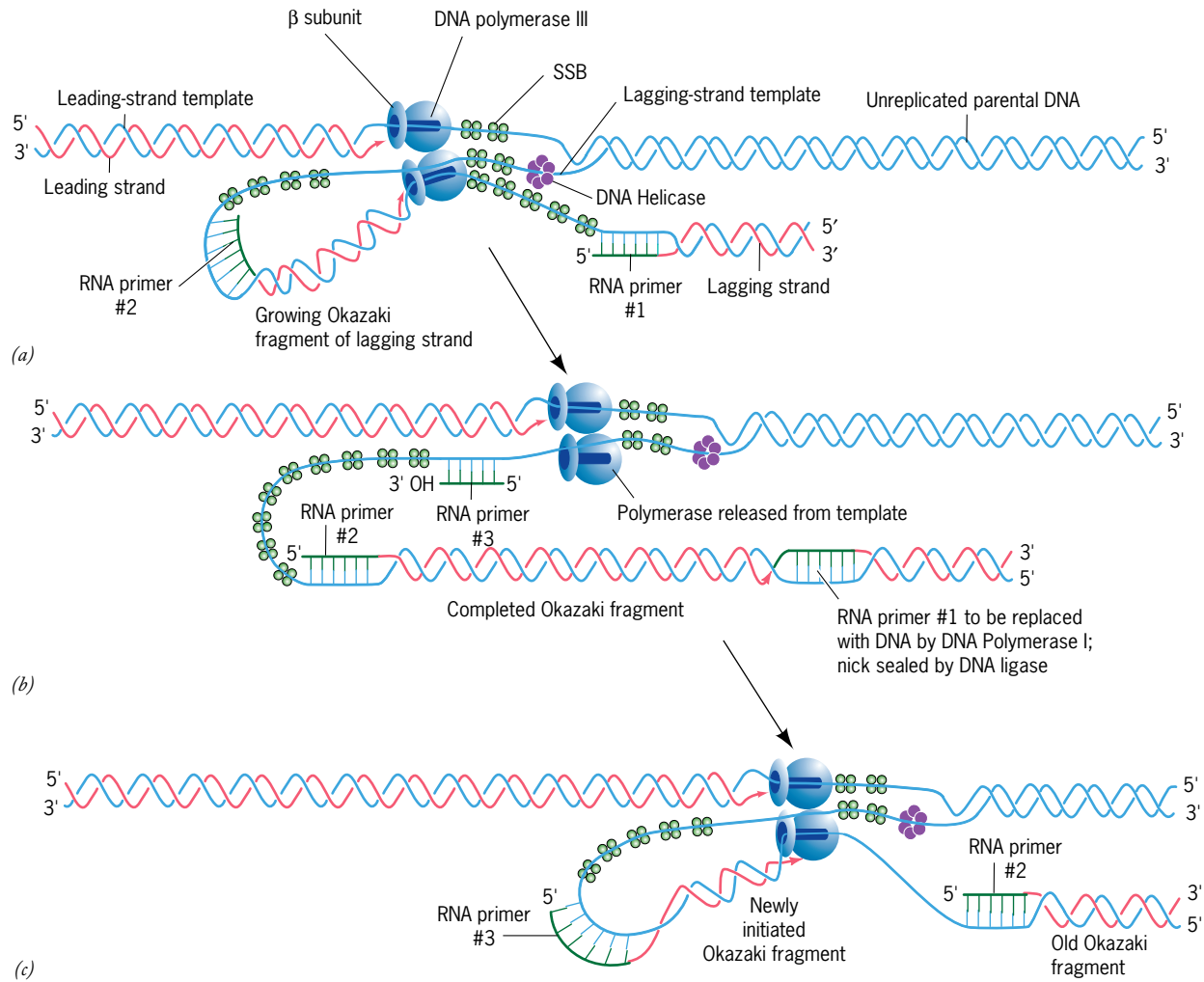
FIGURE 13.13 Replication of the leading and lagging strands in *E. coli* is accomplished by two DNA polymerases working together as part of a single complex. (*a*) The two DNA polymerase III molecules travel together, even though they are moving toward the opposite ends of their respective templates. This is accomplished by causing the lagging-strand template to form a loop. (*b*) The polymerase releases the lagging-strand template when it encounters the previously synthesized Okazaki fragment.

(*c*) The polymerase that was involved in the assembly of the previous Okazaki fragment has now rebound the lagging-strand template farther along its length and is synthesizing DNA onto the end of RNA primer #3 that has just been constructed by the primase. (AFTER D. VOET AND J. G. VOET, BIOCHEMISTRY, 2D ED.; COPYRIGHT © 1995, JOHN WILEY AND SONS, INC. REPRINTED BY PERMISSION OF JOHN WILEY AND SONS, INC.)

stretches if they are to synthesize a continuous complementary strand, and (2) they must be attached loosely enough to the template to move from one nucleotide to the next. These contrasting properties are provided by the doughnut-shaped β

FIGURE 13.14 Schematic representation of DNA polymerase III holoenzyme. The holoenzyme contains ten different subunits organized into several distinct components. Included as part of the holoenzyme are (1) two core polymerases which replicate the DNA, (2) two or more β clamps, which allow the polymerase to remain associated with the DNA, and (3) a clamp loading (γ) complex, which loads each sliding clamp onto the DNA. The clamp loader of an active replication fork contains two τ subunits, which hold the core polymerases in the complex and also bind the helicase. Another term, the *replisome,* is often used to refer to the entire complex of proteins that are active at the replication fork, including the DNA polymerase III holoenzyme, the helicase, SSBs, and primase. (BASED ON DRAWINGS BY M. O'DONNELL.)
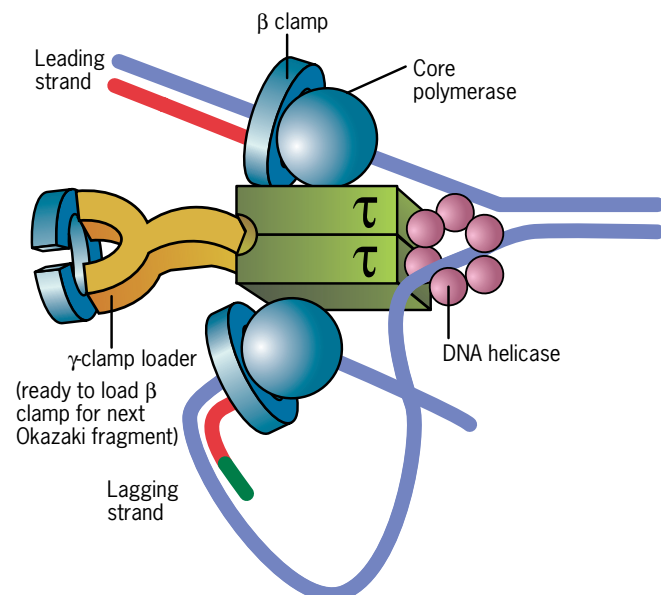
clamp that encircles the DNA (Figure 13.15a) and slides freely along it. As long as it is attached to a β "sliding clamp," a DNA polymerase can move processively from one nucleotide to the next without diffusing away from the template. The polymerase on the leading-strand template remains tethered to a single β clamp during replication. In contrast, when the polymerase on the lagging-strand template completes the synthesis of an Okazaki fragment, it disengages from the β clamp and is cycled to a new β clamp that has been assembled at an RNA primer–DNA template junction located closer to the replication fork (Figure 13.15b). But how does a highly elongated DNA molecule get inside of a ring-shaped clamp as in Figure 13.15a? The assembly of the β clamp around the DNA requires a multisubunit *clamp loader* that is also part of the DNA polymerase III holoenzyme (Figures 13.14, 13.15c). In the ATP-bound state, the clamp loader binds to a primer-template junction while holding the β clamp in an open conformation as illustrated in Figure 13.15c. Once the DNA has squeezed

through the opening in the clamp wall, the ATP bound to the clamp loader is hydrolyzed, causing the release of the clamp, which closes around the DNA. The β clamp is then ready to bind polymerase III as depicted in Figure 13.15b.

DNA polymerase I, which consists of only a single subunit, is involved primarily in DNA repair, a process by which damaged sections of DNA are corrected (page 552). DNA polymerase I also removes the RNA primers at the 5′ end of each Okazaki fragment during replication and replaces them with DNA. The enzyme's ability to accomplish this feat is discussed in the following section.

Exonuclease Activities of DNA Polymerases Now that we have explained several of the puzzling properties of DNA polymerase I, such as the enzyme's inability to initiate strand synthesis, we can consider another curious observation. Kornberg found that DNA polymerase I preparations always contained exonuclease activities; that is, they were able to degrade DNA

(a)

(b)

(c)
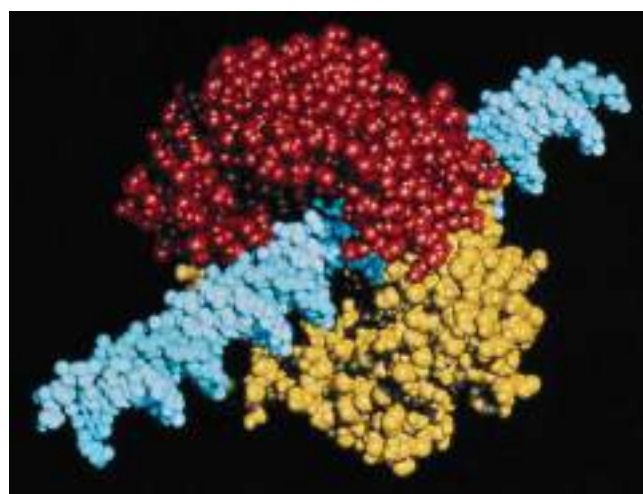
FIGURE 13.15 **The β sliding clamp and clamp loader.** (*a*) Space-filling model showing the two subunits that make up the doughnut-shaped β sliding clamp in *E. coli*. Double-stranded DNA is shown in blue within the β clamp. (*b*) Schematic diagram of polymerase cycling on the lagging strand. The polymerase is held to the DNA by the β sliding clamp as it moves along the template strand and synthesizes the complementary strand. Following completion of the Okazaki fragment, the enzyme disengages from its β clamp and cycles to a recently assembled clamp "waiting" at an upstream RNA primer–DNA template junction. The original β clamp is left behind for a period on the finished Okazaki fragment, but it is eventually disassembled and reutilized. (*c*) A model of a complex between a sliding clamp and a clamp loader from an archaean prokaryote based on electron microscopic image analysis. The clamp loader (shown with red and green subunits) is bound to the sliding clamp (blue), which is held in an open, spiral conformation resembling a lock-washer. The DNA has squeezed through the gap in the clamp. The primer strand of the DNA terminates within the clamp loader whereas the template strand extends through an opening at the top of the protein. The clamp loader has been described as a "screw-cap" that fits onto the DNA in such a way that the subdomains of the protein form a spiral that can thread onto the helical DNA backbone. (A: FROM JOHN KURIYAN, CELL, 69:427, 1992; © CELL PRESS; B: AFTER P. T. STUKENBERG, J. TURNER, AND M. O'DONNELL, CELL 78:878, 1994; BY PERMISSION OF CELL PRESS; C: FROM T. MIYATA ET AL., PROC. NAT'L. ACAD. SCI. U.S.A. 102:13799, 2005; COURTESY OF K. MORIKAWA)
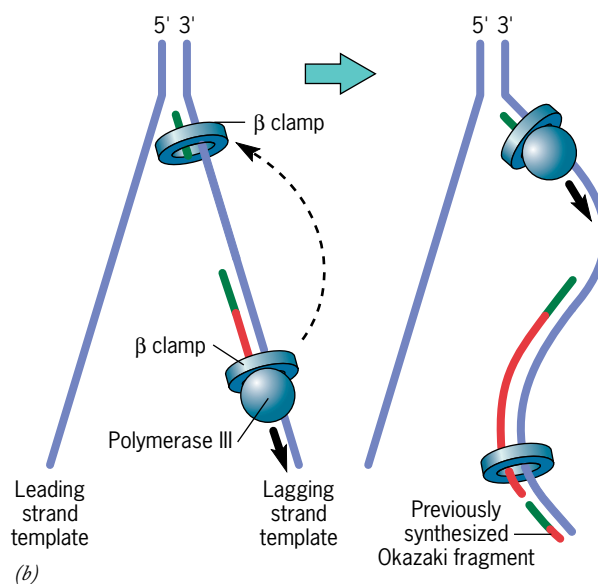
polymers by removing one or more nucleotides from the end of the molecule. At first, Kornberg assumed this activity was due to a contaminating enzyme because the action of exonucleases is so dramatically opposed to that of DNA synthesis. Nonetheless, the exonuclease activity could not be removed from the polymerase preparation and was, in fact, a true activity of the polymerase molecule. It was subsequently shown that all of the bacterial DNA polymerases possess exonuclease activity. Exonucleases can be divided into $5' \rightarrow 3'$ and $3' \rightarrow 5'$ exonucleases, depending on the direction in which the strand is degraded. DNA polymerase I has both $3' \rightarrow 5'$ and $5' \rightarrow 3'$ exonuclease activities, in addition to its polymerizing activity (Figure 13.16). These three activities are found in different domains of the single polypeptide. Thus, remarkably, DNA polymerase I is three different enzymes in one. The two exonuclease activities have entirely different roles in replication. We will consider the $5' \rightarrow 3'$ exonuclease activity first.

Most nucleases are specific for either DNA or RNA, but the $5' \rightarrow 3'$ exonuclease of DNA polymerase I can degrade either type of nucleic acid. Initiation of Okazaki fragments by the primase leaves a stretch of RNA at the $5'$ end of each fragment (see RNA primer #1 of Figure 13.13b), which is removed by the $5' \rightarrow 3'$ exonuclease activity of DNA polymerase I (Figure 13.16a). As the enzyme removes ribonu-

cleotides of the primer, its polymerase activity simultaneously fills the resulting gap with deoxyribonucleotides. The last deoxyribonucleotide incorporated is subsequently joined covalently to the $5'$ end of the previously synthesized DNA fragment by DNA ligase. The role of the $3' \rightarrow 5'$ exonuclease activity will be apparent in the following section.

Ensuring High Fidelity during DNA Replication   The survival of an organism depends on the accurate duplication of the genome. A mistake made in the synthesis of a messenger RNA molecule by an RNA polymerase results in the synthesis of defective proteins, but an mRNA molecule is only one short-lived template among a large population of such molecules; therefore, little lasting damage results from the mistake. In contrast, a mistake made during DNA replication results in a permanent mutation and the possible elimination of that cell's progeny. In *E. coli*, the chance that an incorrect nucleotide will be incorporated into DNA during replication and remain there is less than 10^{-9}, or fewer than 1 out of 1 billion nucleotides. Because the genome of *E. coli* contains approximately 4×10^6 nucleotide pairs, this error rate corresponds to fewer than 1 nucleotide alteration for every 100 replication cycles. This represents the *spontaneous mutation rate* in this bacterium. Humans are thought to have a similar spontaneous mutation rate for replication of protein-coding sequences.

Incorporation of a particular nucleotide onto the end of a growing strand depends on the incoming nucleoside triphosphate being able to form an acceptable base pair with the nucleotide of the template strand (see Figure 13.8b). Analysis of the distances between atoms and bond angles indicates that A-T and G-C base pairs have nearly identical geometry (i.e., size and shape). Any deviation from those pairings results in a different geometry, as shown in Figure 13.17. At each site along the template, DNA polymerase must discriminate

FIGURE 13.16 **The exonuclease activities of DNA polymerase I.** (*a*) The $5' \rightarrow 3'$ exonuclease function removes nucleotides from the $5'$ end of a single-strand nick. This activity plays a key role in removing the RNA primers. (*b*) The $3' \rightarrow 5'$ exonuclease function removes mispaired nucleotides from the $3'$ end of the growing DNA strand. This activity plays a key role in maintaining the accuracy of DNA synthesis. (FROM D. VOET AND J. G. VOET, BIOCHEMISTRY, 2D ED.; COPYRIGHT © 1995, JOHN WILEY AND SONS, INC. REPRINTED BY PERMISSION OF JOHN WILEY AND SONS, INC.)

FIGURE 13.17 **Geometry of proper and mismatched base pairs.** (FROM H. ECHOLS AND M. F. GOODMAN, REPRODUCED WITH PERMISSION FROM THE ANNUAL REVIEW OF BIOCHEMISTRY, VOLUME 60; © 1991, BY ANNUAL REVIEWS INC.)

among four different potential precursors as they move in and out of the active site. Among the four possible incoming nucleoside triphosphates, only one forms a proper geometric fit with the template, producing either an A-T or a G-C base pair that can fit into a binding pocket within the enzyme. This is only the first step in the discrimination process. If the incoming nucleotide is "perceived" by the enzyme as being correct, a conformational change occurs in which the "fingers" of the polymerase rotate toward the "palm" (Figure 13.18a), gripping the incoming nucleotide. This is an example of an induced fit as discussed on page 98. If the newly formed base pair exhibits improper geometry, the active site cannot achieve the conformation required for catalysis and the incorrect nucleotide is not incorporated. In contrast, if the base pair exhibits proper geometry, the incoming nucleotide is covalently linked to the end of the growing strand.

On occasion, the polymerase incorporates an incorrect nucleotide, resulting in a mismatched base pair, that is, a base pair other than A-T or G-C. It is estimated that an incorrect pairing of this sort occurs once for every 10^5–10^6 nucleotides incorporated, a frequency that is 10^3–10^4 times greater than the spontaneous mutation rate of approximately 10^{-9}. How is the mutation rate kept so low? Part of the answer lies in the second of the two exonuclease activities mentioned above, the $3' \rightarrow 5'$ activity (Figure 13.16b). When an incorrect nucleotide is incorporated by DNA polymerase I, the enzyme stalls and the end of the newly synthesized strand has an increased tendency to separate from the template and form a single-stranded $3'$ terminus. When this occurs, the frayed end of the newly synthesized strand is directed into the $3' \rightarrow 5'$ exonuclease site (Figure 13.18), which removes the mismatched nucleotide. This job of "proofreading" is one of the most remarkable of all enzymatic activities and illustrates the sophistication to which biological molecular machinery has evolved. The $3' \rightarrow 5'$ exonuclease activity removes approximately 99 out of every 100 mismatched bases, raising the fidelity to about 10^{-7}–10^{-8}. In addition, bacteria possess a mechanism called mismatch repair that operates after replication (page 554) and corrects nearly all of the mismatches that escape the proofreading step. Together these processes reduce the overall observed error rate to about 10^{-9}. Thus the fidelity of DNA replication can be traced to three distinct activities: (1) accurate selection of nucleotides, (2) immediate proofreading, and (3) postreplicative mismatch repair.

Another remarkable feature of bacterial replication is its rate. The replication of an entire bacterial chromosome in approximately 40 minutes at 37°C requires that each replication fork move about 1000 nucleotides per second, which is equivalent to the length of an entire Okazaki fragment. Thus the entire process of Okazaki fragment synthesis, including formation of an RNA primer, DNA elongation and simultaneous proofreading by the DNA polymerase, excision of the RNA, its replacement with DNA, and strand ligation, occurs within a few seconds. Although it takes *E. coli* approximately 40 minutes to replicate its DNA, a new round of replication can begin before the previous round has been completed. Consequently, when these bacteria are growing at their maximal rate, they double their numbers in about 20 minutes.
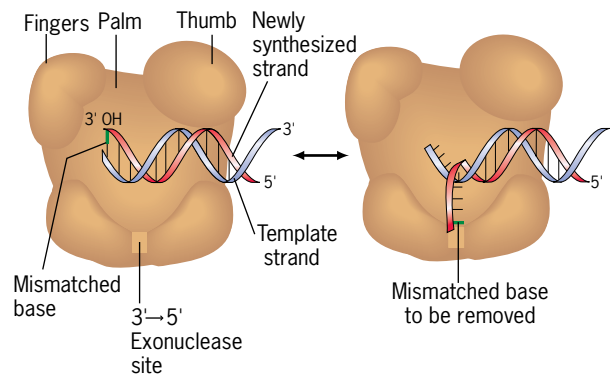
(a)
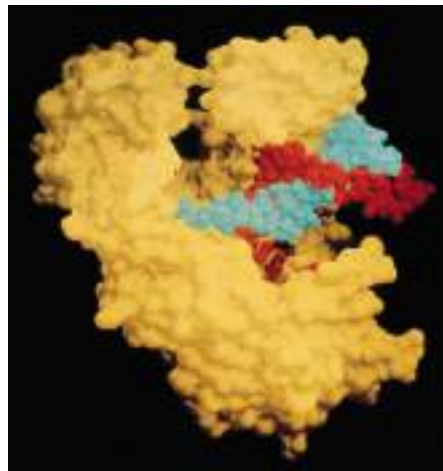
(b)

FIGURE 13.18 Activation of the $3' \rightarrow 5'$ exonuclease of DNA polymerase I. (a) A schematic model of a portion of DNA polymerase I known as the Klenow fragment, which contains the polymerase and $3' \rightarrow 5'$ exonuclease active sites. The $5' \rightarrow 3'$ exonuclease activity is located in a different portion of the polypeptide, which is not shown here. The regions of the Klenow fragment are often likened to the shape of a partially opened right hand, hence the portions labeled as "fingers," "palm," and "thumb." The catalytic site for polymerization is located in the central "palm" subdomain. The $3'$ terminus of the growing strand can be shuttled between the polymerase and exonuclease active sites. Addition of a mismatched base to the end of the growing strand produces a frayed (single-stranded) $3'$ end that enters the exonuclease site, where it is removed. (The polymerase and exonuclease sites of polymerase III operate similarly but are located on different subunits.) (b) A molecular model of the Klenow fragment complexed to DNA. The template DNA strand being copied is shown in blue, and the primer strand to which the next nucleotides would be added is shown in red. (A: AFTER T. A. BAKER AND S. P. BELL, CELL 92:296, 1998; AFTER A DRAWING BY C. M. JOYCE AND T. A. STEITZ, BY PERMISSION OF CELL PRESS; B: COURTESY OF THOMAS A. STEITZ.)

Replication in Eukaryotic Cells

As noted in Chapter 10, the nucleotide letters of the human genome sequence would fill a book roughly one million pages in length. While it took several years for hundreds of researchers to sequence the human genome, a single cell nucleus of approximately 10 μm diameter can copy all of this DNA within a few hours. Given the fact that eukaryotic cells have large genomes

and complex chromosomal structure, our understanding of replication in eukaryotes has lagged behind that in bacteria. This imbalance has been addressed by the development of eukaryotic experimental systems that parallel those used for decades to study bacterial replication. These include

- The isolation of mutant yeast and animal cells unable to produce specific gene products required for various aspects of replication.

- Analysis of the structure and mechanism of action of replication proteins from archaeal species (as in Figure 13.15*c*). Replication in these prokaryotes begins at multiple origins and requires proteins that are homologous to those of eukaryotic cells but are less complex and easier to study.

- The development of in vitro systems where replication can occur in cellular extracts or mixtures of purified proteins. The most valuable of these systems has utilized *Xenopus*, an aquatic frog that begins life as a huge egg stocked with all of the proteins required to carry it through a dozen or so very rapid rounds of cell division. Extracts can be prepared from these frog eggs that will replicate any added DNA, regardless of sequence. Frog egg extracts will also support the replication and mitotic division of mammalian nuclei, which has made this a particularly useful cell-free system. Antibodies can be used to deplete the extracts of particular proteins, and the replication ability of the extract can then be tested in the absence of the affected protein.

Initiation of Replication in Eukaryotic Cells Replication in *E. coli* begins at only one site along the single, circular chromosome (Figure 13.5). Cells of higher organisms may have a thousand times as much DNA as this bacterium, yet their polymerases incorporate nucleotides into DNA at much slower rates. To accommodate these differences, eukaryotic cells replicate their genome in small portions, termed *replicons*. Each replicon has its own origin from which replication forks proceed outward in both directions (see Figure 13.24*a*). In a human cell, replication begins at about 10,000 to 100,000 different replication origins. The existence of replicons was first demonstrated in autoradiographic experiments in which single DNA molecules were shown to be replicated simultaneously at several sites along their length (Figure 13.19).

Approximately 10 to 15 percent of replicons are actively engaged in replication at any given time during the S phase of the cell cycle (see Figure 14.1). Replicons located close together in a given chromosome tend to undergo replication simultaneously (as evident in Figure 13.19). Moreover, those replicons active at a particular time during one round of DNA synthesis tend to be active at a comparable time in succeeding rounds. In mammalian cells, the timing of replication of a chromosomal region is roughly correlated with the activity of the genes in the region and/or its state of compaction. The presence of acetylated histones, which is closely correlated with gene transcription (page 516), is a likely factor in determining the early replication of active gene loci. The most highly compacted, least acetylated regions of the chromosome are packaged into heterochromatin (page 485), and they are the last regions to be replicated. This difference in timing of replica-
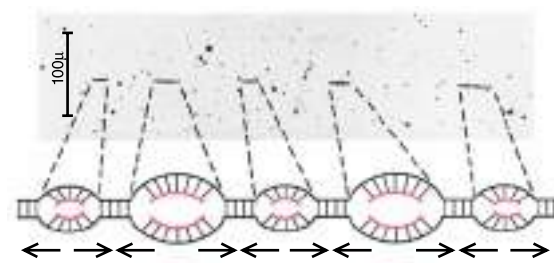
FIGURE 13.19 **Experimental demonstration that replication in eukaryotic chromosomes begins at many sites along the DNA.** Cells were incubated in [³H] thymidine for a brief period before preparation of DNA fibers for autoradiography. The lines of black silver grains indicate sites that had incorporated the radioactive DNA precursor during the labeling period. It is evident that synthesis is occurring at separated sites along the same DNA molecule. As indicated in the accompanying line drawing, initiation begins in the center of each site of thymidine incorporation, forming two replication forks that travel away from each other until they meet a neighboring fork. (MICROGRAPH COURTESY OF JOEL HUBERMAN.)

tion is not related to DNA sequence because the inactive, heterochromatic X chromosome in the cells of female mammals (page 486) is replicated late in S phase, whereas the active, euchromatic X chromosome is replicated at an earlier stage.

The mechanism by which replication is initiated in eukaryotes has been a focus of research over the past decade. The greatest progress in this area has been made with budding yeast because the origins of replication can be removed from the yeast chromosome and inserted into bacterial DNA molecules, conferring on them the ability to replicate either within a yeast cell or in cellular extracts containing the required eukaryotic replication proteins. Because these sequences promote replication of the DNA in which they are contained, they are referred to as *autonomous replicating sequences* (*ARSs*). Those ARSs that have been isolated and analyzed share several distinct elements. The core element of an ARS consists of a conserved sequence of 11 base pairs, which functions as a specific binding site for an essential multiprotein complex called the *origin recognition complex* (*ORC*) (see Figure 13.20). If the ARS is mutated so that it is unable to bind the ORC, initiation of replication cannot occur.

Replication origins have proven more difficult to study in vertebrate cells than in yeast. Part of the problem stems from the fact that virtually any type of purified, naked DNA is suitable for replication using extracts from frog eggs. These studies suggested that, unlike yeast, vertebrate DNA does not possess specific sequences (e.g., ARSs) at which replication is initiated. However, studies of replication of intact mammalian chromosomes in vivo suggest that replication does begin within defined regions of the DNA, rather than by random selection as occurs in the amphibian egg extract. It is thought that a DNA molecule contains many sites where DNA replication *can* be initiated, but only a subset of these sites are actually used at a given time in a given cell. Cells that reproduce via shorter cell cycles, such as those of early amphibian embryos, utilize a greater number of sites as origins of replication than cells with longer cell cycles. The actual selection of sites for initiation of replication is thought to be governed by local epigenetic factors

(page 496), such as the positions of nucleosomes, the types of histone modifications, the state of DNA methylation, the degree of supercoiling, and the level of transcription.

Restricting Replication to Once Per Cell Cycle   It is essential that each portion of the genome is replicated once, and only once, during each cell cycle. Consequently, some mechanism must exist to prevent the reinitiation of replication at a site that has already been duplicated. The initiation of replication at a particular origin requires passage of the origin through several distinct states. Some of the steps that occur at an origin of replication in a yeast cell are illustrated in Figure 13.20. Similar steps requiring homologous proteins take place in plants and animals, suggesting that the basic mechanism of initiation of replication is conserved among eukaryotes.

1. In step 1 (Figure 13.20), the origin of replication is bound by an ORC protein complex, which in yeast cells remains associated with the origin throughout the cell cycle. The ORC has been described as a "molecular landing pad" because of its role in binding the proteins required in subsequent steps.

2. Proteins referred to as "licensing factors" bind to the ORC (step 2, Figure 13.20) to assemble a protein–DNA complex, called the *prereplication complex* (*pre-RC*), that is "licensed" (competent) to initiate replication. Studies of the molecular nature of the licensing factors have focused on a set of six related Mcm proteins (Mcm2–Mcm7). The Mcm proteins are loaded onto the replication origin at a late stage of mitosis, or soon thereafter. Studies indicate that the Mcm2–Mcm7 proteins are capable of associating into a

ring-shaped complex that possesses helicase activity (as in step 4, Figure 13.20). Most evidence suggests that the Mcm2–Mcm7 complex is the eukaryotic replicative helicase; that is, the helicase responsible for unwinding DNA at the replication fork (analogous to DnaB in *E. coli*).

3. Just before the beginning of S phase of the cell cycle, the activation of key protein kinases leads to the activation of the Mcm2–Mcm7 helicase and the initiation of replication (step 3, Figure 13.20). One of these protein kinases is a cyclin-dependent kinase (Cdk) whose function is discussed at length in Chapter 14. Cdk activity remains high from S phase through mitosis, which suppresses the formation of new prereplication complexes. Consequently,
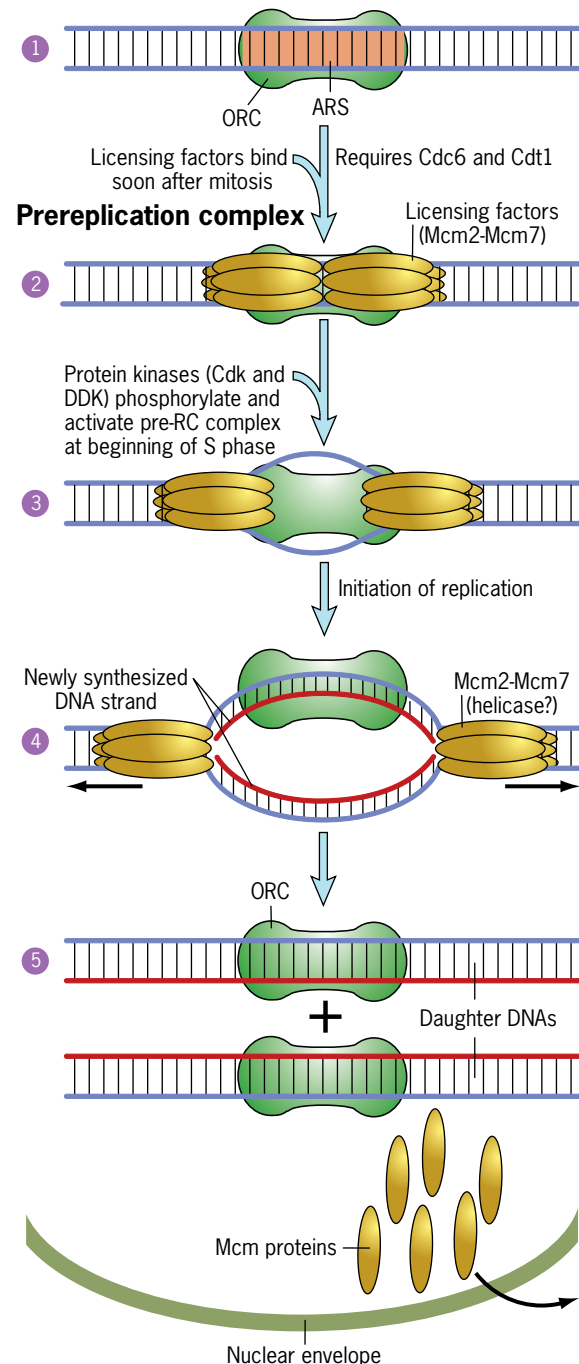
FIGURE 13.20 Steps leading to the replication of a yeast replicon. Yeast origins of replication contain a conserved sequence (ARS) that binds the multisubunit origin recognition complex (ORC) (step 1). The presence of the bound ORC is required for initiation of replication. The ORC is bound to the origin throughout the yeast cell cycle. In step 2, licensing factors (identified as Mcm proteins) bind to the origin during or following mitosis, establishing a prereplication complex that is competent to initiate replication, given the proper stimulus. Loading of Mcm proteins at the origin requires additional proteins (Cdc6 and Cdt1, not shown). In step 3, DNA replication is initiated following the activation of specific protein kinases, including a cyclin-dependent kinase (Cdk). Step 4 shows a stage where replication has proceeded a short distance in both directions from the origin. In this model, the Mcm proteins form a replicative DNA helicase that unwinds DNA of the oppositely directed replication forks. The other proteins required for replication are not shown in this illustration but are indicated in the next figure. In step 5, the two strands of the original duplex have been replicated, an ORC is present at both origins, and the replication proteins, including the Mcm helicases, have been displaced from the DNA. In yeast, the Mcm proteins are exported from the nucleus, and reinitiation of replication cannot occur until the cell has passed through mitosis. [In vertebrate cells, several events appear to prevent reinitiation of replication, including (1) continued Cdk activity from S phase into mitosis, (2) phosphorylation of Cdc6 and its subsequent export from the nucleus, and (3) inactivation of Cdt1 by a bound inhibitor.]

TABLE 13.1 Some of the Proteins Required for Replication

E. coli protein	Eukaryotic protein	Function
DnaA	ORC proteins	Recognition of origin of replication
Gyrase	Topoisomerase I/II	Relieves positive supercoils ahead of replication fork
DnaB	Mcm	DNA helicase that unwinds parental duplex
DnaC	Cdc6, Cdt1	Loads helicase onto DNA
SSB	RPA	Maintains DNA in single-stranded state
γ-complex	RFC	Subunits of the DNA polymerase holoenzyme that load the clamp onto the DNA
pol III core	pol δ/ε	Primary replicating enzymes; synthesize entire leading strand and Okazaki fragments; have proofreading capability
β clamp	PCNA	Ring-shaped subunit of DNA polymerase holoenzyme that clamps replicating polymerase to DNA; works with pol III in *E. coli* and pol δ or ε in eukaryotes
Primase	Primase	Synthesizes RNA primers
———	pol α	Synthesizes short DNA oligonucleotides as part of RNA–DNA primer
DNA ligase	DNA ligase	Seals Okazaki fragments into continuous strand
pol I	FEN-1	Removes RNA primers; pol I of *E. coli* also fills gap with DNA

each origin can only be activated once per cell cycle. Cessation of Cdk activity at the end of mitosis permits the assembly of a pre-RC for the next cell cycle.

4. Once replication is initiated at the beginning of S phase, the Mcm proteins move with the replication fork (step 4) and are essential for completion of replication of a replicon. The fate of the Mcm proteins after replication depends on the species studied. In yeast, the Mcm proteins are displaced from the chromatin and exported from the nucleus (step 5). In contrast, the Mcm proteins in mammalian cells are displaced from the DNA but apparently remain in the nucleus. Regardless, Mcm proteins cannot reassociate with an origin of replication that has already "fired."

The Eukaryotic Replication Fork   Overall, the activities that occur at replication forks are quite similar, regardless of the type of genome being replicated—whether viral, bacterial, archaeal, or eukaryotic. The various proteins in the replication "tool kit" of eukaryotic cells are listed in Table 13.1 and depicted in Figure 13.21. All replication systems require helicases, single-stranded DNA-binding proteins, topoisomerases, primase, DNA polymerase, sliding clamp and clamp

(a)   *(b)*
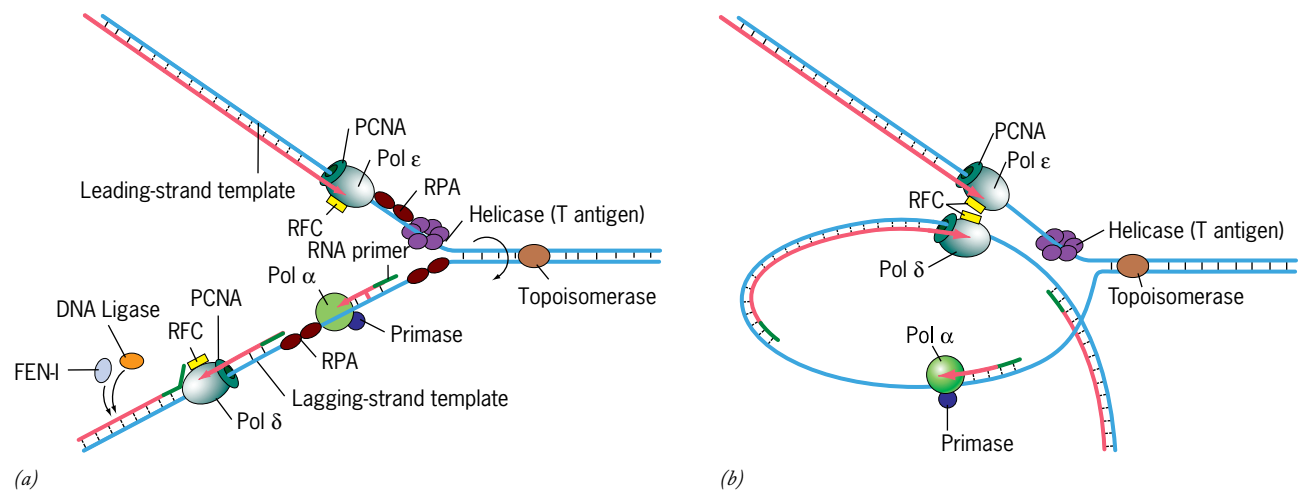
FIGURE 13.21 A schematic view of the major components at the eukaryotic replication fork. (*a*) The proteins required for eukaryotic replication. The viral T antigen is drawn as the replicative helicase in this figure because it is prominently employed in in vitro studies of DNA replication. DNA polymerases δ and ε are thought to be the primary DNA synthesizing enzymes of the lagging and leading strands, respectively. PCNA acts as a sliding clamp for both polymerases δ and ε. The sliding clamp is loaded onto the DNA by a protein called RFC (replication factor C), which is similar in structure and function to the γ-clamp loader of *E. coli*. RPA is a trimeric single-stranded DNA-binding protein comparable in function to that of SSB utilized in *E. coli* replication. The RNA-DNA primers of the lagging strand that are synthesized by the polymerase α-primase complex are displaced by the continued movement of polymerase δ, generating a flap of RNA-DNA that is removed by the FEN-1 endonuclease. The gap is sealed by a DNA ligase. As in *E. coli*, a topoisomerase is required to remove the positive supercoils that develop ahead of the replication fork. (*b*) A proposed version of events at the replication fork illustrating how the replicative polymerases on the leading- and lagging-strand templates might act together as part of a replisome. To date there is no firm evidence that the leading and lagging strands are replicated by a single replicative complex as in *E. coli*.

loader, and DNA ligase. When studying the initiation of eukaryotic replication in vitro, researchers often combine mammalian replication proteins with a viral helicase called the *large T antigen*, which is encoded by the SV40 genome. The large T antigen induces strand separation at the SV40 origin of replication and unwinds the DNA as the replication fork progresses (as in Figure 13.21*a*).

As in bacteria, the DNA of eukaryotic cells is synthesized in a semidiscontinuous manner, although the Okazaki fragments of the lagging strand are considerably smaller than in bacteria, averaging about 150 nucleotides in length. Like DNA polymerase III of *E. coli*, the eukaryotic replicative DNA polymerase is present as a dimer, suggesting that the leading and lagging strands are synthesized in a coordinate manner by a single replicative complex, or *replisome* (Figure 13.21*b*).

To date, five "classic" DNA polymerases have been isolated from eukaryotic cells, and they are designated α, β, γ, δ, and ε. Of these enzymes, polymerase γ replicates mitochondrial DNA, and polymerase β functions in DNA repair. The other three polymerases have replicative functions. Polymerase α is tightly associated with the primase, and together they initiate the synthesis of each Okazaki fragment. Primase initiates synthesis by assembly of a short RNA primer, which is then extended by the addition of about 20 deoxyribonucleotides by polymerase α. Polymerase δ is thought to be the primary DNA-synthesizing enzyme during replication of the lagging strand, whereas polymerase ε is thought to be the primary DNA-synthesizing enzyme during replication of the leading strand. Like the major replicating enzyme of *E. coli*, both polymerase δ and ε require a "sliding clamp" that tethers the enzyme to the DNA, allowing it to move processively along a template. The sliding clamp of eukaryotic cells is very similar in structure and function to the β clamp of *E. coli* polymerase III illustrated in Figure 13.14. In eukaryotes, the sliding clamp is called PCNA. The clamp loader that loads PCNA onto the DNA is called RFC and is analogous to the *E. coli* polymerase III clamp loader complex. After synthesizing an RNA-DNA primer, polymerase α is replaced at each template–primer junction by the PCNA–polymerase δ complex, which completes synthesis of the Okazaki fragment. When polymerase δ reaches the 5′ end of the previously synthesized Okazaki fragment, the polymerase continues along the lagging-strand template, displacing the primer (shown as a green flap in Figure 13.21*a*). The displaced primer is cut from the newly synthesized DNA strand by an endonuclease (FEN-1) and the resulting nick in the DNA is sealed by a DNA ligase. FEN-1 and DNA ligase are thought to be recruited to the replication fork through an interaction with the PCNA sliding clamp. In fact, PCNA is thought to play a major role in orchestrating events that occur during DNA replication, repair, and recombination. Because of its ability to bind a diverse array of proteins, PCNA has been referred to as a "molecular toolbelt."

Like bacterial polymerases, all of the eukaryotic polymerases elongate DNA strands in the 5′ → 3′ direction by the addition of nucleotides to a 3′ hydroxyl group, and none of them is able to initiate the synthesis of a DNA chain without a primer. Polymerases γ, δ, and ε possess a 3′ → 5′ exonuclease, whose proofreading activity ensures that replication oc-

curs with very high accuracy. Several other DNA polymerases (including η, κ and ι) have a specialized function that allows cells to replicate damaged DNA as described on page 558.

Replication and Nuclear Structure Up to this point in the chapter, the illustrations of replication depict a replicating polymerase moving like a locomotive along a stationary DNA track. But the replication apparatus consists of a huge complex of proteins that operates within the confines of a structured nucleus. Considerable evidence suggests that the replication machinery is present in association with both the nuclear lamina (page 477) and the nuclear matrix (page 499). When cells are given very short pulses of radioactive DNA precursors, over 80 percent of the incorporated label is associated with the nuclear matrix. If, instead of fixing the cells immediately after a pulse, the cells are allowed to incorporate *unlabeled* DNA precursors for an hour or so before fixation, most of the radioactivity is chased from the matrix into the surrounding DNA loops. This latter finding suggests that rather than remaining stationary, replicating DNA moves like a conveyer belt through an immobilized replication apparatus (Figure 13.22).

Further studies suggest that replication forks that are active at a given time are not distributed randomly throughout the cell nucleus, but instead are localized within 50 to 250 sites, called **replication foci** (Figure 13.23). It is estimated that each of the bright red regions indicated in Figure 13.23 contains approximately 40 replication forks incorporating nucleotides into DNA strands simultaneously. The clustering of replication forks may provide a mechanism for coordinating
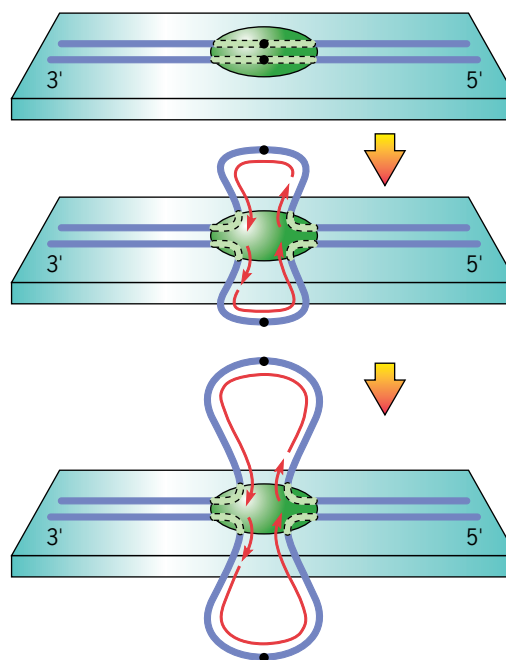
FIGURE 13.22 **The involvement of the nuclear matrix in DNA replication.** The origins of replication are indicated by the black dots, and the arrows indicate the direction of elongation of growing strands. According to this schematic model, it isn't the replication machinery that moves along stationary DNA tracks, but the DNA that is spooled through the replication apparatus, which is firmly attached to the nuclear matrix.
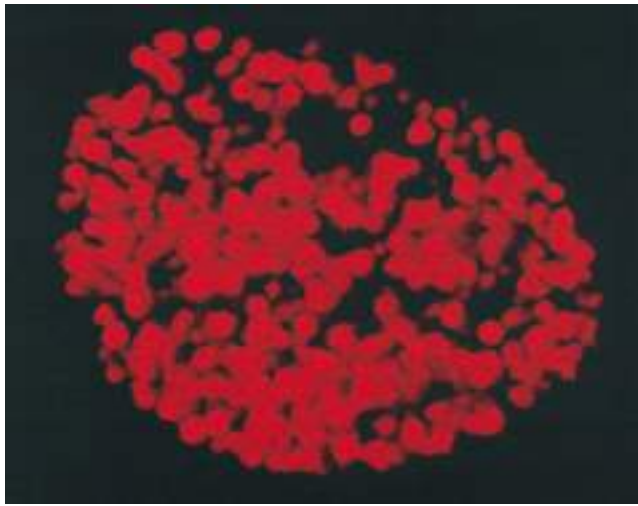
FIGURE 13.23 Demonstration that replication activities do not occur randomly throughout the nucleus but are confined to distinct sites. Prior to the onset of DNA synthesis at the start of S phase, various factors required for the initiation of replication are assembled at discrete sites within the nucleus, forming prereplication centers. These sites appear as discrete red objects in the micrograph, which has been stained with a fluorescent antibody against replication factor A (RPA), which is a single-stranded DNA-binding protein required for the initiation of replication. Other replication factors, such as PCNA and the polymerase–primase complex, are also localized to these foci. (FROM YASUHISA ADACHI AND ULRICH K. LAEMMLI, EMBO J. VOL. 13, COVER NO. 17, 1994.)

the replication of adjacent replicons on individual chromosomes (as in Figure 13.19).

Chromatin Structure and Replication   The chromosomes of eukaryotic cells consist of DNA tightly complexed to regular arrays of histone proteins that are present in the form of nucleosomes (page 481). Movement of the replication machinery along the DNA is thought to displace nucleosomes that reside in its path. Yet, examination of a replicating DNA molecule with the electron microscope reveals nucleosomes on both daughter duplexes very near the replication fork (Figure 13.24a), indicating that the reassembly of nucleosomes is a very rapid event. Collectively, the nucleosomes that form during the replication process are comprised of a roughly equivalent mixture of histone molecules that are inherited from parental chromosomes and histone molecules that have been newly synthesized. Recall from page 482 that the core histone octamer of a nucleosome consists of an $(H3H4)_2$ tetramer together with a pair of H2A/H2B dimers. The way in which parental nucleosomes are distributed during replication has been an area of recent debate. According to one line of research, the $(H3H4)_2$ tetramers present prior to replication remain intact and are distributed randomly between the two daughter duplexes. As a result, old and new $(H3H4)_2$ tetramers are thought to be intermixed on each daughter DNA molecule as indicated in the model shown in Figure 13.24b. According to this model, the two H2A/H2B dimers of each parental nucleosome fail to remain together as the replication fork moves through the chromatin. Instead, the H2A-H2B dimers of a nucleosome separate from one another and bind randomly to the new and old $(H3H4)_2$ tetramers already present on the daughter duplexes (Figure 13.24b). According to another viewpoint, the $(H3H4)_2$ tetramer from parental nucleosomes can be split into two H3-H4 dimers, each

(b)
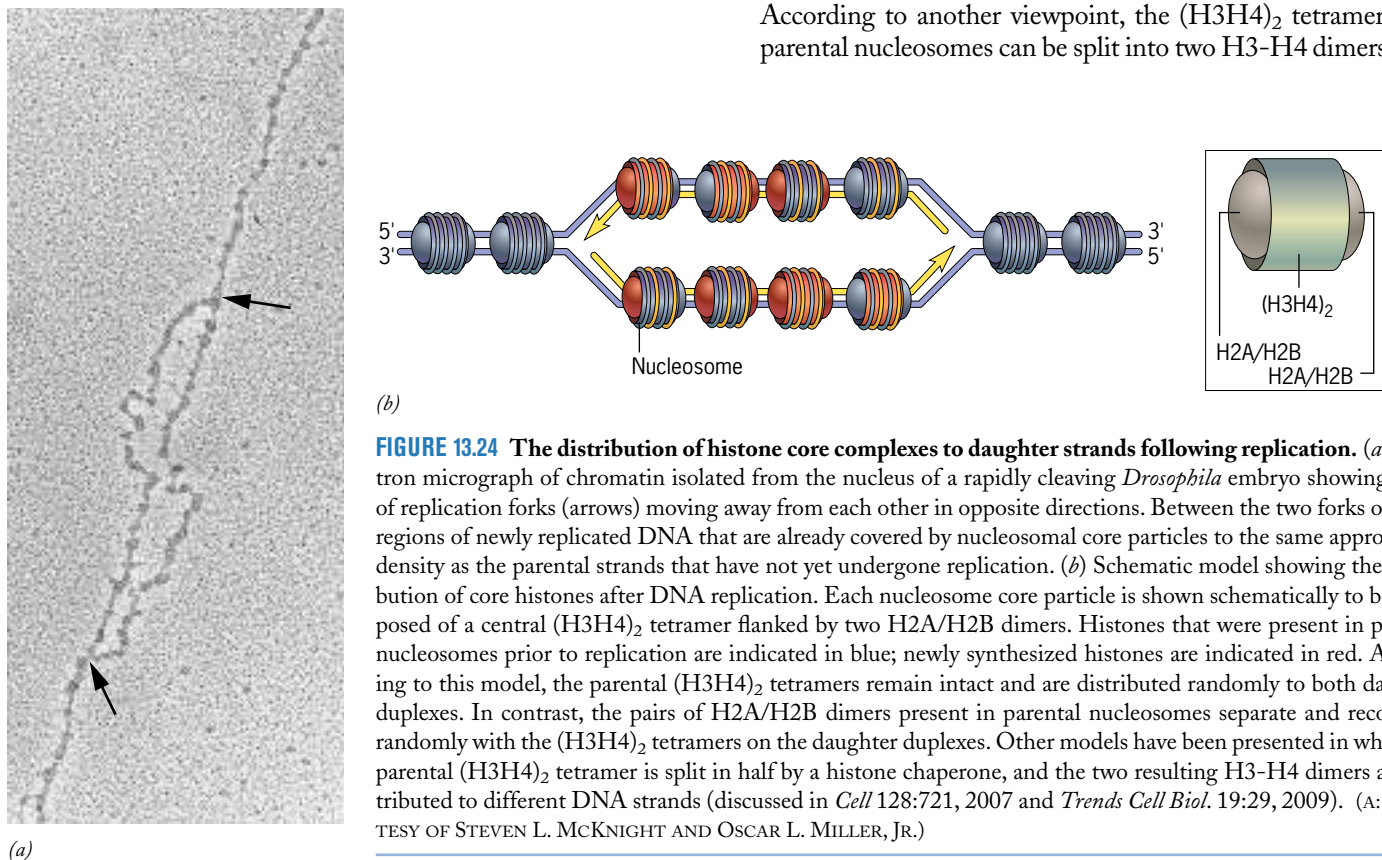
FIGURE 13.24 The distribution of histone core complexes to daughter strands following replication. (a) Electron micrograph of chromatin isolated from the nucleus of a rapidly cleaving *Drosophila* embryo showing a pair of replication forks (arrows) moving away from each other in opposite directions. Between the two forks one sees regions of newly replicated DNA that are already covered by nucleosomal core particles to the same approximate density as the parental strands that have not yet undergone replication. (b) Schematic model showing the distribution of core histones after DNA replication. Each nucleosome core particle is shown schematically to be composed of a central $(H3H4)_2$ tetramer flanked by two H2A/H2B dimers. Histones that were present in parental nucleosomes prior to replication are indicated in blue; newly synthesized histones are indicated in red. According to this model, the parental $(H3H4)_2$ tetramers remain intact and are distributed randomly to both daughter duplexes. In contrast, the pairs of H2A/H2B dimers present in parental nucleosomes separate and recombine randomly with the $(H3H4)_2$ tetramers on the daughter duplexes. Other models have been presented in which the parental $(H3H4)_2$ tetramer is split in half by a histone chaperone, and the two resulting H3-H4 dimers are distributed to different DNA strands (discussed in *Cell* 128:721, 2007 and *Trends Cell Biol.* 19:29, 2009). (A: COURTESY OF STEVEN L. MCKNIGHT AND OSCAR L. MILLER, JR.)

(a)

of which may combine with a newly synthesized H3-H4 dimer to form a "mixed" (H3H4)$_2$ tetramer, which then assembles with H2A-H2B dimers. Regardless of the pattern by which it occurs, the stepwise assembly of nucleosomes and their orderly spacing along the DNA is facilitated by a network of accessory proteins. Included among these proteins are a number of histone chaperones that are able to accept either newly synthesized or parental histones and transfer them to the daughter strands. The best studied of these histone chaperones, CAF-1, is recruited to the advancing replication fork through an interaction with the sliding clamp PCNA.

REVIEW

?

1. The original Watson-Crick proposal for DNA replication envisioned the continuous synthesis of DNA strands. How and why has this concept been modified over the intervening years?

2. What does it mean that replication is semiconservative? How was this feature of replication demonstrated in bacterial cells? in eukaryotic cells?

3. Why are there no heavy bands in the top three centrifuge tubes of Figure 13.3*a*?

4. How is it possible to obtain mutants whose defects lie in genes that are required for an essential activity such as DNA replication?

5. Describe the events that occur at an origin of replication during the initiation of replication in yeast cells. What is meant by replication being bidirectional?

6. Why do the DNA molecules depicted in Figure 13.7*a* fail to stimulate the polymerization of nucleotides by DNA polymerase I? What are the properties of a DNA molecule that allow it to serve as a template for nucleotide incorporation by DNA polymerase I?

7. Describe the mechanism of action of DNA polymerases operating on the two template strands and the effect this has on the synthesis of the lagging versus the leading strand.

8. Contrast the role of DNA polymerases I and III in bacterial replication.

9. Describe the role of the DNA helicase, the SSBs, the β clamp, the DNA gyrase, and the DNA ligase during replication in bacteria.

10. What is the consequence of having the DNA of the lagging-strand template looped back on itself as in Figure 13.13*a*?

11. How do the two exonuclease activities of DNA polymerase I differ from one another? What are their respective roles in replication?

12. Describe the factors that contribute to the high fidelity of DNA replication.

13. What is the major difference between bacteria and eukaryotes that allows a eukaryotic cell to replicate its DNA in a reasonable amount of time?

13.2 DNA REPAIR

Life on Earth is subject to a relentless onslaught of destructive forces that originate in both the internal and external environments of an organism. Of all the molecules in a cell, DNA is placed in the most precarious position. On one hand, it is essential that the genetic information remain mostly unchanged as it is passed from cell to cell and individual to individual. On the other hand, DNA is one of the molecules in a cell that is most susceptible to environmental damage. When struck by ionizing radiation, the backbone of a DNA molecule is often broken; when exposed to a variety of reactive chemicals, many of which are produced by a cell's own metabolism, the bases of a DNA molecule may be altered structurally; when subjected to ultraviolet radiation, adjacent pyrimidines on a DNA strand have a tendency to interact with one another to form a covalent complex, that is, a dimer (Figure 13.25). Even the absorption of thermal energy generated by metabolism is sufficient to split adenine and guanine bases from their attachment to the sugars of the DNA backbone. The magnitude of these spontaneous alterations, or *lesions*, can be appreciated from the estimate that each cell of a warm-blooded mammal loses approximately 10,000 bases per day! Failure to repair such lesions produces permanent alterations, or mutations, in the DNA. If the mutation occurs in a cell destined to become a gamete, the genetic alteration may be passed on to the next generation. Mutations also have effects in somatic cells (i.e., cells that are not in the germ line): they can interfere with transcription and replication, lead to the malignant transformation of a cell, or speed the process by which an organism ages.

Considering the potentially drastic consequences of alterations in DNA molecules and the high frequency at which they occur, it is essential that cells possess mechanisms for repairing DNA damage. In fact, cells have a bewildering arsenal of repair

FIGURE 13.25   A pyrimidine dimer that has formed within a DNA duplex following UV irradiation.

systems that correct virtually any type of damage to which a DNA molecule is vulnerable. It is estimated that less than one base change in a thousand escapes a cell's repair systems. The existence of these systems provides an excellent example of the molecular mechanisms that maintain cellular homeostasis. The importance of DNA repair can be appreciated by examining the effects on humans that result from DNA repair deficiencies, a subject discussed in the Human Perspective on page 556.

Both prokaryotic and eukaryotic cells possess a variety of proteins that patrol vast stretches of DNA, searching for subtle chemical modifications or distortions of the DNA duplex. In some cases, damage can be repaired directly. Humans, for example, possess enzymes that can directly repair damage from cancer-producing alkylating agents. Most repair systems, however, require that a damaged section of the DNA be *excised*, that is, selectively removed. One of the great virtues of the DNA duplex is that each strand contains the information required for constructing its partner. Consequently, if one or more nucleotides is removed from one strand, the complementary strand can serve as a template for reconstruction of the duplex. The repair of DNA damage in eukaryotic cells is complicated by the relative inaccessibility of DNA within the folded chromatin fibers of the nucleus. As in the case of transcription, DNA repair involves the participation of chromatin-reshaping machines, such as the histone modifying enzymes and nucleosome remodeling complexes discussed on page 517. Although presumably important in DNA repair, the roles of these proteins will not be considered in the following discussion.

Nucleotide Excision Repair

Nucleotide excision repair (NER) operates by a cut-and-patch mechanism that removes a variety of bulky lesions, including pyrimidine dimers and nucleotides to which various chemical groups have become attached. Two distinct NER pathways can be distinguished:

1. A *transcription–coupled pathway* in which the template strands of genes that are being actively transcribed are preferentially repaired. Repair of a template strand is thought to occur as the DNA is being transcribed, and the presence of the lesion may be signaled by a stalled RNA polymerase. This preferential repair pathway ensures that those genes of greatest importance to the cell, which are the genes the cell is actively transcribing, receive the highest priority on the "repair list."

2. A slower, less efficient *global genomic pathway* that corrects DNA strands in the remainder of the genome.

Although recognition of the lesion is probably accomplished by different proteins in the two NER pathways (step 1, Figure 13.26), the steps that occur during repair of the lesion are thought to be very similar, as indicated in steps 2–6 of Figure 13.26. One of the key components of the NER repair machinery is TFIIH, a huge protein that also participates in the initiation of transcription. The discovery of the involvement of TFIIH established a crucial link between transcription and DNA repair, two processes that were previously assumed to be independent of one another (discussed in the Experimental Pathways, which can be accessed

on the Web at www.wiley.com/college/karp). Included among the various subunits of TFIIH are two subunits (XPB and XPD) that possess helicase activity; these enzymes separate the two strands of the duplex (step 2, Figure 13.26) in preparation for removal of the lesion. The damaged strand is then cut on both sides of the lesion by a pair of endonucleases (step 3), and the segment of DNA between the incisions is released (step 4). Once excised, the gap is filled by a DNA polymerase (step 5), and the strand is sealed by DNA ligase (step 6).
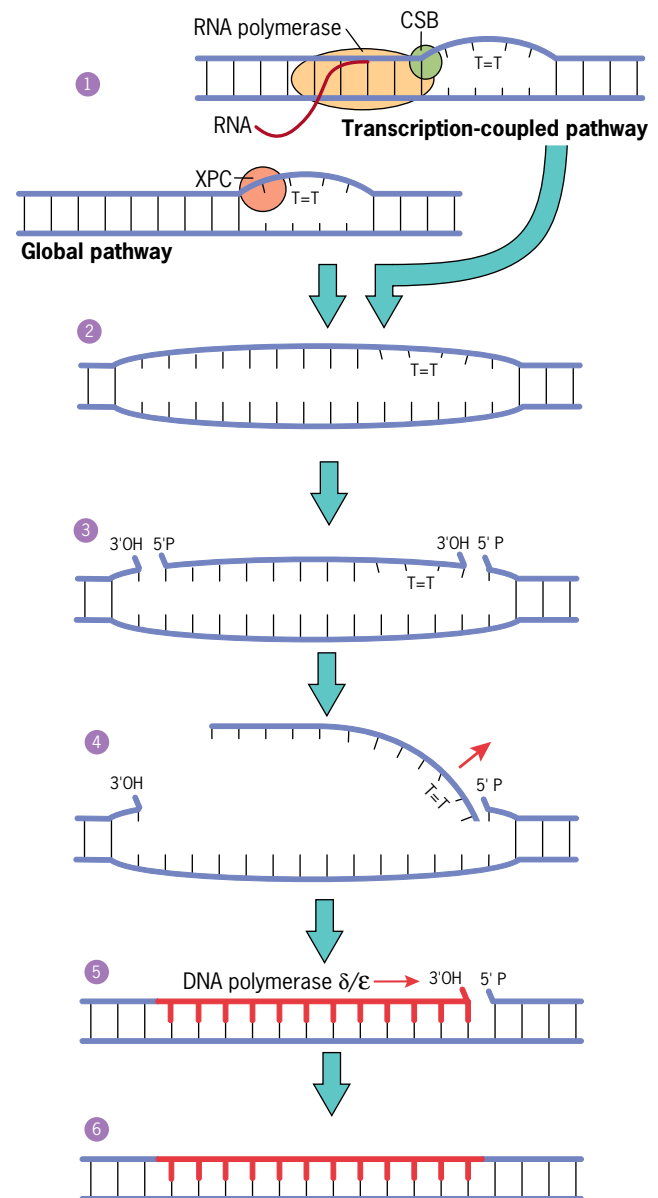
FIGURE 13.26 Nucleotide excision repair. The following steps are depicted in the drawing and discussed in the text: (1) damage recognition in the global pathway is mediated by an XPC-containing protein complex, whereas damage recognition in the transcription-coupled pathway is thought to be mediated by a stalled RNA polymerase in conjunction with a CSB protein; (2) DNA strand separation (by XPB and XPD proteins, two helicase subunits of TFIIH); (3) incision (by XPG on the 3′ side and the XPF–ERCC1 complex on the 5′ side); (4) excision, (5) DNA repair synthesis (by DNA polymerase δ and/or ε); and (6) ligation (by DNA ligase I).

Base Excision Repair

A separate excision repair system operates to remove altered nucleotides generated by reactive chemicals present in the diet or produced by metabolism. The steps in this repair pathway in eukaryotes, which is called **base excision repair** (**BER**), are shown in Figure 13.27. BER is initiated by a *DNA glycosylase* that recognizes the alteration (step 1, Figure 13.27) and removes the base by cleavage of the glycosidic bond holding the base to the deoxyribose sugar (step 2). A number of different DNA glycosylases have been identified, each more-or-less specific for a particular type of altered base, including uracil (formed by the hydrolytic removal of the amino group of cytosine), 8-oxoguanine (caused by damage from oxygen free radicals, page 34), and 3-methyladenine (produced by transfer of a methyl group from a methyl donor, page 431).

Structural studies of the DNA glycosylase that removes the highly mutagenic 8-oxoguanine (oxoG) indicate that this enzyme diffuses rapidly along the DNA "inspecting" each of the G-C base pairs within the DNA duplex (Figure 13.28, step 1). In step 2, the enzyme has come across an oxoG-C base pair. When this occurs, the enzyme inserts a specific amino acid side chain into the DNA helix, causing the nucleotide to rotate ("flip") 180 degrees out of the DNA helix and into the body of the enzyme (step 2). If the nucleotide does, in fact, contain an oxoG, the base fits into the active site of the enzyme (step 3) and is cleaved from its associated sugar. In contrast, if the extruded nucleotide contains a normal guanine, which only differs in structure by two atoms from oxoG, it is unable to fit into the enzyme's active site (step 4) and it is returned to its appropriate position within the stack of bases. Once an altered purine or pyrimidine is removed by a glycosylase, the "beheaded" deoxyribose phosphate remaining in the site is excised by the combined action of a specialized (AP) endonuclease and a DNA polymerase. AP endonuclease cleaves the DNA backbone (Figure 13.27, step 3) and a phosphodiesterase activity of polymerase β removes the sugar–phosphate remnant that had been attached to the excised base (step 4). Polymerase β then fills the gap by inserting a nucleotide complementary to the undamaged strand (step 5), and the strand is sealed by DNA ligase III (step 6).

The fact that cytosine can be converted to uracil may explain why natural selection favored the use of thymine, rather than uracil, as a base in DNA, even though uracil was presumably present in RNA when it served as genetic material during the early evolution of life (page 448). If uracil had been retained as a DNA base, it would have caused difficulty for repair systems to distinguish between a uracil that "belonged" at a particular site and one that resulted from an alteration of cytosine.

Mismatch Repair

It was noted earlier that cells can remove mismatched bases that are incorporated by the DNA polymerase and escape the enzyme's proofreading exonuclease. This process is called **mismatch repair** (**MMR**). A mismatched base pair causes a distortion in the geometry of the double helix that can be recognized by a repair enzyme. But how does the enzyme "recognize"
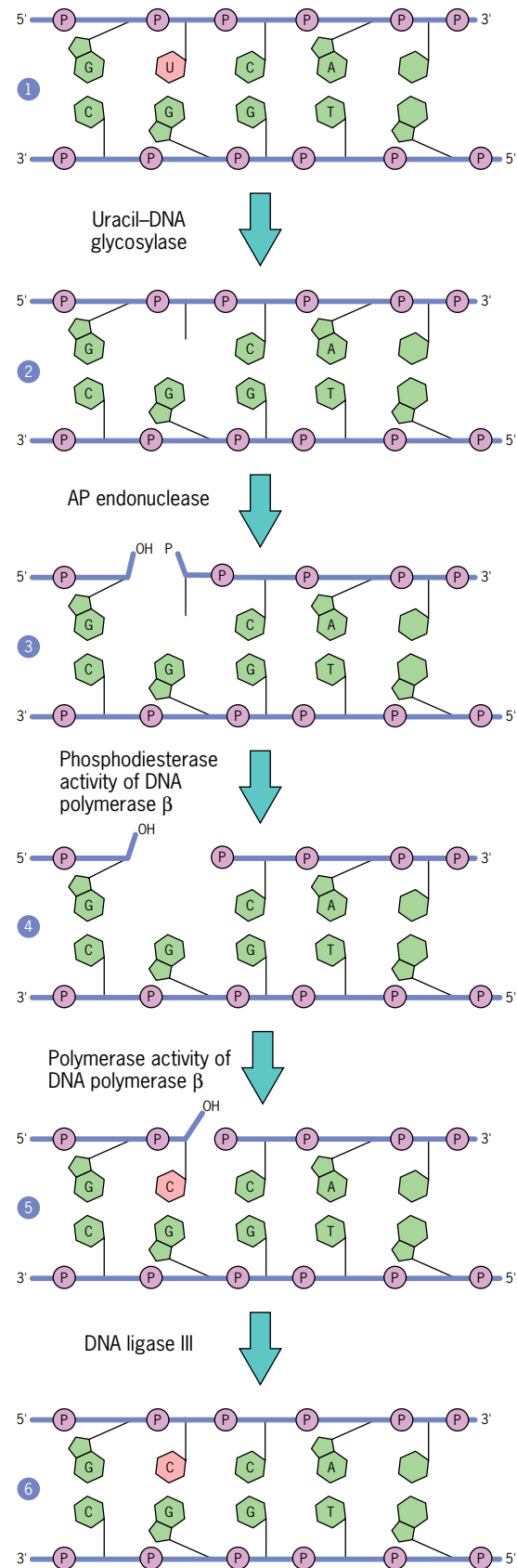
FIGURE 13.27 Base excision repair. The steps are described in the text. Other pathways for BER are known, and BER also has been shown to have distinct transcription-coupled and global repair pathways.
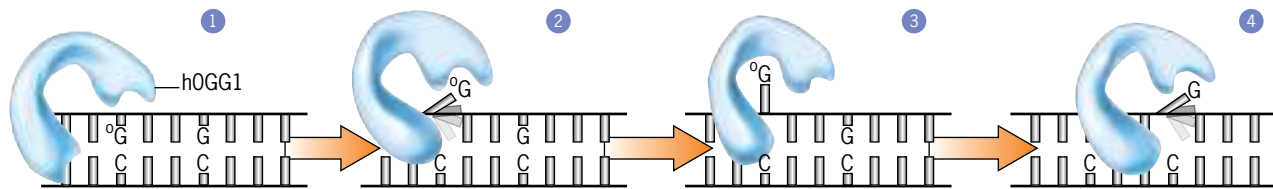
FIGURE 13.28 Detecting damaged bases during BER. In step 1, a DNA glycosylase (named hOGG1) is inspecting a nucleotide that is paired to a cytosine. In step 2, the nucleotide is flipped out of the DNA duplex. In this case, the base is an oxidized version of guanine, 8-oxoguanine, and it is able to fit into the active site of the enzyme (step 3) where it is cleaved from its attached sugar. The subsequent steps in BER were shown in Figure 13.27. In step 4, the extruded base is a normal guanine, which is unable to fit into the active site of the glycosylase and is returned to the base stack. Failure to remove oxoG would have resulted in a G-to-T mutation. (BASED ON S. S. DAVID, WITH PERMISSION FROM NATURE 434:569, 2005; © COPYRIGHT 2005, BY MACMILLAN MAGAZINES LIMITED.)

which member of the mismatched pair is the incorrect nucleotide? If it were to remove one of the nucleotides at random, it would make the wrong choice 50 percent of the time, creating a permanent mutation at that site. Thus, for a mismatch to be repaired after the DNA polymerase has moved past a site, it is important that the repair system distinguish the newly synthesized strand, which contains the incorrect nucleotide, from the parental strand, which contains the correct nucleotide. In *E. coli*, the two strands are distinguished by the presence of methylated adenosine residues on the parental strand. DNA methylation does not appear to be utilized by the MMR system in eukaryotes, and the mechanism of identification of the newly synthe-

sized strand remains unclear. Several different MMR pathways have been identified and will not be discussed.

Double-Strand Breakage Repair

X-rays, gamma rays, and particles released by radioactive atoms are all described as *ionizing radiation* because they generate ions as they pass through matter. Millions of gamma rays pass through our bodies every minute. When these forms of radiation collide with a fragile DNA molecule, they often break both strands of the double helix. **Double-strand breaks** (**DSBs**) can also be caused by certain chemicals, including several (e.g., bleomycin) used in cancer chemotherapy, and free radicals produced by normal cellular metabolism (page 34). DSBs are also introduced during replication of damaged DNA. A single double-strand break can cause serious chromosome abnormalities, which can have grave consequences for the cell. DSBs can be repaired by several alternate pathways. The predominant pathway in mammalian cells is called *nonhomologous end joining* (*NHEJ*), in which a complex of proteins bind to the broken ends of the DNA duplex and catalyze a series of reactions that rejoin the broken strands. The major steps that occur during NHEJ are shown in Figure 13.29*a* and described in the accompanying legend. Figure 13.29*b* shows the nuclei of human fi-

(a)

(b)
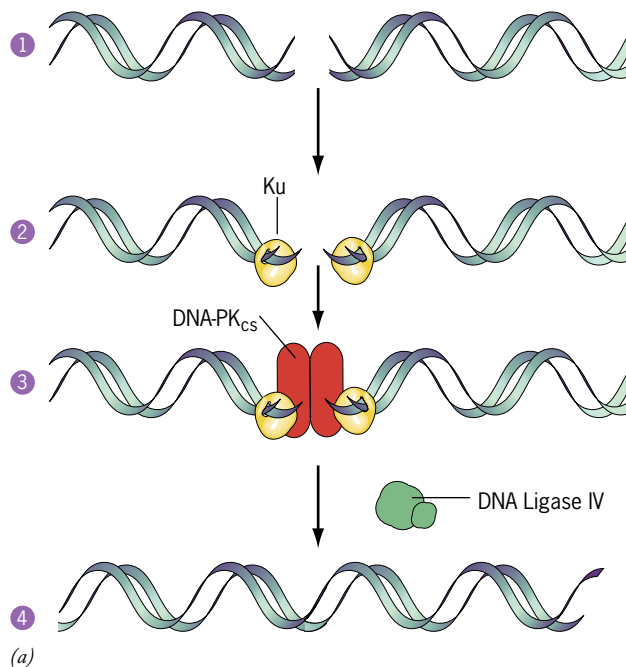
FIGURE 13.29 Repairing double-strand breaks (DSBs) by nonhomologous end joining. (*a*) In this simplified model of double-strand break repair, the lesion (step 1) is detected by a heterodimeric, ring-shaped protein called Ku, that binds to the broken ends of the DNA (step 2). The DNA-bound Ku recruits another protein, called DNA-PK$_{cs}$, which is the catalytic subunit of a DNA-dependent protein kinase (step 3). Most of the substrates phosphorylated by this protein kinase have not been identified. These proteins bring the ends of the broken DNA together in such a way that they can be joined by DNA ligase IV to regenerate an intact DNA duplex (step 4). The NHEJ pathway may also involve the activities of nucleases and polymerases (not shown) and is more error prone than is the homologous recombination pathway of DSB repair. (*b*) Time course analysis of Ku localization at sites of DSB formation induced by laser microbeam irradiation at a site indicated by the arrowheads. The NHEJ protein Ku becomes localized at the damage site immediately following irradiation but remains there just briefly as the damage is presumably repaired. Micrographs were taken (1) immediately, (2) 2 hours, and (3) 8 hours after irradiation. (B: FROM JONG-SOO KIM ET AL, COURTESY OF KYOKO YOKOMORI, J. CELL BIOL. 170:344, 2005; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)
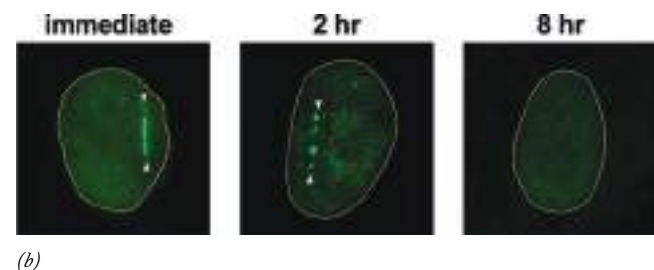
broblasts that had been treated with a laser to induce a localized cluster of double-strand breaks and then stained for the presence of the protein Ku at various times after laser treatment. This NHEJ repair protein is seen to localize at the site of the DSBs immediately following their appearance. Another DSB repair pathway known as *homologous recombination* requires a homologous chromosome to serve as a template for repair of the broken strand. The steps that occur during homologous recombination are similar to those of genetic recombination depicted in Figure 14.47. Defects in both repair pathways have been linked to increased cancer susceptibility.

REVIEW ?

1. Contrast the events of nucleotide excision repair and base excision repair.

2. Why is it important in mismatch repair that the cell distinguish the parental strands from the newly synthesized strands? How is this accomplished?

THE HUMAN PERSPECTIVE

The Consequences of DNA Repair Deficiencies

We owe our lives to light from the sun, which provides the energy captured during photosynthesis. But the sun also emits a constant stream of ultraviolet rays that ages and mutates the cells of our skin. The hazardous effects of the sun are most dramatically illustrated by the rare recessive genetic disorder, *xeroderma pigmentosum* (*XP*). Patients with XP possess a deficient nucleotide excision repair system that cannot remove segments of DNA damaged by ultraviolet radiation. As a result, persons with XP are extremely sensitive to sunlight; even very limited exposure to the direct rays of the sun can produce large numbers of dark-pigmented spots on exposed areas of the body (Figure 1) and a greatly elevated risk of developing disfiguring and fatal skin cancers. Some help for XP patients may be on the way in the form

FIGURE 1 Darkly pigmented regions of the skin are evident in this boy with xeroderma pigmentosum. The area of skin below the chin, which is protected from the sun, is relatively devoid of the lesions. (KEN GREER/ VISUALS UNLIMITED.)

of a skin cream (Dimericine) that contains a bacterial DNA repair enzyme. The enzyme is contained in liposomes that can apparently penetrate the outer layer of the skin and participate in DNA repair.

XP is not the only genetic disorder characterized by nucleotide excision repair deficiency. Cockayne syndrome (CS) is an inherited disorder characterized by acute sensitivity to light, neurological dysfunction due to demyelination of neurons, and dwarfism, but no evident increase in the frequency of skin cancer. Cells from persons with CS are deficient in the pathway by which transcriptionally active DNA is repaired (page 553). The remainder of the genome is repaired at the normal rate, presumably accounting for the normal levels of skin cancer. But why are persons with a defective repair mechanism subject to specific abnormalities such as dwarfism? Most cases of CS can be traced to a mutation in one of two genes, either *CSA* or *CSB*, which are thought to be involved in coupling transcription to DNA repair (see Figure 13.26). Mutations in these genes, in addition to impacting DNA repair, may also disturb the transcription of certain genes, leading to growth retardation and abnormal development of the nervous system. This possibility is strengthened by the finding that, in rare cases, the symptoms of CS can also occur in persons with XP who carry specific mutations in the *XPD* gene. As noted on page 553, *XPD* encodes a subunit of the transcription factor TFIIH required for transcription initiation. Mutations in *XPD* could lead to defects in both DNA repair and transcription. Certain other mutations in the *XPD* gene are responsible for another disease, trichothiodystrophy (TTD), which also combines symptoms suggestive of both DNA repair and transcription defects. Like CS patients, individuals with TTD exhibit increased sun sensitivity without the increased risk of development of cancer. TTD patients have additional symptoms, including brittle hair and scaly skin. These findings indicate that three distinct disorders—XP, CS, and TTD—can be caused by defects in a single gene, with the particular disease outcome determined by the specific mutation present in that gene. Structural studies of mutant XPD molecules suggest that these different mutations affect different functions of the protein.

Elsewhere in this text, we have described circumstances that lead to premature (or accelerated) aging in humans or animal models: as the result of (1) increased free radicals (page 34), (2) increased mitochondrial DNA mutations (page 202), and (3) mutations in a protein of the nuclear envelope (page 477). In 2006, a 15-year-old boy who suffered from frequent sunburns and certain characteristics of prema-

ture aging came to the attention of clinical researchers. Genetic analysis determined that the boy carried a mutation in the *XPF* gene whose encoded protein makes one of the cuts during the NER pathway (Figure 13.26). Patients with mild mutations in *XPF* develop XP and have impaired NER. This individual had a more severe mutation in the *XPF* gene, causing his cells to be unable to repair covalent cross-links that form occasionally between the two strands of a DNA duplex. Studies on the cells of this individual, and on mice with a corresponding mutation, suggested that the unrepaired cross-links lead to increased cell death (apoptosis), which either directly or indirectly promotes premature aging. According to one hypothesis, defects in DNA repair systems that result primarily in an increased mutation rate in the body's cells are associated with an increased susceptibility to cancer, whereas defects in DNA repair systems that result primarily in cell death are associated with accelerated aging.[a] Whether any of these premature-aging syndromes provides insight into the mechanisms of normal aging remains a matter of debate.

Persons with DNA-repair disorders are not the only individuals who should worry about exposure to the sun. Even in a skin cell whose repair enzymes are functioning at optimal levels, a small fraction of the lesions fail to be excised and replaced. Alterations in DNA lead to mutations that can cause a cell to become malignant. Thus, one of the consequences of the failure to correct UV-induced damage is the risk of skin cancer. Consider the following statistics: more than one million persons develop one of three forms of skin cancer every year in the United States, and most of these cases are attributed to overexposure to the sun's ultraviolet rays. Fortunately, the two most common forms of skin cancer—basal cell carcinoma and squamous cell carcinoma—rarely spread to other parts of the body and can usually be excised in a doctor's office. Both of these types of cancer originate from the skin's epithelial cells.

---

[a]It has not been mentioned in this discussion, for a number of reasons, that two of the genes most often responsible for premature aging syndromes encode members of a particular type of DNA helicase family called RecQ helicases. The genes in question are *WRN* and *BLM* which, when mutated, are responsible for the inherited diseases Werner Syndrome and Bloom Syndrome, respectively, which are characterized by both increased cancer risk and features of accelerated aging. It is suggested that these helicases are involved in certain types of base excision and DSB repair pathways. They appear to be particularly important in resolving situations where a replicative DNA polymerase becomes stalled at a lesion and the replication fork "collapses" (disassembles). The subject is discussed in *Trends Biochem. Sci.* 33:609, 2008.

However, malignant melanoma, the third type of skin cancer, is a potential killer. Unlike the others, melanomas develop from pigment cells (melanocytes) in the skin. The number of cases of melanoma diagnosed in the United States is climbing at the alarming rate of 4 percent per year due to the increasing amount of time people have spent in the sun over the past few decades. Studies suggest that one of the greatest risk factors to developing melanoma as an adult is the occurrence of a severe, blistering sunburn as a child or adolescent. Individuals at greatest risk are Caucasians with extremely light skin. Many of these individuals have pigment cells whose surfaces lack a functioning receptor (called MC1R) for a hormone that is secreted by nearby epithelial cells of the skin in response to ultraviolet radiation. Melanocytes respond to MC1R activation by producing the dark pigment melanin, thereby providing the individual with a tan. Tanned skin is more protected from UV rays than is light, untanned skin, even though it is UV radiation that is responsible for triggering the tanning response. What if it were possible to develop a tanned skin without having to suffer UV exposure? A number of research groups are working on such an approach by using various means other than exposure to UV-containing sunlight to stimulate the tanning response in pigment cells. Whether any of these approaches will prove safe and effective remains to be seen.

Skin cancer is not the only disease that is promoted by deficient or overworked DNA repair systems. It is estimated that up to 15 percent of colon cancer cases can be attributed to mutations in the genes that encode the proteins required for mismatch repair. Mutations that cripple the mismatch repair system inevitably lead to a higher mutation rate in other genes because mistakes made during replication are not corrected.

Cancer is also one of the consequences of double-strand DNA breaks that have either gone unrepaired or been repaired incorrectly. Breaks in DNA can be caused by a variety of environmental agents to which we are commonly exposed, including X-rays, gamma rays, and radioactive emissions. The most serious environmental hazard in this regard is probably radon (specifically ^{222}Rn), a radioactive isotope formed during the disintegration of uranium. Some areas of the planet contain relatively high levels of uranium in the soil, and houses built in these regions can contain dangerous levels of radon gas. When the gas is breathed into the lungs, it can lead to double-strand DNA breaks that increase the risk of lung cancer. A significant fraction of lung-cancer deaths in nonsmokers is probably due to radon exposure.

13.3 BETWEEN REPLICATION AND REPAIR

The human perspective describes an inherited disease—xeroderma pigmentosum (XP)—that leaves patients with an inability to repair certain lesions caused by exposure to ultraviolet radiation. Patients described as having the "classical" form of XP have a defect in one of seven different genes involved in nucleotide excision repair (page 553). These genes are designated *XPA, XPB, XPC, XPD, XPE, XPF*, and *XPG*, and some of their roles in NER are indicated in the legend of Figure 13.26. Another group of patients were identified that, like those with XP, were highly susceptible to developing skin cancer as the result of sun exposure. However, unlike the cells from XP sufferers, cells from these patients were capable of

nucleotide excision repair and were only slightly more sensitive to UV light than normal cells. This heightened UV sensitivity revealed itself during replication, as these cells often produced fragmented daughter strands following UV irradiation. Patients in this group were classified as having a variant form of XP, designated XP-V. We will return to the basis of the XP-V defect in a moment.

We have seen in the previous section that cells can repair a great variety of DNA lesions. On occasion, however, a DNA lesion is not repaired by the time that segment of DNA is scheduled to undergo replication. On these occasions, the replication machinery arrives at the site of damage on the template strand and becomes stalled there. When this happens, some type of signal is emitted that leads to the recruit-

ment of a specialized polymerase that is able to bypass the lesion.[3] Suppose the lesion in question is a thymidine dimer (Figure 13.25) in a skin cell that was caused by exposure to UV radiation. When the replicative polymerase (pol δ) reaches the obstacle, the enzyme is temporarily replaced by a "specialized" DNA polymerase designated pol η, which is able to insert two A residues into the newly synthesized strand across from the two T residues that are covalently linked as part of the dimer. Once this "damage bypass" is accomplished, the cell switches back to the normal replicative polymerase and DNA synthesis continues without leaving any trace that a serious problem had been resolved. As you might have guessed from the juxtaposition of topics, patients afflicted with XP-V have

a mutation in the gene encoding pol η and thus have difficulty replicating past thymidine dimers.

Discovered in 1999, polymerase η is a member of a family of DNA polymerases in which each member is specialized for incorporating nucleotides opposite particular types of DNA lesions in the template strand. The polymerases of this family are said to engage in *translesion synthesis* (*TLS*). X-ray crystallographic studies reveal that the TLS polymerases have an unusually spacious active site that is able to physically accommodate altered nucleotides that would not fit in the active site of a replicative polymerase. These TLS polymerases are only capable of incorporating one to a few nucleotides into a DNA strand (they lack processivity); they have no proofreading capability; and they are much more likely to incorporate an incorrect (i.e., noncomplementary) nucleotide when copying undamaged DNA than the classic polymerases.

---

[3]A cell has other options to deal with a stalled replication fork, but they are more complex and poorly understood, and will not be discussed.

SYNOPSIS

DNA replication occurs semiconservatively, which indicates that one intact strand of the parent duplex is transmitted to each of the daughter cells during cell division. This mechanism of replication was first suggested by Watson and Crick as part of their model of DNA structure. They suggested that replication occurred by gradual separation of the strands by means of hydrogen bond breakage, so that each strand could serve as a template for the synthesis of a complementary strand. This model was soon confirmed in both bacterial and eukaryotic cells by showing that cells transferred to labeled media for one generation produce daughter cells whose DNA has one labeled strand and one unlabeled strand. (p. 534)

The mechanism of replication is best understood in bacterial cells. Replication begins at a single origin on the circular bacterial chromosome and proceeds outward in both directions as a pair of replication forks. Replication forks are sites where the double helix is unwound and nucleotides are incorporated into both newly synthesized strands. (p. 537)

DNA synthesis is catalyzed by a family of DNA polymerases. The first of these enzymes to be characterized was DNA polymerase I of *E. coli*. To catalyze the polymerization reaction, the enzyme requires all four deoxyribonucleoside triphosphates, a template strand to copy, and a primer containing a free 3′ OH to which nucleotides can be added. The primer is required because the enzyme cannot initiate the synthesis of a DNA strand. Rather, it can only add nucleotides to the 3′ hydroxyl terminus of an existing strand. Another unexpected characteristic of DNA polymerase I is that it only polymerizes a strand in a 5′ → 3′ direction. It had been presumed that the two new strands would be synthesized in opposite directions by polymerases moving in opposite directions along the two parental template strands. This finding was explained when it was shown that the two strands were synthesized quite differently. (p. 538)

One of the newly synthesized strands (the leading strand) grows toward the replication fork and is synthesized continuously. The other newly synthesized strand (the lagging strand) grows away from the fork and is synthesized discontinuously. In bacterial cells, the lagging strand is synthesized as fragments approximately 1000 nucleotides long, called Okazaki fragments, that are covalently joined to one another by a DNA ligase. In contrast, the leading strand is synthesized as a single continuous strand. Neither the con-

tinuous strand nor any of the Okazaki fragments can be initiated by the DNA polymerase but instead begin as a short RNA primer that is synthesized by a type of RNA polymerase called primase. After the RNA primer is assembled, DNA polymerase continues to synthesize the strand or fragment as DNA. The RNA is subsequently degraded, and the gap is filled in as DNA. (p. 540)

Events at the replication fork require a variety of different types of proteins having specialized functions. These include a DNA gyrase, which is a type II topoisomerase required to relieve the tension that builds up ahead of the fork as a result of DNA unwinding; a DNA helicase that unwinds the DNA by separating the strands; single-stranded DNA-binding proteins that bind selectively to single-stranded DNA and prevent reassociation; a primase that synthesizes the RNA primers; and a DNA ligase that seals the fragments of the lagging strand into a continuous polynucleotide. DNA polymerase III is the primary DNA-synthesizing enzyme that adds nucleotides to each RNA primer, whereas DNA polymerase I is responsible for removing the RNA primers and replacing them with DNA. Two molecules of DNA polymerase III are thought to move together as a complex along their respective template strands. This is accomplished as the lagging-strand template loops back on itself. (p. 541)

DNA polymerases possess separate catalytic sites for polymerization and degradation of nucleic acid strands. Most DNA polymerases possess both 5′ → 3′ and 3′ → 5′ exonuclease activities. The first acts to degrade the RNA primers that begin each Okazaki fragment, and the second removes inappropriate nucleotides following their mistaken incorporation, thus contributing to the fidelity of replication. It is estimated that approximately one in 10^9 nucleotides is incorporated incorrectly during replication in *E. coli*. (p. 544)

Replication in eukaryotic cells follows a similar mechanism and employs similar proteins to those of prokaryotes. All of the DNA polymerases involved in replication elongate DNA strands in the 5′ → 3′ direction. None of them initiates the synthesis of a chain without a primer. Most possess a 3′ → 5′ exonuclease activity, ensuring that replication occurs with high fidelity. Unlike bacteria, replication in eukaryotes is initiated simultaneously at many sites along a chromosome, with replication forks proceeding outward in both directions from each site of initiation. Studies on yeast indicate that

origins of replication contain a specific binding site for an essential multiprotein complex called ORC. Events at the origin ensure that replication of each DNA segment occurs once and only once per cell cycle. (p. 546)

Replication in eukaryotic cells is intimately associated with nuclear structures. Evidence indicates that much of the machinery required for replication is associated with the nuclear matrix. In addition, replication forks that are active at any given time are localized within about 50 to 250 sites called replication foci. Newly synthesized DNA is rapidly associated with nucleosomes. According to one model, $(H3H4)_2$ tetramers present prior to replication remain intact and are passed on to the daughter duplexes, whereas H2A/H2B dimers separate from one another and bind randomly to new and old $(H3H4)_2$ tetramers on the daughter duplexes. (p. 550)

DNA is subject to damage by many environmental influences, including ionizing radiation, common chemicals, and ultraviolet radiation. Cells possess a variety of systems to recognize and repair the resulting damage. It is estimated that less than one base change in a thousand escapes a cell's repair systems. Four major types of DNA repair systems are discussed. Nucleotide excision repair (NER) systems operate by removing a small section of a DNA strand containing a bulky lesion, such as a pyrimidine dimer. During NER, the strands of DNA containing the lesion are separated by a helicase;

paired incisions are made by endonucleases; the gap is filled by a DNA polymerase; and the strand is sealed by a DNA ligase. The template strands of genes that are actively transcribed are preferentially repaired by NER. Base excision repair removes a variety of altered nucleotides that produce minor distortions in the DNA helix. Cells possess a variety of glycosylases that recognize and remove various types of altered bases. Once the base is removed, the remaining portion of the nucleotide is removed by an endonuclease, the gap is enlarged by a phosphodiesterase activity, and the gap is filled and sealed by a polymerase and ligase. Mismatch repair is responsible for removing incorrect nucleotides incorporated during replication that escape the proofreading activity of the polymerase. In bacteria, the newly synthesized strand is selected for repair by virtue of its lack of methyl groups compared to the parental strand. Double-strand breaks are repaired as proteins bind to the broken strands and join the ends together. (p. 552)

In addition to the classic DNA polymerases involved in DNA replication and repair, cells also possess an array of DNA polymerases that facilitate replication at sites of DNA lesions or misalignments. These polymerases, which engage in translesion synthesis, lack processivity and proofreading capability and are more error-prone than classic polymerases. (p. 557)

ANALYTIC QUESTIONS

1. Suppose that Meselson and Stahl had carried out their experiment by growing cells in medium with ^{14}N and then transferring the cells to medium containing ^{15}N. How would the bands within the centrifuge tubes have appeared if replication were semiconservative? If replication were conservative? If replication were dispersive?

2. Suppose you isolated a mutant strain of yeast that replicated its DNA more than once per cell cycle. In other words, each gene in the genome was replicated several times between successive cell divisions. How might you explain such a phenomenon?

3. How would the chromosomes from the experiment on eukaryotic cells depicted in Figure 13.4 have appeared if replication occurred by a conservative or a dispersive mechanism?

4. We have seen that cells possess a special enzyme to remove uracil from DNA. What do you suppose would happen if the uracil groups were not removed? (You might consider the information presented in Figure 11.44 on the pairing properties of uracil.)

5. Draw a partially double-stranded DNA molecule that would not serve as a template for DNA synthesis by DNA polymerase I.

6. Some temperature-sensitive bacterial mutants stop replication immediately following elevation of temperature, whereas others continue to replicate their DNA for a period of time before they cease this activity, and still others continue until a round of replication is completed. How might these three types of mutants differ?

7. Suppose the error rate during replication in human cells were the same as that of bacteria (about 10^{-9}). How would this impact the two cells differently?

8. Figure 13.19 shows the results from an experiment in which cells were incubated with [^{3}H]thymidine for less than 30 minutes prior to fixation. How would you expect this photograph to appear after a one-hour labeling period? Can you conclude that the entire genome is replicated within an hour? If not, why not?

9. Origins of replication tend to have a region that is very rich in A-T base pairs. What function do you suppose these sections might serve?

10. What advantages might you expect for DNA replication to occur in conjunction with the nuclear matrix as opposed to the nucleoplasm? What are the advantages of replication occurring in a small number of replication foci?

11. What are some of the reasons you might expect human cells to have more efficient repair systems than those of a frog?

12. Suppose you were to compare autoradiographs of two cells that had been exposed to [^{3}H]thymidine, one that was engaged in DNA replication (S phase) and another that was not. How would you expect autoradiographs of these cells to differ?

13. Construct a model that would explain how transcriptionally active DNA is repaired preferentially over transcriptionally silent DNA.

14
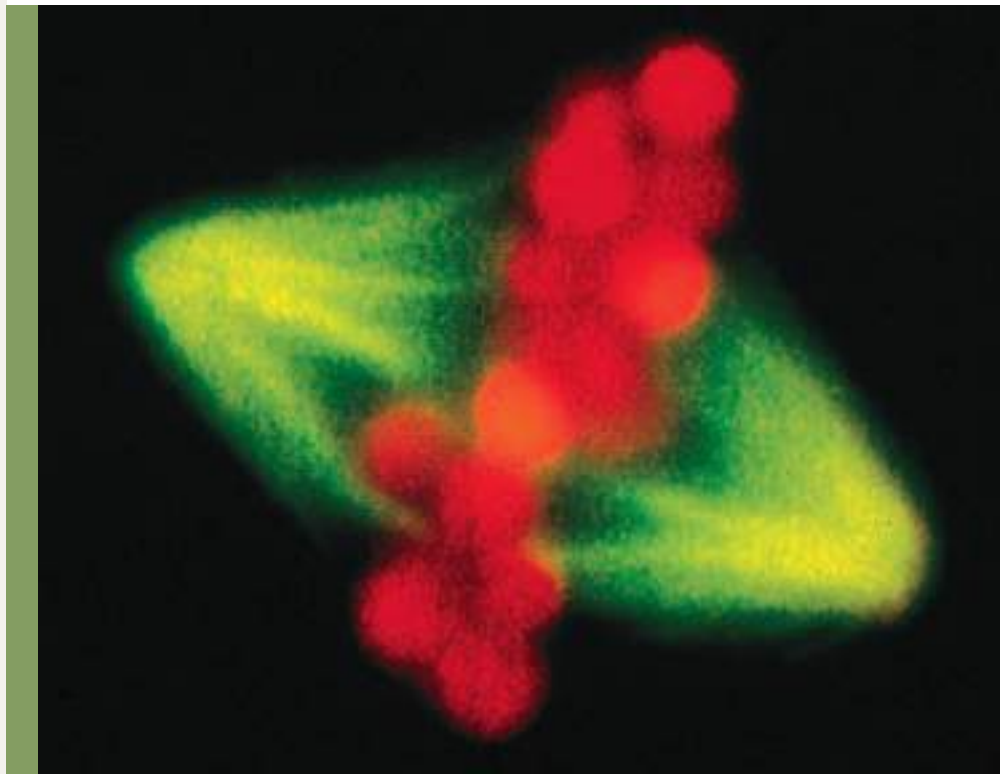
Cellular Reproduction

A ccording to the third tenet of the cell theory, new cells originate only from other living cells. The process by which this occurs is called **cell division**. For a multicellular organism, such as a human or an oak tree, countless divisions of a single-celled zygote produce an organism of astonishing cellular complexity and organization. Cell division does not stop with the formation of the mature organism but continues in certain tissues throughout life. Millions of cells residing within the marrow of your bones or the lining of your intestinal tract are undergoing division at this very moment. This enormous output of cells is needed to replace cells that have aged or died.

Although cell division occurs in all organisms, it takes place very differently in prokaryotes and eukaryotes. We will restrict discussion to the eukaryotic version. Two distinct types of eukaryotic cell division will be discussed in this chapter. Mitosis leads to production of cells that are genetically identical to their parent, whereas meiosis leads to production of cells with half the genetic content of the parent. Mitosis serves as the basis for producing new cells, meiosis as the basis for producing new sexually reproducing organisms. Together, these two types of cell division form the links in the chain between parents and their offspring and, in a broader sense, between living species and the earliest eukaryotic life forms present on Earth. ■

Fluorescence micrograph of a mitotic spindle that had assembled in a cell-free extract prepared from frog eggs, which are cells that lack a centrosome. The red spheres consist of chromatin-covered beads that were added to the extract. It is evident from this micrograph that a bipolar spindle can assemble in the absence of both chromosomes and centrosomes. In this experiment, the chromatin-covered beads served as nucleating sites for the assembly of the microtubules that subsequently formed this spindle. The mechanism by which cells construct mitotic spindles in the absence of centrosomes is discussed on page 575. (REPRINTED WITH PERMISSION FROM R. HEALD ET AL., NATURE VOL. 382, COVER OF 8/1/96; © COPYRIGHT 1996, MACMILLAN MAGAZINES LIMITED.)

14.1 THE CELL CYCLE

In a population of dividing cells, whether inside the body or in a culture dish, each cell passes through a series of defined stages, which constitutes the **cell cycle** (Figure 14.1). The cell cycle can be divided into two major phases based on cellular activities readily visible with a light microscope: M phase and interphase. **M phase** includes (1) the process of **mitosis**, during which duplicated chromosomes are separated into two nuclei, and (2) **cytokinesis**, during which the entire cell divides into two daughter cells. **Interphase**, the period between cell divisions, is a time when the cell grows and engages in diverse metabolic activities. Whereas M phase usually lasts only an hour or so in mammalian cells, interphase may extend for days, weeks, or longer, depending on the cell type and the conditions.

Although M phase is the period when the contents of a cell are actually divided, numerous preparations for an upcoming mitosis occur during interphase, including replication of the cell's DNA. One might guess that a cell engages in replication throughout interphase. However, studies in the early 1950s on asynchronous cultures (i.e., cultures whose cells are randomly distributed throughout the cell cycle) showed that this is not the case. As described in Chapter 13, DNA replication can be monitored by the incorporation of [3H]thymidine into newly synthesized DNA. If [3H]thymidine is given to a culture of cells for a short period (e.g., 30 minutes) and a sample of the cell population is fixed, dried onto a slide, and exam-

ined by autoradiography, only a fraction of the cells are found to have radioactive nuclei. Among cells that were engaged in mitosis at the time of fixation (as evidenced by their compacted chromosomes) none is found to have a radioactively labeled nucleus. These mitotic cells have unlabeled chromosomes because they were not engaged in DNA replication during the labeling period.

If labeling is allowed to continue for one or two hours before the cells are sampled, there are still no cells with labeled mitotic chromosomes (Figure 14.2). We can conclude from these results that there is a definite period of time between the end of DNA synthesis and the beginning of M phase. This period is termed G_2 (for second gap). The duration of G_2 is revealed as one continues to take samples of cells from the culture until labeled mitotic chromosomes are observed. The first cells whose mitotic chromosomes are labeled must have been at the last stages of DNA synthesis at the start of the incubation with [3H]thymidine. The length of time between the start of the labeling period and the appearance of cells with labeled mitotic figures corresponds to the duration of G_2.

DNA replication occurs during a period of the cell cycle termed **S phase**. S phase is also the period when the cell synthesizes the additional histones that will be needed as the cell doubles the number of nucleosomes in its chromosomes (see Figure 13.24). The length of S phase can be determined directly. In an asynchronous culture, the percentage of cells engaged in a particular activity is an approximate measure of the percentage of time that this activity occupies in the lives of
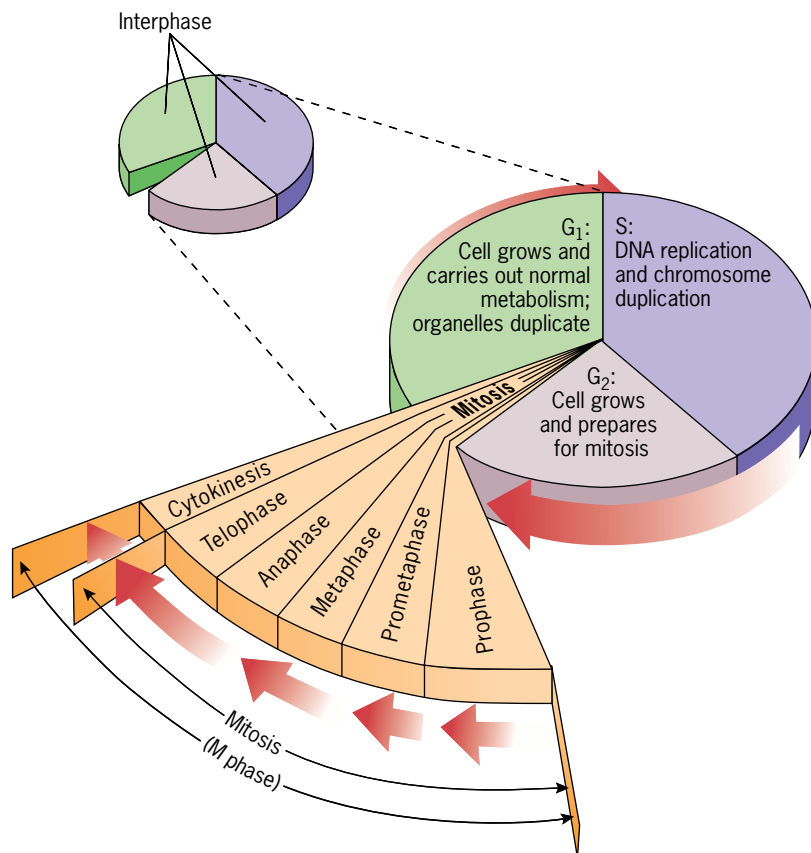
FIGURE 14.1 An overview of the eukaryotic cell cycle. This diagram of the cell cycle indicates the stages through which a cell passes from one division to the next. The cell cycle is divided into two major phases: M phase and interphase. M phase includes the successive events of mitosis and cytokinesis. Interphase is divided into G_1, S, and G_2 phases, with S phase being equivalent to the period of DNA synthesis. The division of interphase into three separate phases based on the timing of DNA synthesis was first proposed in 1953 by Alma Howard and Stephen Pelc of Hammersmith Hospital, London, based on their experiments on plant meristem cells.
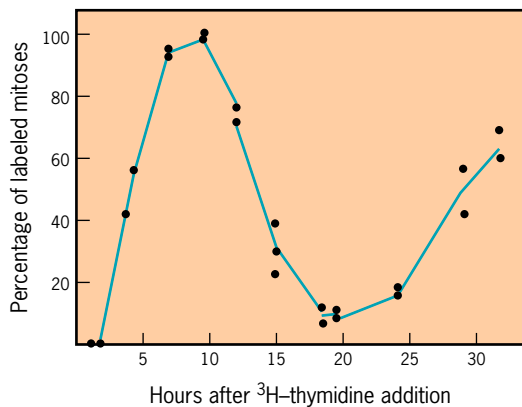
FIGURE 14.2 **Experimental results demonstrating that replication occurs during a defined period of the cell cycle.** HeLa cells were cultured for 30 minutes in medium containing [³H]thymidine and then incubated (chased) for various times in unlabeled medium before being fixed and prepared for autoradiography. Each culture dish was scanned for cells that were in mitosis at the time they were fixed, and the percentage of those mitotic cells whose chromosomes were labeled was plotted as shown. (FROM A STUDY BY R. BASERGA AND F. WIEBEL.)

cells. Thus, if we know the length of the entire cell cycle, the length of S phase can be calculated directly from the percentage of the cells whose nuclei are radioactively labeled during a brief pulse with [³H]thymidine. Similarly, the length of M phase can be calculated from the percentage of cells in the population that are seen to be engaged in mitosis or cytokinesis. When one adds up the periods of $G_2 + S + M$, it is apparent that there is an additional period in the cell cycle yet to be accounted for. This other phase, termed G_1 (for first gap), is the period following mitosis and preceding DNA synthesis.

Cell Cycles in Vivo

One of the properties that distinguishes various types of cells within a multicellular plant or animal is their capacity to grow and divide. We can recognize three broad categories of cells:

1. *Cells, such as nerve cells, muscle cells, or red blood cells, that are highly specialized and lack the ability to divide.* Once these cells have differentiated, they remain in that state until they die.

2. *Cells that normally do not divide but can be induced to begin DNA synthesis and divide when given an appropriate stimulus.* Included in this group are liver cells, which can be induced to proliferate by the surgical removal of part of the liver, and lymphocytes, which can be induced to proliferate by interaction with an appropriate antigen.

3. *Cells that normally possess a relatively high level of mitotic activity.* Included in this category are stem cells of various adult tissues, such as hematopoietic stem cells that give rise to red and white blood cells (Figure 17.6) and stem cells at the base of numerous epithelia that line the body cavities and the body surface (Figure 7.1). The relatively unspecialized cells of apical meristems located near the

tips of plant roots and stems also exhibit rapid and continual cell division. Stem cells have an important property that is not shared by most cells; they are able to divide asymmetrically. An **asymmetric cell division** is one in which the two daughter cells have different properties or fates. The asymmetric division of a stem cell produces one daughter cell that remains an uncommitted stem cell like its parent and another daughter cell that has taken a step towards becoming a differentiated cell of that tissue. In other words, asymmetric divisions allow stem cells to engage in both self-renewal and the formation of differentiated tissue cells. Some types of nonstem cells can also engage in asymmetric (or unequal) cell divisions, as illustrated by the formation of oocytes and polar bodies in Figure 14.41*b* and the division of the T cell in the Chapter 17 opening photo.

Cell cycles can range in length from as short as 30 minutes in a cleaving frog embryo, whose cell cycles lack both G_1 and G_2 phases, to several months in slowly growing tissues, such as the mammalian liver. With a few notable exceptions, cells that have stopped dividing, whether temporarily or permanently, whether in the body or in culture, are present in a stage preceding the initiation of DNA synthesis. Cells that are arrested in this state—which includes the majority of cells in the body—are said to be in the G_0 state to distinguish them from the typical G_1-phase cells that may soon enter S phase. A cell must receive a growth-promoting signal to proceed from G_0 into G_1 phase and thus reenter the cell cycle.

Control of the Cell Cycle

The study of the cell cycle is not only important in basic cell biology, but also has enormous practical implications in combating cancer, a disease that results from a breakdown in a cell's ability to regulate its own division. In 1970, a series of cell fusion experiments carried out by Potu Rao and Robert Johnson of the University of Colorado helped open the door to understanding how the cell cycle is regulated.

Rao and Johnson wanted to know whether the cytoplasm of cells contains regulatory factors that affect cell cycle activities. They approached the question by fusing mammalian cells that were in different stages of the cell cycle. In one experiment, they fused mitotic cells with cells in other stages of the cell cycle. The mitotic cell always induced compaction of the chromatin in the nucleus of the nonmitotic cell (Figure 14.3). If a G_1-phase and an M-phase cell were fused, the chromatin of the G_1-phase nucleus underwent *premature chromosomal compaction* to form a set of elongated compacted chromosomes (Figure 14.3*a*). If a G_2-phase and M-phase cell were fused, the G_2 chromosomes also underwent premature chromosome compaction, but unlike those of a G_1 nucleus, the compacted G_2 chromosomes were visibly doubled, reflecting the fact that replication had already occurred (Figure 14.3*c*). If a mitotic cell was fused with an S-phase cell, the S-phase chromatin also became compacted (Figure 14.3*b*). However, replicating DNA is especially sensitive to damage,

(a)   *(b)*   *(c)*
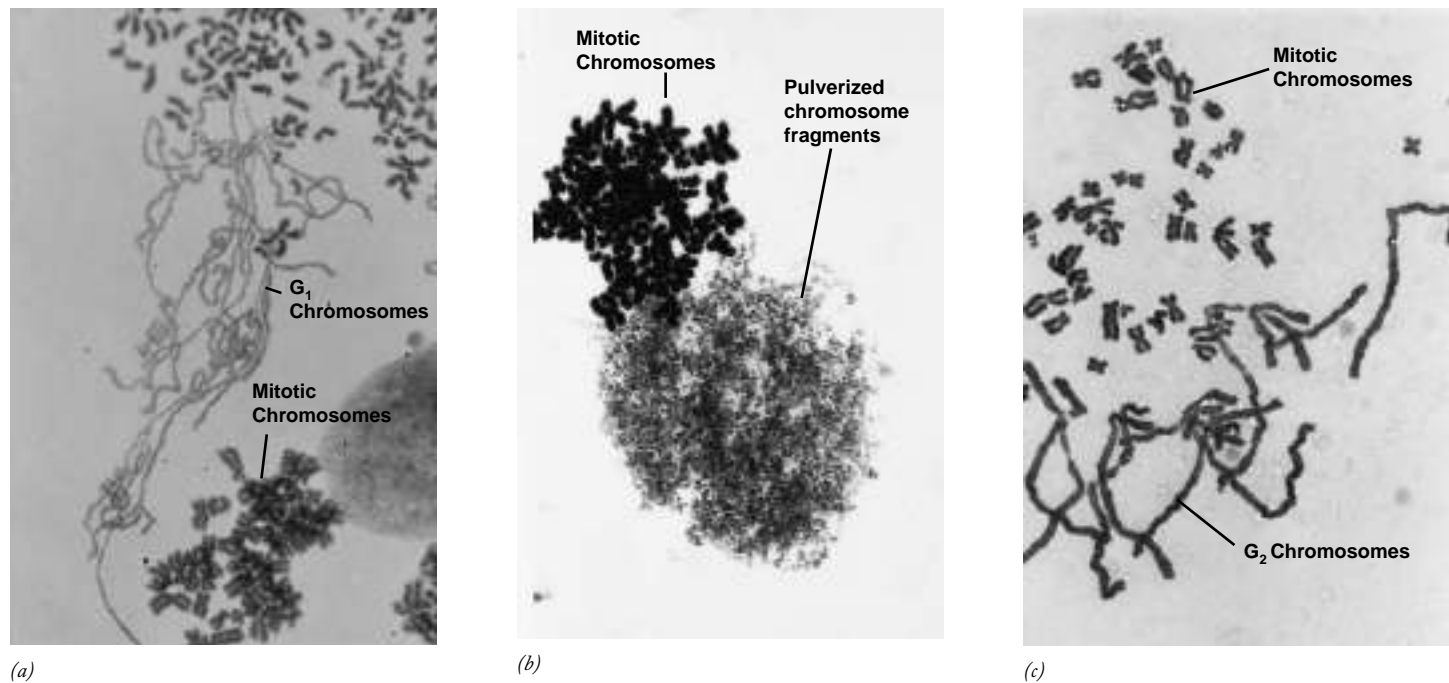
FIGURE 14.3 Experimental demonstration that cells contain factors that stimulate entry into mitosis. The photographs show the results of the fusion of an M-phase HeLa cell with a rat kangaroo PtK2 cell that had been in (*a*) G_1 phase, (*b*) S phase, or (*c*) G_2 phase at the time of cell fusion. As described in the text, the chromatin of the G_1-phase and G_2-phase PtK2 cells undergoes premature compaction, whereas that of the S-phase cell becomes pulverized. The elongated chromatids of the G_2-phase cell in *c* are doubled in comparison with those of the G_1 cell in *a*. (FROM KARL SPERLING AND POTU N. RAO, HUMANGENETIK 23:437, 1974.)

so that compaction in the S-phase nucleus led to the formation of "pulverized" chromosomal fragments rather than intact, compacted chromosomes. The results of this experiment suggested that the cytoplasm of a mitotic cell contained diffusible factors that could induce mitosis in a nonmitotic (i.e., interphase) cell. This finding suggested that the transition from G_2 to M was under positive control; that is, the transition was induced by the presence of some stimulatory agent.

The Role of Protein Kinases   While the cell-fusion experiments revealed the existence of factors that regulated the cell cycle, they provided no information about the biochemical properties of these factors. Insights into the nature of the agents that promote entry of a cell into mitosis (or meiosis) were first gained in a series of experiments on the oocytes and early embryos of frogs and invertebrates. These experiments are described in the Experimental Pathways at the end of this chapter. To summarize here, it was shown that entry of a cell into M phase is initiated by a protein called *maturation-promoting factor* (*MPF*). MPF consists of two subunits: (1) a subunit with kinase activity that transfers phosphate groups from ATP to specific serine and threonine residues of specific protein substrates and (2) a regulatory subunit called *cyclin*. The term *cyclin* was coined because the concentration of this regulatory protein rises and falls in a predictable pattern with each cell cycle (Figure 14.4). When the cyclin concentration is low, the kinase lacks the cyclin subunit and, as a result, is inactive. When the cyclin concentration rises, the kinase is
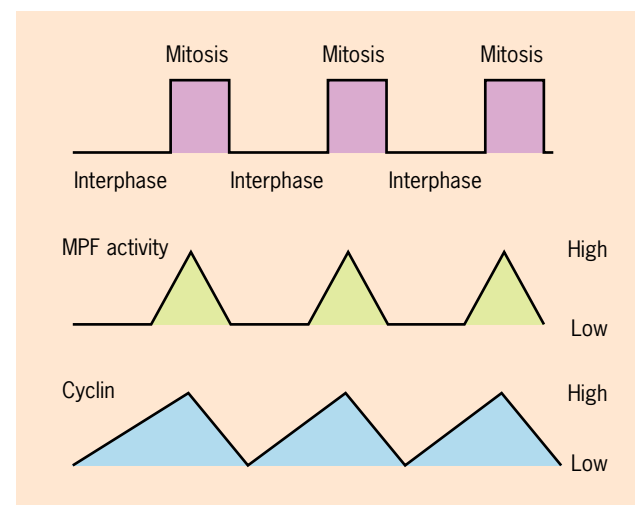
FIGURE 14.4 Fluctuation of cyclin and MPF levels during the cell cycle. This drawing depicts the cyclical changes that occur during early frog development when mitotic divisions occur rapidly and synchronously in all cells of the embryo. The top tracing shows the alternation between periods of mitosis and interphase, the middle tracing shows the cyclical changes in MPF kinase activity, and the lower tracing shows the cyclical changes in the concentrations of cyclins that control the relative activity of the MPF kinase. (REPRINTED WITH PERMISSION FROM A. W. MURRAY AND M. W. KIRSCHNER, SCIENCE 246:616, 1989; © COPYRIGHT 1989; AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

activated, causing the cell to enter M phase. These results suggested that (1) progression of cells into mitosis depends on an enzyme whose sole activity is to phosphorylate other proteins, and (2) the activity of this enzyme is controlled by a subunit whose concentration varies from one stage of the cell cycle to another.

Over the past two decades, a large number of laboratories have focused on MPF-like enzymes, called **cyclin-dependent kinases** (**Cdks**). It has been found that Cdks are not only involved in M phase but are the key agents that orchestrate activities throughout the cell cycle. Yeast cells have been particularly useful in studies of the cell cycle, at least in part because of the availability of temperature-sensitive mutants whose abnormal proteins affect various cell cycle processes. As discussed on page 537, temperature-sensitive mutants can be grown in a relatively normal manner at a lower (permissive) temperature and then shifted to a higher (restrictive) temperature to study the effect of the mutant gene product. Researchers studying the genetic control of the cell cycle have focused on two distantly related yeast species, the budding yeast *Saccharomyces cerevisiae*, which reproduces through the formation of buds at one end of the cell (see Figure 1.18*b*), and the fission yeast, *Schizosaccharomyces pombe*, which reproduces by elongating itself and then splitting into two equal-sized cells (see Figure 14.6*b*). The molecular basis of cell cycle regulation has been remarkably conserved throughout the evolution of eukaryotes. Once a gene involved in cell cycle control has been identified in one of the two yeast species, homologues are sought—and usually found—in the genomes of higher eukaryotes, including humans. By combining genetic, biochemical, and structural analyses, investigators have gained a comprehensive understanding of the major activities that allow a cell to grow and reproduce in a laboratory culture dish.

Research into the genetic control of the cell cycle in yeast began in the 1970s in two laboratories, initially that of Leland Hartwell at the University of Washington working on budding yeast and subsequently that of Paul Nurse at the University of Oxford working on fission yeast. Both laboratories identified a gene that, when mutated, would cause the growth of cells at elevated temperature to stop at certain points in the cell cycle. The product of this gene, which was called *cdc2* in fission yeast (and *CDC28* in budding yeast), was eventually found to be homologous to the catalytic subunit of MPF; in other words, it was a cyclin-dependent kinase. Subsequent research on yeast as well as many different vertebrate cells has supported the concept that the progression of a eukaryotic cell through its cell cycle is regulated at distinct stages. One of the primary stages of regulation occurs near the end of G_1 and another near the end of G_2. These stages represent points in the cell cycle where a cell becomes committed to beginning a crucial event—either initiating replication or entering mitosis.

We will begin our discussion with fission yeast, which has the least complex cell cycle. In this species, the same Cdk (cdc2) is responsible for passage through both points of commitment, though in partnership with different cyclins. A simplified representation of cell cycle regulation in fission yeast is
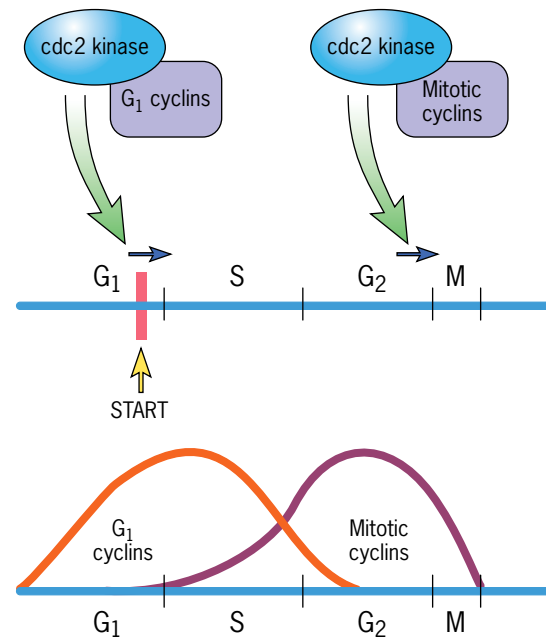
FIGURE 14.5 A simplified model for cell cycle regulation in fission yeast. The cell cycle is controlled primarily at two points, START and the G_2–M transition. In fission yeast, cyclins can be divided into two groups, G_1 cyclins and mitotic cyclins. Passage of a cell through these two critical junctures requires the activation of the same cdc2 kinase by a different type of cyclin. A third major transition occurs at the end of mitosis and is triggered by a rapid drop in concentration of mitotic cyclins. (*Note:* cdc 2 is also known as Cdk1.)

shown in Figure 14.5. The first transition point, which is called START, occurs in late G_1. Once a cell has passed START, it is irrevocably committed to replicating its DNA and, ultimately, completing the cell cycle.[1] Passage through START requires the activation of cdc2 by one or more G_1 *cyclins*, whose levels rise during late G_1 (Figure 14.5). Activation of cdc2 by these cyclins leads to the initiation of replication at sites where pre-replication complexes had previously assembled (see step 3, Figure 13.20).

Passage from G_2 to mitosis requires activation of cdc2 by a different group of cyclins—the *mitotic cyclins*. Cdks containing a mitotic cyclin (e.g., MPF described on page 601) phosphorylate substrates that are required for the cell to enter mitosis. Included among the substrates are proteins required for the dynamic changes in organization of both the chromosomes and cytoskeleton that characterize the shift from interphase to mitosis. Cells make a third commitment during the middle of mitosis, which determines whether they will complete cell division and reenter G_1 of the next cycle. Exit from

[1]Mammalian cells pass through a comparable point during G_1, referred to as the *restriction point*, at which time they become committed to DNA replication and ultimately to completing mitosis. Prior to the restriction point, mammalian cells require the presence of growth factors in their culture medium if they are to progress in the cell cycle. After they have passed the restriction point, these same cells will continue through the remainder of the cell cycle without external stimulation.

mitosis and entry into G_1 depends on a rapid decrease in Cdk activity that results from a plunge in concentration of the mitotic cyclins (Figure 14.5), an event that will be discussed on page 580 in conjunction with other mitotic activities.

Cyclin-dependent kinases are often described as the "engines" that drive the cell cycle through its various stages. The activities of these enzymes are regulated by a variety of "brakes" and "accelerators" that operate in combination with one another. These include:

Cyclin Binding   When a cyclin is present in the cell, it binds to the catalytic subunit of the Cdk, causing a major change in the conformation of the catalytic subunit. X-ray crystallographic structures of various cyclin–Cdk complexes indicate that cyclin binding causes the movement of a flexible loop of the Cdk polypeptide chain away from the opening to the enzyme's active site, allowing the Cdk to phosphorylate its protein substrates.

Cdk Phosphorylation/dephosphorylation   We have already seen in other chapters that many events that take place in a cell are regulated by the addition and removal of phosphate groups from proteins. The same is true for the events that lead to the onset of mitosis. We can see from Figure 14.5 that the level of mitotic cyclins rises through S and G_2. The mitotic cyclins present in a yeast cell during this period bind to the Cdk to form a cyclin–Cdk complex, but the complex shows little evi-

dence of kinase activity. Then, late in G_2, the cyclin–Cdk becomes activated and mitosis is triggered. To understand this change in Cdk activity, we have to look at the activity of three other regulatory enzymes—two kinases and a phosphatase. The role of these enzymes in the fission yeast cycle, which is illustrated in Figure 14.6a, was revealed through a combination of genetic and biochemical analyses. In step 1, one of the kinases, called CAK (Cdk-activating kinase), phosphorylates a critical threonine residue (Thr 161 of cdc2 in Figure 14.6a). Phosphorylation of this residue is necessary, but not sufficient, for the Cdk to be active. A second protein kinase shown in step 1, called Wee1, phosphorylates a key tyrosine residue in the ATP-binding pocket of the enzyme (Tyr 15 of cdc2 in Figure 14.6a). If this residue is phosphorylated, the enzyme is inactive, regardless of the phosphorylation state of any other residue. In other words, the effect of Wee1 overrides the effect of CAK, keeping the Cdk in an inactive state. Line 2 of Figure 14.6b shows the phenotype of cells with a mutant *wee1* gene. These mutants cannot maintain the Cdk in an inactive state and divide at an early stage in the cell cycle producing smaller cells, hence the name "wee." In normal (wild-type) cells, Wee1 keeps the Cdk inactive until the end of G_2. Then, at the end of G_2, the inhibitory phosphate at Tyr 15 is removed by the third enzyme, a phosphatase named Cdc25 (step 2, Figure 14.6a). Removal of this phosphate switches the stored cyclin–Cdk molecules into the active state, allowing it to phosphorylate key substrates and drive the yeast cell into
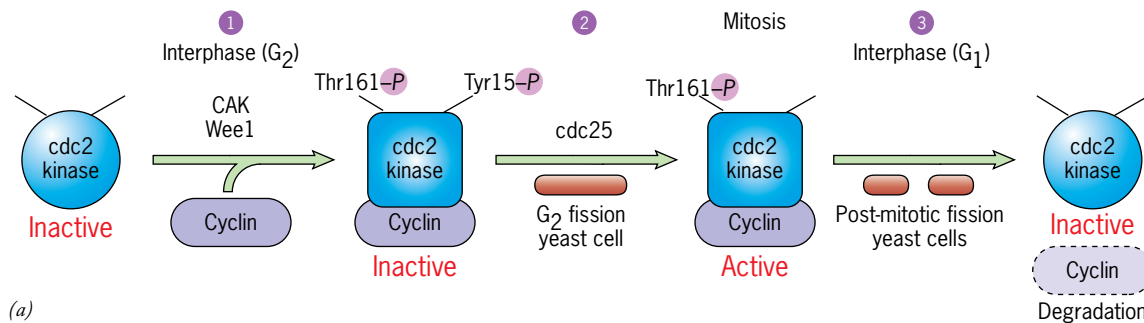
(a)

FIGURE 14.6 Progression through the fission yeast cell cycle requires the phosphorylation and dephosphorylation of critical cdc2 residues. (a) During G_2, the cdc2 kinase interacts with a mitotic cyclin but remains inactive as the result of phosphorylation of a key tyrosine residue (Tyr 15 in fission yeast) by Wee1 (step 1). A separate kinase, called CAK, transfers a phosphate to another residue (Thr 161), which is required for cdc2 kinase activity later in the cell cycle. When the cell reaches a critical size, an enzyme called Cdc25 phosphatase is activated, which removes the inhibitory phosphate on the Tyr 15 residue. The resulting activation of the cdc2 kinase drives the cell into mitosis (step 2). By the end of mitosis (step 3), the stimulatory phosphate group is removed from Thr 161 by another phosphatase. The free cyclin is subsequently degraded, and the cell begins another cycle. (The mitotic Cdk in mammalian cells is phosphorylated and dephosphorylated in a similar manner.) (b) Identification of Wee1 kinase and Cdc25 phosphatase was made by studying mutants that behaved as shown in this figure. Line 1 shows the G_2 and M stages of a wild-type cell. Line 2 shows the effect of a mutant *wee1* gene; the cell divides prematurely, forming small (wee) cells. Line 3 shows the
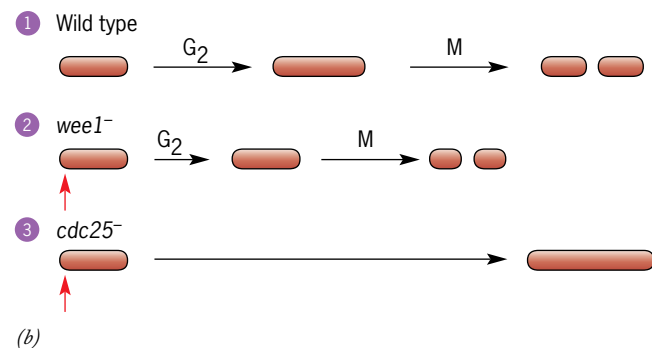
effect of a mutant *cdc25* gene; the cell does not divide but continues to grow. The red arrow marks the time when the temperature is raised to inactivate the mutant protein. (A: AFTER T. R. COLEMAN AND W. G. DUNPHY, CURR. OPIN. CELL BIOL. 6:877, 1994.)

mitosis. Line 3 of Figure 14.6*b* shows the phenotype of cells with a mutant *cdc25* gene. These mutants cannot remove the inhibitory phosphate from the Cdk and cannot enter mitosis. The balance between Wee1 kinase and Cdc25 phosphatase activities, which normally determines whether the cell will remain in G_2 or progress into mitosis, is regulated by still other kinases and phosphatases. As we will see shortly, these pathways can stop the cell from entering mitosis under conditions that might lead to an abnormal cell division.

Cdk Inhibitors   Cdk activity can be blocked by a variety of inhibitors. In budding yeast, for example, a protein called Sic1 acts as a Cdk inhibitor during G_1. The degradation of Sic1 allows the cyclin–Cdk that is present in the cell to initiate DNA replication. The role of Cdk inhibitors in mammalian cells is discussed on page 568.

Controlled Proteolysis   It is evident from Figures 14.4 and 14.5 that cyclin concentrations oscillate during each cell cycle, which leads to changes in the activity of Cdks. Cells regulate the concentration of cyclins, and other key cell cycle proteins, by adjusting both the rate of synthesis and the rate of destruction of the molecule at different points in the cell cycle. Degradation is accomplished by means of the ubiquitin–proteasome pathway described on page 530. Unlike other mechanisms that control Cdk activity, degradation is an irreversible event that helps drive the cell cycle in a single direction. Regulation of the cell cycle requires two classes of multisubunit complexes (SCF and APC complexes) that function as *ubiquitin ligases*. These complexes recognize proteins to be degraded and link these proteins to a polyubiquitin chain, which ensures their destruction in a proteasome. The SCF complex is active from late G_1 through early mitosis (see Figure 14.26*a*) and mediates the destruction of G_1 cyclins, Cdk inhibitors, and other cell cycle proteins. These proteins become targets for an SCF after they are phosphorylated by the protein kinases (i.e., Cdks) that regulate the cell cycle. Mutations that inhibit SCFs from mediating proteolysis of key proteins, such as G_1 cyclins or the Cdk inhibitor Sic1 mentioned above, can prevent cells from entering S phase and replicating their DNA. The APC complex acts in mitosis and degrades a number of key mitotic proteins, including the mitotic cyclins. Destruction of the mitotic cyclins allows a cell to exit mitosis and enter a new cell cycle (page 580).

Subcellular Localization   Cells contain a number of different compartments in which regulatory molecules can either be united with or separated from the proteins they interact with. Subcellular localization is a dynamic phenomenon in which cell cycle regulators are moved into different compartments at different stages. For example, one of the major mitotic cyclins in animal cells (cyclin B1) shuttles between the nucleus and cytoplasm until G_2, when it accumulates in the nucleus just prior to the onset of mitosis (Figure 14.7). Nuclear accumulation of cyclin B1 is facilitated by phosphorylation of a cluster of serine residues that reside in its nuclear export signal (NES, page 481). Phosphorylation presumably blocks subsequent

(a)   *(b)*
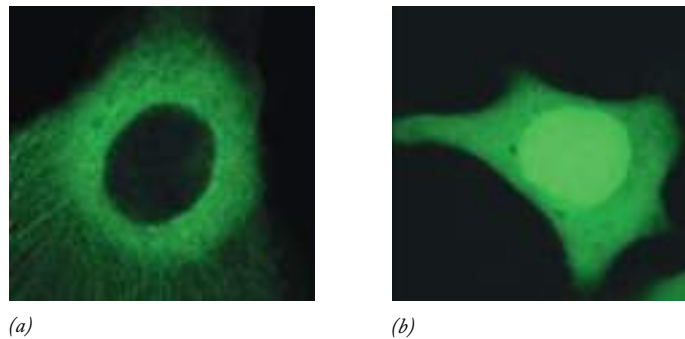
FIGURE 14.7 Experimental demonstration of subcellular localization during the cell cycle. Micrographs of a living HeLa cell that has been injected with cyclin B1 linked to the green fluorescent protein (page 267). The cell shown in *a* is in the G_2 phase of its cell cycle, and the fluorescently labeled cyclin B1 is localized almost entirely in the cytoplasm. The micrograph in *b* shows the cell in prophase of mitosis, and the labeled cyclin B1 is concentrated in the cell nucleus. The basis for this change in localization is discussed in the text. (FROM PAUL CLUTE AND JONATHAN PINES, NATURE CELL BIOL. 1:83, 1999; © COPYRIGHT 1999, MACMILLAN MAGAZINES LIMITED.)

export of the cyclin back to the cytoplasm. If nuclear accumulation of the cyclin is blocked, cells fail to initiate cell division.

As noted above, the proteins and processes that control the cell cycle are remarkably conserved among eukaryotes. As in yeast, successive waves of synthesis and degradation of different cyclins play a key role in driving mammalian cells from one stage to the next. Unlike yeast cells, which have a single Cdk, mammalian cells produce several different versions of this protein kinase. Different cyclin–Cdk complexes target different groups of substrates at different points within the cell cycle. The pairing between individual cyclins and Cdks is specific, and only certain combinations are found (Figure 14.8). In mammalian cells, for example, the activity of a cyclin E–Cdk2 complex drives the cell into S phase, whereas activity of a cyclin B1–Cdk1 complex (the mammalian MPF) drives the cell into mitosis. Cdks do not always stimulate activities, but can also inhibit inappropriate events. For example, cyclin B1–Cdk1 activity during G_2 prevents a cell from rereplicating DNA that has already been replicated earlier in the cell cycle (page 548). This helps ensure that each region of the genome is replicated once and only once per cell cycle.

The roles of the various cyclin–Cdk complexes shown in Figure 14.8 have been determined by a wide range of biochemical studies carried out on mammalian cells for more than two decades. Over the past few years, the roles of these proteins have been reexamined in knockout mice, with some surprising results. As expected, the phenotype of a particular knockout mouse depends on the gene that has been eliminated. Mice that are unable to synthesize Cdk1, cyclin B1, or cyclin A2, die as early embryos, suggesting that the proteins encoded by these three genes are essential for a normal cell cycle. In contrast, a mouse embryo that lacks the genes encoding *all* four of the other cell cycle Cdks (namely, Cdks 2, 3, 4, and
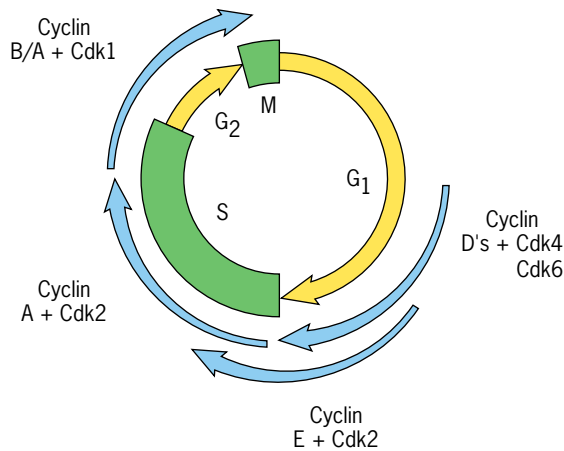
FIGURE 14.8 Combinations between various cyclins and cyclin-dependent kinases at different stages in the mammalian cell cycle. Cdk activity during early G_1 is very low, which promotes the formation of prereplication complexes at the origins of replication (see Figure 13.20). By mid-G_1, Cdk activity is evident due to the association of Cdk4 and Cdk6 with the D-type cyclins (D1, D2, and D3). Among the substrates for these Cdks is an important regulatory protein called pRb (Section 16.3, Figure 16.11). The phosphorylation of pRb leads to the transcription of a number of genes, including those that code for cyclins E and A, Cdk1, and proteins involved in replication. The G_1–S transition, which includes the initiation of replication, is driven by the activity of the cyclin E–Cdk2 and cyclin A–Cdk2 complexes. The transition from G_2 to M is driven by the sequential activity of cyclin A–Cdk1 and cyclin B1–Cdk1 complexes, which phosphorylate such diverse substrates as cytoskeletal proteins, histones, and proteins of the nuclear envelope. (The mammalian Cdk1 kinase is equivalent to the fission yeast cdc2 kinase, and its inhibition and activation are similar to that indicated in Figure 14.6.) (AFTER C. J. SHERR, CELL 73:1060, 1993; BY PERMISSION OF CELL PRESS AND H. A. COLLER, NATURE REVS. MOL. CELL BIOL. 8:667, 2007.)

6) is capable of developing to a stage with fully formed organs, although the animal does not survive to birth. Cells taken from such embryos are capable of proliferating in culture, though more slowly than normal cells. This finding indicates that, as in yeast, Cdk1 is the only Cdk *required* to drive a mammalian cell through all of the stages of the cell cycle. In other words, even though the other Cdks are normally expressed at specific times during the mammalian cell cycle, Cdk1 is able to "cover" for their absence, ensuring that all of the required substrates are phosphorylated at each stage of the cell cycle. This is a classical case of *redundancy*, in which a protein is able to carry out functions that it would not normally perform. Still, the absence of one of these "nonessential" cyclins or Cdks typically results in distinct cell cycle abnormalities, at least in certain types of cells. Mice lacking a gene for cyclin D1, for example, are smaller than control animals, which stems from a reduction in the level of cell division throughout the body. In addition, cyclin D1-deficient animals display a particular lack of cell proliferation during development of the retina. Mice lacking Cdk4 develop without insulin-producing cells in their pancreas. Mice lacking Cdk2 appear to develop normally but exhibit specific defects during

meiosis, which reinforces the important differences in the regulation of mitotic and meiotic divisions.

Checkpoints, Kinase Inhibitors, and Cellular Responses

Ataxia-telangiectasia (*AT*) is an inherited recessive disorder characterized by a host of diverse symptoms, including a greatly increased risk for certain types of cancer.[2] During the late 1960s—following the deaths of several individuals undergoing radiation therapy—it was discovered that patients with AT are extremely sensitive to ionizing radiation (page 555). So too are cells from these patients, which lack a crucial protective response found in normal cells. When normal cells are subjected to treatments that damage DNA, such as ionizing radiation or DNA-altering drugs, their progress through the cell cycle stops while the damage is repaired. If, for example, a normal cell is irradiated during the G_1 phase of the cell cycle, it delays progression into S phase. Similarly, cells irradiated in S phase delay further DNA synthesis, whereas cells irradiated in G_2 delay entry into mitosis.

Studies of this type carried out in yeast gave rise to a concept, formulated by Leland Hartwell and Ted Weinert in 1988, that cells possess **checkpoints** as part of their cell cycle. Checkpoints are surveillance mechanisms that halt the progress of the cell cycle if (1) any of the chromosomal DNA is damaged, or (2) certain critical processes, such as DNA replication during S phase or chromosome alignment during M phase, have not been properly completed. Checkpoints ensure that each of the various events that make up the cell cycle occurs accurately and in the proper order. Many of the proteins of the checkpoint machinery have no role in normal cell cycle events and are only called into action when an abnormality appears. In fact, the genes encoding several checkpoint proteins were first identified in mutant yeast cells that continued their progress through the cell cycle, despite suffering DNA damage or other abnormalities that caused serious defects.

Checkpoints are activated throughout the cell cycle by a system of sensors that recognize DNA damage or cellular abnormalities. If a sensor detects the presence of a defect, it triggers a response that temporarily arrests further cell cycle progress. The cell can then use the delay to repair the damage or correct the defect rather than continuing to the next stage. This is especially important because mammalian cells that undergo division with genetic damage run the risk of becoming transformed into a cancer cell. If the DNA is damaged beyond repair, the checkpoint mechanism can transmit a signal that leads either to (1) the death of the cell or (2) its conversion to a state of permanent cell cycle arrest (known as senescence).

We have seen in numerous places in this text where the study of a rare human disease has led to a discovery of basic importance in cell and molecular biology. The cell's DNA

---

[2]Other symptoms of the disease include unsteady posture (ataxia) resulting from degeneration of nerve cells in the cerebellum, permanently dilated blood vessels (telangiectasia) in the face and elsewhere, susceptibility to infection, and cells with an abnormally high number of chromosome aberrations. The basis for the first two symptoms has yet to be determined.

damage response provides another example of this path to discovery. The gene responsible for ataxia-telangiectasia (the *ATM* gene) encodes a protein kinase that is activated by certain DNA lesions, particularly double-stranded breaks (page 555). Remarkably, the presence of a single break in one of the cell's DNA molecules is sufficient to cause rapid, large-scale activation of ATM molecules, causing cell cycle arrest. A related protein kinase called ATR is activated by other types of lesions, including those resulting from incompletely replicated DNA or UV irradiation. Both ATM and ATR are part of multiprotein complexes capable of binding to chromatin that contains damaged DNA. Once bound, ATM and ATR can phosphorylate a remarkable variety of proteins that participate in cell cycle checkpoints and DNA repair.

How does a cell stop its progress from one stage of the cell cycle to the next? We will briefly examine two well-studied pathways available to mammalian cells to arrest their cell cycle in response to DNA damage.

1. If a cell preparing to enter mitosis is subjected to UV irradiation, ATR kinase is activated and the cell arrests in G_2. ATR kinase molecules are thought to be recruited to sites of protein-coated, single-stranded DNA, such as those present as UV-damaged DNA is repaired (Figure 13.26). ATR phosphorylates and activates a checkpoint kinase, called Chk1 (step 1, Figure 14.9), which in turn phosphorylates Cdc25 on a particular serine residue (step 2), making the Cdc25 molecule a target for a special adaptor protein that binds to Cdc25 in the cytoplasm (step 4). This interaction inhibits Cdc25's phosphatase activity and prevents it from being reimported into the nucleus. As discussed on page 565, Cdc25 normally plays a key role in the G_2/M transition by removing inhibitory phosphates from Cdk1. Thus, the absence of Cdc25 from the nucleus leaves the Cdk in an inactive state (step 5) and the cell arrested in G_2.

2. Damage to DNA also leads to the synthesis of proteins that directly inhibit the cyclin–Cdk complex that drives the cell cycle. For example, cells exposed to ionizing radiation in G_1 synthesize a protein called p21 (molecular mass of 21 kDa) that inhibits the kinase activity of the G_1 Cdk. This prevents the cells from phosphorylating key substrates and from entering S phase. ATM is involved in this checkpoint mechanism. In this particular DNA-damage response, the breaks in DNA that are caused by ionizing radiation serve as sites for the recruitment of a protein complex termed MRN. MRN can be considered as a sensor of DNA breaks. MRN recruits ATM, which phosphorylates and activates another checkpoint kinase called Chk2 (step a, Figure 14.9). Chk2 in turn phosphorylates a transcription factor (p53) (step b), which leads to the transcription and translation of the *p21* gene (steps c and d) and subsequent inhibition of Cdk (step e). Approximately 50 percent of all human tumors show evidence of mutations in the gene that encodes p53, which reflects its importance in the control of cell growth. The role of p53 is discussed at length in Chapter 16.
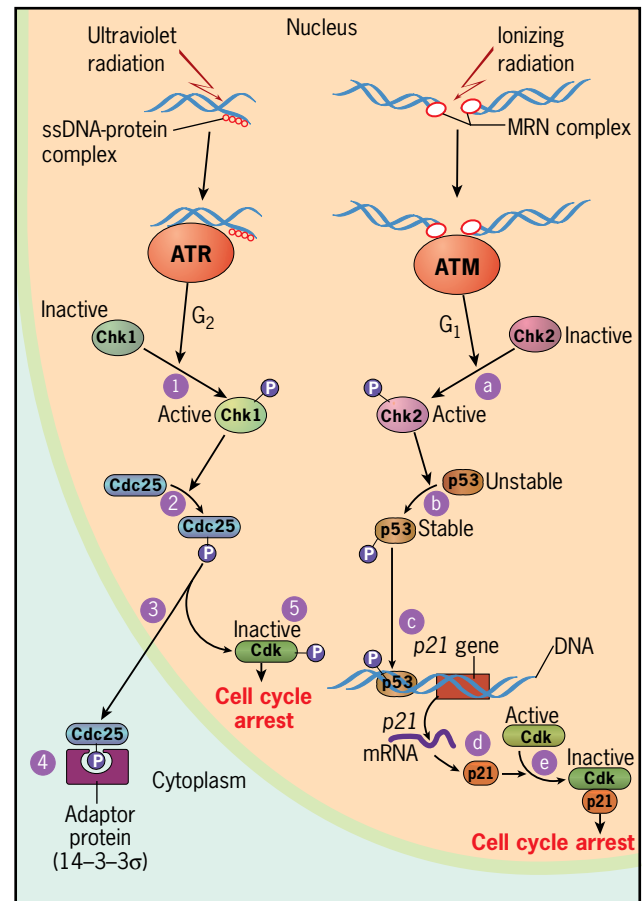
FIGURE 14.9 Models for the mechanism of action of two DNA-damage checkpoints. ATM and ATR are protein kinases that become activated following specific types of DNA damage. Each of these proteins acts through checkpoint signaling pathways that lead to cell cycle arrest. ATM becomes activated in response to double-strand breaks, which are detected by the MRN protein complex. ATR, on the other hand, becomes activated by protein-coated ssDNA that forms when replication forks become stalled or the DNA is being repaired after various types of damage. In the G_2 pathway shown here, ATR phosphorylates and activates the checkpoint kinase Chk1 (step 1), which phosphorylates and inactivates the phosphatase Cdc25 (step 2), which normally shuttles between the nucleus and cytoplasm (step 3). Once phosphorylated, Cdc25 is bound by an adaptor protein in the cytoplasm (step 4) and cannot be reimported into the nucleus, which leaves the Cdk in its inactivated, phosphorylated state (step 5). In the G_1 pathway shown here, ATM phosphorylates and activates the checkpoint kinase Chk2 (step a), which phosphorylates p53 (step b). p53 is normally very short-lived, but phosphorylation by Chk2 stabilizes the protein, enhancing its ability to activate *p21* transcription (step c). Once transcribed and translated (step d), p21 directly inhibits the Cdk (step e). (Numerous other G_1, S, and G_2 checkpoint pathways involving ATM and ATR have been identified.)

p21 is only one of at least seven known Cdk inhibitors. The interaction between a related Cdk inhibitor (p27) and one of the cyclin–Cdk complexes is shown in Figure 14.10a. In this model, the p27 molecule drapes itself across both subunits of the cyclin A–Cdk2 complex, changing the conforma-

(a) (b) (c)

FIGURE 14.10 p27: a Cdk inhibitor that arrests cell cycle progression. (*a*) Three-dimensional structure of a complex between p27 and cyclin A–Cdk2. Interaction with p27 alters the conformation of the Cdk catalytic subunit, inhibiting its protein kinase activity. (*b*) A pair of littermates at 12 weeks of age. In addition to possessing different genes for coat color, the mouse with dark fur has been genetically engineered to lack both copies of the *p27* gene (denoted as *p27*$^{-/-}$), which accounts for its larger size. (*c*) Comparison of the thymus glands from a normal (left) and a *p27*$^{-/-}$ mouse (right). The gland from the *p27* knockout mouse is much larger owing to an increased number of cells. (A: Reprinted with permission from Alicia A. Russo et al., Nature 382:327, 1995, courtesy of Nikola Pavletich; © copyright 1995 Macmillan Magazines Limited; b,c: from Keiko Nakayama et al., courtesy of Kei-ichi Nakayama, Cell 85:710, 711, 1996; by permission of Cell Press.)

tion of the catalytic subunit and inhibiting its kinase activity. In many cells, p27 must be phosphorylated and then degraded before progression into S phase can occur.

Cdk inhibitors, such as p21 and p27, are also active in cell differentiation. Just before cells begin to differentiate—whether into muscle cells, liver cells, blood cells, or some other type—they typically withdraw from the cell cycle and stop dividing. Cdk inhibitors are thought to either allow or directly induce cell cycle withdrawal. Just as the functions of specific Cdks and cyclins have been studied in knockout mice, so too have their inhibitors. Knockout mice that lack the *p27* gene show a distinctive phenotype: they are larger than normal (Figure 14.10*b*), and certain organs, such as the thymus gland and spleen, contain a significantly greater number of cells than those of a normal animal (Figure 14.10*c*). In normal mice, the cells of these particular organs synthesize relatively high levels of p27, and it is presumed that the absence of this protein in the p27-deficient animals allows the cells to divide several more times before they differentiate.

REVIEW ?

1. What is the cell cycle? What are the stages of the cell cycle? How does the cell cycle vary among different types of cells?

2. Describe how [^{3}H]thymidine and autoradiography can be used to determine the length of the various periods of the cell cycle.

3. What is the effect of fusing a G$_1$-phase cell with one in M; of fusing a G$_2$- or S-phase cell with one in M?

4. How does the activity of MPF vary throughout the cell cycle? How is this correlated with the concentration of cyclins? How does the cyclin concentration affect MPF activity?

5. What are the respective roles of CAK, Wee1, and Cdc25 in controlling Cdk activity in fission yeast cells? What is the effect of mutations in the *wee1* or *cdc25* genes in these cells?

6. What is meant by a cell cycle checkpoint? What is its importance? How does a cell stop its progress at one of these checkpoints?

14.2 M PHASE: MITOSIS AND CYTOKINESIS

Whereas our understanding of cell cycle regulation rests largely on genetic studies in yeast, our knowledge of M phase is based on more than a century of microscopic and biochemical research on animals and plants. The name "mitosis" comes from the Greek word *mitos*, meaning "thread." The name was coined in 1882 by the German biologist Walther Flemming to describe the threadlike chromosomes that mysteriously appeared in animal cells just before they divided in two. The beauty and precision of cell division is best appreciated by watching a time-lapse video of the process (e.g., www.bio.unc.edu/faculty/salmon/lab/mitosis/mitosismovies.html) rather than reading about it in a textbook.

Mitosis is a process of nuclear division in which the replicated DNA molecules of each chromosome are faithfully segregated into two nuclei. Mitosis is usually accompanied by **cytokinesis**, a process by which a dividing cell splits in two, partitioning the cytoplasm into two cellular packages. The two daughter cells resulting from mitosis and cytokinesis possess a genetic content identical to each other and to the mother cell from which they arose. Mitosis, therefore, maintains the chromosome number and generates new cells for the

Prophase

1. Chromosomal material condenses to form compact mitotic chromosomes. Chromosomes are seen to be composed of two chromatids attached together at the centromere.

2. Cytoskeleton is disassembled, and mitotic spindle is assembled.

3. Golgi complex and ER fragment. Nuclear envelope disperses.

Prometaphase

1. Chromosomal microtubules attach to kinetochores of chromosomes.

2. Chromosomes are moved to spindle equator.

Metaphase

1. Chromosomes are aligned along metaphase plate, attached by chromosomal microtubules to both poles.

Anaphase

1. Centromeres split, and chromatids separate.

2. Chromosomes move to opposite spindle poles.

3. Spindle poles move farther apart.

Telophase

1. Chromosomes cluster at opposite spindle poles.

2. Chromosomes become dispersed.

3. Nuclear envelope assembles around chromosome clusters.

4. Golgi complex and ER reforms.

5. Daughter cells formed by cytokinesis.

FIGURE 14.11 **The stages of mitosis in an animal cell (left drawings) and a plant cell (right photos).**
(Micrographs from Andrew Bajer.)

growth and maintenance of an organism. Mitosis can take place in either haploid or diploid cells. Haploid mitotic cells are found in fungi, plant gametophytes, and a few animals (including male bees known as drones). Mitosis is a stage of the cell cycle when the cell devotes virtually all of its energy to a single activity—chromosome segregation. As a result, most metabolic activities of the cell, including transcription and translation, are curtailed during mitosis, and the cell becomes relatively unresponsive to external stimuli.

We have seen in previous chapters how much can be learned about the factors responsible for a particular process by studying that process outside of a living cell (page 270). Our understanding of the biochemistry of mitosis has been greatly aided by the use of extracts prepared from frog eggs. These extracts contain stockpiles of all the materials (histones, tubulin, etc.) necessary to support mitosis. When chromatin or whole nuclei are added to the egg extract, the chromatin is compacted into mitotic chromosomes, which are segregated by a mitotic spindle that assembles spontaneously within the cell-free mixture. In many experiments, the role of a particular protein in mitosis can be studied by removing that protein from the egg extract by addition of an antibody (immunodepletion) and determining whether the process can continue in the absence of that substance (see Figure 14.21 for an example).

Mitosis is generally divided into five stages (Figure 14.11), *prophase, prometaphase, metaphase, anaphase*, and *telophase*, each characterized by a particular series of events. Keep in mind that each of these stages represents a segment of a continuous process; the division of mitosis into arbitrary phases is done only for the sake of discussion and experimentation.

Prophase

During the first stage of mitosis, that of **prophase**, the duplicated chromosomes are prepared for segregation, and the mitotic machinery is assembled.

Formation of the Mitotic Chromosome   The nucleus of an interphase cell contains tremendous lengths of chromatin fibers. The extended state of interphase chromatin is ideally suited for the processes of transcription and replication but not for segregation into two daughter cells. Before segregating its chromosomes, a cell converts them into much shorter, thicker structures by a remarkable process of **chromosome compaction** (or *chromosome condensation*), which occurs during early prophase (Figures 14.11 and 14.12).

As described on page 483, the chromatin of an interphase cell is organized into fibers approximately 30 nm in diameter. Mitotic chromosomes are composed of similar types of fibers as seen by electron microscopic examination of whole chromosomes isolated from mitotic cells (Figure 14.13*a*). It appears that chromosome compaction does not alter the nature of the chromatin fiber, but rather the way that the chromatin fiber is packaged. Treatment of mitotic chromosomes with solutions that solubilize the histones and the majority of the nonhistone proteins reveals a structural framework or scaf-
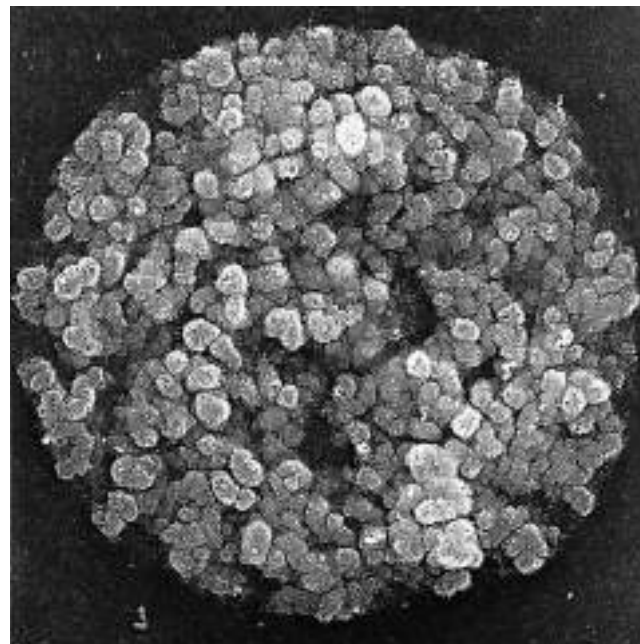
5 µm

FIGURE 14.12 **The chromosomes of this early prophase nucleus** have begun the process of compaction that converts them into short, rod-like mitotic chromosomes that separate at a later stage in mitosis. (FROM A. T. SUMNER, CHROMOSOMA 100:411, 1991.)

fold that retains the basic shape of the intact chromosome (Figure 14.13*b*). Loops of DNA are attached at their base to the nonhistone proteins that make up the chromosome scaffold (shown at higher magnification in Figure 12.11). During interphase, the proteins of the chromosome scaffold are dispersed within the nucleus, possibly forming part of the nuclear matrix (page 499).

In recent years, research on chromosome compaction has focused on an abundant multiprotein complex called *condensin*. The proteins of condensin were discovered by incubating nuclei in frog egg extracts and identifying those proteins that associated with the chromosomes as they underwent compaction. Removal of condensin from the extracts prevented normal chromosome compaction. How is condensin involved in such dramatic changes in chromatin architecture? There is very little data available from in vivo studies to answer this question, but there is considerable speculation.

Supercoiled DNA occupies a much smaller volume than relaxed DNA (see Figure 10.12), and studies suggest that DNA supercoiling plays a key role in compacting a chromatin fiber into the tiny volume occupied by a mitotic chromosome. In the presence of a topoisomerase and ATP, condensin is able to bind to DNA in vitro and curl the DNA into positively supercoiled loops. This finding fits nicely with the observation that chromosome compaction at prophase requires topoisomerase II, which along with condensin is present as part of the mitotic chromosome scaffold (Figure 14.13*b*). A speculative model for condensin action is shown

(a)

(b)

3 µm
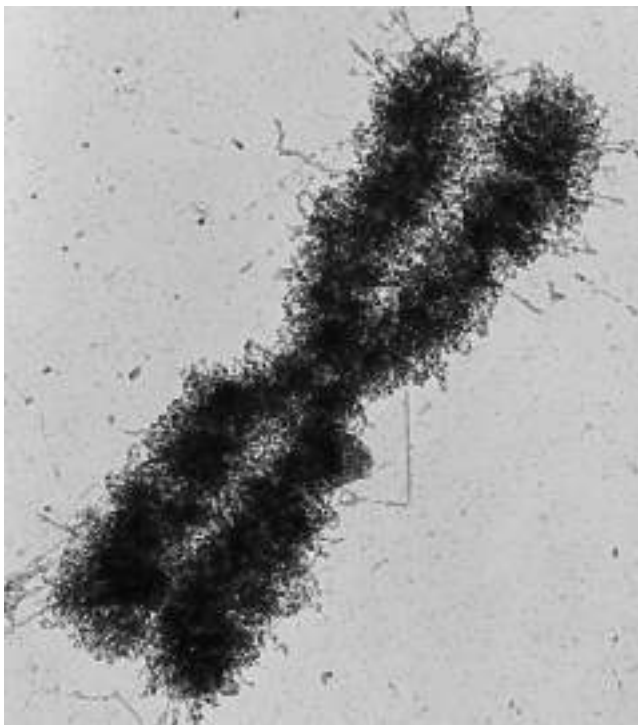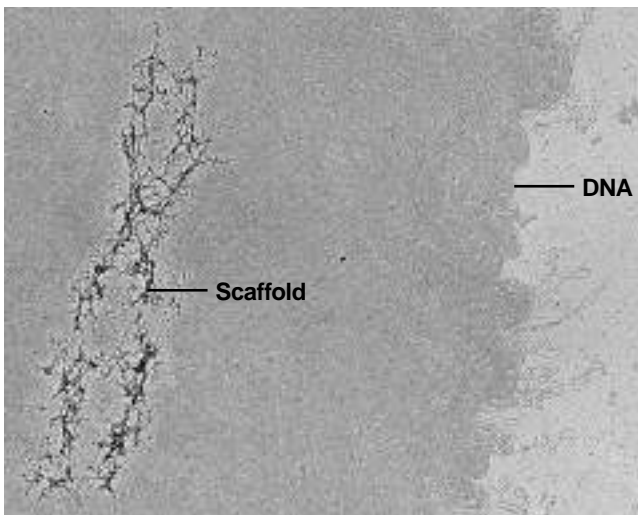
FIGURE 14.13 **The mitotic chromosome.** (*a*) Electron micrograph of a whole-mount preparation of a human mitotic chromosome. The structure is seen to be composed of a knobby fiber 30 nm in diameter, which is similar to that found in interphase chromosomes. (*b*) Appearance of a mitotic chromosome after the histones and most of the nonhistone proteins have been removed. The residual proteins form a scaffold from which loops of DNA are seen to emerge (the DNA loops are shown more clearly in Figure 12.11). (A: COURTESY OF GUNTHER F. BAHR; B: FROM JAMES R. PAULSON AND ULRICH K. LAEMMLI, CELL 12:820, 1977; BY PERMISSION OF CELL PRESS.)

FIGURE 14.14 **Model for the roles of condensin and cohesin in the formation of mitotic chromosomes.** Just after replication, the DNA helices of a pair of sister chromatids would be held in association by cohesin molecules that encircled the sister DNA helices, as shown at the top of the drawing. As the cell entered mitosis, the compaction process would begin, aided by condensin molecules, as shown in the lower part of the drawing. In this model, condensin brings about chromosome compaction by forming a ring around supercoiled loops of DNA within chromatin. Cohesin molecules would continue to hold the DNA of sister chromatids together. It is proposed (but not shown in this drawing), that cooperative interactions between condensin molecules would then organize the supercoiled loops into larger coils, which are then folded into a mitotic chromosome fiber. The top and bottom insets show the subunit structure of an individual cohesin and condensin complex, respectively. Both complexes are built around a pair of SMC subunits. Each of the SMC polypeptides folds back on itself to form a highly elongated antiparallel, coiled coil with an ATP-binding globular domain where the N- and C-termini come together. Cohesin and condensin also have two or three non-SMC subunits that complete the ring-like structure of these proteins.

in Figure 14.14. Condensin is activated at the onset of mitosis by phosphorylation of several of its subunits by the cyclin–Cdk responsible for driving cells from G_2 into mitosis.

Thus condensin is presumably one of the targets through which Cdks are able to trigger cell cycle activities. The subunit structure of a V-shaped condensin molecule is shown in the lower inset of Figure 14.14.

As the result of compaction, the chromosomes of a mitotic cell appear as distinct, rod-like structures. Close examination of mitotic chromosomes reveals each of them to be composed of two mirror-image, "sister" **chromatids** (Figure 14.13*a*). Sister chromatids are a result of replication in the previous interphase.

Prior to replication, the DNA of each interphase chromosome becomes associated at sites along its length with a multiprotein complex called **cohesin** (Figure 14.14). Following replication, cohesin holds the two sister chromatids together through G_2 and into mitosis when they are ultimately separated. As indicated in the insets of Figure 14.14, con-

(a)   1 µm   (b)
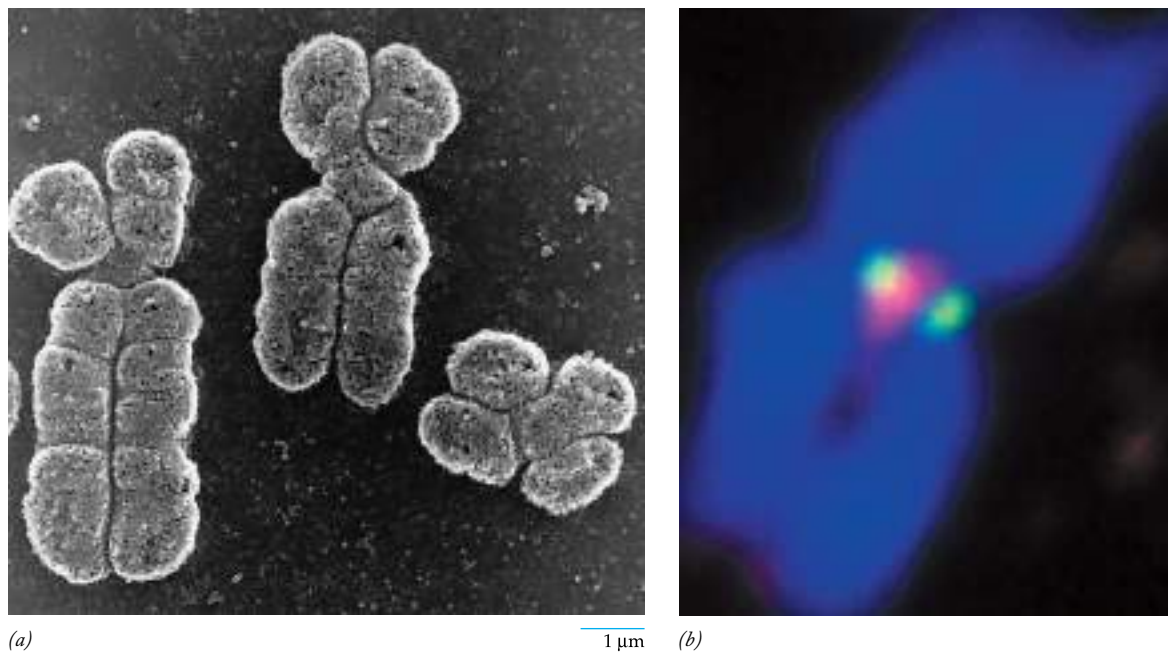
FIGURE 14.15 **Each mitotic chromosome is comprised of a pair of sister chromatids connected to one another by the protein complex cohesin.** (*a*) Scanning electron micrograph of several human metaphase chromosomes showing the paired identical chromatids associated loosely along their length and joined tightly at the centromere. The chromatids are not split apart from one another until anaphase. (*b*) Fluorescence micrograph of a metaphase chromosome in a cultured human cell. The DNA is stained blue, the kinetochores are green, and cohesin is red. At this stage of mitosis, cohesin has been lost from the arms of the sister chromatids but remains concentrated at the centromeres where the two sisters are tightly joined. (A: FROM A. T. SUMNER, CHROMOSOMA 100:415, 1991; B: BY S. HAUF AND JAN-MICHAEL PETERS, REPRINTED WITH PERMISSION FROM T. J. MITCHISON AND E. D. SALMON., NATURE CELL BIOL. 3:E17, 2001; © COPYRIGHT 2001, MACMILLAN MAGAZINES LIMITED.)

densin and cohesin have a similar structural organization. A number of experiments support the hypothesis that the cohesin ring encircles two sister DNA molecules as shown in both the upper and lower portions of Figure 14.14.

In vertebrates, cohesin is released from the chromosomes in two distinct stages. Most of the cohesin dissociates from the arms of the chromosomes as they become compacted during prophase. Dissociation is induced by phosphorylation of cohesin subunits by two important mitotic enzymes called Polo-like kinase and Aurora B kinase. In the wake of this event, the chromatids of each mitotic chromosome are held relatively loosely along their extended arms, but much more tightly at their centromeres (Figure 14.13*a* and Figure 14.15). Cohesin is thought to remain at the centromeres because of the presence there of a phosphatase that removes any phosphate groups added to the protein by the kinases. Release of cohesin from the centromeres is normally delayed until anaphase as described on page 580. If the phosphatase is experimentally inactivated, sister chromatids separate from one another prematurely prior to anaphase.

Centromeres and Kinetochores   The most notable landmark on a mitotic chromosome is an indentation or *primary constriction*, which marks the position of the centromere (Figure 14.13*a*). The centromere is the residence of highly repeated DNA sequences (see Figure 10.19) that serve as the binding sites for specific proteins. Examination of sections through a mitotic chromosome reveals the presence of a proteinaceous, button-like structure, called the **kinetochore**, at the outer surface of the centromere of each chromatid (Figure 14.16*a,b*). The kinetochore assembles at the centromere during prophase. As will be apparent shortly, the kinetochore functions as (1) the site of attachment of the chromosome to the dynamic microtubules of the mitotic spindle (as in Figure 14.30), (2) the residence of several motor proteins involved in chromosome motility (Figure 14.16*c*), and (3) a key component in the signaling pathway of an important mitotic checkpoint (see Figure 14.31).

A question of great interest to scientists studying kinetochores is how these structures are able to maintain their attachment to microtubules that are continually growing and shrinking at their plus end. To maintain this type of "floating grip," the coupler would have to move with the end of the microtubule as subunits were added or removed. Figure 14.16*c* depicts two types of proteins that have been implicated as possible linkers of a kinetochore to dynamic microtubules, namely, motor proteins and a rod-shaped protein called Ndc80. Ndc80 is an essential kinetochore protein that forms fibrils that appear to reach out and bind the surface of the adjacent microtubule. Among the various motor proteins associated with the kinetochore, CENP-E, which is a member of the kinesin superfamily, has been the most strongly implicated as a potential microtubule coupler. A third candidate in yeast cells, called Dam1, is illustrated in Figure 14.30.

(a)   0.2 m

(b)

(c)
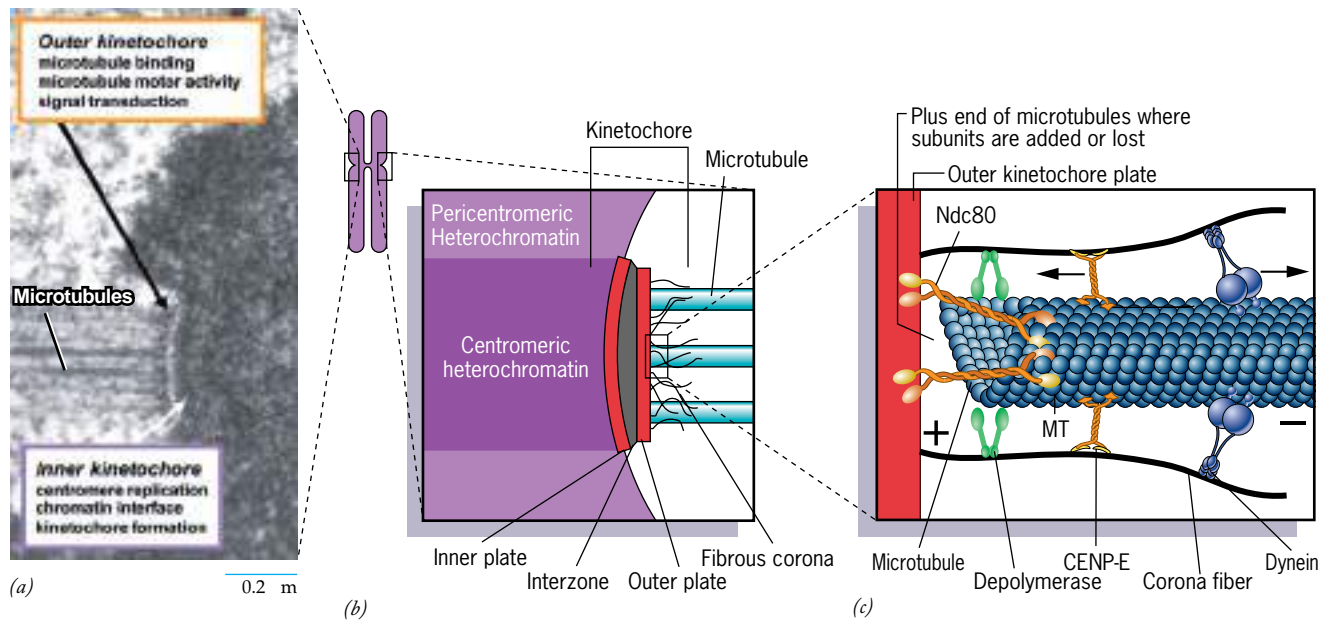
FIGURE 14.16   The kinetochore. (*a*) Electron micrograph of a section through one of the kinetochores of a mammalian metaphase chromosome, showing its three-layered (trilaminar) structure. Microtubules of the mitotic spindle can be seen to terminate at the kinetochore. (*b*) Schematic representation of the kinetochore, which contains an electron-dense inner and outer plate separated by a lightly staining interzone. Proposed functions of the inner and outer plates are indicated in part *a*. The inner plate contains a variety of proteins attached to the centromeric heterochromatin of the chromosome. Associated with the outer plate is the fibrous corona, which binds motor proteins involved in chromosome movement. (*c*) A schematic model showing a proposed disposition of several of the proteins found at the outer surface of the kinetochore. Among the motor proteins associated with the kinetochore, cytoplasmic dynein moves toward the minus end of a microtubule, whereas CENP-E moves toward the plus end. These motors may also play a role in tethering the microtubule to the kinetochore. The protein labeled "depolymerase" is a member of the kinesin superfamily that functions in depolymerization of microtubules rather than motility. In this drawing, the depolymerases are in an inactive state (the microtubule is not depolymerizing). Ndc80 is a protein complex consisting of four different protein subunits that forms a 50 nm-long, rod-shaped molecule extending outward from the body of the kinetochore. These Ndc80 fibrils have been implicated as couplers of the kinetochore to the plus end of a dynamic microtubule. (A: COURTESY OF KEVIN SULLIVAN, FROM DON W. CLEVELAND ET AL., CELL 112:408, 2003; BY PERMISSION OF CELL PRESS.)

Formation of the Mitotic Spindle   We discussed in Chapter 9 how microtubule assembly in animal cells is initiated by a special microtubule-organizing structure, the **centrosome** (page 333). As a cell progresses past G_2 and into mitosis, the microtubules of the cytoskeleton undergo sweeping disassembly in preparation for their reassembly as components of a complex, micro-sized machine called the **mitotic spindle**. The rapid disassembly of the interphase cytoskeleton is thought to be accomplished by the inactivation of proteins that stabilize microtubules (e.g., microtubule-associated proteins, or MAPs) and the activation of proteins that destabilize these polymers.

To understand the formation of the mitotic spindle, we need to first examine the *centrosome cycle* as it progresses in concert with the cell cycle (Figure 14.17*a*). When an animal cell exits mitosis, the cytoplasm contains a single centrosome containing two centrioles situated at right angles to one another. Even before cytokinesis has been completed, the two centrioles of each daughter cell lose their close association to one another (they are said to "disengage"). This event is triggered by the enzyme separase, which becomes activated late in mitosis (page 580). Later, as DNA replication begins in the nucleus at the onset of S phase, each centriole of the centrosome initiates its "replication" in the cytoplasm. The process begins with the appearance of a small *procentriole* next to each preexisting (mater-

nal) centriole and oriented at right angles to it (Figure 14.17*b*). Subsequent microtubule elongation converts each procentriole into a full-length daughter centriole. At the beginning of mitosis, the centrosome splits into two adjacent centrosomes, each containing a pair of mother–daughter centrioles. The initiation of centrosome duplication at the G_1–S transition is normally triggered by phosphorylation of a centrosomal protein by Cdk2, the same agent responsible for the onset of DNA replication (Figure 14.8). Centrosome duplication is a tightly controlled process so that each mother centriole produces only one daughter centriole during each cell cycle. The formation of additional centrioles can lead to abnormal cell division and contribute to the development of cancer (Figure 14.17*c*).

The first stage in the formation of the mitotic spindle in a typical animal cell is the appearance of microtubules in a "sunburst" arrangement, or **aster** (Figure 14.18), around each centrosome during early prophase. As discussed in Chapter 9, microtubules grow by addition of subunits to their plus ends, while their minus ends remain associated with the centrosome. The process of aster formation is followed by separation of the centrosomes from one another and their subsequent movement around the nucleus toward opposite ends of the cell. Centrosome separation is driven by motor proteins associated with the adjacent microtubules. As the centrosomes separate, the micro-

(a)

(b)   0.3 μm

(c)
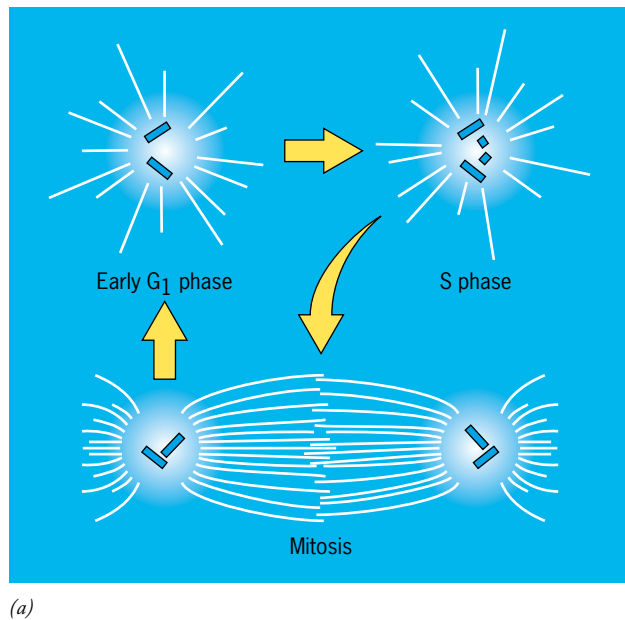
FIGURE 14.17 The centrosome cycle of an animal cell. (*a*) During G$_1$, the centrosome contains a single pair of centrioles that are no longer as tightly associated as they were during mitosis. During S phase, daughter procentrioles form adjacent to maternal centrioles so that two pairs of centrioles become visible within the centrosome (see *b*). The daughter procentrioles continue to elongate during G$_2$ phase, and at the beginning of mitosis, the centrosome splits, with each pair of centrioles becoming part of its own centrosome. As they separate, the centrosomes organize the microtubule fibers that make up the mitotic spindle. (*b*) The centrosome of this cell is seen to contain two pairs of mother–daughter centrioles. The shorter procentrioles are indicated by the arrows. (*c*) This mouse mammary cancer cell contains more than the normal complement of two centrosomes (red) and has assembled a multipolar spindle apparatus (green). Additional centrosomes lead to chromosome missegregation and abnormal numbers of chromosomes (blue), which are characteristic of malignant cells. (A: AFTER D. R. KELLOGG ET AL., REPRODUCED WITH PERMISSION FROM THE ANNUAL REVIEW OF BIOCHEMISTRY, VOL. 63; © 1994, BY ANNUAL REVIEWS INC.; B: FROM J. B. RATTNER AND STEPHANIE G. PHILLIPS, J. CELL BIOL. 57:363, 1973; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS; C: COURTESY OF THEA GOEPFERT AND W. R. BRINKLEY, BAYLOR COLLEGE OF MEDICINE, HOUSTON, TX.)

20 μm

FIGURE 14.18 Formation of the mitotic spindle. During prophase, as the chromosomes are beginning to condense, the centrosomes move apart from one another as they organize the bundles of microtubules that form the mitotic spindle. This micrograph shows a cultured newt lung cell in early prophase that has been stained with fluorescent antibodies against tubulin, which reveals the distribution of the cell's microtubules (green). The microtubules of the developing mitotic spindle are seen to emanate as asters from two sites within the cell. These sites correspond to the locations of the two centrosomes that are moving toward opposite poles at this stage of prophase. The centrosomes are situated above the cell nucleus, which appears as an unstained dark region. (FROM JENNIFER WATERS SHULER, RICHARD W. COLE, AND CONLY L. RIEDER, J. CELL BIOL. 122:364, 1993; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

tubules stretching between them increase in number and elongate (Figure 14.18). Eventually, the two centrosomes reach points opposite one another, thus establishing the two poles of a *bipolar* mitotic spindle (as in Figure 14.17*a*). Following mitosis, one centrosome will be distributed to each daughter cell.

A number of different types of animal cells (including those of the early mouse embryo) lack centrosomes, as do the cells of higher plants, yet all of these cells construct a bipolar mitotic spindle and undergo a relatively typical mitosis. Functional mitotic spindles can even form in mutant *Drosophila* cells that lack centrosomes or in mammalian cells in which the centrosome has been experimentally removed. In all of these cases, the microtubules of the mitotic spindle are nucleated near the chromosomes rather than at the poles where centrosomes would normally reside. Then, once they have polymerized, the minus ends of the microtubules are brought together

FIGURE 14.19 Formation of a spindle pole in the absence of centrosomes. In this model, each motor protein has multiple heads, which are bound to different microtubules. The movement of these motor proteins causes the minus ends of the microtubules to converge to form a distinct spindle pole. This type of mechanism is thought to facilitate the formation of spindle poles in the absence of centrosomes but may also play a role when centrosomes are present. (FROM A. A. HYMAN AND E. KARSENTI, CELL 84:406, 1996; BY PERMISSION OF CELL PRESS.)

(i.e., focused) at each spindle pole through the activity of motor proteins (Figure 14.19). The chapter-opening photograph on page 560 shows a bipolar spindle that has formed in a frog egg extract through the activity of microtubule motors. These types of experiments suggested that cells possess two fundamentally different mechanisms—one centrosome-dependent and the other centrosome-independent—to achieve the same end result. Recent studies have indicated that both pathways to spindle formation operate simultaneously in the same cell and that even cells with functional centrosomes nucleate a significant fraction of their spindle microtubules at the chromosomes.

The Dissolution of the Nuclear Envelope and Partitioning of Cytoplasmic Organelles In most eukaryotic cells, the mitotic spindle is assembled in the cytoplasm and the chromosomes are compacted in the nucleoplasm. Interaction between the spindle and chromosomes is made possible by the breakdown of the nuclear envelope at the end of prophase. The three major components of the nuclear envelope—the nuclear pore complexes, nuclear lamina, and nuclear membranes—are disassembled in separate processes. All of these processes are thought to be initiated by phosphorylation of key substrates by mitotic kinases, particularly cyclin B–Cdk1. The nuclear pore complexes are disassembled as the interactions between nucleoporin complexes are disrupted and the proteins dissociate into the surrounding medium. The nuclear lamina is disassembled by depolymerization of the lamin filaments. The integrity of the nuclear membranes is first disrupted mechanically as holes are torn into the nuclear envelope by cytoplasmic dynein molecules associated with the outer nuclear membrane. The subsequent fate of the membranous portion of the nuclear envelope has been the subject of controversy. According to the classical view, the nuclear membranes are fragmented into a population of small vesicles that disperse throughout the mitotic cell. Alternatively, the membranes of the nuclear envelope may be absorbed into the membranes of the ER.

Some of the membranous organelles of the cytoplasm remain relatively intact through mitosis; these include mitochondria, lysosomes, and peroxisomes, as well as the chloro-plasts of a plant cell. Considerable debate has been generated in recent years over the mechanism by which the Golgi complex and endoplasmic reticulum are partitioned during mitosis. According to one view, the contents of the Golgi complex become incorporated into the ER during prophase, and the Golgi complex ceases to exist briefly as a distinct organelle. According to an alternate view, the Golgi membranes become fragmented to form a distinct population of small vesicles that are partitioned between daughter cells. A third view based primarily on studies in algae and protists has the entire Golgi complex splitting in two, with each daughter cell receiving half of the original structure. Ultimately, we may learn that different types of cells or organisms utilize different mechanisms of Golgi inheritance. Our ideas about the fate of the ER have also changed. Recent studies on living, cultured mammalian cells suggest that the ER network remains relatively intact during mitosis. This view challenges earlier studies performed largely on eggs and embryos that suggested the ER undergoes extensive fragmentation during prophase.

Prometaphase

The dissolution of the nuclear envelope marks the start of the second phase of mitosis, **prometaphase**, during which mitotic spindle assembly is completed and the chromosomes are moved into position at the center of the cell. The following discussion provides a generalized picture of the steps of prometaphase; many variations on these events have been reported.

At the beginning of prometaphase, compacted chromosomes are scattered throughout the space that was the nuclear region (Figure 14.20). As the microtubules of the spindle penetrate into the central region of the cell, the free (plus) ends of the microtubules are seen to grow and shrink in a dynamic fashion, as if they were "searching" for a chromosome. It is not
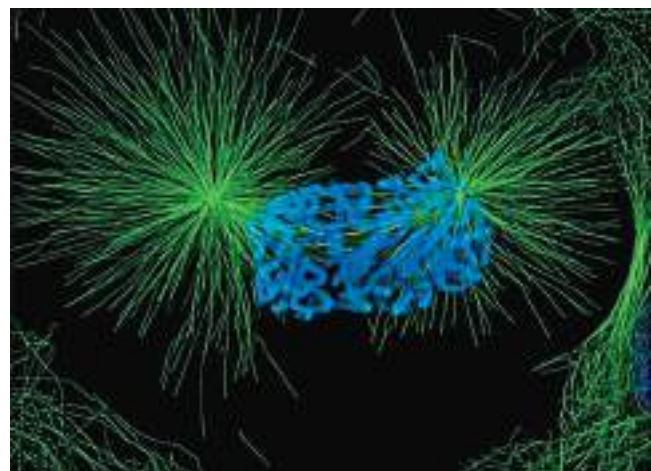
FIGURE 14.20 Prometaphase. Fluorescence micrograph of a cultured newt lung cell at the early prometaphase stage of mitosis, just after the nuclear envelope has broken. The microtubules of the mitotic spindle are now able to interact with the chromosomes. The mitotic spindle appears green after labeling with a monoclonal antibody against tubulin, whereas the chromosomes appear blue after labeling with a fluorescent dye. (COURTESY OF ALEXEY KHODJAKOV.)

certain whether searching is entirely random, as evidence suggests that microtubules may grow preferentially toward a site containing chromatin. Those microtubules that contact a kinetochore are "captured" and stabilized.

A kinetochore typically makes initial contact with the sidewall of a microtubule rather than its end. Once initial contact is made, some chromosomes move actively along the wall of the microtubule, powered by motor proteins located in the kinetochore. Soon, however, the kinetochore tends to become stably associated with the plus end of one or more spindle microtubules from one of the spindle poles. Eventually, the unattached kinetochore on the sister chromatid captures its own microtubules from the opposite spindle pole. It has also been reported that unattached kinetochores may serve as nucleating sites for the assembly of microtubules that grow toward the opposite spindle pole. Regardless of how it occurs, the two sister chromatids of each mitotic chromosome ultimately become connected by their kinetochores to microtubules that extend from opposite poles.

Observations in living cells indicate that prometaphase chromosomes associated with spindle microtubules are not moved directly to the center of the spindle but rather oscillate back and forth in both a poleward and antipoleward direction. Ultimately, the chromosomes of a prometaphase cell are moved by a process called **congression** toward the center of the mitotic spindle, midway between the poles. The forces required for chromosome movements during prometaphase are generated by motor proteins associated with both the kinetochores and arms of the chromosomes (depicted in Figure 14.33a and discussed in the legend). Figure 14.21 shows the consequences of a deficiency of a chromosomal motor protein whose activity pushes chromosomes away from the poles.

Microtubule dynamics also play a key role in facilitating chromosome movements during prometaphase. As the chromosomes congress toward the center of the mitotic spindle, the longer microtubules attached to one kinetochore are shortened, while the shorter microtubules attached to the sister kinetochore are elongated. These changes in microtubule length are thought to be governed by differences in pulling force (tension) on the two sister kinetochores. Shortening and elongation of microtubules occur primarily by loss or gain of subunits at the plus end of the microtubule (Figure 14.22). Remarkably, this dynamic activity occurs while the plus end of each microtubule remains attached to a kinetochore.

FIGURE 14.21 **The consequence of a missing motor protein on chromosome alignment during prometaphase.** The top micrograph shows a mitotic spindle that has assembled in a complete frog egg extract. The lower micrograph shows a mitotic spindle that has assembled in a frog egg extract that has been depleted of a particular kinesin-like protein called Kid that is present along the arms of prometaphase chromosomes. In the absence of this motor protein, the chromosomes fail to align at the center of the spindle and instead are found stretched along spindle microtubules and clustered near the poles. Kid normally provides force for moving chromosomes away from the poles (see Figure 14.33a). (FROM CELIA ANTONIO ET AL., COURTESY OF ISABELLE VERNOS, CELL VOL. 102, COVER #4, 2000; BY PERMISSION OF CELL PRESS.)

FIGURE 14.22 **Microtubule behavior during formation of the metaphase plate.** Initially, the chromosome is connected to microtubules from opposite poles that may be very different in length. As prometaphase continues, this imbalance is corrected as the result of the shortening of microtubules from one pole, due to the rapid loss of tubulin subunits at the kinetochore, and the lengthening of microtubules from the opposite pole, due to the rapid addition of tubulin subunits at the kinetochore. These changes are superimposed over a much slower polymerization and depolymerization (lower drawing) that occur continually during prometaphase and metaphase, causing the subunits of the microtubule to move toward the poles in a process known as microtubule flux.

10 μm
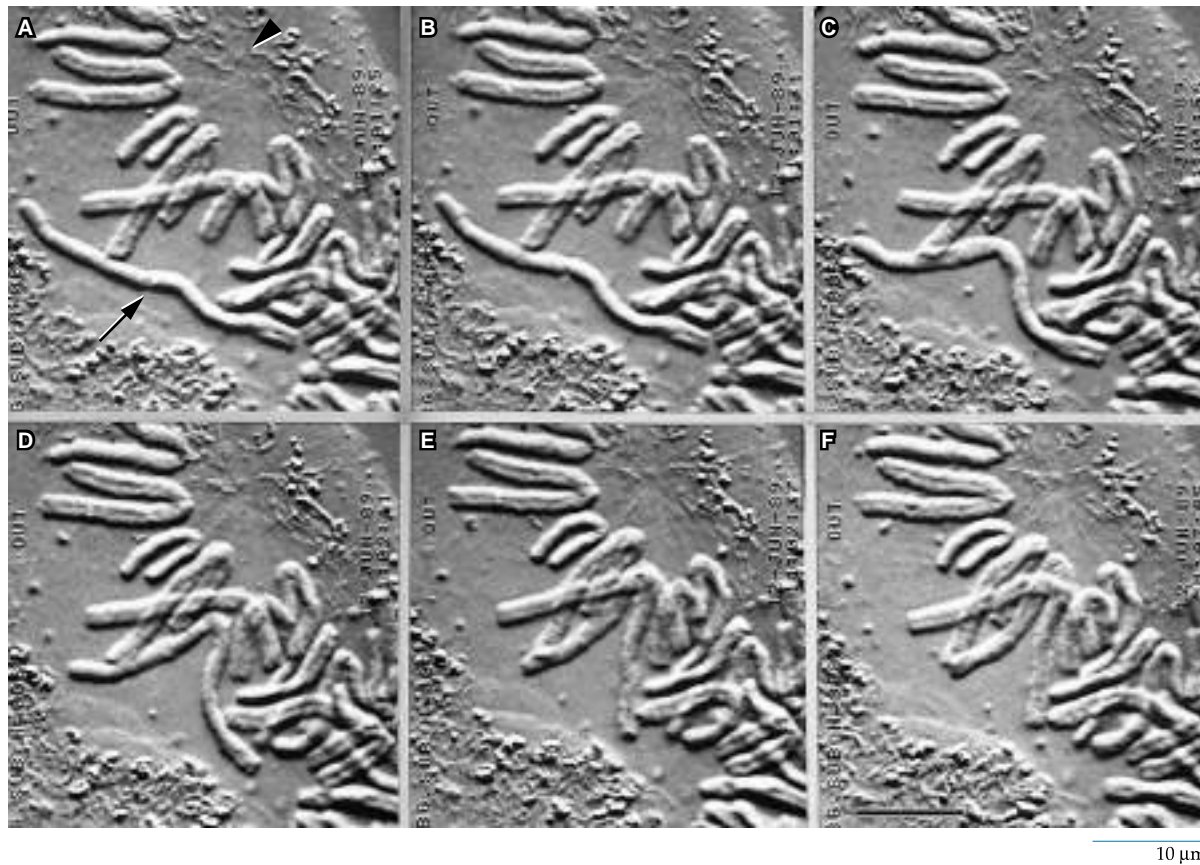
FIGURE 14.23   The engagement of a chromosome during prometaphase and its movement to the metaphase plate. This series of photographs taken from a video recording shows the movements of the chromosomes of a newt lung cell over a period of 100 seconds during prometaphase. Although most of the cell's chromosomes were nearly aligned at the metaphase plate at the beginning of the sequence, one of the chromosomes (arrow) had failed to become attached to spindle fibers from both poles. The wayward chromosome has become attached to spindle fibers from opposite poles in B and then moves toward the spindle equator with variable velocity until it reaches a stable position in F. The position of one pole is indicated by the arrowhead in A. (FROM STEPHEN P. ALEXANDER AND CONLY L. RIEDER, J. CELL BIOL. 113:807, 1991; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

Eventually, each chromosome moves into position along a plane at the center of the spindle, so that microtubules from each pole are equivalent in length. The movement of a wayward chromosome from a peripheral site near one of the poles to the center of the mitotic spindle during prometaphase is shown in the series of photos in Figure 14.23.

Metaphase

Once all of the chromosomes have become aligned at the spindle equator—with one chromatid of each chromosome connected by its kinetochore to microtubules from one pole and its sister chromatid connected by its kinetochore to microtubules from the opposite pole—the cell has reached the stage of **metaphase** (Figure 14.24). The plane of alignment of the chromosomes at metaphase is referred to as the *metaphase plate*. The mitotic spindle of the metaphase cell contains a highly organized array of microtubules that is ideally suited for the task of separating the duplicated chromatids positioned at the center of the cell. Functionally and spatially, the microtubules of the metaphase spindle of an animal cell can be divided into three groups (Figure 14.24):

1. *Astral microtubules* that radiate outward from the centrosome into the region outside the body of the spindle. They help position the spindle apparatus in the cell and may help determine the plane of cytokinesis.

2. *Chromosomal* (or *kinetochore*) *microtubules* that extend between the centrosome and the kinetochores of the chromosomes. In mammalian cells, each kinetochore is attached to a bundle of 20–30 microtubules, which forms a spindle *fiber*. During metaphase, the chromosomal microtubules exert a pulling force on the kinetochores. As a result, the chromosomes are maintained in the equatorial plane by a "tug-of-war" between balanced pulling forces exerted by chromosomal spindle fibers from opposite poles. During anaphase, chromosomal microtubules are required for the movement of the chromosomes toward the poles.

3. *Polar* (or *interpolar*) *microtubules* that extend from the centrosome past the chromosomes. Polar microtubules from one centrosome overlap with their counterparts from the opposite centrosome. The polar microtubules form a structural basket that maintains the mechanical integrity of the spindle.
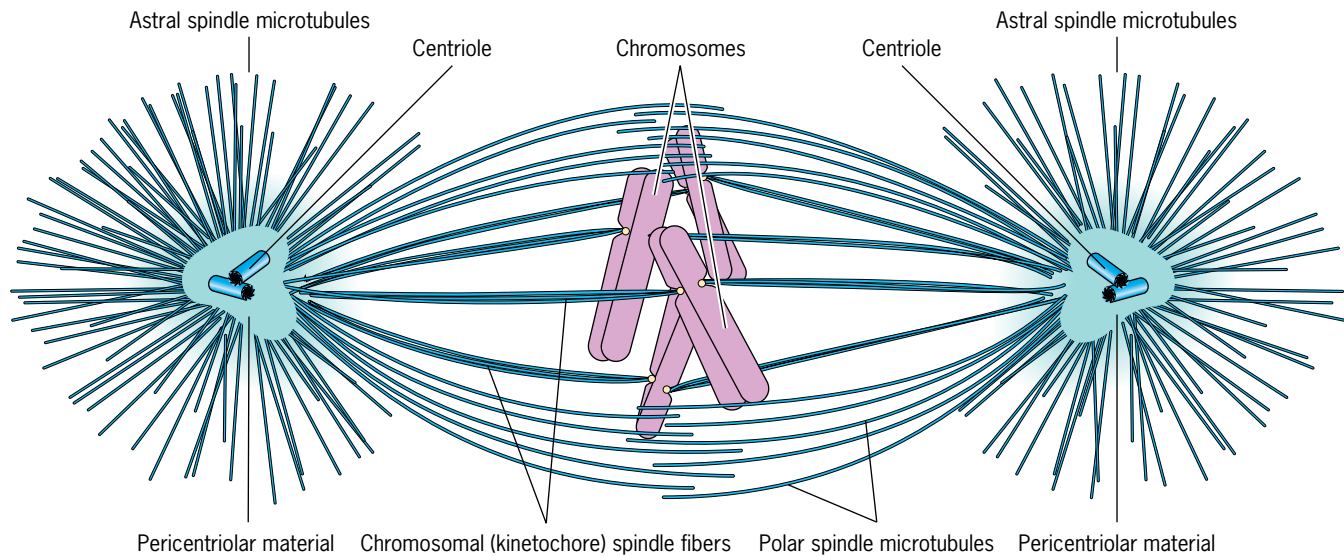
FIGURE 14.24 **The mitotic spindle of an animal cell.** Each spindle pole contains a pair of centrioles surrounded by amorphous pericentriolar material at which the microtubules are nucleated. Three types of spindle microtubules—astral, chromosomal, and polar spindle microtubules— are evident, and their functions are described in the text. All of the spindle microtubules, which can number in the thousands, have their minus ends pointed toward the centrosome.

As one watches films or videos of mitosis, metaphase appears as a stage during which the cell pauses for a brief period, as if all mitotic activities suddenly come to a halt. However, experimental analysis reveals that metaphase is a time when important events occur.

Microtubule Flux in the Metaphase Spindle Even though there is no obvious change in length of the chromosomal microtubules as the chromosomes are aligned at the metaphase plate, studies using fluorescently labeled tubulin indicate that the microtubules exist in a highly dynamic state. Subunits are rapidly lost and added at the plus ends of the chromosomal microtubules, even though these ends are attached to the kinetochore. Thus, the kinetochore does not act like a cap at the end of the microtubule, blocking the entry or exit of terminal subunits, but rather it is the site of dynamic activity. Because more subunits are added to the plus end than are lost, there is a net addition of subunits at the kinetochore. Meanwhile, the minus ends of the microtubules experience a net loss, and thus subunits move along the chromosomal microtubules from the kinetochore toward the pole. This *poleward flux* of tubulin subunits in a mitotic spindle is indicated in the experiment illustrated in Figure 14.25. Loss of tubulin subunits at the poles is likely aided by a member of the kinesin-13 family of motor proteins whose function is to promote microtubule depolymerization rather than movement (page 330).

Anaphase

Anaphase begins when the sister chromatids of each chromosome split apart and start their movement toward opposite poles.
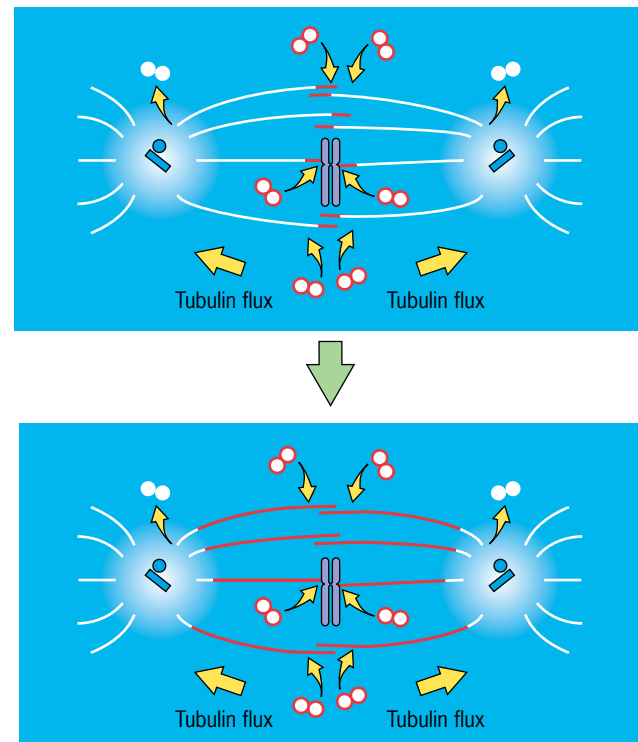
FIGURE 14.25 **Tubulin flux through the microtubules of the mitotic spindle at metaphase.** Even though the microtubules appear stationary at this stage, injection of fluorescently labeled tubulin subunits indicates that the components of the spindle are in a dynamic state of flux. Subunits are incorporated preferentially at the kinetochores of the chromosomal microtubules and the equatorial ends of the polar microtubules, and they are lost preferentially from the minus ends of the microtubules in the region of the poles. Tubulin subunits move through the microtubules of a metaphase spindle at a rate of about 1 μm/min.

The Role of Proteolysis in Progression Through Mitosis

Genetic and biochemical approaches have produced a large body of information on the mechanism responsible for initiation of anaphase. It was pointed out earlier that two distinct multiprotein complexes, SCF and APC, add ubiquitin to proteins at different stages of the cell cycle, targeting them for destruction by a proteasome. The periods during the cell cycle in which the SCF and APC complexes are active is shown in Figure 14.26a. As illustrated in Figure 14.26a, SCF acts primarily during interphase. In contrast, the *anaphase promoting complex*, or *APC*, plays a key role in regulating events that occur during mitosis. The APC contains about a dozen core subunits, in addition to an "adaptor protein" that plays a key role in determining which proteins serve as the APC substrate. Two alternate versions of this adaptor protein—Cdc20 and Cdh1—play an important role in substrate selection during mitosis. APC complexes containing one or the other of these adaptors are known as APC^{Cdc20} or APC^{Cdh1} (Figure 14.26a).

APC^{Cdc20} becomes activated prior to metaphase (Figure 14.26a) and ubiquitinates a major anaphase inhibitor called *securin*—so named because it secures the attachment between sister chromatids. The ubiquitination and destruction of securin at the end of metaphase release an active protease called *separase*. Separase then cleaves a key subunit of the cohesin molecules that hold sister chromatids together (Figure 14.26b). Cleavage of cohesin triggers the separation of sister chromatids to mark the onset of anaphase.

Near the end of mitosis, Cdc20 is inactivated, and the alternate adaptor, Cdh1, takes control of the APC's substrate selection (Figure 14.26a). When Cdh1 is associated with the APC, the enzyme completes the ubiquitination of cyclin B. Destruction of the cyclin leads to a precipitous drop in activity of the mitotic Cdk (cyclin B–Cdk1) and progression of the cell out of mitosis and into the G_1 phase of the next cell cycle. The importance of protein degradation in regulating the events of mitosis and the reentry of cells into G_1 is best revealed with the use of inhibitors (Figure 14.27). If the destruction of cyclin B is prevented with an inhibitor of the proteasome, cells remain arrested in a late stage of mitosis. If such cells that are arrested in mitosis are subsequently treated with a compound that inhibits the kinase activity of Cdk1, the cell will return to its normal activities and continue through mitosis and cytokinesis. The completion of mitosis clearly requires the cessation of activity of Cdk1 (either by the normal destruction of its cyclin activator or by experimental inhibition). Perhaps the most striking finding of all is obtained when the proteasome-inhibited and Cdk1-inhibited cells that have already exited mitosis are washed free of the Cdk1 inhibitor (Figure 14.27). Such cells, which now contain both cyclin B and active Cdk1, actually progress in the reverse direction and head back into mitosis. This reversal is characterized by compaction of the chromosomes, breakdown of the nuclear envelope, assembly of a mitotic spindle, and movement of the chromosomes back to the metaphase plate, as shown in Figure 14.27. All of these events are triggered by the inappropriate reactivation of Cdk1 activity by removal of its inhibitor. This finding dramatically illustrates the importance of proteolysis in moving the normal cell cycle in a single, irreversible direction.

(a)

(b)
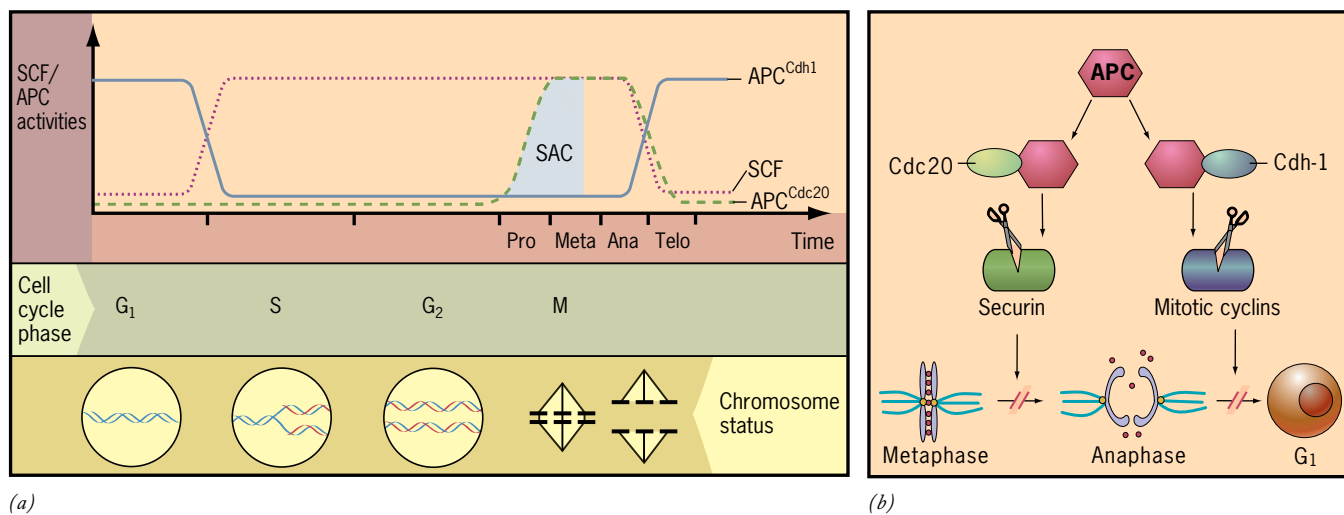
FIGURE 14.26 SCF and APC activities during the cell cycle. SCF and APC are multisubunit complexes that ubiquitinate substrates, leading to their destruction by proteasomes. (a) SCF is active primarily during interphase, whereas APC (anaphase promoting complex) is active during mitosis and G_1. Two different versions of APC are indicated. These two APCs differ in containing either a Cdc20 or a Cdh1 adaptor protein, which alters the substrates recognized by the APC. APC^{Cdc20} is active earlier in mitosis than is APC^{Cdh1}. The label SAC stands for spindle assembly checkpoint, which is discussed on page 584. The SAC prevents APC^{Cdc20} from triggering anaphase until all the chromosomes are properly aligned at the metaphase plate. (b) APC^{Cdc20} is responsible for destroying proteins, such as securin, that inhibit anaphase. Destruction of these substrates promotes the metaphase–anaphase transition. APC^{Cdh1} is responsible for ubiquitinating proteins, such as mitotic cyclins, that inhibit exit from mitosis. Destruction of these substrates promotes the mitosis–G_1 transition. APC^{Cdh1} activity during early G_1 helps maintain the low cyclin–Cdk activity (Figure 14.8) required to assemble prereplication complexes at the origins of replication (Figure 13.20). (A: AFTER J.-M. PETERS, CURR. OPIN. CELL BIOL. 10:762, 1998; COPYRIGHT 1998, WITH PERMISSION FROM ELSEVIER SCIENCE. SEE ALSO NATURE REVS. MOL. CELL BIOL. 7:650, 2006.)
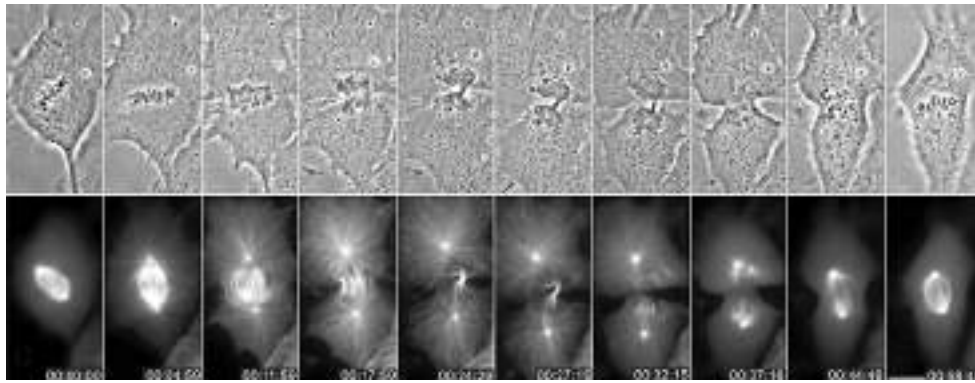
FIGURE 14.27 Experimental demonstration of the importance of proteolysis in a cell's irreversible exit from mitosis. This illustration shows frames from a video of a cell that had been arrested in mitosis by the presence of a proteasome inhibitor (MG132). At time 0, a Cdk1 inhibitor (flavopiridol) was added to the medium, which caused the cell to complete mitosis and initiate cytokinesis. At 25 min, the cell was washed free of the Cdk1 inhibitor. Because the cell still contained cyclin B (which would normally have been destroyed by proteasomes), the cell reentered mitosis and progressed to metaphase as seen in the last five frames of the video. The upper row shows phase-contrast images of the cell at various times, and the lower row shows the corresponding fluorescence micrographs with times indicated. Video 3, from which these frames were taken, can be seen at the online version of Tamara A. Potapova et al. *Nature* 440:954, 2006. Bar in lower right image equals 10 μm. (COURTESY OF GARY J. GORBSKY, © COPYRIGHT 2006, MACMILLAN MAGAZINES LIMITED.)

The Events of Anaphase   All the chromosomes of the metaphase plate are split in synchrony at the onset of anaphase, and the chromatids (now referred to as chromosomes, because they are no longer attached to their sisters) begin their poleward migration (see Figure 14.11). As the chromosome moves during anaphase, its centromere is seen at its leading edge with the arms of the chromosome trailing behind (Figure 14.28*a*). The movement of chromosomes toward opposite poles is very slow relative to other types of cellular movements, proceeding at approximately 1 μm per minute, a

(a)

(b)
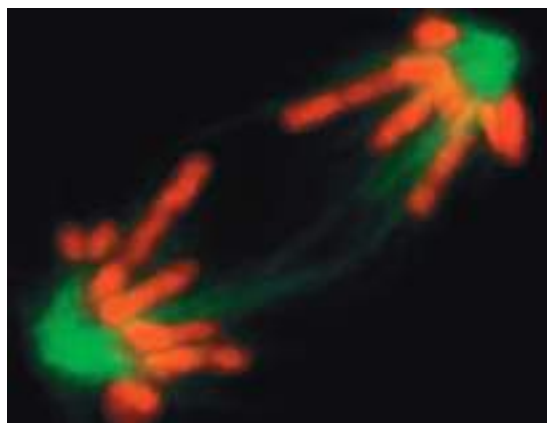
FIGURE 14.28 The mitotic spindle and chromosomes at anaphase. (*a*) Fluorescence micrograph of a cell at late anaphase, showing the highly compacted arms of the chromosomes at this stage. The arms are seen to trail behind as the kinetochores lead the way toward the respective poles. The chromosomal spindle fibers at this late anaphase stage are extremely short and are no longer evident between the forward edges of the chromosomes and the poles. The polar spindle microtubules, however, are clearly visible in the interzone between the separating chromosomes. Relative movements of the polar microtubules are thought to be responsible for causing the separation of the poles that occurs during anaphase B. (*b*) Microtubule dynamics during anaphase. Tubulin subunits are lost from both ends of the chromosomal microtubules, resulting in shortening of chromosomal fibers and movement of the chromosomes toward the poles during anaphase A. Meanwhile, tubulin subunits are added to the plus ends of polar microtubules, which also slide across one another, leading to separation of the poles during anaphase B. (A: FROM FELIPE MORA-BERMÚDAEZ, DANIEL GERLICH, AND JAN ELLENBERG, COVER OF NATURE CELL BIOL. 9, #7, 2007; © COPYRIGHT 2007, MACMILLAN MAGAZINES LIMITED.)
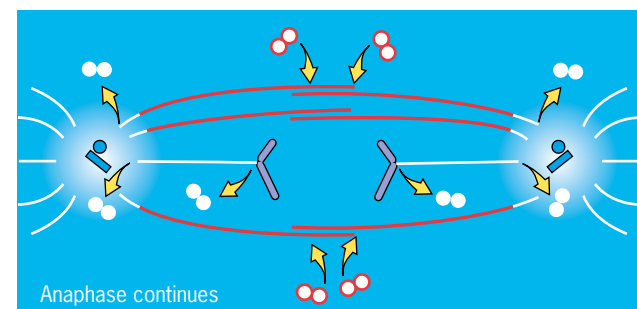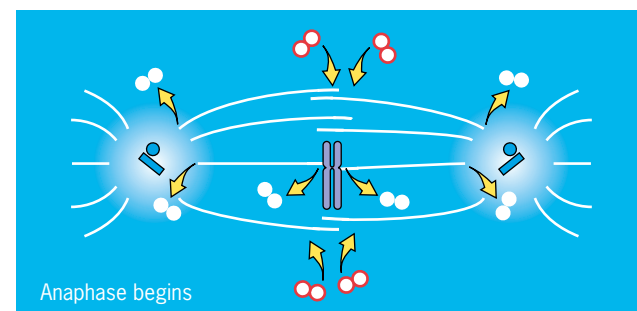
value calculated by one mitosis researcher to be equivalent to a trip from North Carolina to Italy that would take approximately 14 million years. The slow rate of chromosome movement ensures that the chromosomes segregate accurately and without entanglement. The forces thought to power chromosome movement during anaphase are discussed in the following section.

The poleward movement of chromosomes is accompanied by obvious shortening of chromosomal microtubules. It has long been appreciated that tubulin subunits are lost from the plus (kinetochore-based) ends of chromosomal microtubules during anaphase (Figure 14.28b). Subunits are also lost from the minus ends of these microtubules as a result of the continued poleward flux of tubulin subunits that occurs during prometaphase and metaphase (Figures 14.22 and 14.25). The primary difference in microtubule dynamics between metaphase and anaphase is that subunits are added to the plus ends of microtubules during metaphase, keeping the length of the chromosomal fibers constant (Figure 14.25), whereas subunits are lost from the plus ends during anaphase, resulting in shortening of the chromosomal fibers (Figure 14.28b). This change in behavior at the microtubule plus ends is thought to be triggered by a change in tension on the kinetochores following separation of the sister chromatids.

The movement of the chromosomes toward the poles is referred to as **anaphase A** to distinguish it from a separate but simultaneous movement, called **anaphase B**, in which the two spindle poles move farther apart. The elongation of the mitotic spindle during anaphase B is accompanied by the net addition of tubulin subunits to the plus ends of the polar microtubules. Thus, subunits can be preferentially added to polar microtubules and removed from chromosomal microtubules at the same time in different regions of the same mitotic spindle (Figure 14.28b).

Forces Required for Chromosome Movements at Anaphase   In the early 1960s, Shinya Inoué of the Marine Biological Laboratory at Woods Hole proposed that the depolymerization of chromosomal microtubules during anaphase was not simply a consequence of chromosome movement but the cause of it. Inoué suggested that depolymerization of the microtubules that comprise a spindle fiber could generate sufficient mechanical force to pull a chromosome forward.

Early experimental support for the disassembly-force model came from studies in which chromosomes underwent considerable movement as the result of the depolymerization of attached microtubules. An example of one of these experiments is shown in Figure 14.29. In this case, the movement of a microtubule-bound chromosome (arrow) occurs in vitro following dilution of the medium. Dilution reduces the concentration of soluble tubulin, which in turn promotes the depolymerization of the microtubules. These types of experiments, as well as direct force-measuring studies, indicate that microtubule depolymerization alone can generate sufficient force to pull chromosomes through a cell.

Figure 14.30a depicts a model of the events proposed to occur during chromosome movement at anaphase in a budding yeast cell. As indicated in Figure 14.28b, the microtubules that comprise the chromosomal spindle fibers undergo depolymerization at both their minus and plus ends during anaphase. This is true in both yeast and multicellular eukaryotes. These combined activities lead to the movement of chromosomes toward the pole. Depolymerization at the microtubule minus ends serves to transport the chromosomes toward the poles due to poleward flux, reminiscent of a person standing on a "moving walkway" in an airport. In contrast, depolymerization at the microtubule plus ends serves to "chew up" the fiber that is towing the chromosomes. Some cells rely more on poleward flux, others more on plus-end depolymer-
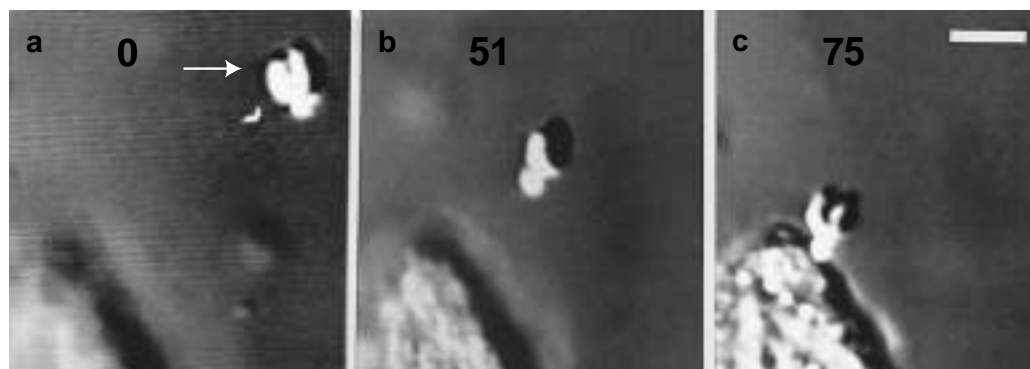
FIGURE 14.29 Experimental demonstration that microtubule depolymerization can move attached chromosomes in vitro. The structure at the lower left is the remnant of a lysed protozoan. In the presence of tubulin, the basal bodies at the surface of the protozoan were used as sites for the initiation of microtubules, which grew outward into the medium. Once the microtubules had formed, condensed mitotic chromosomes were introduced into the chamber and allowed to bind to the ends of the microtubules. The arrow shows a chromosome attached to the end of a bundle of microtubules. The concentration of soluble tubulin within the chamber was then decreased by dilution, causing the depolymerization of the microtubules. As shown in this video sequence, the shrinkage of the microtubules was accompanied by the movement of the attached chromosome. Bar equals 5 μm. (FROM MARTINE COUE, VIVIAN A. LOMBILLO, AND J. RICHARD MCINTOSH, J. CELL BIOL. 112:1169, 1991; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

(a)
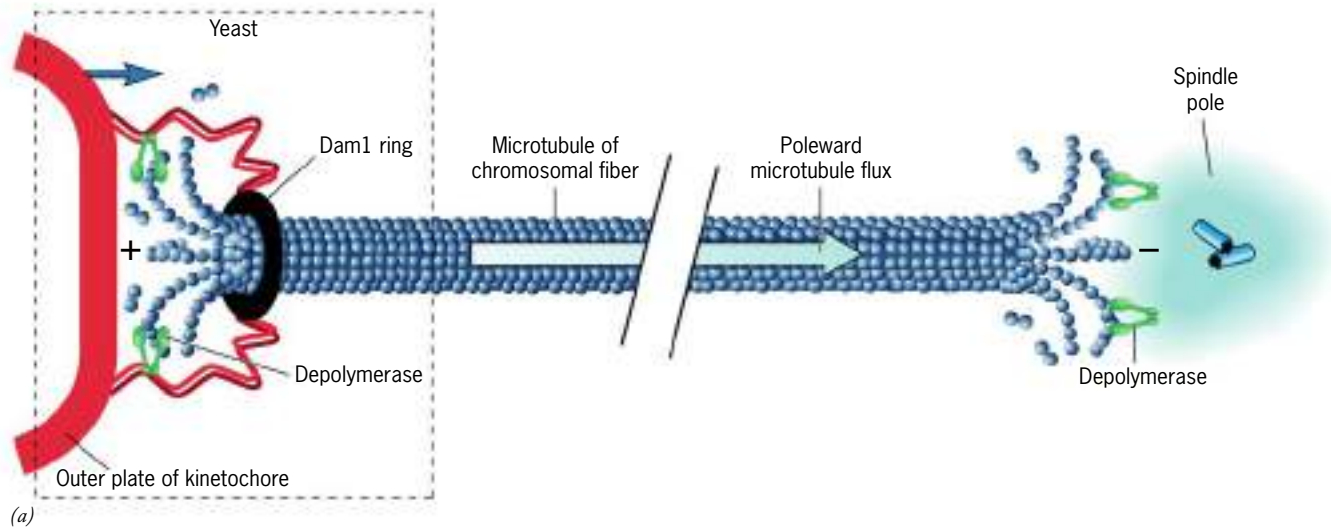
FIGURE 14.30 Proposed mechanism for the movement of chromosomes during anaphase in budding yeast and animal cells. In the model depicted here, chromosome movement toward the poles is accomplished by a combination of poleward flux, which moves the body of each microtubule toward one of the poles, and simultaneous depolymerization of the microtubule at both ends. Depolymerizing kinesins of the kinesin-13 family have been localized at both the plus (kinetochore) and minus (polar) ends of chromosomal microtubules and are postulated to be responsible for depolymerization at their respective sites. (*a*) In this model for budding yeast, the chromosome is able to remain associated with the plus end of the microtubule as it depolymerizes by the presence of the Dam1 ring, which encircles the plus end of the microtubule at the kinetochore. The force required for chromosome movement is provided by the release of strain energy as the microtubule depolymerizes. The released energy is utilized by the curled ends of the depolymerizing protofilaments to slide the Dam1 ring along the microtubule toward the pole. (*b*) Electron micrograph of a negatively stained microtubule encircled by a ring of protein consisting of the yeast Dam1 complex, which is made up of 10 different polypeptides. The ring shown here had assembled around the microtubule during an in vitro incubation with the purified Dam1 subunits. (*c*) Dam1 proteins have not been found in animal cells, and there is no indication that a comparable ring-like structure acts to couple kinetochores to depolymerizing microtubules. In this drawing, the Dam1 ring seen in the boxed region of part *a* is replaced by a different type of proposed coupling device, the Ndc80 protein complex of the outer kinetochore plate. (B: FROM STEFAN WESTERMANN ET AL, NATURE 440:434, 2006, COURTESY OF GEORJANA BARNES; © COPYRIGHT 2006, MACMILLAN MAGAZINES LIMITED.)
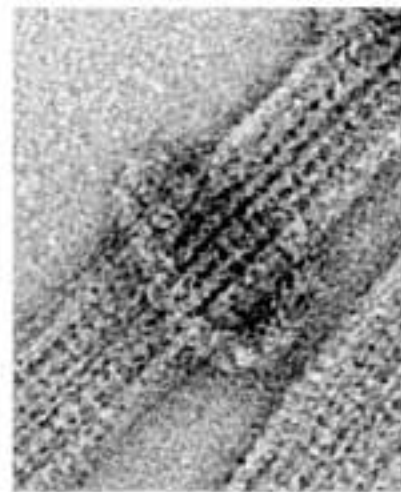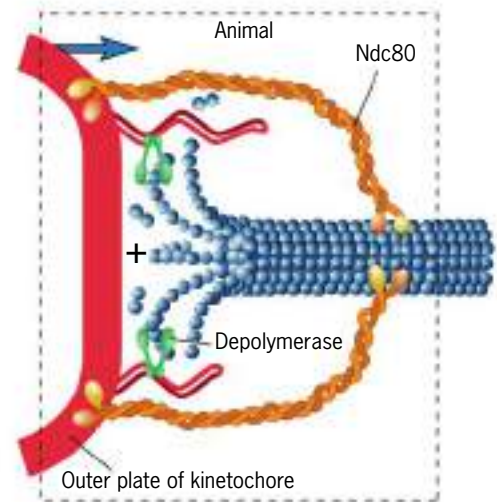
ization. Studies of animal cells in anaphase have revealed that both the plus and minus ends of chromosomal fibers are sites where depolymerizing kinesins (members of the kinesin-13 family, page 330) are localized. These depolymerases are indicated at the opposite ends of the microtubule in Figure 14.30*a*. If either of these microtubule "depolymerases" is specifically inhibited, chromosome segregation during anaphase is at least partially disrupted. These findings suggest that ATP-dependent, kinesin-mediated depolymerization forms the basis for chromosome segregation during mitosis.

It was mentioned on page 573 that one of the questions of greatest interest in the field of mitosis is the mechanism by which kinetochores are able to hold on to the plus ends of mi-

crotubules that are losing tubulin subunits. This question has been most closely studied in budding yeast, as depicted in Figure 14.30*a*, where each kinetochore is thought to contain a ring-shaped protein complex called Dam1 whose inner diameter of 32 nm is large enough to comfortably surround a microtubule. In the presence of microtubules, Dam1 assembles to form rings and helices that encircle the microtubules (Figure 14.30*b*). If these encircled microtubules are induced to depolymerize in vitro, the rings are seen to slide along the microtubules for distances of several μms, just behind the depolymerizing tip.

A proposed role of the Dam1 ring in anaphase is illustrated in Figure 14.30*a*. According to this model, the energy released by the protofilaments as they peel away from the mi-

crotubule (page 337) is utilized to push the Dam1 ring toward the opposite end of the microtubule. Because the ring is attached to the kinetochore of the anaphase chromosome, the entire chromosome is moved toward the spindle pole. But there are problems with this model. Vertebrates lack close homologues of the Dam1 proteins, and no evidence for the presence of Dam1-like rings is seen in high-resolution electron microscopic studies of the kinetochores of dividing cells. Such studies do, however, identify other kinetochore components as potential microtubule couplers. As discussed on page 573, the Ndc80 complexes of the kinetochore are seen as fibrils that reach out to make contact with the plus ends of the microtubule. If these Ndc80 complexes were able to move along the microtubule as it shortens, as suggested in Figure 14.30c, these complexes could serve a similar role to that proposed for the Dam1 rings of yeast. Other proteins present at the kinetochores have also been suggested as the elusive coupling device.

The Spindle Assembly Checkpoint

As discussed on page 567, cells possess checkpoint mechanisms that monitor the status of events during the cell cycle. One of these checkpoints operates at the transition between metaphase and anaphase. The **spindle assembly checkpoint** (**SAC**), as it is called, is best revealed when a chromosome fails to become aligned properly at the metaphase plate. When this happens, the checkpoint mechanism delays the onset of anaphase until the misplaced chromosome has assumed its proper position along the spindle equator. If a cell were not able to postpone chromosome segregation, it would greatly elevate the risk of the daughter cells receiving an abnormal number of chromosomes (aneuploidy). This expectation has been confirmed with the identification of a number of children with inherited deficiencies in one of the spindle checkpoint proteins. These individuals exhibit a disorder (named MVA), which is characterized by a high percentage of aneuploid cells and a greatly increased risk of developing cancer.

How does the cell determine whether or not all of the chromosomes are properly aligned at the metaphase plate? Let's consider a chromosome that is only attached to microtubules from one spindle pole, which is probably the circumstance of the wayward chromosome indicated by the arrow in Figure 14.23a. Unattached kinetochores contain a complex of proteins, the best studied of which is called Mad2, that mediate the spindle assembly checkpoint. The presence of these proteins at an *unattached* kinetochore sends a "wait" signal to the cell cycle machinery that prevents the cell from continuing on into anaphase. Once the wayward chromosome becomes attached to spindle fibers from both spindle poles and becomes properly aligned at the metaphase plate, the signaling complex leaves the kinetochore, which turns off the "wait" signal and allows the cell to progress into anaphase.

Figure 14.31 shows the mitotic spindle of a cell that is arrested prior to metaphase due to a single unaligned chromosome. Unlike all of the other kinetochores in this cell, only the unaligned chromosome is seen to still contain the Mad2 protein. As long as the cell contains unaligned chromosomes,
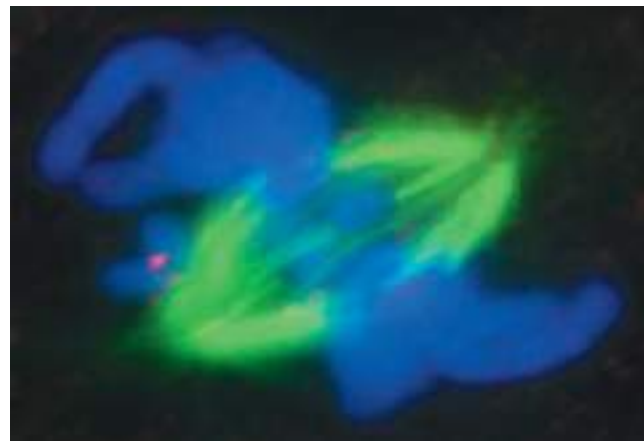
FIGURE 14.31 **The spindle assembly checkpoint.** Fluorescence micrograph of a mammalian cell in late prometaphase labeled with antibodies against the spindle checkpoint protein Mad2 (pink) and tubulin of the microtubules (green). The chromosomes appear blue. Only one of the chromosomes of this cell is seen to contain Mad2, and this chromosome has not yet become aligned at the metaphase plate. The presence of Mad2 on the kinetochore of this chromosome is sufficient to prevent the initiation of anaphase. (FROM JENNIFER WATERS SHULER, REY-HUEI CHEN, ANDREW W. MURRAY, AND E. D. SALMON, J. CELL BIOL. 141, COVER #5, 1998; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

Mad2 molecules are able to inhibit cell cycle progress. According to a favored model, inhibition is achieved through direct interaction between Mad2 and the APC activator Cdc20. During the period that Cdc20 is bound to Mad2, APC complexes would be unable to ubiquitinate the anaphase inhibitor securin, thus keeping all of the sister chromatids attached to one another by their cohesin "glue."

It is well established that the spindle assembly checkpoint is activated by the presence of an unattached kinetochore, but there are other chromosomal abnormalities that arise during the progression to metaphase that also require corrective measures. For example, on occasion, the two kinetochores of sister chromatids will become attached to microtubules from the same spindle pole, a condition referred to as a syntelic attachment. If not corrected, a syntelic attachment is very likely to lead to the movement of both sister chromatids to one of the daughter cells, leaving the other daughter devoid of this chromosome.

Cells are able to correct syntelic attachments (and other types of abnormal microtubule connections) through the action of an enzyme called Aurora B kinase, which is part of a mobile protein complex that resides at the kinetochores during prometaphase and metaphase. Among the substrates of Aurora B kinase are several of the proteins thought to be involved in kinetochore–microtubule attachment, including members of both the Dam1 complex and the Ndc80 complex and the kinesin depolymerase of Figure 14.30. Studies suggest that Aurora B kinase molecules of an incorrectly attached chromosome phosphorylate these protein substrates, which destabilizes microtubule attachment to both kinetochores.

Once freed of their bonds, the kinetochores of each sister chromatid have a fresh opportunity to become attached to microtubules from opposite spindle poles. Inhibition of Aurora B kinase in cells or extracts leads to misalignment and missegregation of chromosomes (see Figure 18.51).

Telophase

As the chromosomes near their respective poles, they tend to collect in a mass, which marks the beginning of the final stage of mitosis, or **telophase** (Figures 14.11 and 14.32). During telophase, daughter cells return to the interphase condition: the mitotic spindle disassembles, the nuclear envelope reforms, and the chromosomes become more and more dispersed until they disappear from view under the microscope. The actual partitioning of the cytoplasm into two daughter cells occurs by a process to be discussed shortly. First, however, let's look back and consider the forces required for several of the major chromosome movements that take place during mitosis.

Nuclear envelope reforming

Furrow (Cytokinesis)

2 μm
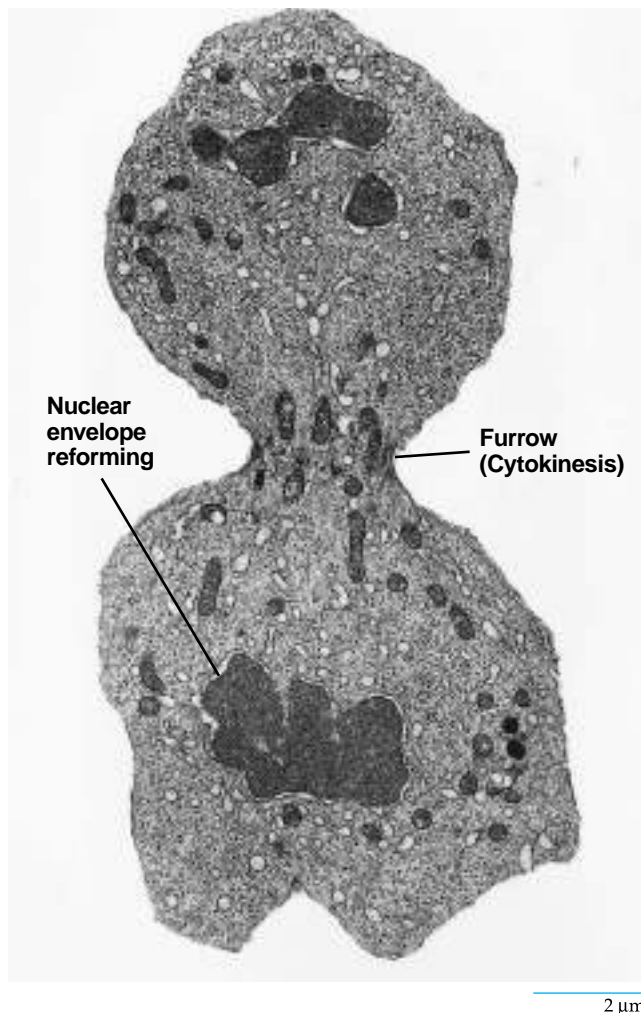
FIGURE 14.32 **Telophase.** Electron micrograph of a section through an ovarian granulosa cell in telophase. (FROM J. A. RHODIN, HISTOLOGY, OXFORD, 1974.)

Forces Required for Mitotic Movements

Mitosis is characterized by extensive movements of cellular structures. Prophase is accompanied by movement of the spindle poles to opposite ends of the cell, prometaphase by movement of the chromosomes to the spindle equator, anaphase A by movement of the chromosomes from the spindle equator to its poles, and anaphase B by the elongation of the spindle. Over the past decade, a variety of different molecular motors have been identified in different locations in mitotic cells of widely diverse species. The motors involved in mitotic movements are primarily microtubule motors, including a number of different kinesin-related proteins and cytoplasmic dynein. Some of the motors move toward the plus end of the microtubule, others toward the minus end. As discussed above, one group of kinesins does not move anywhere, but promotes microtubule depolymerization. Motor proteins have been localized at the spindle poles, along the spindle fibers, and within both the kinetochores and arms of the chromosomes.

The roles of specific motor proteins have been gleaned primarily from three types of studies: (1) analysis of the phenotype of cells that lack the motor because they either carry a mutation in a gene that encodes part of the motor protein or have been treated with an siRNA that blocks synthesis of the protein, (2) injection of antibodies or inhibitors against specific motor proteins into cells at various stages of mitosis, and (3) depletion of the motor protein from cell extracts in which mitotic spindles are formed. Although firm conclusions about the functions of specific motor proteins cannot yet be drawn, a general picture of the roles of these molecules is suggested (Figure 14.33):

- Motor proteins located along the polar microtubules probably contribute by keeping the poles apart (Figure 14.33*a,b*).
- Motor proteins residing on the chromosomes are probably important in the movements of the chromosomes during prometaphase (Figure 14.33*a*), in maintaining the chromosomes at the metaphase plate (Figure 14.33*b*), and in separating the chromosomes during anaphase (Figure 14.33*c*).
- Motor proteins situated along the overlapping polar microtubules in the region of the spindle equator are probably responsible for cross-linking antiparallel microtubules and sliding them over one another, thus elongating the spindle during anaphase B (Figure 14.33*c*).

Cytokinesis

Mitosis accomplishes the segregation of duplicated chromosomes into daughter nuclei, but the cell is divided into two daughter cells by a separate process called **cytokinesis**. The first hint of cytokinesis in most animal cells appears during anaphase as an indentation of the cell surface in a narrow band around the cell. As time progresses, the indentation deepens to form a furrow that completely encircles the cell. The plane of the furrow lies in the same plane previously occupied by the chromosomes of the metaphase plate, so that the two sets of chromosomes are ultimately partitioned into different cells (as

FIGURE 14.33 **Proposed activity of motor proteins during mitosis.** (*a*) Prometaphase. The two halves of the mitotic spindle are moving apart from one another to opposite poles, which is thought to result from the action of plus-end-directed motors that cause polar microtubules from opposite poles to slide relative to one another (step 1). (Additional motors associated with the centrosomes and cortex are not shown.) Meanwhile, the chromosomes have become attached to the chromosomal microtubules and can be seen oscillating back and forth along the microtubules. Ultimately, the chromosomes are moved to the center of the spindle, midway between the poles. Poleward chromosome movements are mediated by minus-end-directed motors (i.e., cytoplasmic dynein) residing at the kinetochore (step 2). Chromosome movements away from the poles are mediated by plus-end-directed motors (i.e., kinesin-like proteins) residing at the kinetochore and especially along the chromosome arms (step 3) (see Figure 14.21). (*b*) Metaphase. The two halves of the spindle maintain their separation as the result of plus-end-directed motor activity associated with the polar microtubules (step 4). The chromosomes are thought to be maintained at the equatorial plane by the balanced activity of motor proteins residing at the kinetochore (step 5). (*c*) Anaphase. The movement of the chromosomes toward the poles is thought to require the activity of kinesin depolymerases that catalyze depolymerization at both the plus and minus ends of microtubules (step 6). The separation of the poles (anaphase B) is thought to result from the continuing activity of the plus-end-directed motors of the polar microtubules (step 7). (AFTER K. E. SAWIN AND J. M. SCHOLEY, TRENDS CELL BIOL. 1:122, 1991.)

(a)

(b)

(c)
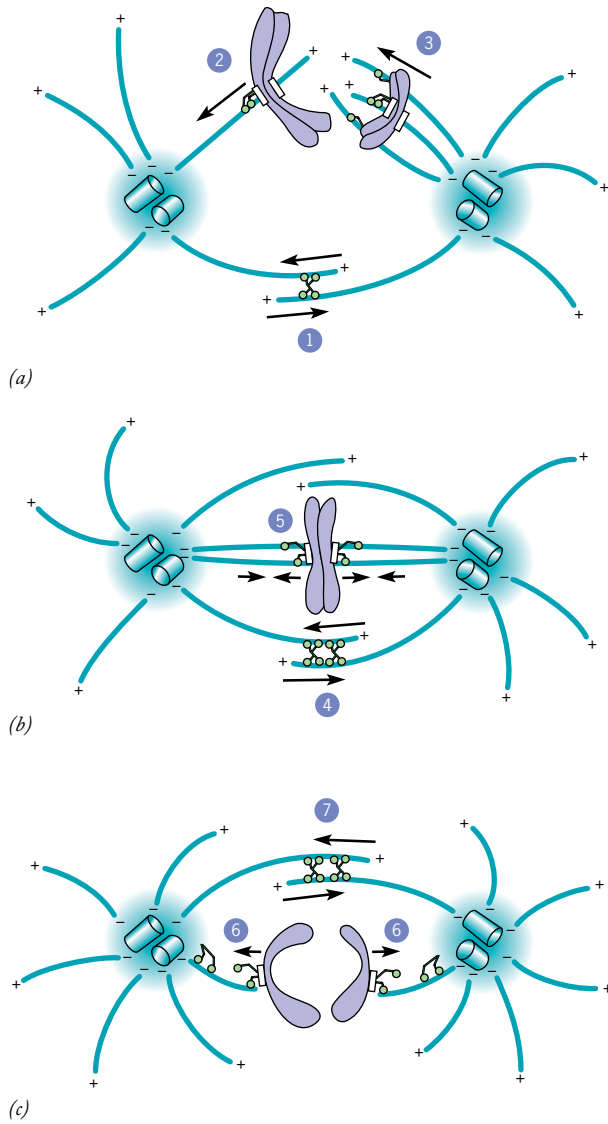
in Figure 14.32). As one cell becomes two cells, additional plasma membrane is delivered to the cell surface via cytoplasmic vesicles that fuse with the advancing cleavage furrow. The furrow continues to deepen as it passes through the tightly packed remnants of the central portion of the mitotic spindle, which forms a cytoplasmic bridge between the daughter cells called the *midbody* (Figure 14.34*a*). Opposing surfaces ultimately fuse with one another in the center of the cell, thus splitting the cell in two (Figure 14.34*b*).
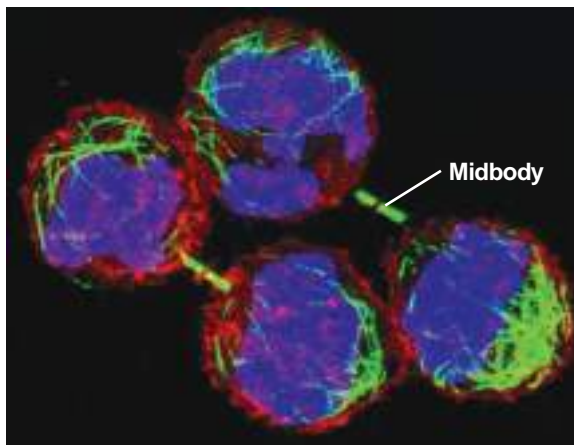
Midbody
(a)

(b)

FIGURE 14.34 **Cytokinesis** (*a*) These cultured mammalian cells are undergoing the final step in cytokinesis, called abscission, in which the cleavage furrow cuts through the midbody, a thin cytoplasmic bridge that is packed with remnants of the central portion of the mitotic spindle. Microtubules are green, actin is red, and DNA is blue. (*b*) This fertilized sea urchin egg has just been split into two cells by cytokinesis. (REPRINTED WITH PERMISSION FROM AHNA R. SKOP ET AL., SCIENCE 305:61, 2004; © COPYRIGHT 2004; AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE; B: COURTESY OF TRYGGVE GUSTAFSON.)

Our present concept of the mechanism responsible for cytokinesis stems from a proposal made by Douglas Marsland in the 1950s known as the *contractile ring theory* (Figure 14.35*a*). Marsland proposed that the force required to cleave a cell is generated in a thin band of contractile cytoplasm located in the *cortex*, just beneath the plasma membrane of the furrow. Microscopic examination of the cortex beneath the furrow of a cleaving cell reveals the presence of large numbers of actin filaments (Figure 14.35*b* and 14.36*a*).

(a)

(b)

FIGURE 14.35 **The formation and operation of the contractile ring during cytokinesis.** (*a*) Actin filaments become assembled in a ring at the cell equator. Contraction of the ring, which requires the action of myosin, causes the formation of a furrow that splits the cell in two. (*b*) Confocal fluorescence micrograph of a fly spermatocyte undergoing cytokinesis at the end of the first meiotic division. Actin filaments, which have been stained with the mushroom toxin phalloidin, are seen to be concentrated in a circular equatorial band within the cleavage furrow: (B: FROM DANIEL SAUL ET AL., J. CELL SCIENCE 117:3893, 2004, COURTESY OF JULIE A. BRILL, BY PERMISSION OF THE COMPANY OF BIOLOGISTS, LTD.)

(c)

FIGURE 14.36 **Experimental demonstration of the importance of myosin in cytokinesis.** (*a,b*) Localization of actin and myosin II in a *Dictyostelium* amoeba during cytokinesis as demonstrated by double-stain immunofluorescence. (*a*) Actin filaments (red) are located at both the cleavage furrow and the cell periphery where they play a key role in cell movement (Section 9.7). (*b*) Myosin II (green) is localized at the cleavage furrow, where it is part of a contractile ring that encompasses the equator. (*c*) A starfish egg that had been microinjected with an antibody against starfish myosin, as observed under polarized light (which causes the mitotic spindles to appear either brighter or darker than the background due to the presence of oriented microtubules). While cytokinesis has been completely suppressed by the antibodies, mitosis (as revealed by the mitotic spindles) continues unaffected. (A,B: COURTESY OF YOSHIO FUKUI; C: FROM DANIEL P. KIEHART, ISSEI MABUCHI, AND SHINYA INOUÉ, J. CELL BIOL. 94:167, 1982; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

Interspersed among the actin filaments are a smaller number of short, bipolar myosin filaments. These filaments are composed of myosin II, as evidenced by their binding of anti-myosin II antibodies (Figure 14.36*b*). The importance of myosin II in cytokinesis is evident from the fact that (1) anti-myosin II antibodies cause the rapid cessation of cytokinesis when injected into a dividing cell (Figure 14.36*c*), and (2) cells lacking a functional myosin II gene carry out nuclear division by mitosis but cannot divide normally into daughter cells. The assembly of the actin-myosin contractile machinery in the plane of the future cleavage furrow is orchestrated by a G protein called RhoA. In its GTP-bound state, RhoA triggers a cascade of events that lead to both the assembly of actin filaments and the activation of myosin II's motor activity. If RhoA is depleted from cells or inactivated, a cleavage furrow fails to develop.

The force-generating mechanism operating during cytokinesis is thought to be similar to the actin and myosin-based contraction of muscle cells. Whereas the sliding of actin filaments of a muscle cell brings about the shortening of the muscle fiber, sliding of the filaments of the contractile ring pulls the cortex and attached plasma membrane toward the center of the cell. As a result, the contractile ring constricts the equatorial region of the cell, much like pulling on a purse string narrows the opening of a purse.

It is generally agreed that the position of the cleavage furrow is determined by the anaphase mitotic spindle, leading to the activation of RhoA in a narrow ring within the cortex. However, there has been considerable debate as to how this particular zone within the cortex is selected. Early pioneering studies on marine invertebrate eggs by Ray Rappaport of Union College in New York demonstrated that the contractile ring forms in a plane midway between the spindle poles, even if one of the poles is displaced by a microneedle inserted into the cell. An example of the relationship between the position of the spindle poles and cleavage plane is shown in the experiment of Figure 14.37. These studies suggest that the site of actin-filament assembly, and thus the plane of cytokinesis, is determined by a signal emanating from the spindle poles. The signal is thought to travel from the spindle poles to the cell cortex along the astral microtubules. When the distance between the poles and the cortex is modified experimentally, the timing of cytokinesis can be dramatically altered (Figure 14.37). In contrast, researchers conducting studies on smaller mammalian cells have found evidence that the site of cleavage furrow formation is defined by a signal that originates in the central part of the mitotic spindle rather than its poles. Researchers have struggled to reconcile these opposing findings. The simplest explanations are that (1) different cell types utilize different mechanisms or (2) both mechanisms operate in the same cell. In fact, recent studies provide evidence for this latter possibility.

Cytokinesis in Plant Cells: Formation of the Cell Plate

Plant cells, which are enclosed by a relatively inextensible cell wall, undergo cytokinesis by a very different mechanism. Unlike animal cells, which are constricted by a furrow that advances inward from the outer cell surface, plant cells must construct an extracellular wall inside a living cell. Wall formation starts in the center of the cell and grows outward to meet
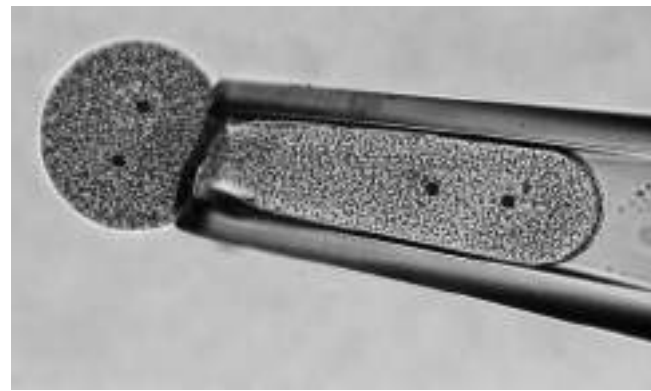
(a)
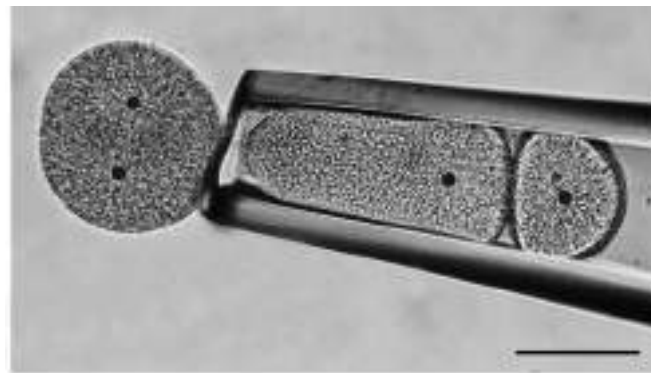
(b)
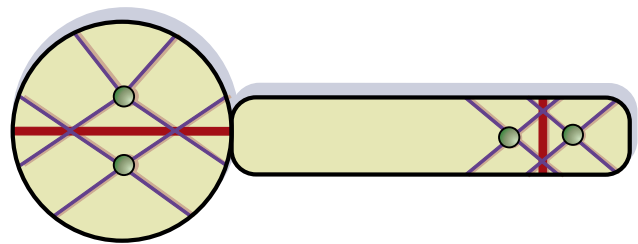
(c)

FIGURE 14.37 The site of formation of the cleavage plane and the time at which cleavage occurs depends on the position of the mitotic spindle. (*a*) This echinoderm egg was allowed to divide once to form a two-cell embryo. Then, once the mitotic spindle appeared in each of the two cells, one of the cells was drawn into a micropipette, causing it to assume a cylindrical shape. The two dark spots in each cell are the spindle poles that have formed prior to the second division of each cell. (*b*) Nine minutes later the cylindrical cell has completed cleavage, while the spherical cell has not yet begun to cleave. These photos indicate that (1) the cleavage plane forms between the spindle poles, regardless of their position, and (2) cleavage occurs more rapidly in the cylindrical cell. Bar equals 80 μm. (*c*) These results can be explained by assuming that (1) the cleavage plane (red bar) forms where the astral microtubules overlap, and (2) cleavage occurs earlier in the cylindrical cell because the distance from the poles (green spheres) to the site of cleavage is reduced, thereby shortening the time it takes the cleavage signal to reach the surface. (A,B: FROM CHARLES B. SHUSTER AND DAVID R. BURGESS, J. CELL BIOL. 146:987, 1999; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

the existing lateral walls. The formation of a new cell wall begins with the construction of a simpler precursor, which is called the **cell plate**.
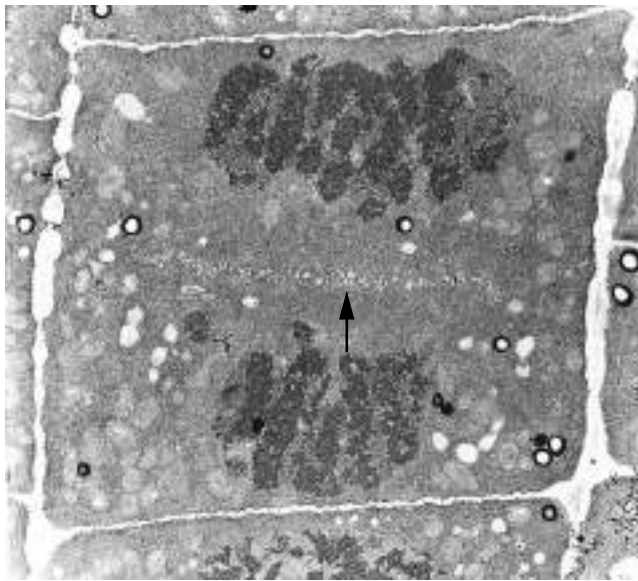
(a)

FIGURE 14.38 The formation of a cell plate between two daughter plant nuclei during cytokinesis. (*a*) A low-magnification electron micrograph showing the formation of the cell plate between the future daughter cells. Secretory vesicles derived from nearby Golgi complexes have become aligned along the equatorial plane (arrow) and are beginning to fuse with one another. The membrane of the vesicles forms the plasma membranes of the two daughter cells, and the contents of the vesicles will provide the material that forms the cell plate separating the cells. (*b*) Steps in the formation of the cell plate as described in the text. (A: DAVID PHILLIPS/PHOTO RESEARCHERS; B: AFTER A. L. SAMUELS, T. H. GIDDINGS, JR., AND L. A. STAEHELIN, J. CELL BIOL. 130:1354, 1995; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

(b)

The plane in which the cell plate forms is perpendicular to the axis of the mitotic spindle but, unlike the case for animal cells, is not determined by the position of the spindle. Rather, the orientation of both the mitotic spindle and cell plate are determined by a belt of cortical microtubules—the *preprophase band*—that forms in late G_2 (see Figure 9.21). Even though the preprophase band has disassembled by prometaphase, it leaves an invisible imprint that determines the future division site. The first sign of cell plate formation is seen in late anaphase with the appearance of the phragmoplast in the center of the dividing cell. The **phragmoplast** consists of clusters of interdigitating microtubules oriented perpendicular to the future plate (Figure 9.21), together with actin filaments, membranous vesicles, and electron-dense material. The microtubules of the phragmoplast, which arise from remnants of the mitotic spindle, serve as tracks for the movement of small Golgi-derived secretory vesicles into the region. The vesicles become aligned along a plane between the daughter nuclei (Figure 14.38*a*). Electron micrographs of rapidly frozen tobacco cells have revealed the steps by which the Golgi-derived vesicles become reorganized into the cell plate. To begin the process (step 1, Figure 14.38*b*), the vesicles send out finger-like tubules that contact and fuse with neighboring vesicles to form an interwoven tubular network in the center of the cell (step 2). Additional vesicles are then directed along microtubules to the lateral edges of the network. The newly arrived vesicles continue the process of tubule formation and fusion, which extends the network in an outward direction (step 2). Eventually, the leading edge of the growing network contacts the parent plasma membrane at the boundary of the cell (step 3). Ultimately, the tubular network loses its cytoplasmic gaps and matures into a continuous, flattened partition. The membranes of the tubular network become the plasma membranes of the two adjacent daughter cells, whereas the secretory products that had been carried within the vesicles contribute to the intervening cell plate. Once the cell plate is completed, cellulose and other materials are added to produce the mature cell wall.
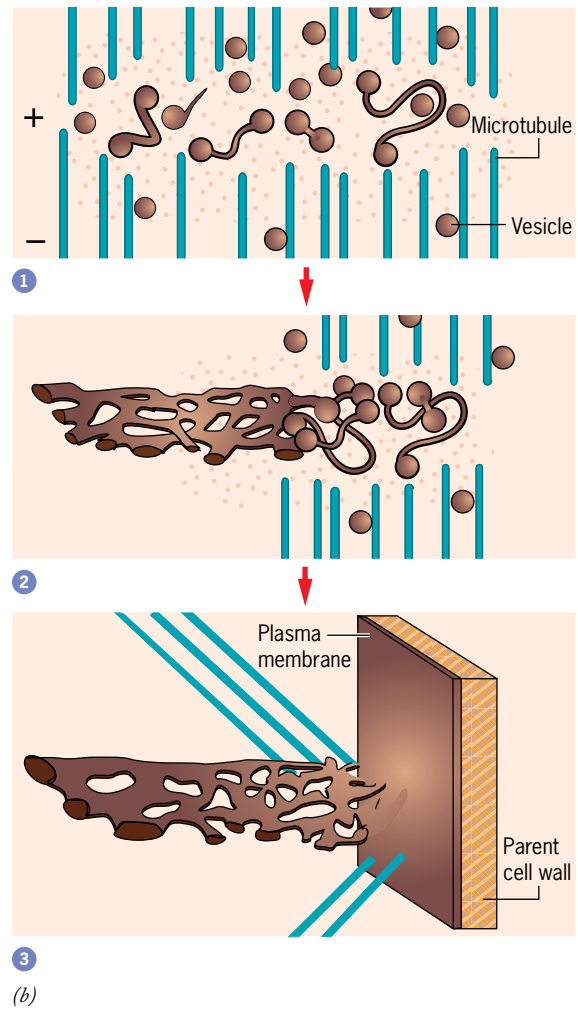
REVIEW ?

1. How do the events of mitotic prophase prepare the chromatids for later separation at anaphase?

2. What are some of the activities of the kinetochore during mitosis?

3. Describe the events that occur in a cell during prometaphase and during anaphase.

4. Describe the similarities and differences in microtubule dynamics between metaphase and anaphase. How are the differences related to anaphase A and B movements?

5. What types of force-generating mechanisms might be responsible for chromosome movement during anaphase?

6. Contrast the events that occur during cytokinesis in typical plant and animal cells.

14.3 MEIOSIS

The production of offspring by sexual reproduction includes the union of two cells, each with a haploid set of chromosomes. As discussed in Chapter 10, the doubling of the chromosome number at fertilization is compensated by an equivalent reduction in chromosome number at a stage prior to formation of the gametes. This is accomplished by **meiosis**, a term coined in 1905 from the Greek word meaning "reduction." Meiosis ensures production of a haploid phase in the life cycle, and fertilization ensures a diploid phase. Without meiosis, the chromosome number would double with each generation, and sexual reproduction would not be possible.

To compare the events of mitosis and meiosis, we need to examine the fate of the chromatids. Prior to both mitosis and meiosis, diploid G_2 cells contain pairs of homologous chromosomes, with each chromosome consisting of two chromatids. During mitosis, the chromatids of each chromosome are split apart and separate into two daughter nuclei in a *single* division. As a result, cells produced by mitosis contain pairs of homologous chromosomes and are genetically identical to their parents. During meiosis, in contrast, the four chromatids of a pair of replicated homologous chromosomes are distributed among four daughter nuclei. Meiosis accomplishes this feat by incorporating two sequential divisions without an intervening round of DNA replication (Figure 14.39). In the first meiotic division, each chromosome (consisting of two chromatids) is separated from its homologue. As a result, each daughter cell contains only one member of each pair of homologous chromosomes. For this to occur, homologous chromosomes are paired during prophase of the first meiotic division (prophase I, Figure 14.39) by an elaborate process that has no counterpart in mitosis. As they are paired, homologous chromosomes engage in a process of genetic recombination that produces chromosomes with new combinations of maternal and paternal alleles (see metaphase I, Figure 14.39). In the second meiotic division, the two chromatids of each chromosome are separated from one another (anaphase II, Figure 14.39).

A survey of various eukaryotes reveals marked differences with respect to the stage within the life cycle at which meiosis occurs and the duration of the haploid phase. The following three groups (Figure 14.40) can be identified on these bases:

1. *Gametic or terminal meiosis.* In this group, which includes all multicellular animals and many protists, the meiotic divisions are closely linked to the formation of the gametes (Figure 14.40, left). In male vertebrates (Figure 14.41*a*), for example, meiosis occurs just prior to the differentiation of the spermatozoa. *Spermatogonia* that are

FIGURE 14.39   **The stages of meiosis.**

FIGURE 14.40 **A comparison of three major groups of organisms based on the stage within the life cycle at which meiosis occurs and the duration of the haploid phase.** (ADAPTED FROM E. B. WILSON, THE CELL IN DEVELOPMENT AND HEREDITY, 3D ED., 1925. REPRINTED WITH PERMISSION OF MACMILLAN PUBLISHING CO., INC.)

when it is fertilized) that the meiotic divisions occur. Vertebrate eggs are typically fertilized at a stage before the completion of meiosis (usually at metaphase II). Meiosis is completed after fertilization, while the sperm resides in the egg cytoplasm.

2. ***Zygotic or initial meiosis.*** In this group, which includes only protists and fungi, the meiotic divisions occur just after fertilization (Figure 14.40, right) to produce haploid spores. The spores divide by mitosis to produce a haploid adult generation. Consequently, the diploid stage of the life cycle is restricted to a brief period after fertilization when the individual is still a zygote.

3. ***Sporic or intermediate meiosis.*** In this group, which includes plants and some algae, the meiotic divisions take place at a stage unrelated to either gamete formation or fertilization (Figure 14.40, center). If we begin the life cycle with the union of a male gamete (the pollen grain) and a female gamete (the egg), the diploid zygote undergoes mitosis and develops into a diploid **sporophyte**. At some stage in the development of the sporophyte, *sporogenesis* (which includes meiosis) occurs, producing spores that germinate directly into a haploid **gametophyte**. The gametophyte can be either an independent stage or, as in the case of seed plants, a tiny structure retained within the ovules. In either case, the gametes are produced from the haploid gametophyte by *mitosis*.

The Stages of Meiosis

As with mitosis, the prelude to meiosis includes DNA replication. The premeiotic S phase generally takes several times longer than a premitotic S phase. Prophase of the first meiotic division (i.e., prophase I) is typically lengthened in extraordinary fashion when compared to mitotic prophase. In the human female, for example, oocytes initiate prophase I of meiosis prior to birth and then enter a period of prolonged arrest. Oocytes resume meiosis just prior to the time they are ovulated, which occurs every 28 days or so after an individual reaches puberty. Consequently, many human oocytes remain arrested in the same approximate stage of prophase for several decades. The first meiotic prophase is also very complex and is customarily divided into several stages that are similar in all sexually reproducing eukaryotes (Figure 14.42).

The first stage of prophase I is *leptotene*, during which the chromosomes become visible in the light microscope. Although the chromosomes have replicated at an earlier stage, there is no indication that each chromosome is actually composed of a pair of identical chromatids. In the electron microscope, however, the chromosomes are revealed to be composed of paired chromatids.

The second stage of prophase I, which is called *zygotene*, is marked by the visible association of homologues with one another. This process of chromosome pairing is called **synapsis** and is an intriguing event with important unanswered questions: On what basis do the homologues recognize one another? How does the pair become so perfectly aligned?

committed to undergo meiosis become *primary spermatocytes*, which then undergo the two divisions of meiosis to produce four relatively undifferentiated *spermatids*. Each spermatid undergoes a complex differentiation to become the highly specialized sperm cell (*spermatozoon*). In female vertebrates (Figure 14.41*b*), *oogonia* become *primary oocytes*, which then enter a greatly extended meiotic prophase. During this prophase, the primary oocyte grows and becomes filled with yolk and other materials. It is only after differentiation of the oocyte is complete (i.e., the oocyte has reached essentially the same state as

(a)

(b)

FIGURE 14.41 The stages of gametogenesis in vertebrates: a comparison between the formation of sperm and eggs. In both sexes, a relatively small population of primordial germ cells present in the embryo proliferates by mitosis to form a population of gonial cells (spermatogonia or oogonia) from which the gametes differentiate. In the male (*a*), meiosis occurs before differentiation, whereas in the female (*b*), both meiotic divisions occur after differentiation. Each primary spermatocyte generally gives rise to four viable gametes, whereas each primary oocyte forms only one fertilizable egg and two or three polar bodies.
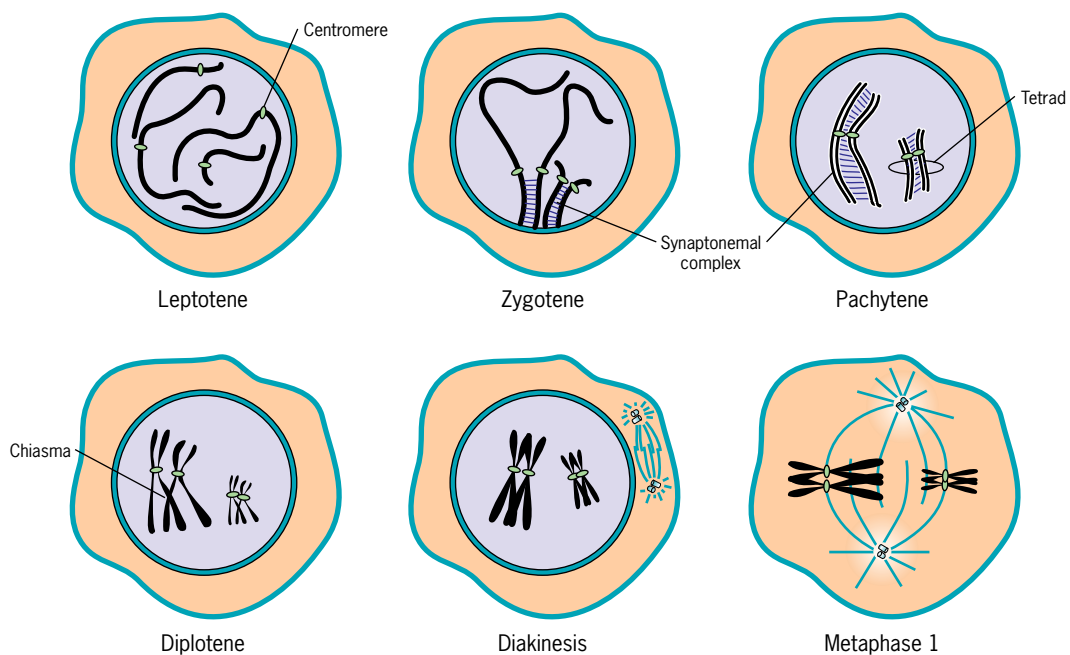
FIGURE 14.42 The stages of prophase I. The events at each stage are described in the text.

When does recognition between homologues first occur? Recent studies have shed considerable light on these questions. It had been assumed for years that interaction between homologous chromosomes first begins as chromosomes initiate synapsis. However, studies on yeast cells by Nancy Kleckner and her colleagues at Harvard University demonstrated that homologous regions of DNA from homologous chromosomes are already associated with one another during leptotene. Chromosome compaction and synapsis during zygotene simply make this arrangement visible under the microscope. As will be discussed below, the first step in genetic recombination is the deliberate introduction of double-stranded breaks in aligned DNA molecules. Studies in both yeast and mice suggest the DNA breaks occur in leptotene, well before the chromosomes are visibly paired.

These findings are supported by studies aimed at locating particular DNA sequences within the nuclei of premeiotic and meiotic cells. We saw on page 497 that individual chromosomes occupy discrete regions within nuclei rather than being randomly dispersed throughout the nuclear space. When yeast cells about to enter meiotic prophase are examined, each pair of homologous chromosomes is found to share a joint territory distinct from the territories shared by other pairs of homologues. This finding suggests that homologous chromosomes are paired to some extent before meiotic prophase begins. The telomeres (terminal segments) of leptotene chromosomes are distributed throughout the nucleus. Then, near the end of leptotene, there is a dramatic reorganization of chromosomes in many species so that the telomeres become localized at the inner surface of the nuclear envelope at one side of the nucleus. Clustering of telomeres at one end of the nuclear envelope occurs in a wide variety of eukaryotic cells and causes the chromosomes to resemble the clustered stems of a bouquet of flowers (Figure 14.43). Whether or not telomere clustering aids in the process of synapsis remains a matter of debate.

Electron micrographs indicate that chromosome synapsis is accompanied by the formation of a complex structure called the synaptonemal complex. The **synaptonemal complex** (**SC**) is a ladder-like structure with transverse protein filaments connecting the two lateral elements (Figure 14.44). The chromatin of each homologue is organized into loops that extend from one of the lateral elements of the SC (Figure 14.44*b*).

FIGURE 14.43 **Association of the telomeres of meiotic chromosomes with the nuclear envelope.** The chromosomes of a stage of meiotic prophase of the male grasshopper in which homologous chromosomes are physically associated as part of a bivalent. The bivalents are arranged in a well-defined "bouquet" with their terminal regions clustered near the inner surface of the nuclear envelope, at the base of the photo. (COURTESY OF BERNARD JOHN, FROM B. JOHN, MEIOSIS. REPRINTED WITH THE PERMISSION OF CAMBRIDGE UNIVERSITY PRESS, 1990.)

FIGURE 14.44 **The synaptonemal complex.** (*a*) Electron micrograph of a human pachytene bivalent showing a pair of homologous chromosomes held in a tightly ordered parallel array. K, kinetochore. (*b*) Schematic diagram of the synaptonemal complex and its associated chromosomal fibers. The dense granules (recombination nodules) seen in the center of the SC (indicated by the arrowhead in part *a*) contain the enzymatic machinery required to complete genetic recombination, which is thought to begin at a much earlier stage in prophase I. Closely paired loops of DNA from the two sister chromatids of each chromosome are depicted. The loops are likely maintained in a paired configuration by cohesin (not shown). Genetic recombination (crossing-over) is presumed to occur between the DNA loops from nonsister chromatids, as shown. (A: COURTESY OF ALBERTO J. SOLARI, CHROMOSOMA 81:330, 1980.)

The lateral elements are composed primarily of cohesin (page 572), which presumably binds together the chromatin of the sister chromatids. For many years, the SC was thought to hold each pair of homologous chromosomes in the proper position to initiate genetic recombination between strands of homologous DNA. It is now evident that the SC is not required for genetic recombination. Not only does the SC form after genetic recombination has been initiated, but mutant yeast cells unable to assemble an SC can still engage in the exchange of genetic information between homologues. It is currently thought that the SC functions primarily as a scaffold to allow interacting chromatids to complete their crossover activities, as described below.

The complex formed by a pair of synapsed homologous chromosomes is called a **bivalent** or a **tetrad**. The former term reflects the fact that the complex contains two homologues, whereas the latter term calls attention to the presence of four chromatids. The end of synapsis marks the end of zygotene and the beginning of the next stage of prophase I, called *pachytene*, which is characterized by a fully formed synaptonemal complex. During pachytene, the homologues are held closely together along their length by the SC. The DNA of sister chromatids is extended into parallel loops (Figure 14.44). Under the electron microscope, a number of electron-dense bodies about 100 nm in diameter are seen within the center of the SC. These structures have been named *recombination nodules* because they correspond to the sites where crossing-over is taking place, as evidenced by the associated synthesis of DNA that occurs during intermediate steps of recombination (page 598). Recombination nodules contain the enzymatic machinery that facilitates genetic recombination, which is completed by the end of pachytene.

The beginning of *diplotene*, the next stage of meiotic prophase I (Figure 14.42), is recognized by the dissolution of the SC, which leaves the chromosomes attached to one another at specific points by X-shaped structures, termed **chiasmata** (singular **chiasma**) (Figure 14.45). Chiasmata are located at sites on the chromosomes where crossing-over between DNA molecules from the two chromosomes had previously occurred. Chiasmata are formed by covalent junctions between a chromatid from one homologue and a nonsister chromatid from the other homologue. These points of attachment provide a striking visual portrayal of the extent of genetic recombination. The chiasmata are made more visible by a tendency for the homologues to separate from one another at the diplotene stage.

In vertebrates, diplotene can be an extremely extended phase of oogenesis during which the bulk of oocyte growth occurs. Thus diplotene can be a period of intense metabolic activity. Transcription during diplotene in the oocyte provides the RNA utilized for protein synthesis during both oogenesis and early embryonic development following fertilization.

During the final stage of meiotic prophase I, called *diakinesis*, the meiotic spindle is assembled, and the chromosomes are prepared for separation. In those species in which the chromosomes become highly dispersed during diplotene, the chromosomes become recompacted during diakinesis. Diakinesis

(a)

(b)

(c)   4 μm
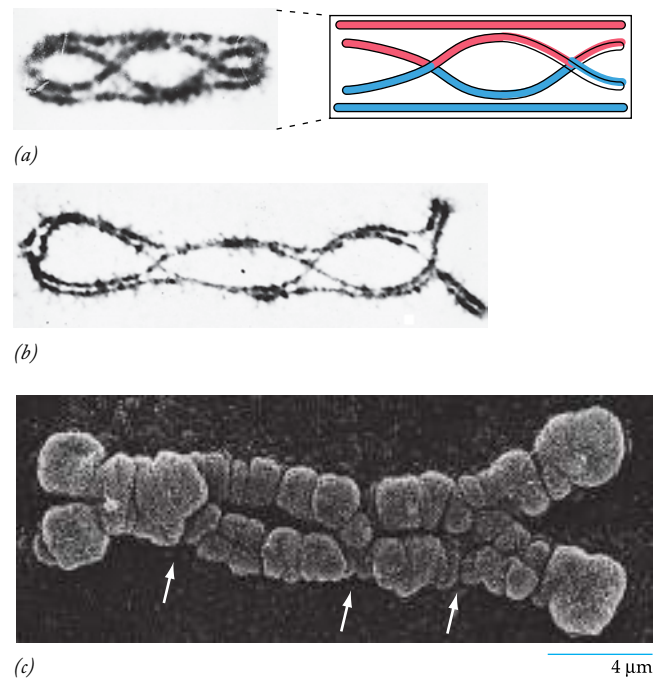
FIGURE 14.45 Visible evidence of crossing-over. (*a,b*) Diplotene bivalents from the grasshopper showing the chiasmata formed between chromatids of each homologous chromosome. The accompanying inset indicates the crossover that has presumably occurred within the bivalent in *a*. The chromatids of each diplotene chromosome are closely apposed except at the chiasmata. (*c*) Scanning electron micrograph of a bivalent from the desert locust with three chiasmata (arrows). (A,B: FROM BERNARD JOHN, MEIOSIS, CAMBRIDGE, 1990; REPRINTED WITH THE PERMISSION OF CAMBRIDGE UNIVERSITY PRESS. C: FROM KLAUS WERNER WOLF, BIOESS. 16:108, 1994.)

ends with the disappearance of the nucleolus, the breakdown of the nuclear envelope, and the movement of the tetrads to the metaphase plate. In vertebrate oocytes, these events are triggered by an increase in the level of the protein kinase activity of MPF (maturation-promoting factor). As discussed in the Experimental Pathways at the end of the chapter, MPF was first identified by its ability to initiate these events, which represent the *maturation* of the oocyte (page 599).

In most eukaryotic species, chiasmata can still be seen in homologous chromosomes aligned at the metaphase plate of meiosis I. In fact, chiasmata are required to hold the homologues together as a bivalent during this stage. In humans and other vertebrates, every pair of homologues typically contains at least one chiasma, and the longer chromosomes tend to have two or three of them. It is thought that some mechanism exists to ensure that even the smallest chromosomes form a chiasma. If a chiasma does not occur between a pair of homologous chromosomes, the chromosomes of that bivalent tend to separate from one another after dissolution of the SC. This premature separation of homologues often results in the formation of nuclei with an abnormal number of chromosomes. The consequences of such an event are discussed in the accompanying Human Perspective.

At metaphase I, the two homologous chromosomes of each bivalent are connected to the spindle fibers from opposite poles (Figure 14.46*a*). In contrast, sister chromatids are connected to microtubules from the same spindle pole, which is made possible by the side-by-side arrangement of their kinetochores as seen in the inset of Figure 14.46*a*. The orientation of the maternal and paternal chromosomes of each bivalent on the metaphase I plate is random; the maternal member of a particular bivalent has an equal likelihood of facing either pole. Consequently, when homologous chromosomes separate during anaphase I, each pole receives a random assortment of maternal and paternal chromosomes (see Figure 14.39). Thus, anaphase I is the cytological event that corresponds to Mendel's law of independent assortment (page 381). As a re-

sult of independent assortment, organisms are capable of generating a nearly unlimited variety of gametes.

Separation of homologous chromosomes at anaphase I requires the dissolution of the chiasmata that hold the bivalents together. The chiasmata are maintained by cohesion between sister chromatids in regions that flank these sites of recombination (Figure 14.46*a*). The chiasmata disappear at the metaphase I–anaphase I transition, as the arms of the chromatids of each bivalent lose cohesion (Figure 14.46*b*). Loss of cohesion between the arms is accomplished by proteolytic cleavage of the cohesin molecules in those regions of the chromosome. In contrast, cohesion between the joined centromeres of sister chromatids remains strong, because the cohesin situated there is protected from proteolytic attack (Figure 14.46*b*). As a result, sister chromatids remain firmly attached to one another as they move together toward a spindle pole during anaphase I.

Telophase I of meiosis I produces less dramatic changes than telophase of mitosis. Although chromosomes often undergo some dispersion, they do not reach the extremely extended state of the interphase nucleus. The nuclear envelope may or may not reform during telophase I. The stage between the two meiotic divisions is called *interkinesis* and is generally short-lived. In animals, cells in this fleeting stage are referred to as *secondary spermatocytes* or *secondary oocytes*. These cells are characterized as being haploid because they contain only one member of each pair of homologous chromosomes. Even though they are haploid, they have twice as much DNA as a haploid gamete because each chromosome is still represented by a pair of attached chromatids. Secondary spermatocytes are said to have a 2C amount of DNA, half as much as a primary spermatocyte, which has a 4C DNA content, and twice as much as a sperm cell, which has a 1C DNA content.

Interkinesis is followed by prophase II, a much simpler prophase than its predecessor. If the nuclear envelope had reformed in telophase I, it is broken down again. The chromosomes become recompacted and line up at the metaphase plate. Unlike metaphase I, the kinetochores of sister chromatids of metaphase II face opposite poles and become attached to opposing sets of chromosomal spindle fibers (Figure 14.46*c*). The progression of meiosis in vertebrate oocytes stops at metaphase II. The arrest of meiosis at metaphase II is brought about by factors that inhibit APC^{Cdc20} activation, thereby preventing cyclin B degradation. As long as cyclin B levels remain high within the oocyte, Cdk activity is maintained, and the cells cannot progress to the next meiotic stage. Metaphase II arrest is released only when the oocyte (now called an egg) is fertilized. Fertilization leads to a rapid influx of Ca^{2+} ions, the activation of APC^{Cdc20} (page 580), and the destruction of cyclin B. The fertilized egg responds to these changes by completing the second meiotic division. Anaphase II begins with the synchronous splitting of the centromeres, which had held the sister chromatids together, allowing them to move toward opposite poles of the cell (Figure 14.46*d*). Meiosis II ends with telophase II, in which the chromosomes are once again enclosed by a nuclear envelope. The products of meiosis are haploid cells with a 1C amount of nuclear DNA.

FIGURE 14.46 Separation of homologous chromosomes during meiosis I and separation of chromatids during meiosis II. (*a*) Schematic diagram of a pair of homologous chromosomes at metaphase I. The chromatids are held together along both their arms and centromeres by cohesin. The pair of homologues are maintained as a bivalent by the chiasmata. Inset micrograph shows that the kinetochores (arrowheads) of sister chromatids are situated on one side of the chromosome, facing the same pole. The black dots are gold particles bound to the motor protein CENP-E of the kinetochores (see Figure 14.16*c*). (*b*) At anaphase I, the cohesin holding the arms of the chromatids is cleaved, allowing the homologues to separate from one another. Cohesin remains at the centromere, holding the chromatids together. (*c*) At metaphase II, the chromatids are held together at the centromere, with microtubules from opposite poles attached to the two kinetochores. Inset micrograph shows the kinetochores of the sister chromatids are now on opposite sides of the chromosome, facing opposite poles. (*d*) At anaphase II, the cohesin holding the chromatids together has been cleaved, allowing the chromosomes to move to opposite poles. (INSETS: FROM JIBAK LEE ET AL., MOL. REPROD. DEVELOP. 56:51, 2000.)

THE HUMAN PERSPECTIVE

Meiotic Nondisjunction and Its Consequences

Meiosis is a complex process, and meiotic mistakes in humans appear to be surprisingly common. Homologous chromosomes may fail to separate from each other during meiosis I, or sister chromatids may fail to come apart during meiosis II. When either of these situations occurs, gametes are formed that contain an abnormal number of chromosomes—either an extra chromosome or a missing chromosome (Figure 1). If one of these gametes happens to fuse with a normal gamete, a zygote with an abnormal number of chromosomes forms, and serious consequences arise. In most cases, the zygote develops into an abnormal embryo that dies at some stage between conception and birth. In a few cases, however, the zygote develops into an infant whose cells have an abnormal chromosome number, a condition known as *aneuploidy*. The consequences of aneuploidy depend on which chromosome or chromosomes are affected.

The normal human chromosome complement is 46: 22 pairs of autosomes and one pair of sex chromosomes. An extra chromosome (producing a total of 47 chromosomes) creates a condition referred

FIGURE 2 The karyotype of a person with Down syndrome. The karyotype shows an extra chromosome 21 (trisomy 21).

to as a *trisomy* (Figure 2). A person whose cells contain an extra chromosome 21, for example, has trisomy 21. A missing chromosome (producing a total of 45 chromosomes) produces a *monosomy*. We will begin by considering the effects of an abnormal number of autosomes.

The absence of one autosomal chromosome, regardless of which chromosome is affected, invariably proves to be lethal at some stage during embryonic or fetal development. Consequently, a zygote containing an autosomal monosomy does not give rise to a fetus that is carried to term. Although one might not expect that possession of an extra chromosome would create a life-threatening condition, the fact is that trisomies do not fare much better than monosomic zygotes. Of the 22 different autosomes in the human chromosome complement, only persons with trisomy 21 can survive beyond the first few weeks or months of life. Most of the other possible trisomies are lethal during development, whereas trisomies for chromosomes 13 and 18 are often born alive but have such severe abnormalities that they succumb soon after birth. More than one-quarter of fetuses that spontaneously abort carry a chromosomal trisomy. It is thought that many more zygotes carrying abnormal chromosome numbers produce embryos that die at an early stage of development before the pregnancy is recognized. For example, for every trisomic zygote formed at fertilization, there is presumably an equal number of monosomic zygotes that fare even less well. It is estimated that approximately 20 to 25 percent of human oocytes are aneuploid, which is much higher than any other species that has been studied. Mouse eggs, for example, typically exhibit an aneuploidy level of 1 to 2 percent. Exposure of mice to an estrogen-like compound—bisphenol A—used in the manufacture of polycarbonate plastics can greatly increase the level of nondisjunction during meiosis in mice. This was the first clear demonstration of a relationship between synthetic compounds present in the environment and meiotic aneuploidy. Whether this or other environmental agents contribute to the high rate of aneuploidy in human oocytes remains unclear. Whatever the reason, meiosis in males occurs with a much lower level of chromosomal abnormalities than in females.

FIGURE 1 Meiotic nondisjunction occurs when chromosomes fail to separate from each other during meiosis. If the failure to separate occurs during the first meiotic division, which is called *primary nondisjunction*, all of the haploid cells have an abnormal number of chromosomes. If nondisjunction occurs during the second meiotic division, which is called *secondary nondisjunction*, only two of the four haploid cells are affected. (More complex types of nondisjunction than shown here can also occur.)

Even though chromosome 21 is the smallest human chromosome with fewer than 400 genes, the presence of an extra copy of this genetic material leads to *Down syndrome*. Persons with Down syndrome exhibit varying degrees of mental impairment, alteration in certain body features, circulatory problems, increased susceptibility to infectious diseases, a greatly increased risk of developing leukemia, and the early onset of Alzheimer's disease. All of these medical problems are thought to result from an abnormal level of expression of genes located on chromosome 21. However, the expression of genes on other chromosomes can also be affected by the excess levels of certain transcription factors encoded by genes on chromosome 21.

The presence of an abnormal number of sex chromosomes is much less disruptive to human development. A zygote with only one X chromosome and no second sex chromosome (denoted as XO) develops into a female with *Turner syndrome*, in which genital development is arrested in the juvenile state, the ovaries fail to develop, and body structure is slightly abnormal. Because a Y chromosome is male determining, persons with at least one Y chromosome develop as males. A male with an extra X chromosome (XXY) develops *Klinefelter syndrome*, which is characterized by mental retardation, underdevelopment of genitalia, and the presence of feminine physical characteristics (such as breast enlargement). A zygote with an extra Y (XYY) develops into a physically normal male who is likely to be taller than average. Considerable controversy developed surrounding claims that XYY males tend to exhibit more aggressive, antisocial, and criminal behavior than do XY males, but this hypothesis has never been substantiated.

The likelihood of having a child with Down syndrome rises dramatically with the age of the mother—from 0.05 percent for mothers 19 years of age to greater than 3 percent for women over the age of 45. Most studies show that no such correlation is found between the age of the father and the likelihood of bearing a child with trisomy 21. Estimates based on comparisons of DNA sequences between the offspring and the parents, indicate that approximately 95 percent of trisomies 21 can be traced to nondisjunction having occurred in the mother.

It was noted above that an abnormal chromosome number can result from nondisjunction at either of the two meiotic divisions (Figure 1). Although these different nondisjunction events produce the same effect in terms of chromosome numbers in the zygote, they can be distinguished by genetic analysis. Primary nondisjunction transmits two homologous chromosomes to the zygote, whereas secondary nondisjunction transmits two sister chromatids (most likely altered by crossing over) to the zygote. Studies indicate that most of the mistakes occur during meiosis I. For example, in one study of 433 cases of trisomy 21 that resulted from maternal nondisjunction, 373 were the result of errors that had occurred during meiosis I and 60 were the result of errors during meiosis II.

Why should meiosis I be more susceptible to nondisjunction than meiosis II? We don't know the precise answer to that question, but it almost certainly reflects the fact that oocytes of older women have remained arrested in meiosis I for a very long period within the ovary. It was noted in the chapter that chiasmata—which are visual indicators of genetic recombination—play an important role in holding a bivalent together during metaphase I. According to one hypothesis, meiotic spindles of older oocytes are less able to hold together weakly constructed bivalents (e.g., bivalents with only one chiasma located near the tip of the chromosome) than those of younger oocytes, increasing the likelihood that homologous chromosomes will missegregate at anaphase I. Another possibility is that sister chromatid cohesion, which prevents the chiasmata from "sliding off" the end of the chromosome, is not fully maintained over an extended period, allowing homologues to separate prematurely.

Genetic Recombination During Meiosis

In addition to reducing the chromosome number as required for sexual reproduction, meiosis increases the genetic variability in a population of organisms from one generation to the next. Independent assortment allows maternal and paternal chromosomes to become shuffled during formation of the gametes, and genetic recombination (crossing-over) allows maternal and paternal alleles on a given chromosome to become shuffled as well (see Figure 14.39). Without genetic recombination, the alleles along a particular chromosome would remain tied together, generation after generation. By mixing maternal and paternal alleles between homologous chromosomes, meiosis generates organisms with novel genotypes and phenotypes on which natural selection can act (see Figure 10.7 for an example from the fruit fly).

Recombination involves the physical breakage of individual DNA molecules and the ligation of the split ends from one DNA duplex with the split ends of the duplex from the homologous chromosome. Recombination is a remarkably precise process that normally occurs without the addition or loss of a single base pair. To occur so faithfully, recombination depends on the complementary base sequences that exist between a single strand from one chromosome and the homologous strand of another chromosome, as discussed below.

The precision of recombination is further ensured by the involvement of DNA repair enzymes that fill gaps that develop during the exchange process.

A simplified model of the proposed steps that occur during recombination in eukaryotic cells is shown in Figure 14.47. In this model, two DNA duplexes that are about to recombine become aligned next to one another as the result of some type of *homology search* in which homologous DNA molecules associate with one another in preparation for recombination. Once they are aligned, an enzyme (Spo11) introduces a double-stranded break into one of the duplexes (step 1, Figure 14.47). The gap is subsequently widened (resected) as indicated in step 2. Resection may occur by the action of a $5' \rightarrow 3'$ exonuclease or by an alternate mechanism. Regardless, the broken strands possess exposed single-stranded tails, each bearing a $3'$ OH terminus. In the model shown in Figure 14.47, one of the single-stranded tails leaves its own duplex and invades the DNA molecule of a nonsister chromatid, hydrogen bonding with the complementary strand in the neighboring duplex (step 3). In *E. coli*, this process in which a single strand invades an intact homologous duplex and displaces the corresponding strand in that duplex is catalyzed by a recombinase enzyme, called the *RecA protein*. The RecA
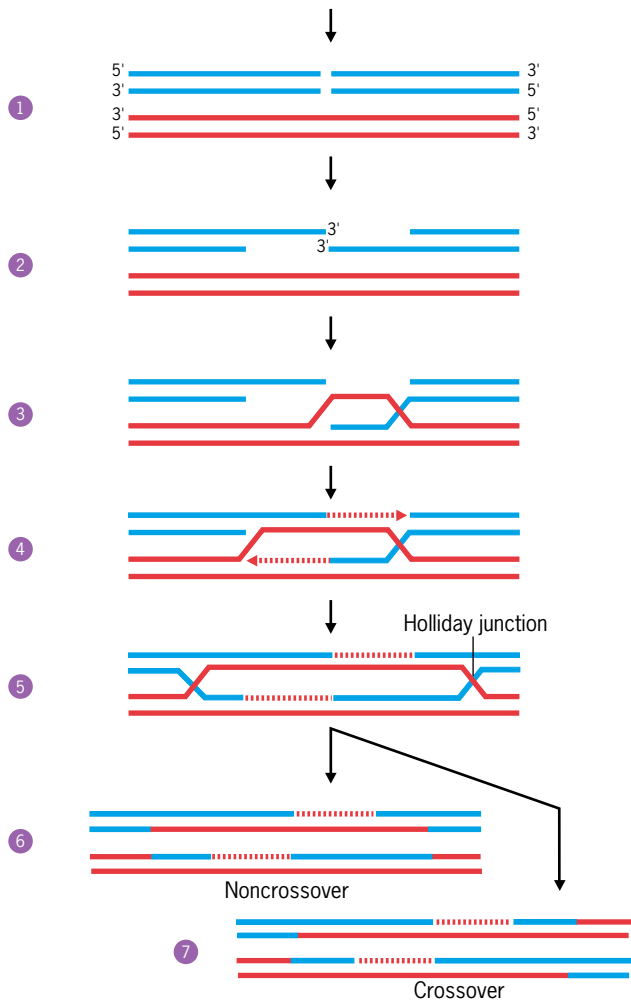
FIGURE 14.47 A proposed mechanism for genetic recombination initiated by double-strand breaks. The steps are described in the text.

protein polymerizes to form a filament that binds along a length of single-stranded DNA. RecA enables the single-stranded DNA to search for and invade an homologous double helix. Eukaryotic cells have homologues of RecA (e.g., Rad51) that are thought to catalyze strand invasion. Strand invasion activates a DNA repair activity (Section 13.3) that fills the gaps as shown in step 4.

As a result of the reciprocal exchange of DNA strands, the two duplexes are covalently linked to one another to form a joint molecule (or *heteroduplex*) that contains a pair of DNA crossovers, or *Holliday junctions*, that flank the region of strand exchange (steps 4 and 5, Figure 14.47). These junctions are named after Robin Holliday, who proposed their existence in 1964. This type of recombination intermediate need not be a static structure because the point of linkage may move in one direction or another (an event known as *branch migration*) by breaking the hydrogen bonds holding the original pairs of strands and reforming hydrogen bonds between strands of the newly joined duplexes (step 5). Formation of Holliday junctions and branch migration occur during pachytene (Figure 14.42).

To resolve the interconnected DNA molecules of the Holliday junctions and restore the DNA back to two separate duplexes, another round of DNA cleavage must occur. Depending on the particular DNA strands that are cleaved and ligated, two alternate products can be generated. In one case, the two duplexes contain only short stretches of genetic exchange, which represents a noncrossover (step 6, Figure 14.47). In the alternate pathway of breakage and ligation, the duplex of one DNA molecule is covalently joined to the duplex of the homologous molecule, creating a site of genetic recombination (i.e., a crossover) (step 7). The decision as to whether a recombinational interaction will result in a crossover or a noncrossover is made long before the stage when the double Holliday junction is actually resolved. Crossovers, which represent the fusion of a maternal and paternal chromosome (step 7), develop into the chiasma required to hold the homologues together during meiosis I (page 594).

REVIEW ?

1. Contrast the overall roles of mitosis and meiosis in the lives of a plant or animal. How do the nuclei formed by these two processes differ from one another?

2. Contrast the events that take place during prophase I and prophase II of meiosis.

3. Contrast the timing of meiosis in spermatogenesis versus oogenesis.

4. What is the role of DNA strand breaks in genetic recombination?

EXPERIMENTAL PATHWAYS

The Discovery and Characterization of MPF

As an amphibian oocyte nears the end of oogenesis, the large nucleus (called a germinal vesicle) moves toward the periphery of the cell. In subsequent steps, the nuclear envelope disassembles, the compacted chromosomes become aligned along a metaphase plate near one end (the animal pole) of the oocyte, and the cell undergoes the first meiotic division to produce a large secondary oocyte and a small polar body. The processes of germinal vesicle breakdown and first meiotic division are referred to as *maturation* and can be induced in fully grown oocytes by treatment with the steroid hormone progesterone. The first sign of maturation in the hormone-treated amphibian oocyte is seen 13–18 hours following progesterone treatment as the germinal vesicle moves near the oocyte surface. Germinal vesicle breakdown soon follows, and the oocyte reaches metaphase of the second meiotic division by about 36 hours after hormone treatment. Progesterone induces maturation only if it is applied to the external medium surrounding the oocyte; if the hormone is injected into the oocyte, the oocyte shows no response.[1] It appears that the hormone acts at the cell surface to trigger secondary changes in the cytoplasm of the oocyte that lead to germinal vesicle breakdown and the other changes associated with maturation.

To learn more about the nature of the cytoplasmic change that was responsible for triggering maturation, Yoshio Masui of the University of Toronto and Clement Markert of Yale University began a series of experiments in which they removed cytoplasm from isolated frog oocytes at various stages following progesterone treatment and injected 40–60 nanoliters (nl) of the donor cytoplasm into fully grown, immature oocytes that had not been treated with the hormone.[2] They found that cytoplasm taken from oocytes during the first 12 hours following progesterone treatment had little or no effect on recipient oocytes. After this period, however, the cytoplasm gained the ability to induce maturation in the recipient oocyte. The cytoplasm from the donor oocyte was maximally effective about 20 hours after progesterone treatment, and its effectiveness declined by 40 hours (Figure 1). However, cytoplasm taken from early embryos continued to show some ability to induce oocyte maturation. Masui and Markert referred to the cytoplasmic substance(s) that induce

maturation in recipient oocytes as "maturation promoting factor," which became known as MPF.

Because it was assumed that MPF was involved specifically in triggering oocyte maturation, relatively little interest was paid at first to the substance or its possible mechanism of action. Then in 1978, William Wasserman and Dennis Smith of Purdue University published a report on the behavior of MPF during early amphibian development.[3] It had been assumed that MPF activity present in early embryos was simply a residue of activity that had been present in the oocyte. But Wasserman and Smith discovered that MPF activity undergoes dramatic fluctuations in cleaving eggs that correlate with changes in the cell cycle. It was found, for example, that cytoplasm taken from cleaving frog eggs within 30–60 minutes after fertilization contains little or no detectable MPF activity, as assayed by injection into immature oocytes (Figure 2). However, if cytoplasm is taken from an egg at 90 minutes after fertilization, MPF activity can again be demonstrated. MPF activity reaches a peak at 120 minutes after fertilization and starts to decline again at 150 minutes (Figure 2). At the time the eggs undergo their first cytokinesis at 180 minutes, no activity is detected in the eggs. Then, as the second cleavage cycle gets underway, MPF activity once again reappears, reaching a peak at 225 minutes postfertilization, and then declines again to a very low level. Similar results were found in *Xenopus* eggs, except that the fluctuations in MPF activity occur more rapidly than in *Rana* and correlate with the more rapid rate of cleavage divisions in the early *Xenopus* embryo. Thus, MPF activity disappears and reappears in both amphibian species on a time scale that correlates with the length of the cell cycle. In both species, the peak of MPF activity corresponds to the time of nuclear membrane breakdown and the entry of the cells into mitosis. These findings suggested that MPF does more than simply control the time of oocyte maturation and, in fact, may play a key role in regulating the cell cycle of dividing cells.

It became apparent about this same time that MPF activity is not limited to amphibian eggs and oocytes but is present in a wide variety of organisms. It was found, for example, that mammalian cells growing in culture also possess MPF activity as assayed by the
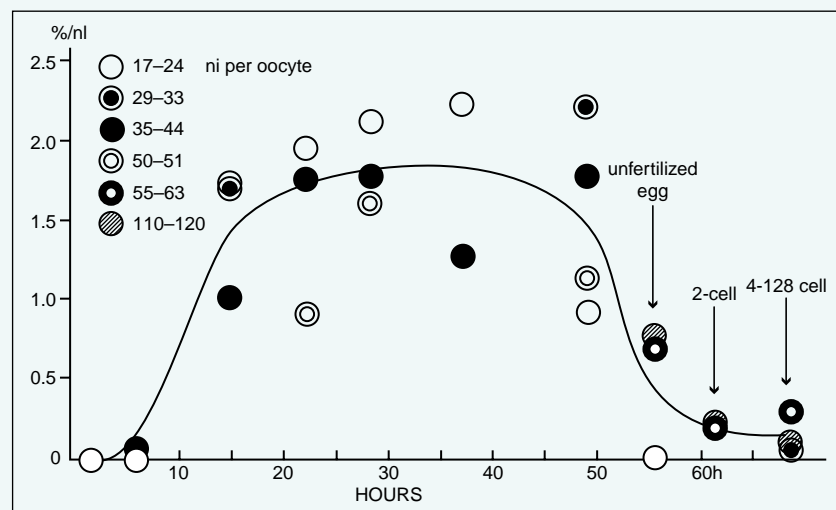
FIGURE 1 Change of activity of the maturation-promoting factor in the oocyte cytoplasm of *Rana pipiens* during the course of maturation and early development. Ordinate: the ratio of frequency of induced maturation to volume of injected cytoplasm. The higher the ratio, the more effective the cytoplasm. nl, nanoliters of injected cytoplasm. Abscissa: age of the donors (hours after administration of progesterone). (FROM Y. MASUI AND C. L. MARKERT, J. EXP. ZOOL. 177:142, 1971.)
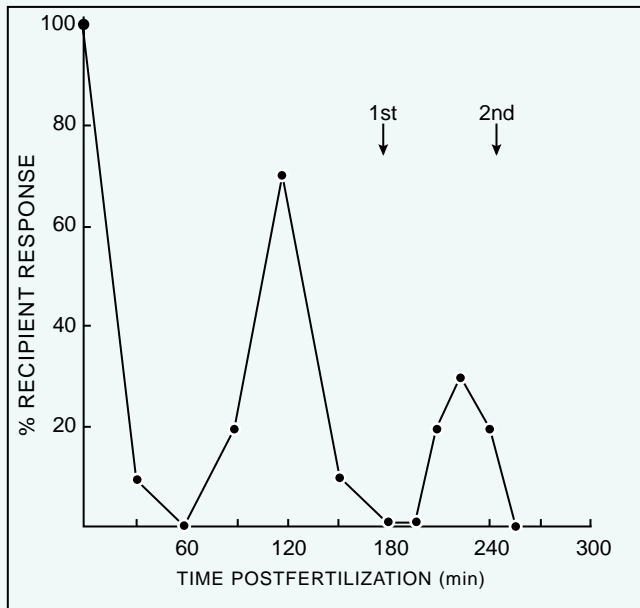
FIGURE 2 Cycling of MPF activity in fertilized *R. pipiens* eggs. Ordinate: percent recipient oocytes undergoing germinal vesicle breakdown in response to 80 nanoliters of cytoplasm from fertilized eggs. Abscissa: time after fertilization when the *Rana* egg cytoplasm was assayed for MPF activity. Arrows indicate time of cleavage divisions. (FROM W. J. WASSERMAN AND L. D. SMITH, J. CELL BIOL. 78:R17, 1978; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

FIGURE 3 Maturation-promoting activity of HeLa cell extracts during different stages of the cell cycle. Because 228 ng of mitotic protein induced germinal vesicle breakdown (GVBD) in 100 percent of the cases, the percent activity for other phases of the cell cycle was normalized to that amount of protein. E, early; M, mid-, and L, late. (FROM P. S. SUNKARA, D. A. WRIGHT, AND P. N. RAO, PROC. NAT'L. ACAD. SCI. U.S.A. 76:2801, 1979.)

ability of mammalian cell extracts to induce germinal vesicle breakdown when injected into amphibian oocytes.[4] MPF activity of mammalian cells fluctuates with the cell cycle, as it does in dividing amphibian eggs. Extracts from cultured HeLa cells prepared from early G_1-, late G_1-, or S-phase cells lack MPF activity (Figure 3). MPF appears in early G_2, rises dramatically in late G_2, and reaches a peak in mitosis.

Another element of the machinery that regulates the cell cycle was discovered in studies on sea urchin embryos. Sea urchin eggs are favorite subjects for studies of cell division because the mitotic divisions following fertilization occur rapidly and are separated by highly predictable time intervals. If sea urchin eggs are fertilized in seawater containing an inhibitor of protein synthesis, the eggs fail to undergo the first mitotic division, arresting at a stage prior to chromosome compaction and breakdown of the nuclear envelope. Similarly, each of the subsequent mitotic divisions can also be blocked if an inhibitor of protein synthesis is added to the medium at a time well before the division would normally occur. This finding had suggested that one or more proteins must be synthesized during each of the early cell cycles if the ensuing mitotic division is to occur. But early studies on cleaving sea urchin eggs failed to reveal the appearance of new species of proteins during this period.

In 1983, Tim Hunt and his colleagues at the Marine Biological Laboratory at Woods Hole reported on several proteins that are synthesized in fertilized sea urchin eggs but not unfertilized eggs.[5] To study these proteins further, they incubated fertilized eggs in seawater containing [^{35}S]methionine and withdrew samples at 10-minute intervals beginning at 16 minutes after fertilization. Crude protein extracts were prepared from the samples and subjected to polyacrylamide gel electrophoresis, and the labeled proteins were located au-

toradiographically. Several prominent bands were labeled in gels from fertilized egg extracts that were not evident in comparable extracts made from unfertilized eggs. One of the bands that appeared strongly labeled at early stages after fertilization virtually disappeared from the gel by 85 minutes after fertilization, suggesting that the protein had been selectively degraded. This same band then reappeared in gels from eggs sampled at later times and disappeared once again in a sample taken at 127 minutes after fertilization. The fluctuations in the amount of this protein are plotted in Figure 4 (protein band A) together with the cleavage index, which indicates the time course of the first two cell divisions. The degradation of the protein occurs at about the same time that the cells undergo the first and second division. A similar protein was found in the eggs of the surf clam, another invertebrate whose eggs are widely studied. Hunt and colleagues named the protein "cyclin" and noted the striking parallel in behavior between the fluctuations in cyclin levels in their investigation and MPF activity in the earlier studies. Subsequent studies showed that there were two distinct cyclins, A and B, which are degraded at different times during the cell cycle. Cyclin A is degraded during a 5–6 minute period beginning just before the metaphase–anaphase transition, and cyclin B is degraded a few minutes after this transition. These studies provided the first indication of the importance of controlled proteolysis (page 566) in the regulation of a major cellular activity.

The first clear link between cyclin and MPF was demonstrated by Joan Ruderman and her colleagues at the Woods Hole Marine Biological Laboratory.[6] In these studies, an mRNA encoding cyclin
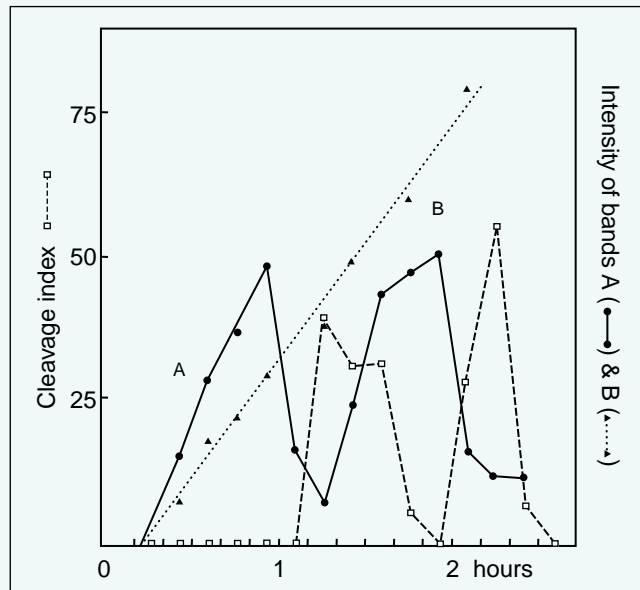
FIGURE 4 Correlation of the level of cyclin with the cell division cycle. A suspension of eggs was fertilized, and after 6 minutes, [^{35}S] methionine was added. Samples were taken for analysis by gel electrophoresis at 10-minute intervals, starting at 16 minutes after fertilization. The autoradiograph of the electrophoretic gel was scanned for label density, and the data were plotted as shown. Protein A, which varies according to the cell cycle and is named cyclin, is shown as solid circles. Protein B (not to be confused with cyclin B), which shows no cell cycle fluctuation, is plotted as solid triangles. The percentage of cells undergoing division at any given time period is given as the cleavage index (open squares). (FROM T. EVANS ET AL., CELL 33:391, 1983; BY PERMISSION OF CELL PRESS.)

FIGURE 5 Kinetics of *Xenopus* oocyte activation by progesterone and cyclin A mRNA. Large, immature oocytes were isolated from fragments of ovaries and were incubated with progesterone or microinjected with varying amounts of cyclin A mRNA. At 3 to 4 hours after injection, obviously damaged oocytes were removed (about 2–4 per starting group of 20), and the remaining ones (which represent the 100% value) were allowed to develop. Germinal vesicle breakdown (GVBD) and oocyte activation were indicated by the formation of a white spot in the region of the animal pole and were confirmed by dissection of the oocytes. (FROM K. I. SWENSON, K. M. FARRELL, AND J. V. RUDERMAN, CELL 47:865, 1986; BY PERMISSION OF CELL PRESS.)

A was transcribed in vitro from a cloned DNA fragment that contained the entire cyclin A coding sequence. The identity of this mRNA was verified by translating it in vitro and finding that it encoded authentic clam cyclin A. When the synthetic cyclin mRNA was injected into *Xenopus* oocytes, the cells underwent germinal vesicle breakdown and chromosome compaction over a time course not unlike that induced by progesterone treatment (Figure 5). These results suggested that the rise in cyclin A, which occurs normally during meiosis and mitosis, has a direct role in promoting entry into M phase. The amount of cyclin A normally drops rapidly and must be resynthesized prior to the next division or the cells cannot reenter M phase.

But what is the relationship between cyclins and MPF? One of the difficulties in answering this question was the use of different organisms. MPF had been studied primarily in amphibians, and cyclins in sea urchins and clams. Evidence indicated that frog oocytes contain a pool of inactive pre-MPF molecules, which are converted to active MPFs during meiosis I. Cyclin, on the other hand, is totally absent from clam oocytes but appears soon after fertilization. Ruderman considered the possibility that cyclin A is an activator of MPF. We will return to this shortly.

Meanwhile, another line of research was initiated to purify and characterize the substance responsible for MPF activity. In 1980, Michael Wu and John Gerhart of the University of California, Berkeley, accomplished a 20- to 30-fold purification of MPF by precipitating the protein in ammonium sulfate and subjecting the redissolved material to column chromatography. In addition to stim-

ulating oocyte maturation, injections of the partially purified MPF stimulated the incorporation of ^{32}P into proteins of the amphibian oocyte.[7] When partially purified MPF preparations were incubated with [^{32}P]ATP in vitro, proteins present within the sample became phosphorylated, suggesting that MPF induced maturation by acting as a protein kinase.

MPF was finally purified in 1988 by a series of six successive chromatographic steps.[8] MPF activity in these purified preparations was consistently associated with two polypeptides, one having a molecular mass of 32 kDa and the other of 45 kDa. The purified MPF preparation possessed a high level of protein kinase activity, as determined by the incorporation of radioactivity from [^{32}P]ATP into proteins. When the purified preparation was incubated in the presence of [^{32}P]ATP, the 45-kDa polypeptide became labeled.

By the end of the 1980s, the efforts to uncover the role of cyclins and MPF had begun to merge with another line of research that had been conducted on fission yeast by Paul Nurse and his colleagues at the University of Oxford.[9] It had been shown that yeast produced a protein kinase with a molecular weight of 34 kDa whose activity was required for these cells to enter M phase (discussed on page 564). The yeast protein was called p34^{cdc2} or simply cdc2. The first evidence of a link between cdc2 and MPF came as the result of a collaboration between yeast and amphibian research groups.[10,11] Recall from the previous study that MPF was found to contain a 32- and 45-kDa protein. Antibodies formed against cdc2 from fission yeast were shown to react specifically with the 32-kDa component of MPF isolated from *Xenopus* eggs. These findings indicate that this

component of MPF is a homologue of the 34-kDa yeast kinase and, therefore, that the machinery controlling the cell cycle in yeast and vertebrates contains evolutionarily conserved components.

A similar study using antibodies against yeast cdc2 showed that the homologous protein in vertebrates does not fluctuate during the cell cycle.[12] This supports the proposal that the 32-kDa protein kinase in vertebrate cells depends on another protein. The modulator was predicted to be cyclin, which rises in concentration during each cell cycle and is then destroyed as cells enter anaphase. This proposal was subsequently verified in a number of studies in which the MPF was purified from amphibians, clams, and starfish, and its polypeptide composition analyzed.[13–15] In all of these cases, it was shown that the active MPF present in M-phase animal cells is a complex consisting of two types of subunits: (1) a 32-kDa subunit that contains the protein kinase active site and is homologous to the yeast cdc2 protein kinase, and (2) a larger subunit (45 kDa) identified as a cyclin whose presence is required for kinase activity. The studies described in this Experimental Pathway provided a unified view of the regulation of the cell cycle in all eukaryotic organisms. In addition, they set the stage for analysis of the numerous factors that control the activity of MPF (cdc2) at various points during yeast and mammalian cell cycles, which have become the focus of attention in the past few years. Many of the most important findings of these recent studies are discussed in the first section of this chapter.

References

1. SMITH, L. D. & ECKER, R. E. 1971. The interaction of steroids with *R. pipiens* oocytes in the induction of maturation. *Dev. Biol.* 25:233–247.

2. MASUI, Y. & MARKERT, C. L. 1971. Cytoplasmic control of nuclear behavior during meiotic maturation of frog oocytes. *J. Exp. Zool.* 177:129–146.

3. WASSERMAN, W. J. & SMITH, L. D. 1978. The cyclic behavior of a cytoplasmic factor controlling nuclear membrane breakdown. *J. Cell Biol.* 78:R15–R22.

4. SUNKARA, P. S., WRIGHT, D. A., & RAO, P. N. 1979. Mitotic factors from mammalian cells induce germinal vesicle breakdown and chromosome condensation in amphibian oocytes. *Proc. Nat'l. Acad. Sci. U.S.A.* 76:2799–2802.

5. EVANS, T., ET AL. 1983. Cyclin: A protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell* 33:389–396.

6. SWENSON, K. I., FARRELL, K. M., & RUDERMAN, J. V. 1986. The clam embryo protein cyclin A induces entry into M phase and the resumption of meiosis in *Xenopus* oocytes. *Cell* 47:861–870.

7. WU, M. & GERHART, J. C. 1980. Partial purification and characterization of the maturation-promoting factor from eggs of *Xenopus laevis. Dev. Biol.* 79:465–477.

8. LOHKA, M. J., HAYES, M. K., & MALLER, J. L. 1988. Purification of maturation-promoting factor, an intracellular regulator of early mitotic events. *Proc. Nat'l. Acad. Sci. U.S.A.* 85:3009–3013.

9. NURSE, P. 1990. Universal control mechanism regulating onset of M-phase. *Nature* 344:503–507.

10. GAUTIER, J., ET AL. 1988. Purified maturation-promoting factor contains the product of a *Xenopus* homolog of the fission yeast cell cycle control gene *cdc2* 1. *Cell* 54:433–439.

11. DUNPHY, W. G., ET AL. 1988. The *Xenopus* cdc2 protein is a component of MPF, a cytoplasmic regulator of mitosis. *Cell* 54:423–431.

12. LABBE, J. C., ET AL. 1988. Activation at M-phase of a protein kinase encoded by a starfish homologue of the cell cycle control gene *cdc2. Nature* 335:251–254.

13. LABBE, J. C., ET AL. 1989. MPF from starfish oocytes at first meiotic metaphase is a heterodimer containing one molecule of cdc2 and one molecule of cyclin B. *EMBO J.* 8:3053–3058.

14. DRAETTA, G., ET AL. 1989. cdc2 protein kinase is complexed with both cyclin A and B: Evidence for proteolytic inactivation of MPF. *Cell* 56:829–838.

15. GAUTIER, J., ET AL. 1990. Cyclin is a component of maturation-promoting factor from *Xenopus. Cell* 60:487–494.

SYNOPSIS

The stages through which a cell passes from one cell division to the next constitute the cell cycle. The cell cycle is divided into two major phases: M phase, which includes the process of mitosis, in which duplicated chromosomes are separated into two nuclei, and cytokinesis, in which the entire cell is physically divided into two daughter cells; and interphase. Interphase is typically much longer than M phase and is subdivided into three distinct phases based on the time of replication, which is confined to a definite period within the cell cycle. G_1 is the period following mitosis and preceding replication; S is the period during which DNA synthesis (and histone synthesis) occurs; and G_2 is the period following replication and preceding the onset of mitosis. The length of the cell cycle, and the phases of which it is made up, vary greatly from one type of cell to another. Certain types of terminally differentiated cells, such as vertebrate skeletal muscle and nerve cells, have lost the ability to divide. (p. 561)

Early studies showed that entry of a cell into M phase is triggered by the activation of a protein kinase called MPF. MPF consists of two subunits: a catalytic subunit that transfers phosphate groups to specific serine and threonine residues of specific protein substrates, and a regulatory subunit consisting of a member of a family of proteins called cyclins. The catalytic subunit is called a cyclin-dependent kinase (Cdk). When the cyclin concentration is low, the kinase lacks the cyclin subunit and is inactive. When the cyclin concentration reaches a sufficient level, the kinase is activated, triggering the entry of the cell into M phase. (p. 563)

The activities that control the cell cycle are focused primarily at two points: the transition between G_1 and S and the transition between G_2 and the entry into mitosis. Passage through each of these points requires the transient activation of a Cdk by a specific cyclin. In yeast, the same Cdk is active at both G_1–S and G_2–M, but it is stimulated by different cyclins. The concentrations of the various cyclins rise and fall during the cell cycle as the result of changes in the rate of synthesis and destruction of the protein molecules. In addition to regulation by cyclins, Cdk activity is also controlled by the state of phosphorylation of the catalytic subunit, which in turn is controlled by at least two kinases (CAK and Wee1) and a phosphatase (Cdc25). In mammalian cells, at least eight different cyclins and a half-dozen different Cdks play a role in cell cycle regulation. (p. 564)

Cells possess surveillance mechanisms that monitor the status of cell cycle events, such as replication and chromosome compaction, and determine whether or not the cycle continues. If a cell is subjected to treatments that damage DNA, cell cycle progression is delayed until the damage is repaired. Arrest of a cell at one of the checkpoints of the cell cycle is accomplished by inhibitors whose

synthesis is stimulated by events such as DNA damage. Once the cell has been stimulated by external agents to traverse the G_1-S checkpoint and initiate replication, the cell generally continues without further external stimulation through the subsequent mitosis. (p. 567)

Mitosis ensures that two daughter nuclei receive a complete and equivalent complement of genetic material. Mitosis is divided into prophase, prometaphase, metaphase, anaphase, and telophase. Prophase is characterized by the preparation of chromosomes for segregation and the assembly of the machinery required for chromosome movement. Mitotic chromosomes are highly compacted, rod-shaped structures. Each mitotic chromosome can be seen to be split longitudinally into two chromatids, which are duplicates of one another formed by replication during the previous S phase. The primary constriction on the mitotic chromosome marks the centromere, which houses a platelike structure, the kinetochore, to which the microtubules of the spindle attach. During spindle formation, the centrosomes move away from one another toward the poles. As this happens, the microtubules that stretch between them increase in number and elongate. Eventually, the two centrosomes reach opposite points in the cell, establishing the two poles. A number of types of cells, including those of plants, assemble a mitotic spindle in the absence of centrosomes. The end of prophase is marked by the rupture of the nuclear envelope. (p. 569)

During prometaphase and metaphase, individual chromosomes are attached to spindle microtubules emanating from both poles and then moved to a plane at the center of the spindle. At the beginning of prometaphase, microtubules from the forming spindle penetrate into the region that had been the nucleus and make attachments to the kinetochores of the condensed chromosomes. The kinetochores soon become stably associated with the plus ends of microtubules from both poles of the spindle. Eventually, each chromosome is moved into position along a plane at the center of the spindle, a process that is accompanied by the shortening of some microtubules due to loss of tubulin subunits and the lengthening of others due to addition of subunits. Once the chromosomes are stably aligned, the cell has reached metaphase. The mitotic spindle of a typical animal cell at metaphase consists of astral microtubules, which radiate out from the centrosome; chromosomal microtubules, which are attached to the kinetochores; and polar microtubules that extend past the chromosomes and form a structural basket that maintains the integrity of the spindle. The microtubules of the metaphase spindle exhibit dynamic activity as demonstrated by the movement of fluorescently labeled subunits. (p. 576)

During anaphase and telophase, sister chromatids are separated from one another into separate regions of the dividing cell, and the chromosomes are returned to their interphase condition. Anaphase begins as the sister chromatids suddenly split away from one another, a process that is triggered by the ubiquitin-mediated destruction of cohesin, a protein complex that holds the sisters together. The separated chromosomes then move toward their respective poles accompanied by the shortening of the attached chromosomal microtubules, which results from the net loss of subunits at both the poles and kinetochore. The movement of the chromosomes, which is called anaphase A, is typically accompanied by the elongation of the mitotic spindle and consequent separation of the poles, which is referred to as anaphase B. Telophase is characterized by the reformation of the nuclear envelope, the dispersal of the chromosomes, and the reformation of membranous cytoplasmic networks. (p. 579)

Cytokinesis, which is the division of the cytoplasm into two daughter cells, occurs by constriction in animal cells and by construction in plant cells. Animal cells are constricted in two by an indentation or furrow that forms at the surface of the cell and moves inward. The advancing furrow contains a band of actin filaments, which slide over one another, driven by small, force-generating myosin II filaments. The site of cytokinesis is thought to be selected by a signal diffusing from the mitotic spindle. Plant cells undergo cytokinesis by building a cell membrane and cell wall in a plane lying between the two poles. The first sign of cell plate formation is the appearance of clusters of interdigitating microtubules and interspersed, electron-dense material. Small vesicles then move into the region and become aligned into a plane. The vesicles fuse with one another to form a membranous network that develops into a cell plate. (p. 585)

Meiosis is a process that includes two sequential nuclear divisions, producing haploid daughter nuclei that contain only one member of each pair of homologous chromosomes, thus reducing the number of chromosomes in half. Meiosis can occur at different stages in the life cycle, depending on the type of organism. To ensure that each daughter nucleus has only one set of homologues, an elaborate process of chromosome pairing occurs during prophase I that has no counterpart in mitosis. The pairing of chromosomes is accompanied by the formation of a proteinaceous, ladder-like structure called the synaptonemal complex (SC). The chromatin of each homologue is intimately associated with one of the lateral bars of the SC. During prophase I, homologous chromosomes engage in genetic recombination that produces chromosomes with new combinations of maternal and paternal alleles. Following recombination, the chromosomes become further condensed, and the homologues remain attached to one another at specific points called chiasmata, which represent sites of recombination. The paired homologues (called bivalents or tetrads) become oriented at the metaphase plate so that both chromatids of one chromosome face the same pole. During anaphase I, homologous chromosomes separate from one another with the maternal and paternal chromosomes of each tetrad segregating independently. The cells, which are now haploid in chromosome content, progress into the second meiotic division during which the sister chromatids of each chromosome separate into different daughter nuclei. (p. 590)

Genetic recombination during meiosis occurs as the result of the breakage and reunion of DNA strands from different homologues of a tetrad. During recombination homologous regions on different DNA strands are exchanged without the addition or loss of a single base pair. In an initial step, two duplexes become aligned next to one another. Once they are aligned, breaks occur in both strands of one of the duplexes. In the following steps, a DNA strand from one duplex invades the other duplex, forming an interconnected structure. Subsequent steps may include the activity of nucleases and polymerases to create and fill gaps in the strands similar to the way in which DNA repair occurs. (p. 597)

ANALYTIC QUESTIONS

1. In what way is cell division a link between humans and the earliest eukaryotic cells?
2. What types of synthetic events would you expect to occur in G_1 that do not occur in G_2?

3. Suppose you are labeling a population of cells growing asynchronously with [^{3}H]thymidine. G_1 is 6 hours, S is 6 hours, G_2 is 5 hours, and M is 1 hour. What percentage of the cells would be labeled after a 15-minute pulse? What percentage of mitotic

cells would be labeled after such a pulse? How long would you have to chase these cells before you saw any labeled mitotic chromosomes? What percentage of the cells would have labeled mitotic chromosomes if you chased the cells for 18 hours?

4. Suppose you take a culture of the same cells used in the previous question, but instead of pulse-labeling them with [^{3}H]thymidine, you labeled them continuously for 20 hours. Plot a graph showing the amount of radioactive DNA that would be present in the culture over the 20-hour period. What would be the minimum amount of time needed to ensure that all cells had some incorporated label? How could you determine the length of the cell cycle in this culture without use of a radioactive label?

5. Fusing G_1- and S-phase cells produces different results from fusing a G_2-phase cell with one in S. What would you expect this difference to be and how can it be explained? (*Hint:* Look back at Figure 13.20, which shows that the initiation of replication requires formation of a prereplication complex, which can only form in G_1.)

6. Figure 14.6 shows the effect on the cell cycle of mutations in the genes that encode Wee1 and Cdc25. The kinase CAK was identified biochemically rather than genetically (i.e., by isolation of mutant cells). What phenotype would you expect in a yeast cell carrying a temperature-sensitive CAK mutation, after raising the temperature of the culture medium in early G_1? in late G_2? Why might it be different depending upon the stage at which the temperature was raised?

7. Give four distinct mechanisms by which a Cdk can be inactivated.

8. A syncytium is a "cell" that contains more than one nucleus; examples are a skeletal muscle fiber and a blastula of a fly embryo. These two types of syncytia arise by very different pathways. What two mechanisms can you envision that could lead to formation of syncytia? What does this tell you about the relationship between mitosis and cytokinesis?

9. How could you determine experimentally if the microtubules of the polar microtubules were in a state of dynamic flux during anaphase? Knowing the events that occur at this stage, what would you expect to see?

10. If you were to add [^{3}H]thymidine to a cell as it underwent replication (S phase) prior to beginning meiosis, what percentage of the chromosomes of the gametes produced would be labeled? If one of these gametes (a sperm) were to fertilize an unlabeled egg, what percentage of the chromosomes of the two-cell stage would be labeled?

11. If the haploid number of chromosomes in humans is 23 and the amount of nuclear DNA in a sperm is 1C, how many chromosomes does a human cell possess in the following stages:

metaphase of mitosis, prophase I of meiosis, anaphase I of meiosis I, prophase II of meiosis, anaphase II of meiosis. How many chromatids does the cell have at each of these stages? How much DNA (in terms of numbers of C) does the cell have at each of these stages?

12. Plot the amount of DNA in the nucleus of a spermatogonia from the G_1 stage prior to the first meiotic division through the completion of meiosis. Label each of the major stages of the cell cycle and of meiosis on the graph.

13. How many centrioles does a cell have at metaphase of mitosis?

14. Suppose you were told that most cases of trisomy result from aging of an egg in the oviduct as it awaits fertilization. What type of evidence could you obtain from examining spontaneously aborted fetuses that would confirm this suggestion? How would this fit with data that has already been collected?

15. Suppose you incubate a meiotic cell in [^{3}H]thymidine between the leptotene and zygotene stages, then you fix the cell during pachytene and prepare an autoradiograph. You find that the chiasmata are sites of concentrations of silver grains. What does this tell you about the mechanism of recombination?

16. What type of phenotype would you expect of a cell whose Cdc20 polypeptide was mutated so that (1) it was unable to bind to Mad2, or (2) it was unable to bind to the other subunits of the APC, or (3) it failed to dissociate from the APC at the end of anaphase?

17. Assume for a moment that crossing-over did not occur. Would you agree that you received half of your chromosomes from each parent? Would you agree that you received one-quarter of your chromosomes from each grandparent? Would the answer to these questions change if you allowed for crossing-over to have occurred?

18. Fetuses whose cells are triploid, that is, contain three full sets of chromosomes, develop to term and die as infants, whereas fetuses with individual chromosome trisomies tend not to fare as well. How can you explain this observation?

19. What type of phenotype would you expect of a fission yeast cell whose Cdk subunit was lacking each of the following residues as the result of mutation: Tyr 15, Thr 161?

20. It was noted on page 574 that centrosome duplication and DNA synthesis are both initiated by cyclin E–Cdk2, which becomes active at the end of G_1. A recent study found that if cyclin E–Cdk2 is activated at an earlier stage, such as the beginning of G_1, that centrosome duplication begins at that point in the cell cycle, but that DNA replication is not initiated until S phase would normally begin. Provide a hypothesis to explain why DNA synthesis does not begin as well. You might look back at Figure 13.20 for further information.

15
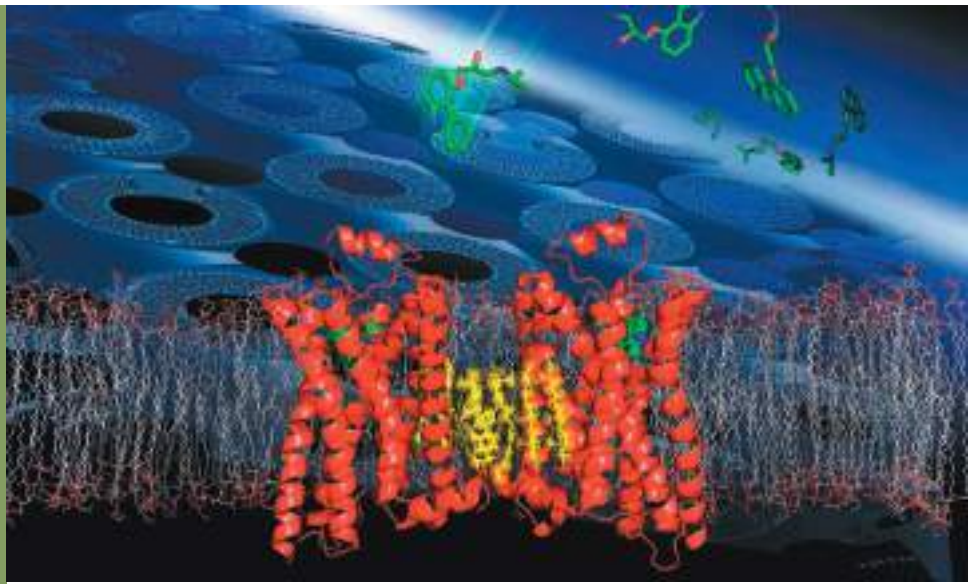
Cell Signaling and Signal Transduction: Communication Between Cells

T he English poet John Donne expressed his belief in the interdependence of humans in the phrase "No man is an island." The same can be said of the cells that make up a complex multicellular organism. Most cells in a plant or animal are specialized to carry out one or more specific functions. Many biological processes require various cells to work together and to coordinate their activities. To make this possible, cells have to communicate with each other, which is accomplished by a process called **cell signaling**. Cell signaling makes it possible for cells to respond in an appropriate manner to a specific environmental stimulus.

Three-dimensional, X-ray crystallographic structure of a β_2-adrenergic receptor (β_2-AR), which is a representative member of the G protein-coupled receptor (GPCR) superfamily. These integral membrane proteins are characterized as containing seven transmembrane helices. As a group, these proteins bind an astonishing array of biological messengers, which constitutes the first step in eliciting many of the body's most basic responses. The β_2-AR is a resident of the plasma membrane of a variety of cells, where it normally binds the ligand epinephrine and mediates such responses as increased heart rate and relaxation of smooth muscle cells. Beta-adrenergic receptors are the targets of a number of important drugs, including β-blockers, which are widely prescribed for the treatment of high blood pressure and heart arrhythmias. GPCRs have been very difficult to crystallize so that high-resolution structures of these important proteins have been lacking. This situation is now changing as the result of recent advances in crystallization technology, and it is hoped that these new high-resolution structures will lead to the development of new classes of structure-designed drugs. The image shown here depicts two β_2-ARs, which were crystallized in the presence of cholesterol and palmitic acid (yellow) and a receptor-binding ligand (green). (FROM VADIM CHEREZOV ET AL., COURTESY OF RAYMOND C. STEVENS, SCIENCE 318:1258, 2007; © COPYRIGHT 2007, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

Cell signaling affects virtually every aspect of cell structure and function, which is one of the primary reasons that this chapter appears near the end of the book. On one hand, an understanding of cell signaling requires knowledge about other types of cellular activity. On the other hand, insights into cell signaling can tie together a variety of seemingly independent cellular processes. Cell signaling is also intimately involved in the regulation of cell growth and division. This makes the study of cell signaling crucially important for understanding how a cell can lose the ability to control cell division and develop into a malignant tumor. ∎

15.1 THE BASIC ELEMENTS OF CELL SIGNALING SYSTEMS

It may be helpful to begin the discussion of this complex subject by describing a few of the general features that are shared by most signaling pathways. Cells usually communicate with each other through **extracellular messenger molecules**. Extracellular messengers can travel a short distance and stimulate cells that are in close proximity to the origin of the message, or they can travel throughout the body, potentially stimulating cells that are far away from the source. In the case of *autocrine* signaling, the cell that is producing the messenger expresses receptors on its surface that can respond to that messenger (Figure 15.1*a*). Consequently, cells releasing the message will stimulate (or inhibit) themselves. During *paracrine* stimulation (Figure 15.1*b*), messenger molecules travel only short distances through the extracellular space to cells that are in close proximity to the cell that is generating the message. Paracrine messenger molecules are usually limited in their ability to travel around the body because they are inherently unstable, or they are degraded by enzymes, or they bind to the extracellular matrix. Finally, during *endocrine* signaling, messenger molecules reach their target cells via passage through the bloodstream (Figure 15.1*c*). Endocrine messengers are also called *hormones*, and they typically act on target cells located at distant sites in the body.

An overview of cellular signaling pathways is depicted in Figure 15.2. Cell signaling is initiated with the release of a messenger molecule by a cell that is engaged in sending messages to other cells in the body (step 1, Figure 15.2). Cells can only respond to an extracellular message if they express **receptors** that specifically recognize and bind that particular messenger molecule (step 2). In most cases, the messenger molecule (or ligand) binds to a receptor at the extracellular surface of the responding cell. This interaction causes a signal to be relayed across the membrane to the receptor's cytoplasmic domain (step 3). Once it has reached the inner surface of the plasma membrane, there are two major routes by which the signal is transmitted into the cell interior, where it elicits the appropriate response. The particular route taken depends on the type of receptor that is activated. In the following discussion, we will focus on these two major routes of signal transduction, but keep in mind there are other ways that extracellular signals can have an impact on a cell. For example, we saw on page 164 how neurotransmitters act by opening plasma membrane ion channels and on page 514 how steroid hormones diffuse through the plasma membrane and bind to intracellular receptors. In the two major routes discussed in this chapter:

- One type of receptor (Section 15.3) transmits a signal from its cytoplasmic domain to a nearby enzyme (step 4), which generates a **second messenger** (step 5). Because it brings about (effects) the cellular response by generating a second messenger, the enzyme responsible is referred to as an **effector**. Second messengers are small substances that typically activate (or inactivate) specific proteins. Depending on its chemical structure, a second messenger may diffuse through the cytosol or remain embedded in the lipid bilayer of a membrane.

- Another type of receptor (Section 15.4) transmits a signal by transforming its cytoplasmic domain into a recruiting station for cellular signaling proteins (step 4a). Proteins interact with one another, or with components of a cellular membrane, by means of specific types of interaction domains, such as the SH3 domain discussed on page 60.

Whether the signal is transmitted by a second messenger or by protein recruitment, the outcome is similar; a protein that is positioned at the top of an intracellular **signaling pathway** is activated (step 6, Figure 15.2). Signaling pathways are the information superhighways of the cell. Each signaling

(a)   (b)   (c)
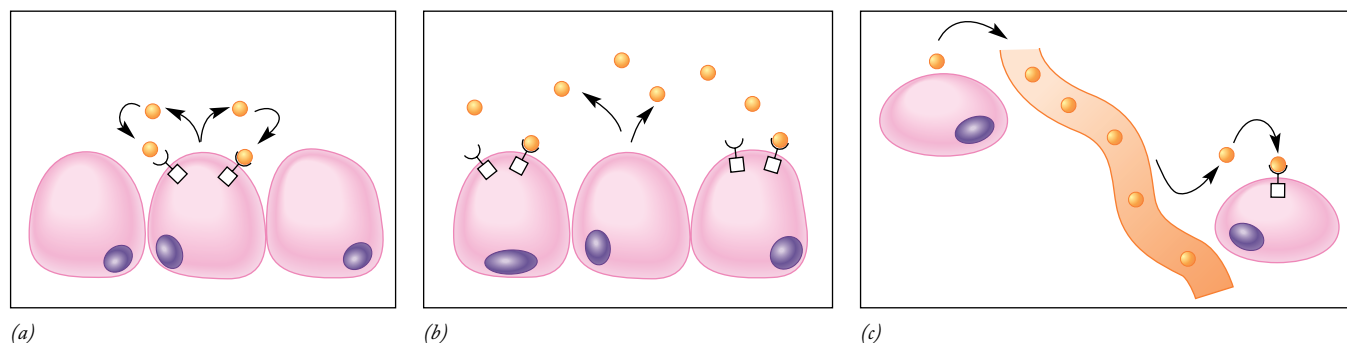
FIGURE 15.1 Autocrine (*a*), paracrine (*b*), and endocrine (*c*) types of intercellular signaling.
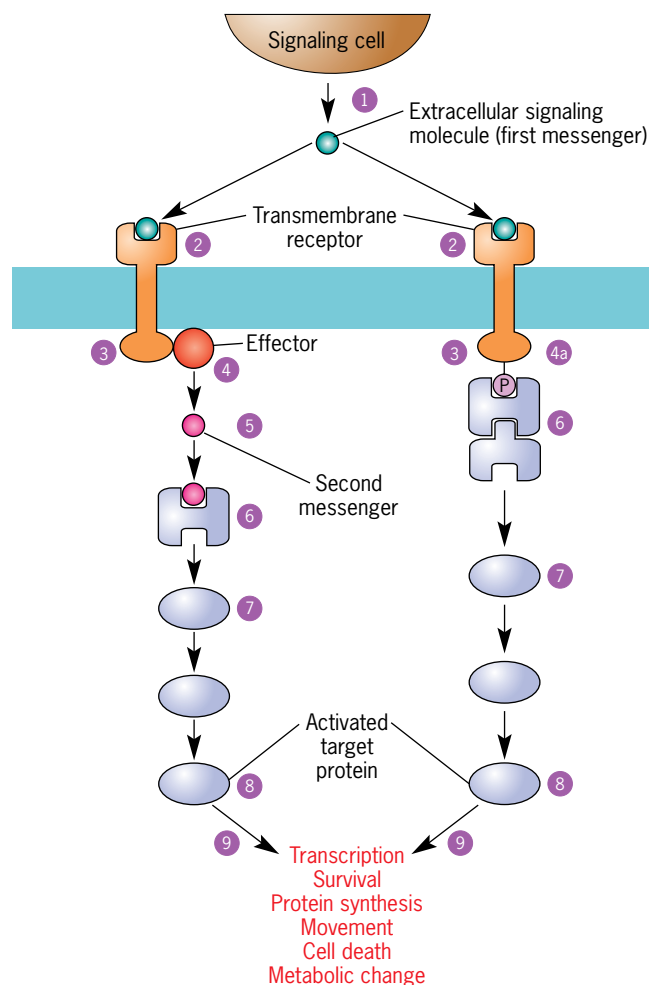
FIGURE 15.2 An overview of the major signaling pathways by which extracellular messenger molecules can elicit intracellular responses. Two different types of signal transduction pathways are depicted, one in which a signaling pathway is activated by a diffusible second messenger and another in which a signaling pathway is activated by recruitment of proteins to the plasma membrane. Most signal transduction pathways involve a combination of these mechanisms. It should also be noted that signaling pathways are not typically linear tracks as depicted here, but are branched and interconnected to form a complex web. The steps are described in the text.

FIGURE 15.3 Signal transduction pathway consisting of protein kinases and protein phosphatases whose catalytic actions change the conformations, and thus the activities, of the proteins they modify. In the example depicted here, protein kinase 2 is activated by protein kinase 1. Once activated, protein kinase 2 phosphorylates protein kinase 3, activating the enzyme. Protein kinase 3 then phosphorylates a transcription factor, increasing its affinity for a site on the DNA. Binding of a transcription factor to the DNA affects the transcription of the gene in question. Each of these activation steps in the pathway is reversed by a phosphatase. Whereas protein kinases typically work as a single subunit, many protein phosphatases contain a key regulatory subunit that helps determine substrate specificity.

pathway consists of a series of distinct proteins that operate in sequence (step 7). Each protein in the pathway typically acts by altering the conformation of the subsequent (or downstream) protein in the series, an event that activates or inhibits that protein (Figure 15.3). It should come as no surprise, after reading about other topics in cell biology, that alterations in the conformation of signaling proteins are often accomplished by protein kinases and protein phosphatases that, respectively, add or remove phosphate groups from other proteins (Figure 15.3). The human genome encodes more than 500 different protein kinases and approximately 150 different protein phosphatases. Most protein kinases transfer phosphate groups to serine or threonine residues of their protein substrates, but a very important group of kinases phosphory-

lates tyrosine residues. Some protein kinases and phosphatases are soluble cytoplasmic proteins; others are integral membrane proteins. It is remarkable that, even though thousands of proteins in a cell contain amino acid residues with the potential of being phosphorylated, each protein kinase or phosphatase is able to recognize only its specific substrates and ignore all of the others. Some protein kinases and phosphatases have numerous proteins as their substrates, whereas others phosphorylate or dephosphorylate only a single amino acid residue of a single protein substrate. Many of the protein substrates of these enzymes are enzymes themselves—most often other kinases and phosphatases—but the substrates also include ion channels, transcription factors, and various types of regulatory proteins. It is thought that at least 50 percent of transmembrane and cytoplasmic proteins are phosphorylated at one or more sites. Protein phosphorylation can change protein behavior in several different ways. Phosphorylation can activate or inactivate an enzyme, it can increase or decrease protein–protein interactions, it can induce a protein to move from one subcellular compartment to another, or it can

act as a signal that initiates protein degradation. Large-scale proteomic approaches have been employed (page 68) to identify the potential substrates of various protein kinases. The primary challenge is to understand the roles of these diverse posttranslational modifications in the activities of different cell types.

Signals transmitted along such signaling pathways ultimately reach *target proteins* (step 8, Figure 15.2) involved in basic cellular processes (step 9). Depending on the type of cell and message, the response initiated by the target protein may involve a change in gene expression, an alteration of the activity of metabolic enzymes, a reconfiguration of the cytoskeleton, an increase or decrease in cell mobility, a change in ion permeability, activation of DNA synthesis, or even the death of the cell. Virtually every activity in which a cell is engaged is regulated by signals originating at the cell surface. This overall process in which information carried by extracellular messenger molecules is translated into changes that occur inside a cell is referred to as **signal transduction**.

Finally, signaling has to be terminated. This is important because cells have to be responsive to additional messages that they may receive. The first order of business is to eliminate the extracellular messenger molecule. To do this, certain cells produce extracellular enzymes that destroy specific extracellular messengers. In other cases, activated receptors are internalized (page 612). Once inside the cell, the receptor may be degraded together with its ligand, which can leave the cell with decreased sensitivity to subsequent stimuli. Alternatively, receptor and ligand may be separated within an endosome, after which the ligand is degraded and the receptor is returned to the cell surface.

REVIEW   ?

1. What is meant by the term *signal transduction*? What are some of the steps by which signal transduction can occur?
2. What is a second messenger? Why do you suppose it is called this?

15.2 A SURVEY OF EXTRACELLULAR MESSENGERS AND THEIR RECEPTORS

A large variety of molecules can function as extracellular carriers of information. These include

- Amino acids and amino acid derivatives. Examples include glutamate, glycine, acetylcholine, epinephrine, dopamine, and thyroid hormone. These molecules act as neurotransmitters and hormones.
- Gases, such as NO and CO.
- Steroids, which are derived from cholesterol. Steroid hormones regulate sexual differentiation, pregnancy, carbohydrate metabolism, and excretion of sodium and potassium ions.

- Eicosanoids, which are nonpolar molecules containing 20 carbons that are derived from a fatty acid named arachidonic acid. Eicosanoids regulate a variety of processes including pain, inflammation, blood pressure, and blood clotting. Several over-the-counter drugs that are used to treat headaches and inflammation inhibit eicosanoid synthesis.
- A wide variety of polypeptides and proteins. Some of these are present as transmembrane proteins on the surface of an interacting cell (page 247). Others are part of, or associate with, the extracellular matrix. Finally, a large number of proteins are excreted into the extracellular environment where they are involved in regulating processes such as cell division, differentiation, the immune response, or cell death and cell survival.

Extracellular signaling molecules are usually, but not always, recognized by specific receptors that are present on the surface of the responding cell. As illustrated in Figure 15.2, receptors bind their signaling molecules with high affinity and translate this interaction at the outer surface of the cell into changes that take place on the inside of the cell. The receptors that have evolved to mediate signal transduction are indicated below.

- G protein-coupled receptors (GPCRs) are a huge family of receptors that contain seven transmembrane α helices. These receptors translate the binding of extracellular signaling molecules into the activation of GTP-binding proteins. **GTP-binding proteins** (or **G proteins**) were discussed in connection with vesicle budding and fusion in Chapter 8, microtubule dynamics in Chapter 9, protein synthesis in Chapters 8 and 11, and nucleocytoplasmic transport in Chapter 12. In the present chapter, we will explore their role in transmitting messages along "cellular information circuits."
- Receptor protein-tyrosine kinases (RTKs) represent a second class of receptors that have evolved to translate the presence of extracellular messenger molecules into changes inside the cell. Binding of a specific extracellular ligand to an RTK usually results in receptor dimerization followed by activation of the receptor's protein-kinase domain, which is present within its cytoplasmic region. Upon activation, these protein kinases phosphorylate specific tyrosine residues of cytoplasmic substrate proteins, thereby altering their activity, their localization, or their ability to interact with other proteins within the cell.
- Ligand-gated channels represent a third class of cell-surface receptors that bind to extracellular ligands. The ability of these proteins to conduct a flow of ions across the plasma membrane is regulated directly by ligand binding. A flow of ions across the membrane can result in a temporary change in membrane potential, which will affect the activity of other membrane proteins, for instance, voltage-gated channels. This sequence of events is the basis for formation of a nerve impulse (page 162). In addition, the influx of certain ions, such as Ca^{2+}, can change the activity of particular cytoplasmic enzymes. As dis-

cussed in Section 4.8, one large group of ligand-gated channels functions as receptors for neurotransmitters.

■ Steroid hormone receptors function as ligand-regulated transcription factors. Steroid hormones diffuse across the plasma membrane and bind to their receptors, which are present in the cytoplasm. Hormone binding results in a conformational change that causes the hormone–receptor complex to move into the nucleus and bind to elements present in the promoters or enhancers of hormone-responsive genes (see Figure 12.43). This interaction gives rise to an increase or decrease in the rate of gene transcription.

■ Finally, there are a number of other types of receptors that act by unique mechanisms. Some of these receptors, for example, the B- and T-cell receptors that are involved in the response to foreign antigens, associate with known signaling molecules such as cytoplasmic protein-tyrosine kinases. We will concentrate in this chapter on the GPCRs and RTKs.

15.3 G PROTEIN-COUPLED RECEPTORS AND THEIR SECOND MESSENGERS

G protein-coupled receptors (**GPCRs**) are so named because they interact with G proteins, as discussed below. Members of the GPCR superfamily are also referred to as seven-transmembrane (7TM) receptors because they contain seven transmembrane helices (Figure 15.4). Thousands of different GPCRs have been identified in organisms ranging from yeast to flowering plants and mammals that together regulate an extraordinary spectrum of cellular processes. In fact, GPCRs constitute the single largest superfamily of proteins encoded by animal genomes. Included among the natural ligands that bind to GPCRs are a diverse array of hormones (both plant and animal), neurotransmitters, opium derivatives, chemoattractants (e.g., molecules that attract phagocytic cells of the immune system), odorants and tastants (molecules detected by olfactory and gustatory receptors elic-

(a)

(b)

(c)

FIGURE 15.4 The membrane-bound machinery for transducing signals by means of a seven transmembrane receptor and a heterotrimeric G protein. (*a*) Receptors of this type, including those that bind epinephrine and glucagon, contain seven membrane-spanning helices. When bound to their ligand, the receptor interacts with a trimeric G protein, which activates an effector, such as adenylyl cyclase. As indicated in the figure, the α and γ subunits of the G protein are linked to the membrane by lipid groups that are embedded in the lipid bilayer. (*Note:* Many GPCRs may be active as complexes of two or more receptor molecules.) (*b*) A model depicting the activation of the GPCR rhodopsin based on recent X-ray crystallographic structures. On the left, rhodopsin is shown in its inactive (dark-adapted) conformation together with an unbound heterotrimeric G protein (called transducin). When the retinal cofactor (shown in red on the left rhodopsin molecule) absorbs a photon, it undergoes an isomerization reaction (from a *cis* to a *trans* form), which leads to the disruption of an ionic linkage between residues on the third and sixth transmembrane helix of the protein. This event in turn leads to a change in conformation of the protein, including the outward movement of the sixth transmembrane helix (red curved arrow), which exposes a binding site for the G_α subunit of the G protein. The rhodopsin molecule on the right is shown in the proposed active conformation with a portion of the G_α subunit (in red) bound to the

receptor's cytoplasmic face. (*c*) Ribbon model showing the structure of the heterotrimeric G protein. The three subunits of the G protein are color coded. (B: FROM THUE W. SCHWARTZ AND WAYNE L. HUBBELL, NATURE 455, 473, 2008; COPYRIGHT © 2008, BY MACMILLAN JOURNALS LIMITED; C: FROM HEIDI E. HAMM, PROC. NAT'L. ACAD. SCI. U.S.A. 98:4819, 2001.)

TABLE 15.1 Examples of Physiologic Processes Mediated by GPCRs and Heterotrimeric G Proteins

Stimulus	Receptor	Effector	Physiologic response
Epinephrine	β-Adrenergic receptor	Adenylyl cyclase	Glycogen breakdown
Serotonin	Serotonin receptor	Adenylyl cyclase	Behavioral sensitization and learning in *Aplysia*
Light	Rhodopsin	cGMP phosphodiesterase	Visual excitation
IgE–antigen complexes	Mast cell IgE receptor	Phospholipase C	Secretion
f-Met Peptide	Chemotactic receptor	Phospholipase C	Chemotaxis
Acetylcholine	Muscarinic receptor	Potassium channel	Slowing of pacemaker activity

Adapted from L. Stryer and H. R. Bourne, reproduced with permission from the *Annual Review of Cell Biology*, vol. 2, © 1986, by Annual Reviews Inc.

iting the senses of smell and taste), and photons. A list of some of the ligands that operate by means of this pathway and the effectors through which they act is provided in Table 15.1.

Signal Transduction by G Protein-Coupled Receptors

Receptors   G protein-coupled receptors normally have the following topology. Their amino-terminus is present on the outside of the cell, the seven α helices that traverse the plasma membrane are connected by loops of varying length, and the carboxyl-terminus is present on the inside of the cell (Figure 15.4). There are three loops present on the outside of the cell that, together, form the ligand-binding site. There are also three loops present on the cytoplasmic side of the plasma membrane that provide binding sites for intracellular signaling proteins. G proteins bind to the third intracellular loop. Arrestins, whose function is described on page 612, also bind to the third intracellular loop and compete with G proteins for binding to the receptor. Finally, there is a growing number of proteins that bind to the carboxyl-termini of GPCRs. Many of these proteins act as molecular scaffolds that link receptors to various signaling proteins and effectors present in the cell.

It is very difficult, for a number of technical reasons, to prepare crystals of GPCRs that are suitable for X-ray crystallographic analysis. For a number of years, rhodopsin was the only member of the superfamily to have its X-ray crystal structure determined. Rhodopsin has an unusually stable structure for a GPCR, owing to the fact that its ligand (a retinal group) is permanently bound to the protein and the protein molecule can only exist in a single conformation in the absence of a stimulus (i.e., in the dark). Beginning in 2007, as a result of years of effort by a number of research groups, a flurry of GPCR crystal structures appeared in the literature. For the most part, these structures revealed the GPCR in the inactive state, but there is also a body of structural and spectroscopic data (of the type described on page 132) that provide insights into some of the conformational changes that occur as a GPCR is activated and becomes bound to a G protein. The inactive conformation is stabilized by noncovalent interactions between specific residues in the transmembrane α helices. Ligand binding disturbs these interactions, thereby causing the receptor to assume an active conformation. This requires rotations and shifts of the transmembrane α helices relative to each other. Because they are attached to the cytoplasmic loops, rotation or movement of these transmembrane α helices causes

changes in the conformation of the cytoplasmic loops. This in turn leads to an increase in the affinity of the receptor for a G protein that is present on the cytoplasmic surface of the plasma membrane (Figure 15.4b). As a consequence, the ligand-bound receptor forms a receptor–G protein complex (Figure 15.5, step 1). The interaction with the receptor induces a conformational change in the α subunit of a G protein, causing the release of GDP, which is followed by binding of GTP (step 2). While in the activated state, a single receptor can activate a number of G protein molecules, providing a means of signal amplification (discussed further on page 620).

G Proteins   Heterotrimeric G proteins were discovered, purified, and characterized by Martin Rodbell and his colleagues at the National Institutes of Health and Alfred Gilman and colleagues at the University of Virginia. Their studies are discussed in the Experimental Pathways, which can be found on the Web at www.wiley.com/college/karp. These proteins are referred to as G proteins because they bind guanine nucleotides, either GDP or GTP. They are described as heterotrimeric because all of them consist of three different polypeptide subunits, called α, β, and γ (Figure 15.4). This property distinguishes them from small, monomeric G proteins, such as Ras, which are discussed later in this chapter. Heterotrimeric G proteins are held at the plasma membrane by lipid chains that are covalently attached to the α and γ subunits (Figure 15.4a).

The guanine nucleotide-binding site is present on the G_α subunit. Replacement of GDP by GTP, following interaction with an activated GPCR, results in a conformational change in the G_α subunit. In its GTP-bound conformation, the G_α subunit has a low affinity for $G_{\beta\gamma}$, leading to its dissociation from the complex. Each dissociated G_α subunit (with GTP attached) is free to activate an effector protein, such as adenylyl cyclase (Figure 15.5, step 3). In this case, activation of the effector leads to the production of the second messenger cAMP (step 4). Other effectors include phospholipase C-β and cyclic GMP phosphodiesterase (see below). Second messengers, in turn, activate one or more cellular signaling proteins.

A G protein is said to be "on" when its α subunit is bound to GTP. G_α subunits can turn themselves off by hydrolysis of GTP to GDP and inorganic phosphate (Pi) (Figure 15.5, step 5). This results in a conformational change causing a decrease in the affinity for the effector and an increase in the affinity for the βγ subunit. Thus, following hydrolysis of

FIGURE 15.5 **The mechanism of receptor-mediated activation (or inhibition) of effectors by means of heterotrimeric G proteins.** In step 1, the ligand binds to the receptor, altering its conformation and increasing its affinity for the G protein to which it binds. In step 2, the G_α subunit releases its GDP, which is replaced by GTP. In step 3, the G_α subunit dissociates from the $G_{\beta\gamma}$ complex and binds to an effector (in this case adenylyl cyclase), activating the effector. The $G_{\beta\gamma}$ dimer may also bind to an effector (not shown), such as an ion channel or an enzyme. In step 4, activated adenylyl cyclase produces cAMP. In step 5, the GTPase activity of G_α hydrolyzes the bound GTP, deactivating G_α. In step 6, G_α reassociates with $G_{\beta\gamma}$, reforming the trimeric G protein, and the effector ceases its activity. In step 7, the receptor has been phosphorylated by a GRK and in step 8 the phosphorylated receptor has been bound by an arrestin molecule, which inhibits the ligand-bound receptor from activating additional G proteins. The receptor bound to arrestin is likely to be taken up by endocytosis.

GTP, the G_α subunit will dissociate from the effector and reassociate with the $\beta\gamma$ subunit to reform the inactive heterotrimeric G protein (step 6). In a sense, heterotrimeric G proteins function as molecular timers. They are turned on by the interaction with an activated receptor and turn themselves off by hydrolysis of bound GTP after a certain amount of time has passed. While they are active, G_α subunits can turn on downstream effectors.

Heterotrimeric G proteins come in four flavors, G_s, G_q, G_i, and $G_{12/13}$. This classification is based on the G_α subunits and the effectors to which they couple. The particular response elicited by an activated GPCR depends on the type of G protein with which it interacts, although some GPCRs can interact with different G proteins and trigger more than one physiologic response. G_s family members couple receptors to adenylyl cyclase. Adenylyl cyclase is activated by GTP-bound G_s subunits. G_q family members contain G_α subunits that activate PLCβ. PLCβ hydrolyzes phosphatidylinositol bisphosphate, producing inositol trisphosphate and diacylglycerol (page 616). Activated G_i subunits function by inhibiting adenylyl cyclase. $G_{12/13}$ members are less well characterized than the other G protein families although their inappropriate activation has been associated with excessive cell proliferation and malignant transformation. Following its dissociation from the G_α subunit, the $\beta\gamma$ complex also has a signaling function and it can couple to at least four different types of effectors: PLCβ, K^+ ion channels, adenylyl cyclase, and PI 3-kinase.

Termination of the Response We have seen that ligand binding results in receptor activation. The activated receptors turn on G proteins, and G proteins turn on effectors. To prevent overstimulation, receptors have to be blocked from continuing to activate G proteins. To regain sensitivity to future stimuli, the receptor, the G protein, and the effector must all be returned to their inactive state. *Desensitization*, the process that blocks active receptors from turning on additional G proteins, takes place in two steps. In the first step, the cytoplasmic domain of the activated GPCR is phosphorylated by a specific type of kinase, called *G protein-coupled receptor kinase* (*GRK*) (Figure 15.5, step 7). GRKs form a small family of serine-threonine protein kinases that specifically recognize activated GPCRs.
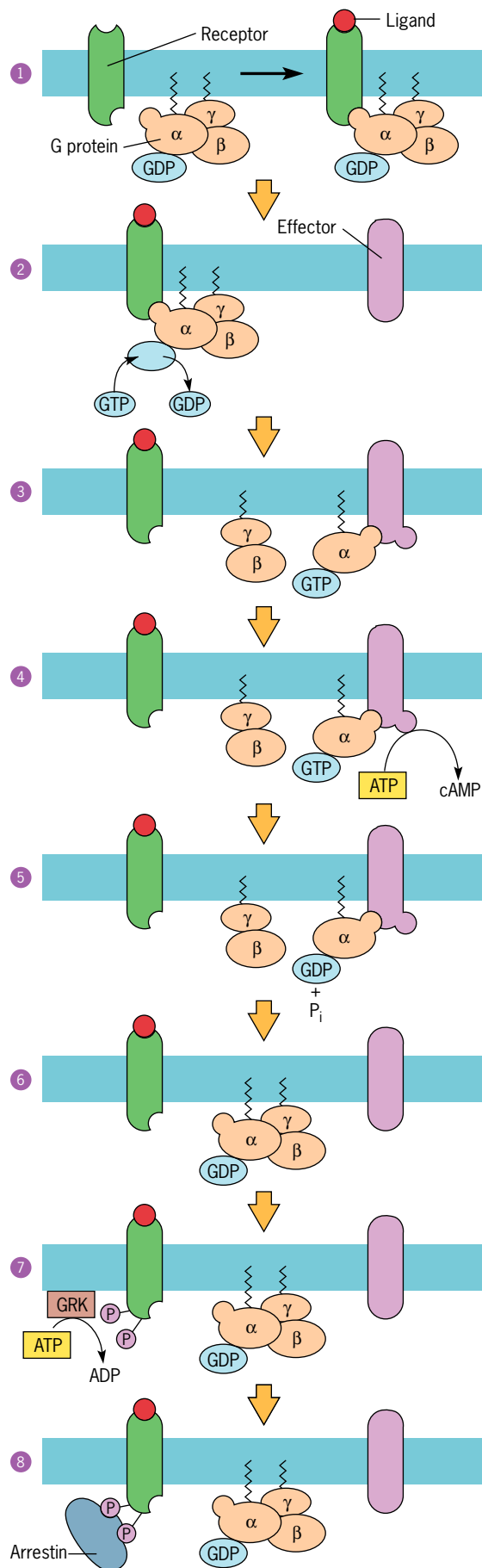
Phosphorylation of the GPCR sets the stage for the second step, which is the binding of proteins, called *arrestins* (Figure 15.5, step 8). Arrestins form a small group of proteins that bind to GPCRs and compete for binding with heterotrimeric G proteins. As a consequence, arrestin binding prevents the further activation of additional G proteins. This action is termed desensitization because the cell stops responding to the stimulus, while that stimulus is still acting on the outer surface of the cell. Desensitization is one of the mechanisms that allows a cell to respond to a change in its environment, rather than continuing to "fire" endlessly in the presence of an unchanging environment. The importance of desensitization is illustrated by the observation that mutations that interfere with phosphorylation of rhodopsin by a GRK lead to the death of the photoreceptor cells in the retina. This type of retinal cell death is thought to be one of the causes of blindness resulting from the disease retinitis pigmentosa.

While they are bound to phosphorylated GPCRs, arrestin molecules are also capable of binding to clathrin molecules that are situated in clathrin-coated pits (page 302). The interaction between bound arrestin and clathrin promotes the uptake of phosphorylated GPCRs into the cell by endocytosis. Depending on the circumstances, receptors that have been removed from the surface by endocytosis may be dephosphorylated and returned to the plasma membrane. Alternatively, internalized receptors are degraded in the lysosomes (see Figure 8.42). If receptors are degraded, the cells lose, at least temporarily, sensitivity for the ligand in question. If receptors are returned to the cell surface, the cells remain sensitive to the ligand.

Signaling by the activated G_α subunit is terminated by a less complex mechanism: the bound GTP molecule is simply hydrolyzed to GDP (step 5, Figure 15.5). Thus, the strength and duration of the signal are determined in part by the rate of GTP hydrolysis by the G_α subunit. G_α subunits possess a weak GTPase activity, which allows them to slowly hydrolyze the bound GTP and inactivate themselves. Termination of the response is accelerated by *regulators of G protein signaling* (*RGSs*). The interaction with an RGS protein increases the rate of GTP hydrolysis by the G_α subunit. Once the GTP is hydrolyzed, the G_α-GDP reassociates with the $G_{\beta\gamma}$ subunits to reform the inactive trimeric complex (step 6) as discussed above. This returns the system to the resting state.

The mechanism for transmitting signals across the plasma membrane by G proteins is of ancient evolutionary origin and is highly conserved. This is illustrated by an experiment in which yeast cells were genetically engineered to express a receptor for the mammalian hormone somatostatin. When these yeast cells were treated with somatostatin, the mammalian receptors at the cell surface interacted with the yeast heterotrimeric G proteins at the inner surface of the membrane and triggered a response leading to proliferation of the yeast cells.

The effects of certain mutations on the function of G protein-coupled receptors are discussed in the accompanying Human Perspective.

THE HUMAN PERSPECTIVE

Disorders Associated with G Protein-Coupled Receptors

GPCRs represent the largest family of genes encoded by the human genome. Their importance in human biology is reflected by the fact that over one-third of all prescription drugs act as ligands that bind to this huge superfamily of receptors. A number of inherited disorders have been traced to defects in both GPCRs (Figure 1) and heterotrimeric G proteins (Table 1). Retinitis pigmentosa (RP) is an inherited disease characterized by progressive degeneration of the retina and eventual blindness. RP can be caused by mutations in the gene that encodes rhodopsin, the visual pigment of the rods. Many of these mutations lead to premature termination or improper folding of the rhodopsin protein and its elimination from the cell before it reaches the plasma membrane (page 282). Other mutations may

FIGURE 1 Two-dimensional representation of a "composite" transmembrane receptor showing the approximate sites of a number of mutations responsible for causing human diseases. Most of the mutations (numbers 1, 2, 5, 6, 7, and 8) result in constitutive stimulation of the effector, but others (3 and 4) result in blockage of the receptor's ability to stimulate the effector. Mutations at sites 1 and 2 are found in the MSH (melanocyte-stimulating hormone) receptor; 3 in the ACTH (adrenocorticotrophic hormone) receptor; 4 in the vasopressin receptor; 5 and 6 in the TSH (thyroid-stimulating hormone) receptor; 7 in the LH (luteinizing hormone) receptor; and 8 in rhodopsin, the light-sensitive pigment of the retina.

TABLE 1 Human Diseases Linked to the G Protein Pathway

Disease	Defective G protein*
Albright's hereditary osteodystrophy and pseudohypoparathyroidisms	$G_{s\alpha}$
McCune–Albright syndrome	$G_{s\alpha}$
Pituitary, thyroid tumors (*gsp* oncogene)	$G_{s\alpha}$
Adrenocortical, ovarian tumors (*gip* oncogene)	$G_{i\alpha}$
Combined precocious puberty and peudohypoparathyroidism	$G_{s\alpha}$

Disease	Defective G protein-coupled receptor
Familial hypocalciuric hypercalcemia	Human analogue of BoPCAR1 receptor
Neonatal severe hyperparathyroidism	Human analogue of BoPCAR1 receptor (homozygous)
Hyperthyroidism (thyroid adenomas)	Thyrotropin receptor
Familial male precocious puberty	Luteinizing hormone receptor
X-linked nephrogenic diabetes insipidus	V2 vasopressin receptor
Retinitis pigmentosa	Rhodopsin receptor
Color blindness, spectral sensitivity variations	Cone opsin receptor
Familial glucocorticoid deficiency and isolated glucocorticoid deficiency	Adrenocorticoptropic hormone (ACTH) receptor

*As described in the text, a G protein with a $G_{s\alpha}$ acts to stimulate the effector, whereas a G protein with a $G_{i\alpha}$ inhibits the effector.
Source: D. E. Clapham, reprinted with permission from *Nature*, vol. 371, p. 109, 1994. © Copyright 1994, by Macmillan Magazines Ltd.

lead to the synthesis of a rhodopsin molecule that cannot activate its G protein and thus cannot pass the signal downstream to the effector.

RP results from a mutation that leads to a loss of function of the encoded receptor. Many mutations that alter the structure of signaling proteins can have an opposite effect, leading to what is described as a "gain of function." In one such case, mutations have been found to cause a type of benign thyroid tumor, called an adenoma. Unlike normal thyroid cells that secrete thyroid hormone only in response to stimulation by the pituitary hormone TSH, the cells of these thyroid adenomas secrete large quantities of thyroid hormone without having to be stimulated by TSH (the receptor is said to act *constitutively*). The TSH receptor in these cells contains an amino acid substitution that affects the structure of the third intracellular loop of the protein (Figure 1, mutations at sites 5 or 6). As a result of the mutation, the TSH receptor constitutively activates a G protein on its inner surface, sending a continual signal through the pathway that leads not only to excessive thyroid hormone secretion but to the excessive cell proliferation that causes the tumor. This conclusion was verified by introducing the mutant gene into cultured cells that normally lack this receptor and demonstrating that the synthesis of the mutant protein and its incorporation into the plasma membrane led to the continuous production of cAMP in the genetically engineered cells.

The mutation that causes thyroid adenomas is not found in the normal portion of a patient's thyroid but only in the tumor tissue, indicating that the mutation was not inherited but arose in one of the cells of the thyroid, which then proliferated to give rise to the tumor. A mutation in a cell of the body, such as a thyroid cell, is called a *somatic mutation* to distinguish it from an inherited mutation that would be present in all of the individual's cells. As will be evident in the following chapter, somatic mutations are a primary cause of human cancer. At least one cancer-causing virus has been shown to encode a protein that acts as a constitutively active GPCR. The virus is a type of herpes virus that is responsible for Kaposi's sarcoma, which causes purplish skin lesions and is prevalent in AIDS patients. The virus genome encodes a constitutively active receptor for interleukin-8, which stimulates signaling pathways that control cell proliferation.

As noted in Table 1, mutations in genes that encode the subunits of heterotrimeric G proteins can also lead to inherited disorders. This is illustrated by a report on two male patients suffering from a rare combination of endocrine disorders: precocious puberty and hypoparathyroidism. Both patients were found to contain a single amino acid substitution in one of the G_α isoforms. The alteration in amino acid sequence caused two effects on the mutant G protein. At temperatures below normal body temperature, the mutant G protein remained in the active state, even in the absence of a bound ligand. In contrast, at normal body temperatures, the mutant G protein was inactive, both in the presence and absence of bound ligand. The testes, which are housed outside of the body's core, have a lower temperature than the body's visceral organs (33°C vs. 37°C). Normally, the endocrine cells of the testes initiate testosterone production at the time of puberty in response to the pituitary hormone LH, which begins to be produced at that time. The circulating LH binds to LH receptors on the surface of the testicular cells, inducing the synthesis of cAMP and subsequent production of the male sex hormone. The testicular cells of the patients bearing the G protein mutation were stimulated to synthesize cAMP in the absence of the LH ligand, leading to premature synthesis of testosterone and precocious puberty. In contrast, the mutation in this same G_α subunit in the cells of the parathyroid glands, which function at a temperature of 37°C, caused the G protein to remain inactive. As a result, the cells of the parathyroid gland could not respond to stimuli that would normally cause them to secrete parathyroid hormone, leading to the condition of hypoparathyroidism. The fact that most of the bodily organs functioned in a normal manner in these patients suggests that this particular G_α isoform is not essential in the activities of most other cells.

Mutations are thought of as rare and disabling changes in the nucleotide sequence of a gene. Genetic polymorphisms, in contrast, are thought of as common, "normal" variations within the population (page 408). Yet it has become clear in recent years that genetic polymorphism may have considerable impact on human disease, causing certain individuals to be more or less susceptible to particular disorders than other individuals. This has been well documented in the case of GPCRs. For example, certain alleles of the gene encoding the β_2 adrenergic receptor have been associated with an increased likelihood of developing asthma or high blood pressure; certain alleles of a dopamine receptor are associated with increased risk of substance abuse or schizophrenia; and certain alleles of a chemokine receptor (*CCR5*) are associated with prolonged survival in HIV-infected individuals. As discussed on page 410, identifying associations between disease susceptibility and genetic polymorphisms is a current focus of clinical research.

Bacterial Toxins   Because G proteins are so important to the normal physiology of multicellular organisms, they provide excellent targets for bacterial pathogens. For example, cholera toxin (produced by the bacterium *Vibrio cholerae*) exerts its effect by modifying G_α subunits and inhibiting their GTPase activity in the cells of the intestinal epithelium. As a result, adenylyl cyclase molecules remain in an activated mode, churning out cAMP, which causes the epithelial cells to secrete large volumes of fluid into the intestinal lumen. The loss of water associated with this inappropriate response often leads to death due to dehydration.

Pertussis toxin is one of several virulence factors produced by *Bordetella pertussis*, a microorganism that causes whooping cough. Whooping cough is a debilitating respiratory tract infection seen in 50 million people worldwide each year, causing death in about 350,000 of these cases annually. Pertussis toxin also inactivates G_α subunits, thereby interfering with the signaling pathway that leads the host to mount a defensive response against the bacterial infection.

Second Messengers

The Discovery of Cyclic AMP, a Prototypical Second Messenger

How does the binding of a hormone to the plasma membrane change the activity of cytoplasmic enzymes, such as glycogen phosphorylase, an enzyme involved in glycogen metabolism? The answer to this question was provided by studies that began in the mid-1950s in the laboratories of Earl Sutherland and his colleagues at Case Western Reserve University and Edwin Krebs and Edmond Fischer at the University of Washington. Sutherland's goal was to develop an in vitro system to study the physiologic responses to hormones. After considerable effort, he was able to activate glycogen phosphorylase in a preparation of *broken* cells that had been incubated with glucagon or epinephrine. This broken-cell preparation could be divided by centrifugation into a particulate fraction consisting primarily of cell membranes and a soluble supernatant fraction. Even though glycogen phosphorylase was present only in the supernatant fraction, the particulate material was required to obtain the hormone response. Subsequent experiments indicated that the response occurred in at least two distinct steps. If the particulate fraction of a liver homogenate was isolated and incubated with the hormone, some substance was released that, when added to the supernatant fraction, activated the soluble glycogen phosphorylase molecules. Sutherland identified the substance released by the membranes of the particulate fraction as cyclic adenosine monophosphate (*cyclic AMP*, or simply *cAMP*). This discovery is heralded as the beginning of the study of signal transduction. As will be discussed below, cAMP stimulates glucose mobilization by activating a protein kinase that adds a phosphate group onto a specific serine residue of the glycogen phosphorylase polypeptide.

Cyclic AMP is a **second messenger** that is capable of diffusing to other sites within the cell. The synthesis of cyclic AMP follows the binding of a first messenger—a hormone or other ligand—to a receptor at the outer surface of the cell. Figure 15.6 shows the diffusion of cyclic AMP within the cytoplasm of a neuron following stimulation by an extracellular messenger molecule. Whereas the first messenger binds exclusively to a single receptor species, the second messenger often stimulates a variety of cellular activities. As a result, second messengers enable cells to mount a large-scale, coordinated response following stimulation by a single extracellular ligand. Other second messengers include Ca^{2+}, phosphoinositides, inositol trisphosphate, diacylglycerol, cGMP, and nitric oxide.

Phosphatidylinositol-Derived Second Messengers

It wasn't very long ago that the phospholipids of cell membranes were considered strictly as structural molecules that made membranes cohesive and impermeable to aqueous solutes. Our appreciation of phospholipids has increased with the realization that these molecules form the precursors of a number of second messengers. Phospholipids of cell membranes
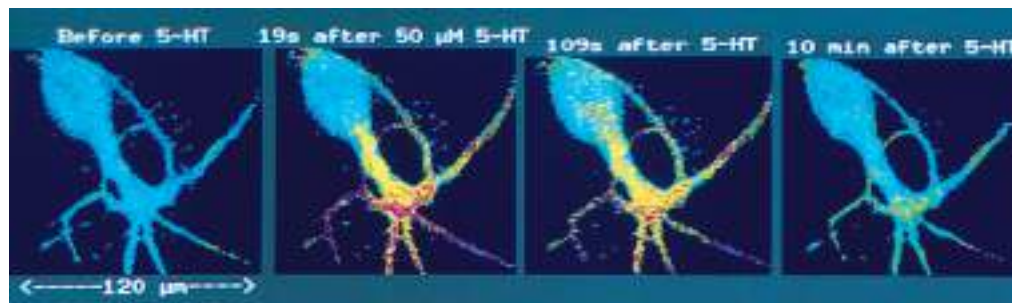
FIGURE 15.6 **The localized formation of cAMP in a live cell in response to the addition of an extracellular messenger molecule.** This series of photographs shows a sensory nerve cell from the sea hare *Aplysia*. The concentration of free cAMP is indicated by the color: blue represents a low cAMP concentration, yellow an intermediate concentration, and red a high concentration. The left image shows the intracellular cAMP level in the unstimulated neuron, and the next three images show the effects of stimulation by the neurotransmitter serotonin (5-hydroxytryptamine) at the times indicated. Notice that the cAMP levels drop by 109s despite the continued presence of the neurotransmitter. (The cAMP level was determined indirectly in this experiment by microinjection of a fluorescently labeled cAMP-dependent protein kinase labeled with both fluorescein and rhodamine on different subunits. Energy transfer between the subunits (see Figure 18.8) provides a measure of cAMP concentration.) (REPRINTED WITH PERMISSION FROM BRIAN J. BACSKAI ET AL., SCIENCE 260:223, 1993; © COPYRIGHT 1993, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

(a)   (b)   (c)
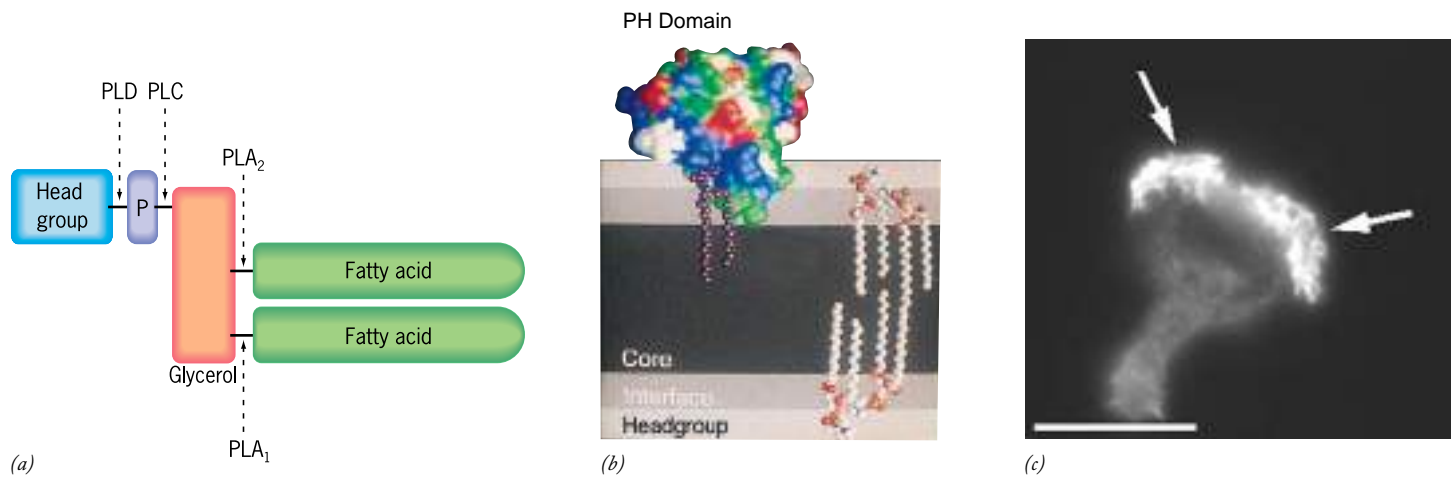
FIGURE 15.7 Phospholipid-based second messengers. (*a*) The structure of a generalized phospholipid (see Figure 2.22). Phospholipids are subject to attack by four types of phospholipases that cleave the molecule at the indicated sites. Of these enzymes, we will focus on PLC, which splits the phosphorylated head group from the diacylglycerol (see Figure 15.8). (*b*) A model showing the interaction between a portion of a PLC enzyme molecule containing a PH domain that binds to the phosphorylated inositol ring of a phosphoinositide. This interaction holds the enzyme to the inner surface of the plasma membrane and may alter its enzymatic activity. (*c*) Fluorescence micrograph of a cell that has been stimulated to move toward a chemoattractant (i.e., a chemical to which the cell is attracted). This cell has been stained with an antibody that binds specifically to PI 3,4,5-trisphosphate (PIP$_3$), which is seen to be localized at the leading edge of the migrating cell (arrows). Bar equals 15 μm. (B: FROM JAMES H. HURLEY AND JAY A. GROBLER, CURR. OPIN. STRUCT. BIOL. 7:559, 1997; C: FROM PAULA RICKERT ET AL., COURTESY OF HENRY R. BOURNE, TRENDS CELL BIOL. 10:470, 2000.)

are converted into second messengers by a variety of enzymes that are regulated in response to extracellular signals. These enzymes include phospholipases (lipid-splitting enzymes), phospholipid kinases (lipid-phosphorylating enzymes), and phospholipid phosphatases (lipid-dephosphorylating enzymes). Phospholipases are enzymes that hydrolyze specific ester bonds that connect the different building blocks that make up a phospholipid molecule. Figure 15.7a shows the cleavage sites within a generalized phospholipid that are attacked by the main classes of phospholipases. All four of the enzyme classes depicted in Figure 15.7a can be activated in response to extracellular signals and the products they produce function as second messengers. In this section, we will focus on the best-studied lipid second messengers, which are derived from phosphatidylinositol, and are generated following the transmission of signals by G protein-coupled receptors and receptor protein-tyrosine kinases. Another group of lipid second messengers derived from sphingomyelin is not discussed.

Phosphatidylinositol Phosphorylation When the neurotransmitter acetylcholine binds to the surface of a smooth muscle cell within the wall of the stomach, the muscle cell is stimulated to contract. When a foreign antigen binds to the surface of a mast cell, the cell is stimulated to secrete histamine, a substance that can trigger the symptoms of an allergy attack. Both of these responses, one leading to contraction and the other to secretion, are triggered by the same second messenger, a substance derived from the compound phosphatidylinositol, a minor component of most cellular membranes (see Figure 4.10).

The first indication that phospholipids might be involved in cellular responses to extracellular signals emerged from studies carried out in the early 1950s by Lowell and Mabel Hokin of Montreal General Hospital and McGill University. These investigators had set out to study the effects of acetylcholine on RNA synthesis in the pancreas. To carry out these studies, they incubated slices of pigeon pancreas in [^{32}P]orthophosphate. The idea was that [^{32}P]orthophosphate would be incorporated into nucleoside triphosphates, which are used as precursors during the synthesis of RNA. Interestingly, they found that treatment of the tissue with acetylcholine led to the incorporation of radioactivity into the phospholipid fraction of the cell. Further analysis revealed that the isotope was incorporated primarily into phosphatidylinositol (PI), which was rapidly converted to other phosphorylated derivatives, which are collectively referred to as **phosphoinositides**. This suggested that inositol-containing lipids can be phosphorylated by specific lipid kinases that are activated in response to extracellular messenger molecules, such as acetylcholine. It is now well established that lipid kinases are activated in response to a large variety of extracellular signals.

Several of the reactions of phosphoinositide metabolism are shown in Figure 15.8. As indicated on the left side of this figure, the inositol ring, which resides at the cytoplasmic surface of the bilayer, has six carbon atoms. Carbon number 1 is involved in the linkage between inositol and diacylglycerol. The 3, 4, or 5 carbons can be phosphorylated by specific phosphoinositide kinases present in cells. For example, transfer of a single phosphate group to the 4-position of the inositol sugar of PI by PI 4-kinase (PI4K) generates PI 4-phosphate (PI(4)P), which can be phosphorylated by PIP 5-kinase (PIP5K) to form PI 4,5-bisphosphate (PI(4,5)P$_2$; Figure 15.8, steps 1 and 2). PI(4,5)P$_2$ can be phosphorylated by PI 3-kinase (PI3K) to form PI(3,4,5)P$_3$ (PIP$_3$) (shown in Figure 15.23c). The phosphorylation of PI(4,5)P$_2$ to form PIP$_3$ is of particular interest because

the PI3K enzymes involved in this process can be controlled by a large variety of extracellular molecules and PI3K overactivity has been associated with human cancers. The formation of PIP_3 during the response to insulin is discussed on page 632.

All of the phospholipid species discussed above remain in the cytoplasmic leaflet of the plasma membrane; they are membrane-bound second messengers. Just as there are lipid kinases to add phosphate groups to phosphoinositides, there are lipid phosphatases to remove them. The activity of these kinases and phosphatases are coordinated so that specific phosphoinositides appear at specific regions of the membrane at specific times after a signal has been received. The role of specific phosphoinositides in membrane trafficking was discussed on page 304.

The phosphorylated inositol rings of phosphoinositides form binding sites for several lipid-binding domains found in proteins. Best known is the **PH domain** (Figure 15.7*b*), which has been identified in over 150 different proteins. Binding of a protein by its PH domain to PIP_2 or PIP_3 typically recruits the protein to the cytoplasmic face of the plasma membrane where it can interact with other membrane-bound proteins, including activators, inhibitors, or substrates. Figure 15.7*c* shows an example where PIP_3 is specifically localized to a particular portion of the plasma membrane of a cell. PIP_3 is produced at the front of the cell by a localized lipid kinase and is subsequently degraded at the rear and sides of the cell by a localized lipid phosphatase. The cell shown in Figure 15.7*c* is engaged in chemotaxis, which is to say that it is moving to-ward an increasing concentration of a particular chemical in the medium that serves as the chemoattractant. This is the mechanism that causes phagocytic cells such as macrophages to move toward bacteria or other targets that they engulf. Chemotaxis depends on the localized production of phosphoinositide messengers, which in turn influence the formation of actin filaments and lamellipodia that are required to move the cell in the direction of the target.

Phospholipase C Not all inositol-containing second messengers remain in the lipid bilayer of a membrane. When acetylcholine binds to a smooth muscle cell, or an antigen binds to a mast cell, the bound receptor activates a heterotrimeric G protein (Figure 15.8, step 3), which, in turn, activates the effector *phosphatidylinositol-specific phospholipase C-β* (PLCβ) (step 4). Like the protein depicted in Figure 15.7*b*, PLCβ is situated at the inner surface of the membrane (Figure 15.8), bound there by the interaction between its PH domain and a phosphoinositide embedded in the bilayer. PLCβ catalyzes a reaction that splits PIP_2 into two molecules, *inositol 1,4,5-trisphosphate* (*IP₃*) and *diacylglycerol* (*DAG*) (step 5, Figure 15.8), both of which play important roles as second messengers in cell signaling. We will examine each of these second messengers separately.

Diacylglycerol Diacylglycerol (Figure 15.8) is a lipid molecule that remains in the plasma membrane following its for-
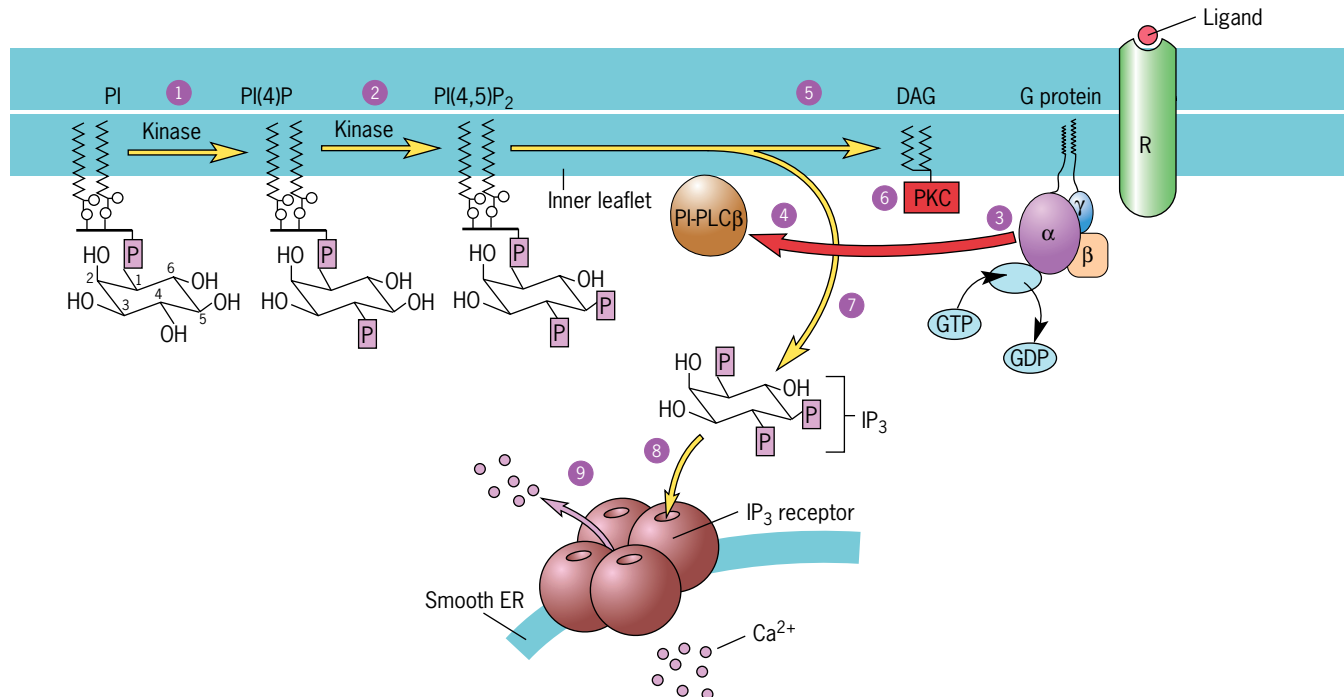
FIGURE 15.8 The generation of second messengers as a result of ligand-induced breakdown of phosphoinositides (PI) in the lipid bilayer. In steps 1 and 2, phosphate groups are added by lipid kinases to phosphatidylinositol (PI) to form PIP_2. When a stimulus is received by a receptor, the ligand-bound receptor activates a heterotrimeric G protein (step 3), which activates the enzyme PI-specific phospholipase C-β(step 4), which catalyzes the reaction in which PIP_2 is split into diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3) (step 5). DAG recruits the protein kinase PKC to the membrane and activates the enzyme (step 6). IP_3 diffuses into the cytosol (step 7), where it binds to an IP_3 receptor and Ca^{2+} channel in the membrane of the SER (step 8). Binding of IP_3 to its receptor causes release of calcium ions into the cytosol (step 9).

mation by PLCβ. There it recruits and activates effector proteins that bear a DAG-binding C1 domain. The best-studied of these proteins is *protein kinase C* (*PKC*) (step 6, Figure 15.8), which phosphorylates serine and threonine residues on a wide variety of target proteins.

Protein kinase C has a number of important roles in cellular growth and differentiation, cellular metabolism, cell death, and transcriptional activation (Table 15.2). The apparent importance of protein kinase C in growth control is seen in studies with a group of powerful plant compounds, called *phorbol esters*, that resemble DAG. These compounds activate protein kinase C in a variety of cultured cells, causing them to lose growth control and behave temporarily as malignant cells. When the phorbol ester is removed from the medium, the cells recover their normal growth properties. In contrast, cells that have been genetically engineered to constitutively express protein kinase C exhibit a permanent malignant phenotype in cell culture and can cause tumors in susceptible mice. Finally, application of phorbol esters to the skin in combination with certain other chemicals will cause the formation of skin tumors.

Inositol 1,4,5–trisphosphate (IP₃)   Inositol 1,4,5-trisphosphate (IP_3) is a sugar phosphate—a small, water-soluble molecule capable of rapid diffusion throughout the interior of the cell. IP_3 molecules formed at the membrane diffuse into the cytosol (step 7, Figure 15.8) and bind to a specific IP_3 receptor located at the surface of the smooth endoplasmic reticulum (step 8). It was noted on page 273, that the smooth endoplasmic reticulum is a site of calcium storage in a variety of cells. The IP_3 receptor also functions as a tetrameric Ca^{2+} channel. Binding of IP_3 opens the channel, allowing Ca^{2+} ions to diffuse into the cytoplasm (step 9). Calcium ions can also be considered as intracellular or second messengers because they bind to various target molecules, triggering specific responses. In the two examples used above, contraction of a smooth muscle cell and exocytosis of histamine-containing secretory granules in a mast cell, both are triggered by elevated calcium levels. So, too, is the response of a liver cell to the hormone vasopressin (the same hormone that has antidiuretic ac-
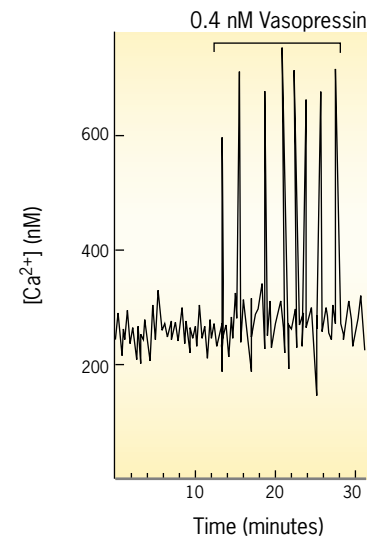
FIGURE 15.9 Experimental demonstration of changes in free calcium concentration in response to hormone stimulation. A single liver cell was injected with aequorin, a protein extracted from certain jellyfish that luminesces when it binds calcium ions. The intensity of the luminescence is proportional to the concentration of free calcium ions. Exposure of the cell to vasopressin leads to controlled spikes in the concentration of free calcium at periodic intervals. Higher concentrations of hormone do not increase the height (amplitude) of the spikes, but they do increase their frequency. (REPRINTED WITH PERMISSION FROM N. M. WOODS, K. S. CUTHBERTSON, AND P. H .COBBOLD, NATURE 319:601, 1986; COPYRIGHT © 1986, BY MACMILLAN JOURNALS LIMITED.)

tivity in the kidney, page 146). Vasopressin binds to its receptor at the liver cell surface and causes a series of IP_3-mediated bursts of Ca^{2+} release, which appear as oscillations of free cytosolic calcium in the recording shown in Figure 15.9. The frequency and intensity of such oscillations may encode information that governs the cell's specific response. A list of some of the responses mediated by IP_3 is indicated in Table 15.3. We will have more discussion about Ca^{2+} ions in Section 15.5.

TABLE 15.2 Examples of Responses Mediated by Protein Kinase C

Tissue	Response
Blood platelets	Serotonin release
Mast cells	Histamine release
Adrenal medulla	Secretion of epinephrine
Pancreas	Secretion of insulin
Pituitary cells	Secretion of GH and LH
Thyroid	Secretion of calcitonin
Testes	Testosterone synthesis
Neurons	Dopamine release
Smooth muscle	Increased contractility
Liver	Glycogen hydrolysis
Adipose tissue	Fat synthesis

Source: U. Kikkawa and Y. Nishizuka, reproduced with permission from the *Annual Review of Cell Biology*, vol. 2, © 1986, by Annual Reviews Inc.

TABLE 15.3 Summary of Cellular Responses Elicited by Adding IP₃ to Either Permeabilized or Intact Cells

Cell type	Response
Vascular smooth muscle	Contraction
Stomach smooth muscle	Contraction
Slime mold	Cyclic GMP formation, actin polymerization
Blood platelets	Shape change, aggregation
Salamander rods	Modulation of light response
Xenopus oocytes	Calcium mobilization, membrane depolarization
Sea urchin eggs	Membrane depolarization, cortical reaction
Lacrimal gland	Increased potassium current

Adapted from M. J. Berridge, reproduced with permission from the *Annual Review of Biochemistry*, vol. 56, © 1987, by Annual Reviews Inc.

The Specificity of G Protein-Coupled Responses

A wide variety of agents, including hormones, neurotransmitters, and sensory stimuli, act by way of GPCRs and heterotrimeric G proteins to transmit information across the plasma membrane, triggering a wide variety of cellular responses. This does not mean that all of the various parts of the signal transduction machinery are identical in every cell type. The receptor for a given ligand can exist in several different versions (isoforms). For example, researchers have identified 9 different isoforms of the adrenergic receptor, which binds epinephrine, and 15 different isoforms of the receptor for serotonin, a powerful neurotransmitter released by nerve cells in parts of the brain. Different isoforms can have different affinities for the ligand or may interact with different types of G proteins. Different isoforms of a receptor may coexist in the same plasma membrane, or they may occur in the membranes of different types of target cells. The heterotrimeric G proteins that transmit signals from receptor to effector can also exist in multiple forms, as can many of the effectors. The human genome encodes at least 16 different G_α subunits, 5 different G_β subunits, and 11 different G_γ subunits, along with 9 isoforms of the effector adenylyl cyclase. Different combinations of specific subunits construct G proteins having different capabilities of reacting with specific isoforms of both receptors and effectors.

As mentioned on page 611, some G proteins act by inhibiting their effectors. The same stimulus can activate a stimulatory G protein (one with a $G_{\alpha s}$ subunit) in one cell and an inhibitory G protein (one with a $G_{\alpha i}$ subunit) in a different cell. For example, when epinephrine binds to a β-adrenergic receptor on a cardiac muscle cell, a G protein with a $G_{\alpha s}$ subunit is activated, which stimulates cAMP production, leading to an increase in the rate and force of contraction. In contrast, when epinephrine binds to an α-adrenergic receptor on a smooth muscle cell in the intestine, a G protein with a $G_{\alpha i}$ subunit is activated, which inhibits cAMP production, producing muscle relaxation. Finally, some adrenergic receptors turn on G proteins with $G_{\alpha q}$ subunits, leading to activation of PLCβ. Clearly, the same extracellular messenger can activate a variety of pathways in different cells.

Regulation of Blood Glucose Levels

Glucose can be utilized as a source of energy by all cell types present in the body. It is oxidized to CO_2 and H_2O by glycolysis and the TCA cycle, providing cells with ATP that can be used to drive energy-requiring reactions. The body maintains glucose levels in the bloodstream within a narrow range. As discussed in Chapter 3, excess glucose is stored in animal cells as glycogen, a large branched polymer composed of glucose monomers that are linked through glycosidic bonds. The hormone glucagon is produced by the alpha cells of the pancreas in response to low blood glucose levels. Glucagon stimulates breakdown of glycogen and release of glucose into the bloodstream, thereby causing glucose levels to rise. The hormone insulin is produced by the beta cells of the pancreas in response to high glucose levels and stimulates glucose uptake and storage as glycogen. Finally, epinephrine—which is sometimes called the "fight-or-flight" hormone—is produced by the adrenal gland in

stressful situations. Epinephrine causes an increase in blood glucose levels to provide the body with the extra energy resources needed to deal with the stressful situation at hand.

Insulin acts through a receptor protein-tyrosine kinase and its signal transduction is discussed on page 631. In contrast, both glucagon and epinephrine act by binding to GPCRs. Glucagon is a small protein that is composed of 29 amino acids, whereas epinephrine is a small molecule that is derived from the amino acid tyrosine. Structurally speaking, these two molecules have nothing in common, yet both of them bind to GPCRs and stimulate the breakdown of glycogen into glucose 1-phosphate (Figure 15.10). In addition, the binding of either of these hormones leads to the inhibition of the enzyme glycogen synthase, which catalyzes the opposing reaction in which glucose units are added to growing glycogen molecules. Thus two different stimuli (glucagon and epinephrine), recognized by different receptors, induce the same response in a single target cell. The two receptors differ from one another primarily in the structure of the ligand-binding pocket on the extracellular surface of the cell, which is specific for one or the other hormone. Following activation by their respective ligands, both receptors activate the same type of heterotrimeric G proteins that cause an increase in the levels of cAMP.
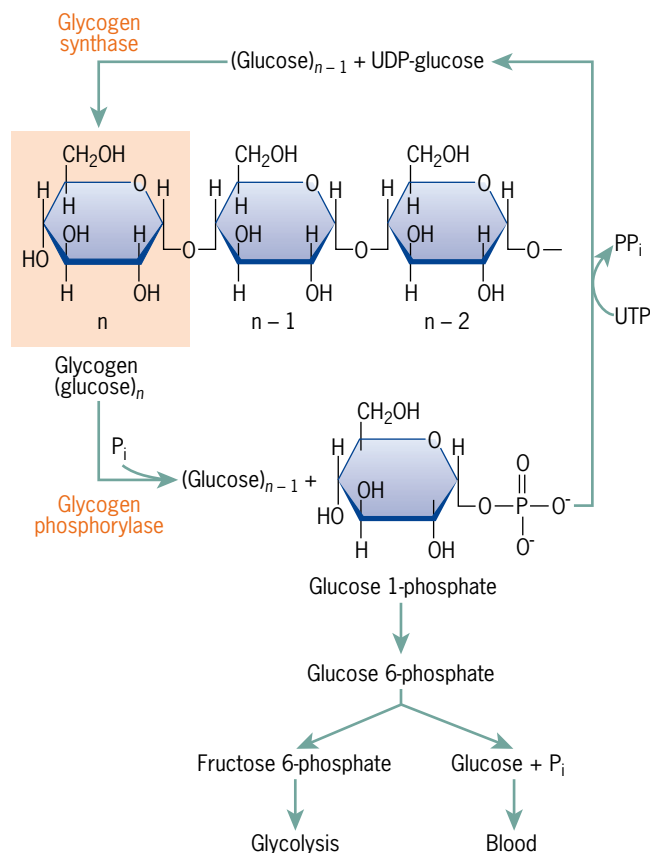
FIGURE 15.10 **The reactions that lead to glucose storage or mobilization.** The activities of two of the key enzymes in these reactions, glycogen phosphorylase and glycogen synthase, are controlled by hormones that act through signal transduction pathways. Glycogen phosphorylase is activated in response to glucagon and epinephrine, whereas glycogen synthase is activated in response to insulin (page 633).

FIGURE 15.11 **Formation of cyclic AMP from ATP is catalyzed by adenylyl cyclase,** an integral membrane protein that consists of two parts, each containing six transmembrane helices (shown here in two dimensions). The enzyme's active site is located on the inner surface of the membrane in a cleft situated between two similar cytoplasmic domains. The breakdown of cAMP (not shown) is accomplished by a phosphodiesterase, which converts the cyclic nucleotide to a 5′ monophosphate.
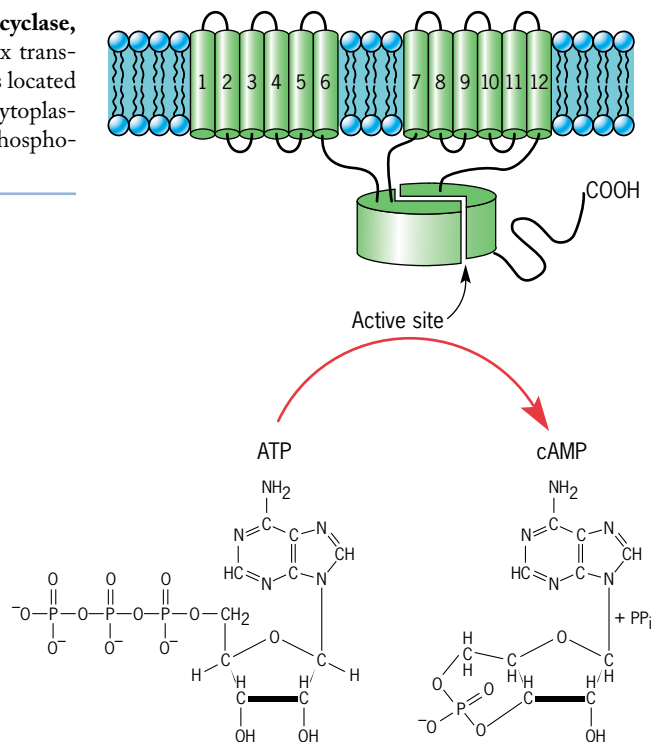
Glucose Mobilization: An Example of a Response Induced by cAMP

cAMP is synthesized by adenylyl cyclase, an integral membrane protein whose catalytic domain resides at the inner surface of the plasma membrane (Figure 15.11). cAMP evokes a response that leads to glucose mobilization by initiating a chain of reactions, as illustrated in Figure 15.12. The first step in this *reaction cascade* occurs as the hormone binds to its receptor, activating a $G_{\alpha s}$ subunit, which activates an adenylyl cyclase effector. The activated enzyme catalyzes the formation of cAMP (steps 1 and 2, Figure 15.12).

Once formed, cAMP molecules diffuse into the cytoplasm where they bind to an allosteric site on a regulatory subunit of a cAMP-dependent protein kinase (*protein kinase A, PKA*) (step 3, Figure 15.12). In its inactive form, PKA is a heterotetramer composed of two regulatory (R) and two catalytic (C) subunits. The regulatory subunits normally inhibit the catalytic activity of the enzyme. cAMP binding causes the dissociation of the
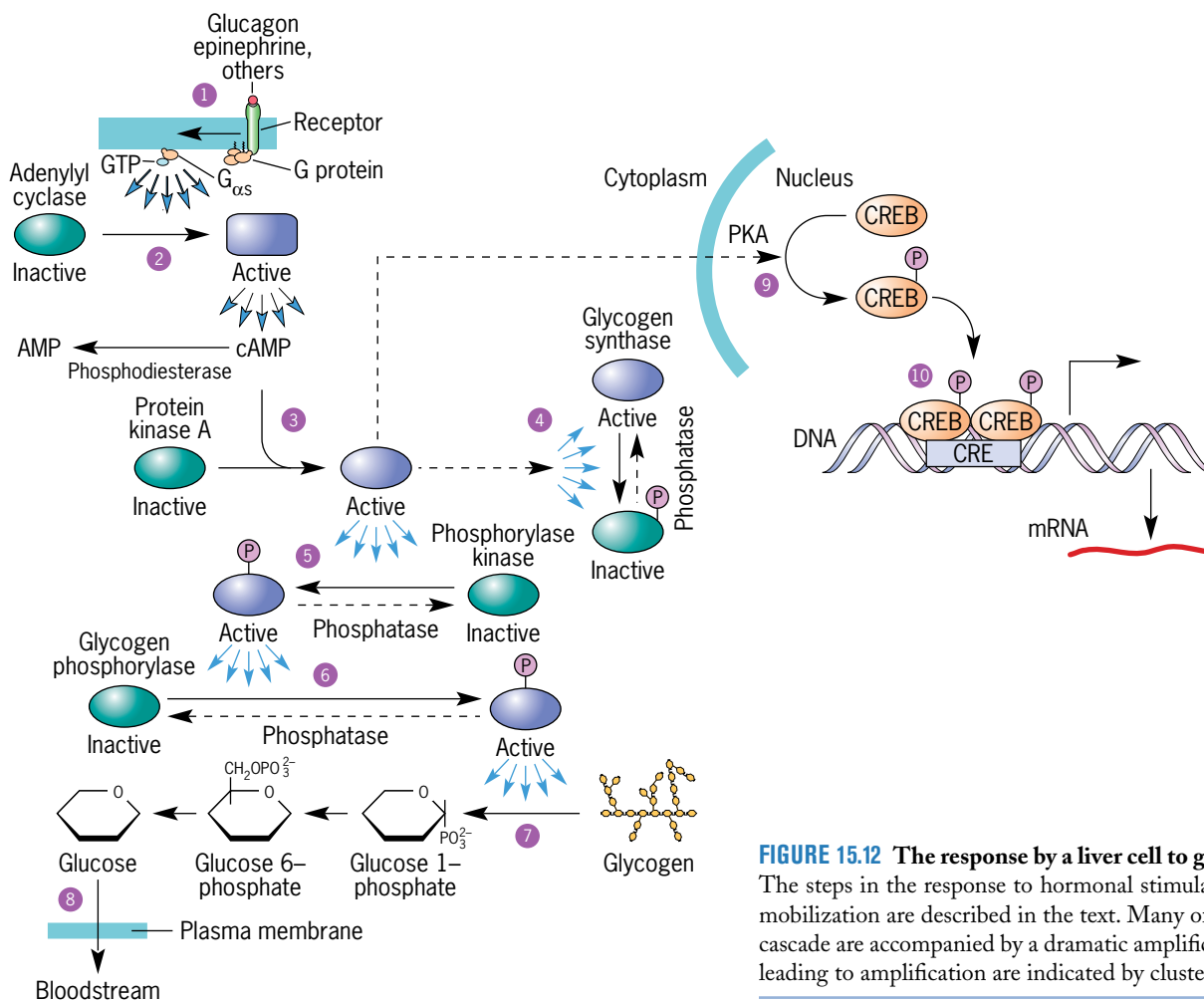
FIGURE 15.12 **The response by a liver cell to glucagon or epinephrine.** The steps in the response to hormonal stimulation that lead to glucose mobilization are described in the text. Many of the steps in the reaction cascade are accompanied by a dramatic amplification of the signal. Steps leading to amplification are indicated by clusters of blue arrows.

regulatory subunits, thereby releasing the active catalytic subunits of PKA. The target substrates of PKA in a liver cell include two enzymes that play a pivotal role in glucose metabolism, glycogen synthase and phosphorylase kinase (steps 4 and 5). Phosphorylation of glycogen synthase inhibits its catalytic activity and thus prevents the conversion of glucose to glycogen. In contrast, phosphorylation of phosphorylase kinase activates the enzyme to catalyze the transfer of phosphate groups to glycogen phosphorylase molecules. As discovered by Krebs and Fischer, the addition of a single phosphate group to a specific serine residue in the glycogen phosphorylase polypeptide activates this enzyme (step 6), stimulating the breakdown of glycogen (step 7). The glucose 1-phosphate formed in the reaction is converted to glucose, which diffuses into the bloodstream and so reaches the other tissues of the body (step 8).

As one might expect, a mechanism must exist to reverse the steps discussed above; otherwise the cell would remain in the activated state indefinitely. Liver cells contain phosphatases that remove the phosphate groups added by the kinases. A particular member of this family of enzymes, protein phosphatase-1, can remove phosphates from all of the phosphorylated enzymes of Figure 15.12: phosphorylase kinase, glycogen synthase, and glycogen phosphorylase. The destruction of cAMP molecules present in the cell is accomplished by the enzyme cAMP phosphodiesterase, which helps terminate the response.

Signal Amplification   The binding of a single hormone molecule at the cell surface can activate a number of G proteins, each of which can activate an adenyl cyclase effector, each of which can produce a large number of cAMP messengers in a short period of time. Thus, the production of a second messenger provides a mechanism to greatly amplify the signal generated from the original message. Many of the steps in the reaction cascade illustrated in Figure 15.12 result in amplification of the signal (these steps are indicated by the blue arrows). cAMP molecules activate PKA. Each PKA catalytic subunit phosphorylates a large number of phosphorylase kinase molecules, which in turn phosphorylate an even larger number of glycogen phosphorylase molecules, which in turn can catalyze the formation of a much larger number of glucose phosphates. Thus, what begins as a barely perceptible stimulus at the cell surface is rapidly transformed into a major mobilization of glucose within the cell.

Other Aspects of cAMP Signal Transduction Pathways   Although the most rapid and best-studied effects of cAMP are produced in the cytoplasm, the nucleus and its genes also participate in the response. A fraction of the activated PKA molecules translocate into the nucleus where they phosphorylate key nuclear proteins (step 9, Figure 15.12), most notably a transcription factor called *CREB* (*cAMP response element–binding protein*). The phosphorylated version of CREB binds as a dimer to sites on the DNA (Figure 15.12, step 10) containing a particular nucleotide sequence (TGACGTCA), known as the *cAMP response element* (*CRE*). Recall from page 514 that response elements are sites in the DNA where transcription factors bind and increase the rate of initiation of transcription. CREs are located in the regulatory regions of genes that play a role in the response to cAMP. In liver cells, for example, several of the enzymes involved in gluconeogenesis, a pathway by which glucose is formed from the intermediates of glycolysis (see Figure 3.31), are encoded by genes that contain nearby CREs. Thus, epinephrine and glucagon not only activate catabolic enzymes involved in glycogen breakdown, they promote the synthesis of anabolic enzymes required to synthesize glucose from smaller precursors.

cAMP is produced in many different cells in response to a wide variety of different ligands (i.e., first messengers). Several of the hormonal responses mediated by cAMP in mammalian cells are listed in Table 15.4. Cyclic AMP pathways have also been implicated in processes occurring in the nervous system, including learning, memory, and drug addiction. Chronic use of opiates, for example, leads to elevated levels of adenylyl cyclase and PKA, which may be partially responsible for the physiologic responses that occur during drug withdrawal. Another cyclic nucleotide, cyclic GMP, also acts as a second messenger in certain cells as illustrated by the induced relaxation of smooth muscle cells discussed on page 641. As discussed in the next section, cyclic GMP also plays a key role in the signaling pathway involved in vision.

Because cAMP exerts most of its effects by activating PKA, the response of a given cell to cAMP is typically determined by the specific proteins phosphorylated by this kinase (Figure 15.13). Although activation of PKA in a liver cell in response to epinephrine leads to the breakdown of glycogen, activation of the same enzyme in a kidney tubule cell in response to vasopressin causes an increase in the permeability of the membrane to water, and activation of the enzyme in a thy-

TABLE 15.4 **Examples of Hormone-Induced Responses Mediated by cAMP**

Tissue	Hormone	Response
Liver	Epinephrine and glucagon	Glycogen breakdown, glucose synthesis (gluconeogenesis), inhibition of glycogen synthesis
Skeletal muscle	Epinephrine	Glycogen breakdown, inhibition of glycogen synthesis
Cardiac muscle	Epinephrine	Increased contractility
Adipose	Epinephrine, ACTH, and glucagon	Triacylglycerol catabolism
Kidney	Vasopressin (ADH)	Increased permeability of epithelial cells to water
Thyroid	TSH	Secretion of thyroid hormones
Bone	Parathyroid hormone	Increased calcium resorption
Ovary	LH	Increased secretion of steroid hormones
Adrenal cortex	ACTH	Increased secretion of glucocorticoids

FIGURE 15.13 Schematic illustration of the variety of processes that can be affected by changes in cAMP concentration. All of these effects are thought to be mediated by activation of the same enzyme, protein kinase A. In fact, the same hormone can elicit very different responses in different cells, even when it binds to the same receptor. Epinephrine, for example, binds to a similar β-adrenergic receptor in liver cells, fat cells, and smooth muscle cells of the intestine, causing the production of cAMP in all three cell types. The responses, however, are quite different: glycogen is broken down in the liver cell, triacylglycerols are broken down in the fat cell, and the smooth muscle cells undergo relaxation. In addition to PKA, cAMP is known to interact with ion channels, phosphodiesterases, and GEFs (page 628).
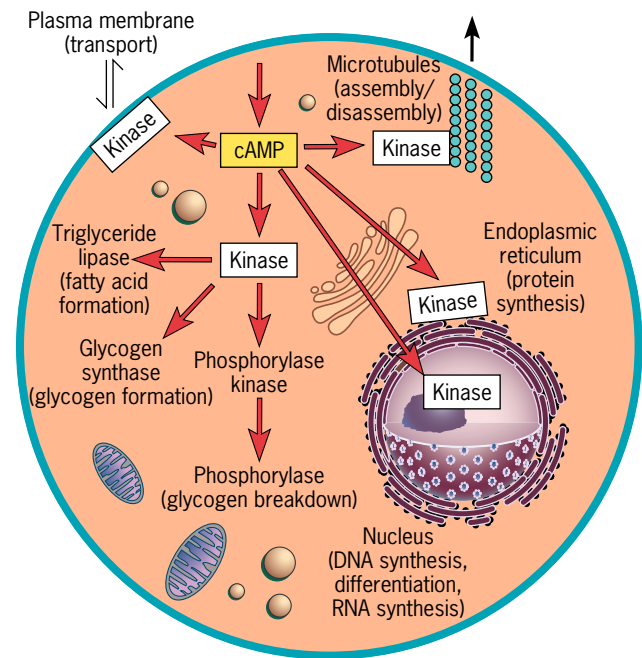
roid cell in response to TSH leads to the secretion of thyroid hormone. Clearly, PKA must phosphorylate different substrates in each of these cell types, thereby linking the increase in cAMP levels induced by epinephrine, vasopressin, and TSH to different physiological responses.

   Over one hundred PKA substrates have been described. Most of these carry out different functions, which brings up the question as to how PKA phosphorylates the appropriate substrates in response to a particular stimulus, in a particular cell type. This question was answered in part by the observation that different cells express different PKA substrates and in part by the discovery of PKA-anchoring proteins or *AKAPs* that function as signaling hubs. The first AKAPs were discov-

ered as proteins that co-purified with PKA. More than 30 AKAPs have been discovered since, several of which are shown in Figure 15.14. As indicated in this figure, AKAPs provide a structural framework for coordinating protein–
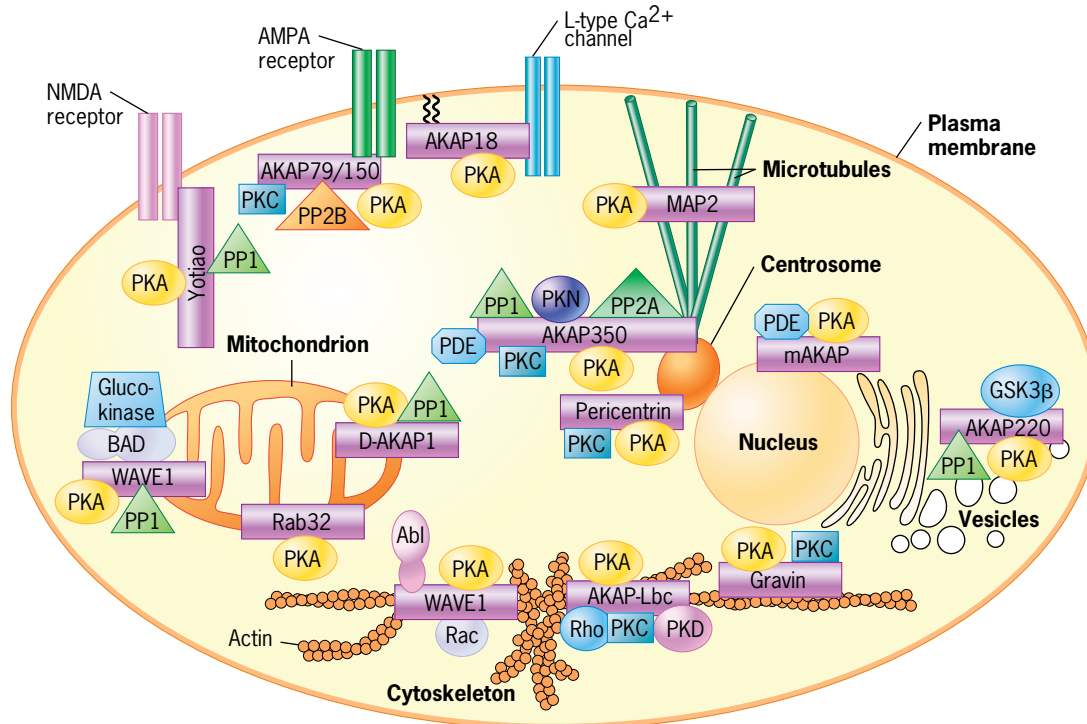
FIGURE 15.14 A schematic representation of AKAP signaling complexes operating in different subcellular compartments. The AKAP in each of these protein complexes is represented by the purple bar. In each case, the AKAP forms a scaffold that brings together a PKA molecule with potential substrates and other proteins involved in the signaling pathway, including phosphatases (green triangles) that can remove the added phosphate groups. The AKAPs shown here target PKA to a number of different compartments, including the plasma membrane, mitochondrion, cytoskeleton, centrosome, and nucleus. (REPRINTED WITH PERMISSION FROM W. WONG AND J. D. SCOTT, NATURE REVIEWS MOL. CELL BIOL. 5:961, 2004; © COPYRIGHT 2004, BY MACMILLAN JOURNALS LIMITED.)

protein interactions by sequestering PKA to specific locations within the cell. As a consequence, PKA accumulates in close proximity to one or more substrates. When cAMP levels rise and PKA is activated, the relevant substrates are present close by and they are the first ones to become phosphorylated. Substrate selection thus is partly a consequence of localization of PKA in the presence of particular substrates. Different cells express different AKAPs, resulting in localization of PKA in the presence of different substrates and consequently phosphorylation of different substrates following an increase in cAMP levels. It is interesting to note that, unlike most proteins with a similar function, AKAPs have a diverse structure, suggesting that evolution has co-opted a variety of different types of proteins to carry out a similar role in cell signaling.

The Role of GPCRs in Sensory Perception

Our ability to see, taste, and smell depends largely on GPCRs. It was mentioned above that rhodopsin, whose structure and activation is depicted in Figure 15.4b, is a GPCR. Rhodopsin is the light-sensitive protein present in the rods of our retina, which are the photoreceptor cells that respond to low light intensity and provide us with a black-and-white picture of our environment at night or in a darkened room. Several closely related GPCRs are present in the cones of the retina, which provide us with color vision under conditions of brighter light. Absorption of a single photon of light induces a conformational change in the rhodopsin molecule, which transmits a signal to a heterotrimeric G protein (called *transducin*), which activates a coupled effector. The effector in this case is the enzyme cGMP phosphodiesterase, which hydrolyzes the cyclic nucleotide cGMP, a second messenger similar in structure to cAMP (Figure 15.11). cGMP plays an important role in visual excitation in the rod cells of the retina. In the dark, cGMP levels remain high and thus capable of binding to cGMP-gated sodium channels in the plasma membrane, keeping the channels in an open configuration. Activation of cGMP phosphodiesterase results in lowered cGMP levels, leading to the closure of the sodium channels. This response, which is unusual in that it is triggered by a decrease in the concentration of a second messenger, may lead to the generation of action potentials along the optic nerve.

Our sense of smell depends on nerve impulses transmitted along olfactory neurons that extend from the epithelium that lines our upper nasal cavity to the olfactory bulb that is located in our brain stem. The distal tips of these neurons, which are located in the nasal epithelium, contain *odorant receptors*, which are GPCRs capable of binding various chemicals that enter our nose. Mammalian odorant receptors were first identified in 1991 by Linda Buck and Richard Axel of Columbia University. It is estimated that humans express roughly 400 different odorant receptors that, taken together, are able to combine with a large variety of different chemical structures (odorants).[1] Each

---

[1]The human genome contains roughly 1000 genes that encode odorant receptors but the majority are present as nonfunctional pseudogenes (page 402). Mice, which depend more heavily than humans on their sense of smell, have more than 1000 of these genes in their genome, and 95 percent of them encode functional receptors.

olfactory neuron is thought to contain only one of the hundreds of different odorant receptors encoded by the genome and, consequently, is only capable of responding to one or a few related chemicals. As a result, activation of different neurons containing different odorant receptors provides us with the perception of different aromas. Mutations in a specific gene encoding a particular odorant receptor can leave a person with the inability to detect a particular chemical in their environment that most other members of the population can perceive. When activated by bound ligands, odorant receptors signal through heterotrimeric G proteins to adenylyl cyclase, resulting in the synthesis of cAMP and the opening of a cAMP-gated cation channel. This response leads to the generation of action potentials that are transmitted to the brain.

Our perception of taste is much less discriminating than our perception of smell. Each taste receptor cell in the tongue transmits a sense of one of only four basic taste qualities, namely: salty, sour, sweet, or bitter. (A fifth type of taste receptor cell responds to the amino acids aspartate and glutamate and to purine nucleotides, generating a perception that a food is "savory." This is the reason that monosodium glutamate and disodium guanylate are commonly added to processed foods to enhance flavor.) The perception that a food or beverage is salty or sour is elicited directly by sodium ions or protons in the food. It is proposed that these ions enter cation channels in the plasma membrane of the taste receptor cell, leading to a depolarization of the membrane (page 161). In contrast, the perception that a food is bitter, sweet, or savory depends on a compound interacting with a GPCR at the surface of the receptor cell. Humans encode a family of about 30 bitter-taste receptors called T2Rs, which are coupled to the same heterotrimeric G protein. As a group, these taste receptors bind a diverse array of different compounds, including plant alkaloids or cyanides, that evoke a bitter taste in our mouths. For the most part, substances that evoke this perception are toxic compounds that elicit a distasteful, protective response that causes us to expel the food matter from our mouth. Unlike olfactory cells that contain a single receptor protein, a single taste-bud cell that evokes a bitter sensation contains a variety of different T2R receptors that respond to unrelated noxious substances. As a result, many diverse substances evoke the same basic taste, which is simply that the food we have eaten is bitter and disagreeable. In contrast, a food that elicits a sweet taste is likely to be one that contains energy-rich carbohydrates. Studies suggest that humans possess only one high affinity sweet-taste receptor (a T1R2-T1R3 heterodimer) and it responds to both sugars and artificial sweeteners. Fortunately, food that is chewed releases odorants that travel via the throat to olfactory receptor cells in our nasal mucosa, allowing the brain to learn much more about the food we have eaten than the relatively simple messages provided by taste receptors. It is this merged input from both olfactory and taste (gustatory) neurons that provides us with our rich sense of taste. The importance of olfactory neurons in our perception of taste becomes more evident when we have a cold that causes us to lose some of our appreciation for the taste of food.

1. What is the role of G proteins in a signaling pathway?

2. Describe Sutherland's experiment that led to the concept of the second messenger.

3. What is meant by the term *amplification* in regard to signal transduction? How does the use of a reaction cascade result in amplification of a signal? How does it increase the possibilities for metabolic regulation?

4. How is it possible that the same first messenger, such as epinephrine, can evoke different responses in different target cells? That the same second messenger, such as cAMP, can also evoke different responses in different target cells? That the same response, such as glycogen breakdown, can be initiated by different stimuli?

5. Describe the steps that lead from the synthesis of cAMP at the inner surface of the plasma membrane of a liver cell to the release of glucose into the bloodstream. How is this process controlled by GRKs and arrestin? By protein phosphatases? By cAMP phosphodiesterase?

6. Describe the steps between the binding of a ligand such as glucagon to a seven transmembrane receptor and the activation of an effector, such as adenylyl cyclase. How is the response normally attenuated?

7. What is the mechanism of formation of the second messenger IP_3? What is the relationship between the formation of IP_3 and an elevation of intracellular $[Ca^{2+}]$?

8. Describe the relationship between phosphatidylinositol, diacylglycerol, calcium ions, and protein kinase C. How do phorbol esters interfere with signal pathways involving DAG?

15.4 PROTEIN-TYROSINE PHOSPHORYLATION AS A MECHANISM FOR SIGNAL TRANSDUCTION

Protein-tyrosine kinases are enzymes that phosphorylate specific tyrosine residues on protein substrates. Protein-tyrosine phosphorylation is a mechanism for signal transduction that appeared with the evolution of multicellular organisms. Over 90 different protein-tyrosine kinases are encoded by the human genome. These kinases are involved in the regulation of growth, division, differentiation, survival, attachment to the extracellular matrix, and migration of cells. Expression of mutant protein-tyrosine kinases that cannot be regulated and are continually active can lead to uncontrolled cell division and the development of cancer. One type of leukemia, for example, occurs in cells that contain an unregulated version of the protein-tyrosine kinase ABL.

Protein-tyrosine kinases can be divided in two groups: **Receptor protein-tyrosine kinases** (**RTKs**), which are integral membrane proteins that contain a single transmembrane helix and an extracellular ligand binding domain, and non-receptor or *cytoplasmic protein-tyrosine kinases*. The human genome encodes nearly 60 RTKs and 32 non-receptor TKs. RTKs are activated directly by extracellular growth and differentiation factors such as epidermal growth factor (EGF) and platelet-derived growth factor (PDGF) or by metabolic regulators such as insulin. Non-receptor protein-tyrosine kinases are regulated indirectly by extracellular signals and they control processes as diverse as the immune response, cell adhesion, and neuronal cell migration. This section of the chapter is focused on signal transduction by RTKs.

Receptor Dimerization An obvious question comes to mind when considering the mechanics of signal transduction: How is the presence of a growth factor on the outside of the cell translated into biochemical changes inside the cell? It has been widely accepted among researchers in the protein-tyrosine kinase field that ligand binding results in the dimerization of the extracellular ligand-binding domains of a pair of receptors. Two mechanisms for receptor dimerization have been recognized: ligand-mediated dimerization and receptor-mediated dimerization (Figure 15.15). Early work suggested that ligands of RTKs contain two receptor-binding sites. This made it possible for a single growth or differentiation factor molecule to bind to two receptors at the same time, thereby causing ligand-mediated receptor dimerization (Figure 15.15*a*). This model was supported by the observation that growth and differentiation factors such as platelet-derived growth factor (PDGF) or colony-stimulating factor-1 (CSF-1) are composed of two similar or identical disulfide-linked subunits, in which each subunit contains a receptor-binding site. However, not all growth factors appeared to conform to this model. More recently, it has been established that some growth factors (e.g., EGF or TGFα) contain only a single receptor-binding site. Structural work now supports a second mechanism (Figure 15.15*b*) in which ligand binding induces a conformational change in the extracellular domain of a receptor, leading to the formation or exposure of a receptor dimerization interface. With this mechanism, ligands act as allosteric regulators that turn on the ability of their receptors to form dimers. Regardless of the mechanism, receptor dimerization results in the juxtapositioning of two protein-tyrosine kinase domains on the cytoplasmic side of the plasma membrane. Bringing two kinase domains in close contact allows for *trans–autophosphorylation*, in which the protein kinase activity of one receptor of the dimer phosphorylates tyrosine residues in the cytoplasmic domain of the other receptor of the dimer, and vice versa (Figure 15.15*a,b*).

Protein Kinase Activation Autophosphorylation sites on RTKs can carry out two different functions: they can regulate the receptor's kinase activity or serve as binding sites for cytoplasmic signaling molecules. Kinase activity is usually controlled by autophosphorylation on tyrosine residues that are present in the *activation loop* of the kinase domain. The activation loop, when unphosphorylated, obstructs the substrate-binding site, thereby preventing ATP from entering. Following
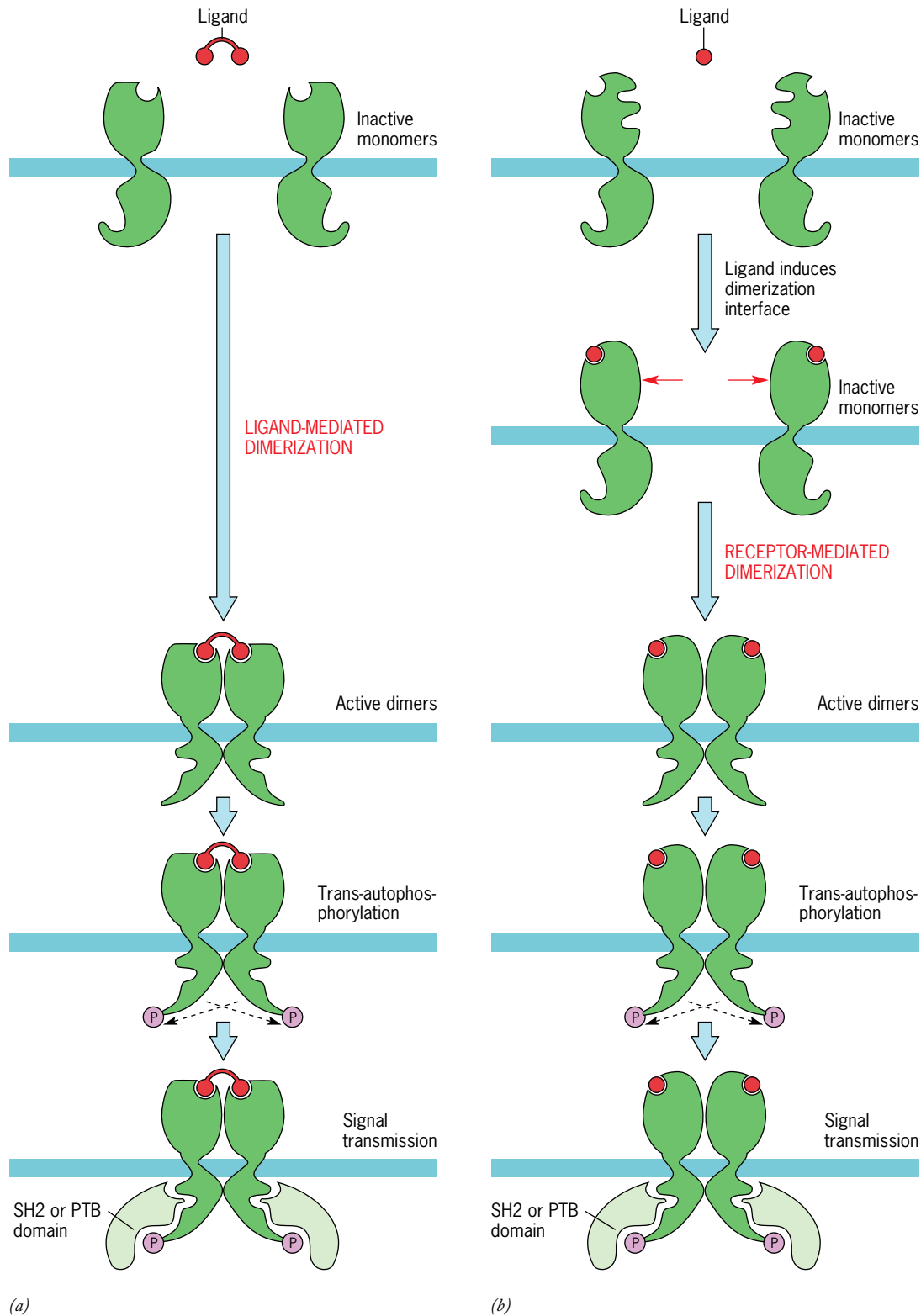
FIGURE 15.15 **Steps in the activation of a receptor protein-tyrosine kinase (RTK).** (*a*) Ligand-mediated dimerization. In the nonactivated state, the receptors are present in the membrane as monomers. Binding of a bivalent ligand leads directly to dimerization of the receptor and activation of its kinase activity, causing it to add phosphate groups to the cytoplasmic domain of the other receptor subunit. The newly formed phosphotyrosine residues of the receptor serve as binding sites for target proteins containing either SH2 or PTB domains. The target proteins become activated as a result of their interaction with the receptor.

(*b*) Receptor-mediated dimerization. The sequence of events are similar to those in part *a*, except that the ligand is monovalent and, consequently, a separate ligand molecule binds to each of the inactive monomers. Binding of each ligand induces a conformational change in the receptor that creates a dimerization interface (red arrows). The ligand-bound monomers interact through this interface to become an active dimer. (BASED ON A DRAWING BY J. SCHLESSINGER AND A. ULLRICH, NEURON 9:384, 1992; BY PERMISSION OF CELL PRESS.)

its phosphorylation, the activation loop is stabilized in a position away from the substrate-binding site, resulting in activation of the kinase domain. Once their kinase domain has been activated, the receptor subunits proceed to phosphorylate each other on tyrosine residues that are present in regions adjacent to the kinase domain. It is these autophosphorylation sites that act as binding sites for cellular signaling proteins.

Phosphotyrosine-Dependent Protein–Protein Interactions
Signaling pathways consist of a chain of signaling proteins that interact with one another in a sequential manner (see Figure 15.3). Signaling proteins are able to associate with activated protein-tyrosine kinase receptors, because such proteins contain domains that bind specifically to phosphorylated tyrosine residues (as in Figure 15.15). Two of these domains have been identified thus far: the *Src-homology 2* (*SH2*) *domain* and the *phosphotyrosine-binding* (*PTB*) *domain*. SH2 domains were initially identified as part of proteins encoded by the genome of tumor-causing (oncogenic) viruses. They are composed of approximately 100 amino acids and contain a conserved binding-pocket that accommodates a phosphorylated tyrosine residue (Figure 15.16). More than 110 SH2 domains are encoded by the human genome. They mediate a large number of phosphorylation-dependent protein–protein interactions. These interactions occur following phosphorylation

of specific tyrosine residues. The specificity of the interactions is determined by the amino acid sequence immediately adjacent to the phosphorylated tyrosine residues. For example, the SH2 domain of the Src protein-tyrosine kinase recognizes pTyr-Glu-Glu-Ile, whereas the SH2 domains of PI 3-kinase bind to pTyr-Met-X-Met (in which X can be any residue). It is interesting to note that the budding-yeast genome encodes only one SH2-domain-containing protein, which correlates with the overall lack of tyrosine-kinase signaling activity in these lower single-celled eukaryotes.

PTB domains were discovered more recently. They can bind to phosphorylated tyrosine residues that are usually present as part of an asparagine-proline-X-tyrosine (Asn-Pro-X-Tyr) motif. The story is more complicated, however, because some PTB domains appear to bind specifically to an unphosphorylated Asn-Pro-X-Tyr motif, whereas others bind specifically to the phosphorylated motif. PTB domains are poorly conserved and different PTB domains possess different residues that interact with their ligands.

Activation of Downstream Signaling Pathways   We have seen that receptor protein-tyrosine kinases (RTKs) are autophosphorylated on one or more tyrosine residues. A variety of signaling proteins with SH2 or PTB domains are present in the cytoplasm. Receptor activation therefore results in formation of signaling complexes, in which SH2- or PTB-containing signaling proteins bind to specific autophosphorylation sites present on the receptor (as in Figure 15.15). We can distinguish several groups of signaling proteins, including adaptor proteins, docking proteins, transcription factors, and enzymes (Figure 15.17).

- Adaptor proteins function as linkers that enable two or more signaling proteins to become joined together as part of a signaling complex (Figure 15.17*a*). Adaptor proteins contain an SH2 domain and one or more additional protein–protein interaction domains. For instance, the adaptor protein Grb2 contains one SH2 and two SH3 (Src-homology 3) domains (Figure 15.18). As shown on page 61, SH3 domains bind to proline-rich sequence motifs. The SH3 domains of Grb2 bind constitutively to other proteins, including Sos and Gab. The SH2 domain binds to phosphorylated tyrosine residues within a Tyr-X-Asn motif. Consequently, tyrosine phosphorylation of the Tyr-X-Asn motif on an RTK results in translocation of Grb2-Sos or Grb2-Gab from the cytosol to a receptor, which is present at the plasma membrane (Figure 15.17*a*).

FIGURE 15.16 The interaction between an SH2 domain of a protein and a peptide containing a phosphotyrosine residue. The SH2 domain of the protein is shown in a cutaway view with the accessible surface area represented by red dots and the polypeptide backbone as a purple ribbon. The phosphotyrosine-containing heptapeptide (Pro-Asn-pTyr-Glu-Glu-Ile-Pro) is shown as a space-filling model whose side chains are colored green and backbone is colored yellow. The phosphate group is shown in light blue. The phosphorylated tyrosine residue and the isoleucine residue (+3) are seen to project into pockets on the surface of the SH2 domain, creating a tightly fitting interaction, but only when the key tyrosine residue is phosphorylated. (FROM GABRIEL WAKSMAN ET AL., COURTESY OF JOHN KURIYAN, CELL 72:783, 1993; BY PERMISSION OF CELL PRESS.)

- Docking proteins, such as IRS, supply certain receptors with additional tyrosine phosphorylation sites (Figure 15.17*b*). Docking proteins contain either a PTB domain or an SH2 domain and a number of tyrosine phosphorylation sites. Binding of an extracellular ligand to a receptor leads to autophosphorylation of the receptor, which provides a binding site for the PTB or SH2 domain of the docking protein. Once bound together, the receptor phosphorylates tyrosine residues present on the docking protein. These phosphorylation sites then act as

(a)

(b)

(c)

(d)
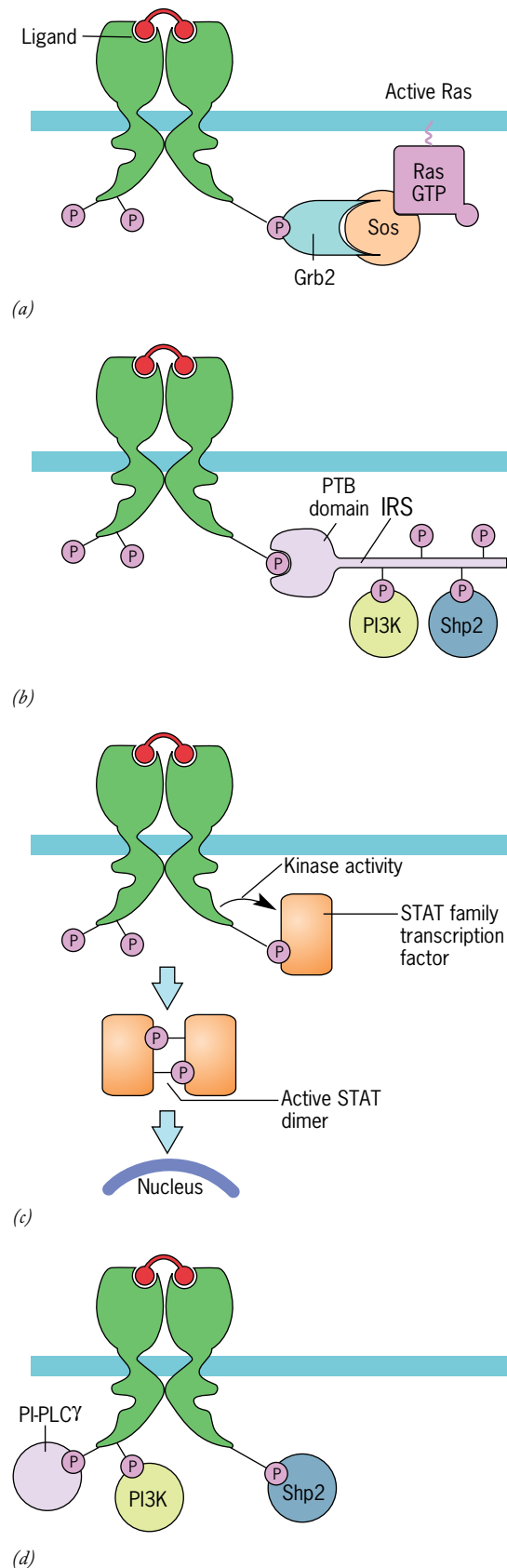
FIGURE 15.17 **A diversity of signaling proteins.** Cells contain numerous proteins with SH2 or PTB domains that bind to phosphorylated tyrosine residues. (*a*) Adaptor proteins, such as Grb2, function as a link between other proteins. As shown here, Grb2 can serve as a link between an activated growth factor RTK and Sos, an activator of a downstream protein named Ras. The function of Ras is discussed later. (*b*) The docking protein IRS contains a PTB domain that allows it to bind to the activated receptor. Once bound, tyrosine residues on the docking protein are phosphorylated by the receptor. These phosphorylated residues function as binding sites for other signaling proteins. (*c*) Certain transcription factors bind to activated RTKs, an event that leads to the phosphorylation and activation of the transcription factor and its translocation to the nucleus. Members of the STAT family of transcription factors (discussed further in Section 17.4) become activated in this manner. (*d*) A wide array of signaling enzymes are activated following binding to an activated RTK. In the case depicted here, a phospholipase (PLC-γ), a lipid kinase (PI3K), and a protein-tyrosine phosphatase (Shp2) have all bound to phosphotyrosine sites on the receptor.
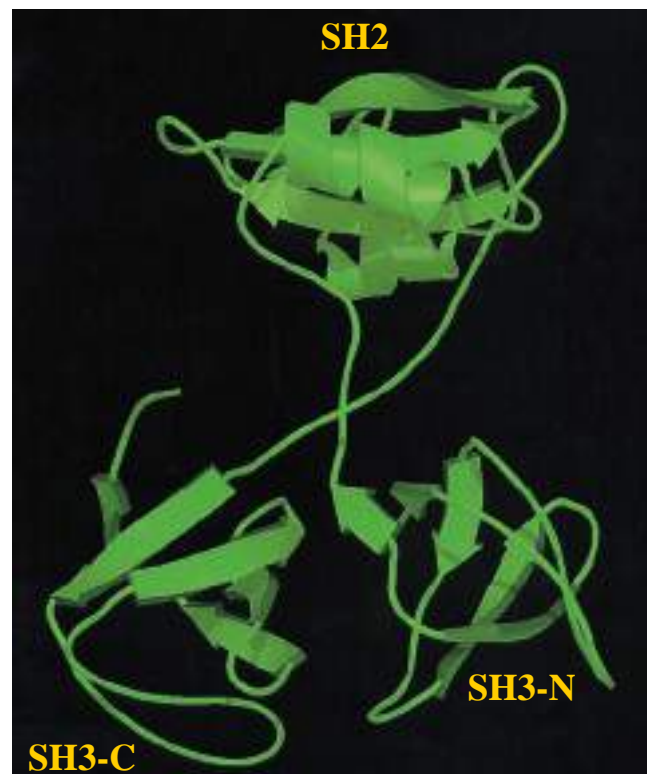
FIGURE 15.18 **Tertiary structure of an adaptor protein, Grb2.** Grb2 consists of three parts: two SH3 domains and one SH2 domain. SH2 domains bind to a protein (e.g., the activated EGF receptor) containing a particular motif that includes a phosphotyrosine residue. SH3 domains bind to a protein (e.g., Sos) that contains a particular motif that is rich in proline residues. Dozens of proteins that bear these domains have been identified. Interactions involving SH3 and SH2 domains are shown in Figures 2.40 and 15.16, respectively. Other adaptor proteins include Nck, Shc, and Crk. (REPRINTED WITH PERMISSION FROM SÉBASTIEN MAIGNAN ET AL. SCIENCE 268:291, 1995; © COPYRIGHT 1995, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE. COURTESY OF ARNAUD DUCRUIX.)

binding sites for additional signaling molecules. Docking proteins provide versatility to the signaling process, because the ability of the receptor to turn on signaling molecules can vary with the docking proteins that are expressed in a particular cell.

■ Transcription factors were discussed at length in Chapter 12. Transcription factors that belong to the STAT family play an important role in the function of the immune system (discussed in Section 17.4). STATs contain an SH2 domain together with a tyrosine phosphorylation site that can act as a binding site for the SH2 domain of another STAT molecule (Figure 15.17c). Tyrosine phosphorylation of STAT SH2 binding sites situated within a dimerized receptor leads to the recruitment of STAT proteins (Figure 15.17c). Upon association with the receptor complex, tyrosine residues in these STAT proteins are phosphorylated. As a result of the interaction between the phosphorylated tyrosine residue on one STAT protein and the SH2 domain on a second STAT protein, and vice versa, these transcription factors will form dimers. Dimers, but not monomers, move to the nucleus where they stimulate the transcription of specific genes involved in an immune response.

■ Signaling enzymes include protein kinases, protein phosphatases, lipid kinases, phospholipases, and GTPase activating proteins. When equipped with SH2 domains, these enzymes associate with activated RTKs and are turned on directly or indirectly as a consequence of this association (Figure 15.17d). Three general mechanisms have been identified by which these enzymes are activated following their association with a receptor. Enzymes can be activated simply as a result of translocation to the membrane, which places them in close proximity to their substrates. Enzymes can also be activated through an allosteric mechanism (page 113), in which binding to phosphotyrosine results in a conformational change in the SH2 domain that causes a conformational change in the catalytic domain, resulting in a change in catalytic activity. Finally, enzymes can be regulated directly by phosphorylation. As will be described below, signaling proteins that associate with activated RTKs initiate cascades of events that lead to the biochemical changes required to respond to the presence of extracellular messenger molecules.

Ending the Response   Signal transduction by RTKs is usually terminated by internalization of the receptor. Exactly what causes receptor internalization remains an area of active research. One mechanism involves a receptor-binding protein named Cbl. When RTKs are activated by ligands, they autophosphorylate tyrosine residues, which can act as a binding site for Cbl, which possesses an SH2 domain. Cbl then associates with the receptor and catalyzes the attachment of a ubiquitin molecule to the receptor. Ubiquitin is a small protein that is linked covalently to other proteins, thereby marking those proteins for internalization (page 305) or degradation (page 530). Binding of the Cbl complex to activated receptors

is followed by receptor ubiquitination, internalization, and in most cases degradation in a lysosome.

Now that we have discussed some of the general mechanisms by which RTKs are able to activate signaling pathways, we can look more closely at a couple of important pathways that are activated downstream of RTKs. First we will discuss the Ras-MAP kinase pathway, which is probably the best characterized signaling cascade that is turned on by activated protein-tyrosine kinases. A different cascade will be described in the context of the insulin receptor.

The Ras-MAP Kinase Pathway

Retroviruses are small viruses that carry their genetic information in the form of RNA. Some of these viruses contain genes, called oncogenes, that enable them to transform normal cells into tumor cells. Ras was originally described as the product of a retroviral oncogene and, only later, determined to be derived from its mammalian host. It was subsequently discovered that approximately 30 percent of all human cancers contain mutant versions of *RAS* genes. At this point it is important to note that Ras proteins are part of a superfamily of more than 150 small G proteins including the Rabs (page 295), Sar1 (page 290), and Ran (page 480). These proteins are involved in the regulation of numerous processes, including cell division, differentiation, gene expression, cytoskeletal organization, vesicle trafficking, and nucleocytoplasmic transport. The principles discussed in connection with Ras apply to many members of the small G-protein superfamily.

Ras is a small GTPase that is anchored at the inner surface of the plasma membrane by a lipid group that is embedded in the inner leaflet of the bilayer (Figure 15.17a). Ras is functionally similar to the heterotrimeric G proteins that were discussed earlier and, like those proteins, Ras also acts as both a switch and a molecular timer. Unlike heterotrimeric G proteins, however, Ras consists of only a single small subunit. Ras proteins are present in two different forms: an active GTP-bound form and an inactive GDP-bound form (Figure 15.19a). Ras-GTP binds and activates downstream signaling proteins. Ras is turned off by hydrolysis of its bound GTP to GDP. Mutations in the human *RAS* gene that lead to tumor formation prevent the protein from hydrolyzing the bound GTP back to the GDP form. As a result, the mutant version of Ras remains in the "on" position, sending a continuous message downstream along the signaling pathway, keeping the cell in the proliferative mode.

The cycling of monomeric G proteins, such as Ras, between active and inactive states is aided by accessory proteins that bind to the G protein and regulate its activity (Figure 15.19b). These accessory proteins include

1. ***GTPase-activating proteins (GAPs).*** Most monomeric G proteins possess some capability to hydrolyze a bound GTP, but this capability is greatly accelerated by interaction with specific GAPs. Because they stimulate hydrolysis of the bound GTP, which inactivates the G protein, GAPs dramatically shorten the duration of a G protein-
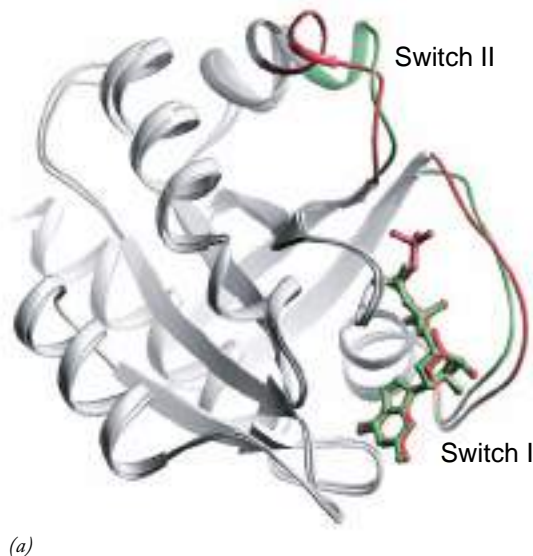
FIGURE 15.19 The structure of a G protein and the G protein cycle.
(*a*) Comparison of the tertiary structure of the active GTP-bound state (red) and inactive GDP-bound state (green) of the small G protein Ras. A bound guanine nucleotide is depicted in the ball-and-stick form. The differences in conformation occur in two flexible regions of the molecule known as switch I and switch II. The difference in conformation shown here affects the molecule's ability to bind to other proteins. (*b*) The G protein cycle. G proteins are in their inactive state when they are bound by a molecule of GDP. If the inactive G protein interacts with a guanine nucleotide dissociation inhibitor (GDI), release of the GDP is inhibited and the protein remains in the inactive state (step 1a). If the inactive G protein interacts with a guanine nucleotide exchange factor (GEF; step 1b), the G protein exchanges its GDP for a GTP (step 2),
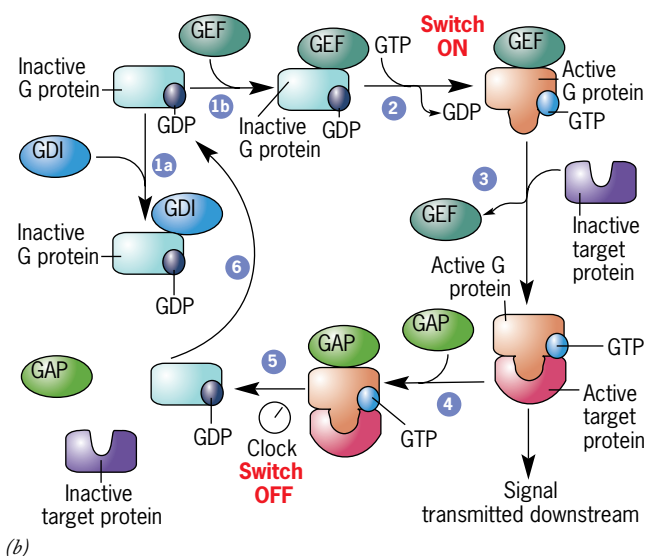
which activates the G protein so that it can bind to a downstream target protein (step 3). Binding to the GTP-bound G protein activates the target protein, which is typically an enzyme such as a protein kinase or a protein phosphatase. This has the effect of transmitting the signal farther downstream along the signaling pathway. G proteins have a weak intrinsic GTPase activity that is stimulated by interaction with a GTPase-activating protein (GAP) (step 4). The degree of GTPase stimulation by a GAP determines the length of time that the G protein is active. Consequently, the GAP serves as a type of clock that regulates the duration of the response (step 5). Once the GTP has been hydrolyzed, the complex dissociates, and the inactive G protein is ready to begin a new cycle (step 6). (A: FROM STEVEN J. GAMBLIN AND STEPHEN J. SMERDON, STRUCT. 7:R200, 1999.)

mediated response. Mutations in one of the Ras-GAP genes (*NF1*) cause neurofibromatosis 1, a disease in which patients develop large numbers of benign tumors (neurofibromas) along the sheaths that line the nerve trunks.

2. *Guanine nucleotide–exchange factors* (*GEFs*). An inactive G protein is converted to the active form when the bound GDP is replaced with a GTP. GEFs are proteins that bind to an inactive monomeric G protein and stimulate dissociation of the bound GDP. Once the GDP is released, the G protein rapidly binds a GTP, which is present at relatively high concentration in the cell, thereby activating the G protein.

3. *Guanine nucleotide–dissociation inhibitors* (*GDIs*). GDIs are proteins that inhibit the release of a bound GDP from a monomeric G protein, thus maintaining the protein in the inactive, GDP-bound state.

The activity and localization of these various accessory proteins are tightly regulated by other proteins, which thus regulates the state of the G protein.

Ras-GTP can interact directly with several downstream targets. Here we will discuss Ras as an element of the **Ras-MAP kinase cascade**. The Ras-MAP kinase cascade is turned on in response to a wide variety of extracellular signals and plays a key role in regulating vital activities such as cell prolif-

eration and differentiation. The pathway relays extracellular signals from the plasma membrane through the cytoplasm and into the nucleus. The overall outline of the pathway is depicted in Figure 15.20. This pathway is activated when a growth factor, such as EGF or PDGF, binds to the extracellular domain of its RTK. Many activated RTKs possess phosphorylated tyrosine residues that act as docking sites for the adaptor protein Grb2. Grb2, in turn, binds to Sos, which is a guanine nucleotide exchange factor (a GEF) for Ras. Creation of a Grb2-binding site on an activated receptor promotes the translocation of Grb2-Sos from the cytoplasm to the cytoplasmic surface of the plasma membrane, placing Sos in close proximity to Ras (as in Figure 15.17a).

Simply bringing Sos to the plasma membrane is sufficient to cause Ras activation. This was illustrated by an experiment with a mutant version of Sos that is permanently tethered to the inner surface of the plasma membrane. Expression of this membrane-bound Sos mutant results in constitutive activation of Ras and transformation of the cell to a malignant phenotype. Interaction with Sos opens the Ras nucleotide-binding site. As a result, GDP is released and is replaced by GTP. Exchange of GDP for GTP in the nucleotide-binding site of Ras results in a conformational change and the creation of a binding interface for a number of proteins, including an important signaling protein called Raf. Raf is then recruited to

the inner surface of the plasma membrane where it is activated by a combination of phosphorylation and dephosphorylation reactions.

Raf is a serine-threonine protein kinase. One of its substrates is the protein kinase MEK (Figure 15.20). MEK, which is activated as a consequence of phosphorylation by Raf, goes on to phosphorylate and activate two MAP kinases named ERK1 and ERK2. Over 160 proteins that can be phosphorylated by these kinases have been identified, including transcription factors, protein kinases, cytoskeletal proteins, apoptotic regulators, receptors, and other signaling proteins. Once activated, the MAP kinase is able to move into the nucleus where it phosphorylates and activates a number of transcription factors and other nuclear proteins. Eventually, the pathway leads to the activation of genes involved in cell proliferation, including cyclin D1, which plays a key role in driving a cell from G1 into S phase (Figure 14.8).
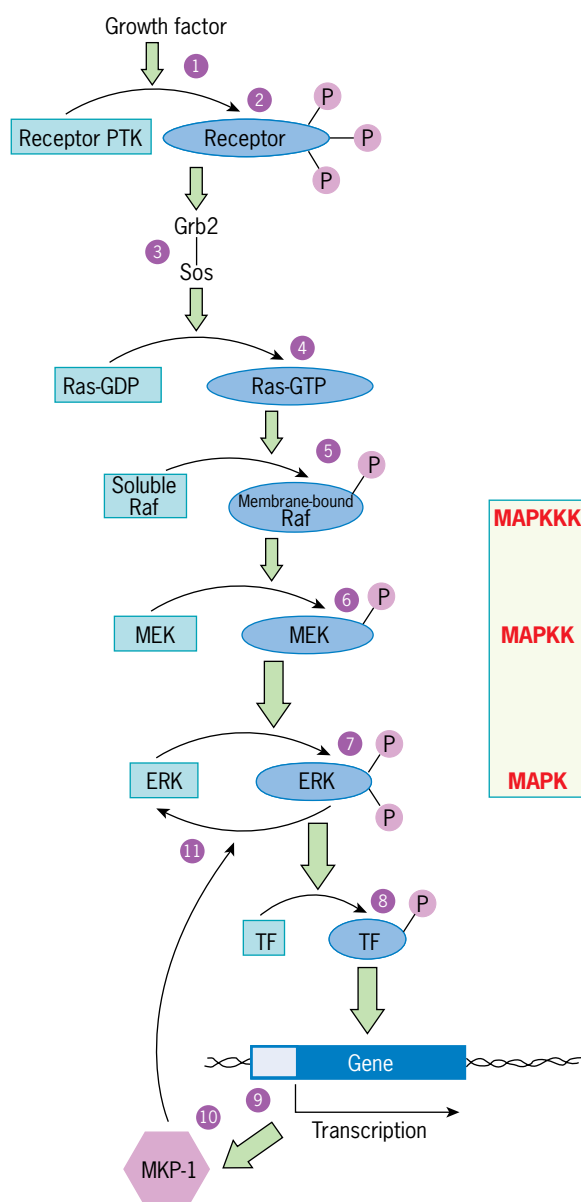
As discussed in the following chapter, oncogenes are identified by their ability to cause cells to become cancerous. Oncogenes are derived from normal cellular genes that have either become mutated or are overexpressed. Many of the proteins that play a part in the Ras signaling pathway were discovered because they were encoded by cancer-causing oncogenes. This includes the genes for Ras, Raf, and a number of the transcriptional factors generated at the end of the pathway (e.g., Fos and Jun). Genes for several of the RTKs situated at the beginning of the pathway, including the receptors for both EGF and PDGF, have also been identified among the several dozen known oncogenes. The fact that so many proteins in this pathway are encoded by genes that can cause cancer when mutated emphasizes the importance of the pathway in the control of cell growth and proliferation.

Adapting the MAP Kinase to Transmit Different Types of Information The same basic pathway from RTKs through Ras to the activation of transcription factors, as illustrated in Figure 15.20, is found in all eukaryotes investigated, from yeast through flies and nematodes to mammals. Evolution has adapted the pathway to meet many different ends. In yeast, for example, the MAP kinase cascade is required for cells to respond to mating pheromones; in fruit flies, the pathway is utilized during the differentiation of the photoreceptors in the compound eye; and in flowering plants, the pathway transmits signals that initiate a defense against pathogens. In each case, the core of the pathway contains a trio of enzymes that act se-

FIGURE 15.20 **The steps of a generalized MAP kinase cascade.** Binding of growth factor to its receptor (step 1) leads to the autophosphorylation of tyrosine residues of the receptor (step 2) and the subsequent recruitment of the Grb2-Sos proteins (step 3). This complex causes the GTP-GDP exchange of Ras (step 4), which recruits the protein Raf to the membrane, where it is phosphorylated and thus activated (step 5). In the pathway depicted here, Raf phosphorylates and activates another kinase named MEK (step 6), which in turn phosphorylates and activates still another kinase termed ERK (step 7). This three-step phosphorylation scheme shown in steps 5–7 is characteristic of all MAP kinase cascades. Because of their sequential kinase activity, Raf is known as a MAPKKK (MAP kinase kinase kinase), MEK as a MAPKK (MAP kinase kinase), and ERK as a MAPK (MAP kinase). MAPKKs are dual-specificity kinases, a term denoting that they can phosphorylate tyrosine as well as serine and threonine residues. All MAPKs have a tripeptide near their catalytic site with the sequence Thr-X-Tyr. MAPKK phosphorylates MAPK on both the threonine and tyrosine residue of this sequence, thereby activating the enzyme (step 7). Once activated, MAPK translocates into the nucleus where it phosphorylates transcription factors (TF, step 8), such as Elk-1. Phosphorylation of the transcription factors increases their affinity for regulatory sites on the DNA (step 9), leading to an increase in the transcription of specific genes (e.g., *Fos* and *Jun*) involved in the growth response. One of the genes whose expression is stimulated encodes a MAPK phosphatase (MKP-1; step 10). Members of the MKP family can remove phosphate groups from both tyrosine and threonine residues of MAPK (step 11), which inactivates MAPK and stops further signaling activity along the pathway. (AFTER H. SUN AND N. K. TONKS, TRENDS BIOCHEM. SCI. 19:484, 1994.)

quentially: a MAP kinase kinase kinase (MAPKKK), a MAP kinase kinase (MAPKK), and a MAP kinase (MAPK) (Figure 15.20). Each of these components is represented by a small family of proteins. To date, 14 different MAPKKKs, 7 different MAPKKs, and 13 different MAPKs have been identified in mammals. By utilizing different members of these protein families, mammals are able to assemble a number of different MAP kinase pathways that transmit different types of extracellular signals. We have already described how mitogenic stimuli are transmitted along one type of MAP kinase pathway that leads to cell proliferation. In contrast, when cells are exposed to stressful stimuli, such as X-rays or damaging chemicals, signals are transmitted along different MAP kinase pathways that cause the cell to withdraw from the cell cycle, rather than progressing through it, as indicated in Figure 15.20. Withdrawal from the cell cycle gives the cell time to repair the damage resulting from the adverse conditions.

Recent studies have focused on the signaling specificity of MAP kinase cascades in an attempt to understand how cells are able to utilize similar proteins as components of pathways that elicit different cellular responses. Studies of amino acid sequences and protein structure suggest that part of the answer lies in selective interactions between enzymes and sub-strates. For example, certain members of the MAPKKK family phosphorylate specific members of the MAPKK family, which in turn phosphorylate specific members of the MAPK family. But many members of these families can participate in more than one MAPK signaling pathway.

Specificity in MAP kinase pathways is also achieved by spatial localization of the component proteins. Spatial localization is accomplished by structural (i.e., nonenzymatic) proteins referred to as *scaffolding proteins*, whose apparent function is to tether the appropriate members of a signaling pathway in a specific spatial orientation that enhances their mutual interactions. The AKAPs depicted in Figure 15.14 are examples of scaffolding proteins, and similar proteins play a role in routing signals through one of various MAP kinase pathways. Scaffolding proteins may also take an active role in signaling events by inducing a change in conformation of bound proteins. In addition to facilitating a particular series of reactions, scaffolding proteins may prevent proteins involved in one signaling pathway from participating in other pathways. It is evident from this section of the chapter that the signal transduction pathways of the cell are very complex. This has made the subject challenging for both students and biologists alike (Figure 15.21).
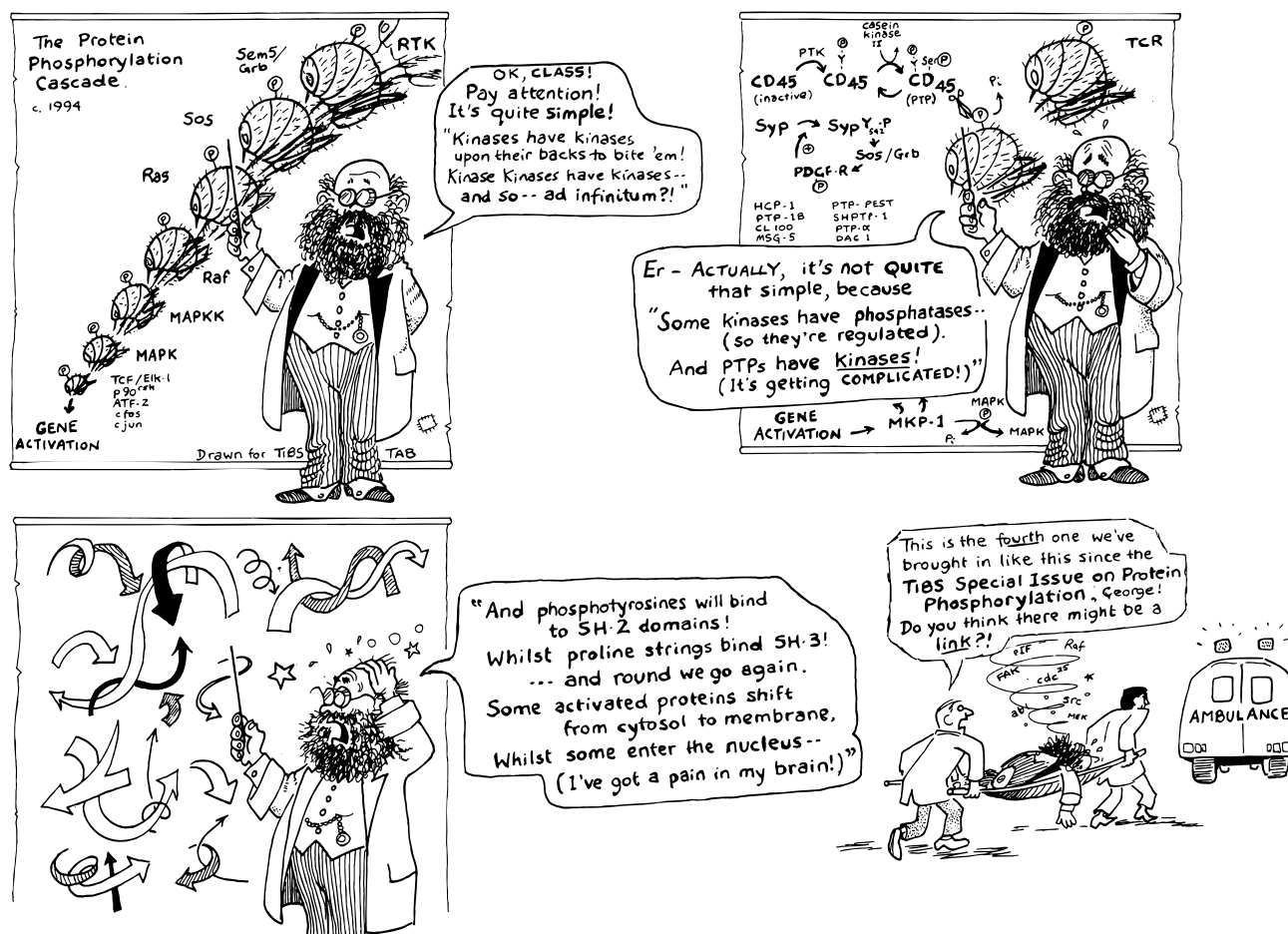
FIGURE 15.21 (ART BY TONY BRAMLEY, FROM TRENDS BIOCHEM. SCI. 19:469, 1994.)

Signaling by the Insulin Receptor

Our bodies spend considerable effort maintaining blood glucose levels within a narrow range. A decrease in blood glucose levels can lead to loss of consciousness and coma, as the central nervous system depends largely on glucose for its energy metabolism. A persistent elevation in blood glucose levels results in a loss of glucose, fluids, and electrolytes in the urine and serious health problems. The levels of glucose in the circulation are monitored by the pancreas. When blood glucose levels fall below a certain level, the alpha cells of the pancreas secrete glucagon. As discussed earlier, glucagon acts through GPCRs and stimulates the breakdown of glycogen resulting in an increase in blood glucose levels. When glucose levels rise, as occurs after a carbohydrate-rich meal, the beta cells of the pancreas respond by secreting insulin. Insulin functions as an extracellular messenger molecule, informing cells that glucose levels are high. Cells that express insulin receptors on their surface, such as cells in the liver, respond to this message by increasing glucose uptake, increasing glycogen and triglyceride synthesis, and decreasing gluconeogenesis.

The Insulin Receptor Is a Protein-Tyrosine Kinase Each insulin receptor is composed of an α and a β chain, which are derived from a single precursor protein by proteolytic processing. The α chain is entirely extracellular and contains the insulin-binding site. The β chain is composed of an extracellular region, a single transmembrane region, and a cytoplasmic region (Figure 15.22). The α and β chains are linked together by disulfide bonds (Figure 15.22). Two of these αβ heterodimers are held together by disulfide bonds between the α chains. Thus, while most RTKs are thought to be present on the cell surface as monomers, insulin receptors are present as stable dimers. Like other RTKs, insulin receptors are inactive in the absence of ligand (Figure 15.22a). Recent work suggests that the insulin receptor dimer binds a single insulin molecule. This causes repositioning of the ligand-binding domains on the outside of the cell, which causes the tyrosine

kinase domains on the inside of the cell to come into close physical proximity (Figure 15.22b). Juxtaposition of the kinase domains leads to trans-autophosphorylation and receptor activation (Figure 15.22c).

Several tyrosine phosphorylation sites have been identified in the cytoplasmic region of the insulin receptor. Three of these phosphorylation sites are present in the activation loop. In the unphosphorylated state, the activation loop assumes a conformation in which it occupies the active site. Upon phosphorylation of the three tyrosine residues, the activation loop assumes a new conformation away from the catalytic cleft. This new conformation requires a rotation of the small and large lobes of the kinase domain with respect to each other, thereby bringing residues that are essential for catalysis closer together. In addition, the activation loop now leaves the catalytic cleft open so that it can bind substrates. Following activation of the kinase domain, the receptor phosphorylates itself on tyrosine residues that are present adjacent to the membrane and in the carboxyl-terminal tail (Figure 15.22c).

Insulin Receptor Substrates 1 and 2 Most RTKs possess autophosphorylation sites that directly recruit SH2 domain-containing signaling proteins (as in Figure 15.17a, c, and d). The insulin receptor is an exception to this general rule, because it associates instead with a small family of docking proteins (Figure 15.17b), called **insulin-receptor substrates** (**IRSs**). The IRSs, in turn, provide the binding sites for SH2 domain-containing signaling proteins. Some of the events that occur during insulin signaling are shown in Figure 15.23. Following ligand binding and kinase activation, the insulin receptor autophosphorylates tyrosine 960, which then forms a binding site for the phosphotyrosine binding (PTB) domains of insulin receptor substrates. As indicated in Figure 15.23a, IRSs are characterized by the presence of an N-terminal PH domain, a PTB domain, and a long tail containing tyrosine phosphorylation sites. The PH domain may interact with phospholipids present at the inside leaflet of the plasma mem-

(a)   *(b)*   *(c)*
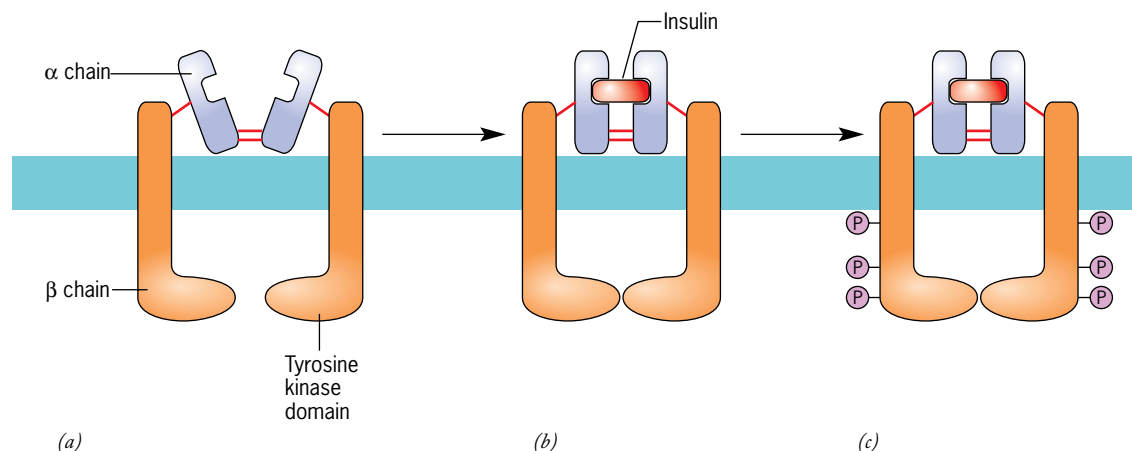
FIGURE 15.22 The response of the insulin receptor to ligand binding. (a) The insulin receptor, shown here in schematic form in the inactive state, is a tetramer consisting of two α and two β subunits. (b) Binding of a single insulin molecule to the α subunits causes a conformational change in the β subunits, which activates the tyrosine kinase activity of the β subunits. (c) The activated β subunits phosphorylate tyrosine residues located on the cytoplasmic domain of the receptor as well as tyrosine residues on several insulin receptor substrates (IRSs) that are discussed below.

(a)

(b)

(c)
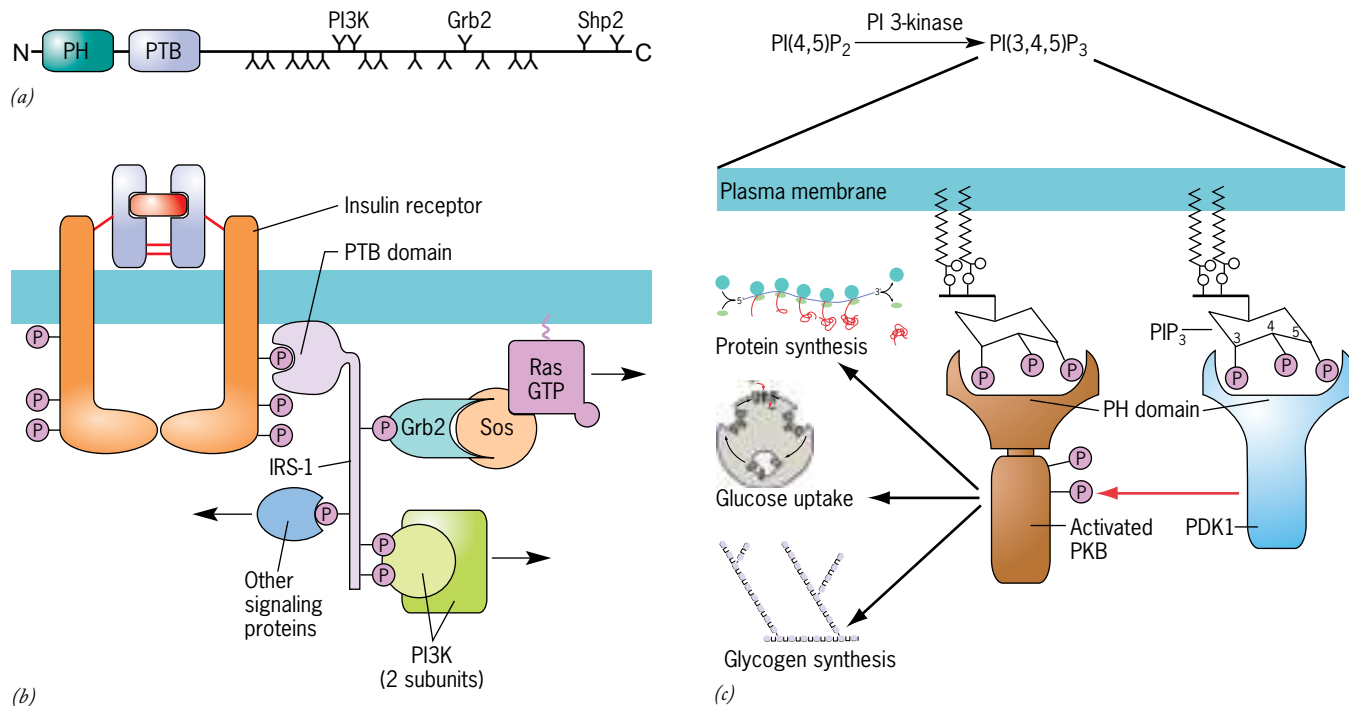
FIGURE 15.23 The role of tyrosine-phosphorylated IRS in activating a variety of signaling pathways. (*a*) Schematic representation of an IRS polypeptide. The N-terminal portion of the molecule contains a PH domain that allows it to bind to phosphoinositides of the membrane and a PTB domain that allows it to bind to a specific phosphorylated tyrosine residue (#960) on the cytoplasmic domain of an activated insulin receptor. Once bound to the insulin receptor, a number of tyrosine residues in the IRS may be phosphorylated (indicated as Y). These phosphorylated tyrosines can serve as binding sites for other proteins, including a lipid kinase (PI3K), an adaptor protein (Grb2), and a protein-tyrosine phosphatase (Shp2). (*b*) Phosphorylation of IRSs by the activated insulin receptor is known to activate PI3K and Ras pathways, both of which are discussed in the chapter. Other pathways that are less well defined are also activated by IRSs. (The IRS is drawn as an extended, two-dimensional

molecule for purposes of illustration.) (*c*) Activation of PI3K leads to the formation of membrane-bound phosphoinositides, including PIP$_3$. One of the key kinases in numerous signaling pathways is PKB (AKT), which interacts at the plasma membrane with PIP$_3$ by means of a PH domain on the protein. This interaction changes the conformation of the PKB molecule, making it a substrate for another PIP$_3$-bound kinase (PDK1), which phosphorylates PKB. The second phosphate shown linked to PKB is added by a second kinase, mostly likely mTOR. Once activated, PKB dissociates from the plasma membrane and moves into the cytosol and nucleus. PKB is a major component of a number of separate signaling pathways that mediate the insulin response. These pathways lead to translocation of glucose transporters to the plasma membrane, synthesis of glycogen, and the synthesis of new proteins in the cell.

brane, the PTB domain binds to tyrosine phosphorylation sites on the activated receptor, and the tyrosine phosphorylation sites provide docking sites for SH2 domain-containing signaling proteins. At least four members of the IRS family have been identified. Based on the results obtained in knock-out experiments in mice, it is thought that IRS-1 and IRS-2 are most relevant to insulin-receptor signaling.

Autophosphorylation of the activated insulin receptor at tyrosine 960 provides a binding site for IRS-1 or IRS-2. Only after stable association with either IRS-1 or IRS-2 is the activated insulin receptor able to phosphorylate tyrosine residues present on these docking proteins (Figure 15.23*b*). Both IRS-1 and IRS-2 contain a large number of potential tyrosine phosphorylation sites that include binding sites for the SH2 domains of PI 3-kinase, Grb2, and Shp2 (Figure 15.23*a*,*b*). These proteins associate with the receptor-bound IRS-1 or IRS-2 and activate downstream signaling pathways.

PI 3-kinase (PI3K) is composed of two subunits, one containing two SH2 domains and the other containing the catalytic domain (Figure 15.23*b*). PI 3-kinase, which is activated

directly as a consequence of binding of its two SH2 domains to tyrosine phosphorylation sites, phosphorylates phosphoinositides at the 3 position of the inositol ring (Figure 15.23*c*). The products of this enzyme, which include PI 3,4-bisphosphate (PIP$_2$) and PI 3,4,5-trisphosphate (PIP$_3$), remain in the cytosolic leaflet of the plasma membrane where they provide binding sites for PH domain-containing signaling proteins such as the serine-threonine kinases PKB and PDK1. As indicated in Figure 15.23*c*, PKB (also known as AKT) plays a role in mediating the response to insulin, as well as to other extracellular signals. Recruitment of PDK1 to the plasma membrane, in close proximity to PKB, provides a setting in which PDK1 can phosphorylate and activate PKB (Figure 15.23*c*). While phosphorylation by PDK1 is essential, it is not sufficient for activation of PKB. Activation of PKB also depends on phosphorylation by a second kinase, most likely mTOR.

Glucose Transport PKB is directly involved in regulating glucose transport and glycogen synthesis. The glucose transporter GLUT4 carries out insulin-dependent glucose trans-
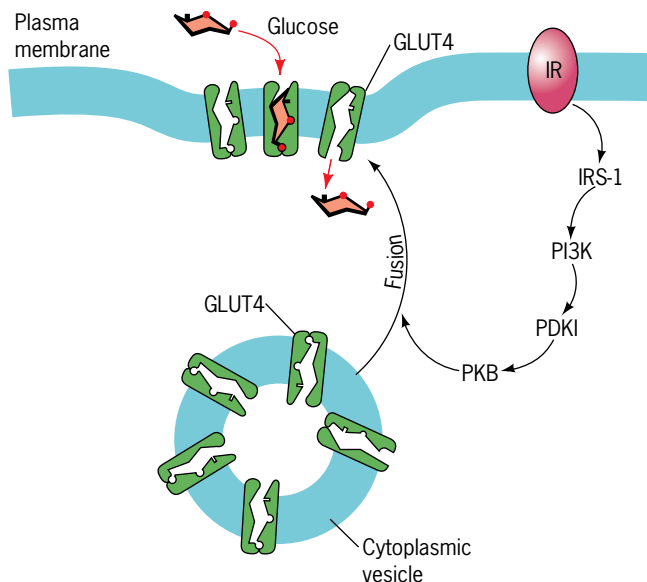
FIGURE 15.24 **Regulation of glucose uptake in muscle and fat cells by insulin.** Glucose transporters are stored in the walls of cytoplasmic vesicles that form by budding from the plasma membrane (endocytosis). When the insulin level increases, a signal is transmitted through the IRS-PI3K-PKB pathway, which triggers the translocation of cytoplasmic vesicles to the cell periphery. The vesicles fuse with the plasma membrane (exocytosis), delivering the transporters to the cell surface where they can mediate glucose uptake. A second pathway leading from the insulin receptor to GLUT4 translocation is not shown (see *Trends Biochem. Sci.* 31:215, 2006). (AFTER D. VOET AND J. G. VOET, BIOCHEMISTRY, 2D ED.; COPYRIGHT © 1995, JOHN WILEY & SONS, INC. REPRINTED BY PERMISSION OF JOHN WILEY & SONS, INC.)

port from the blood (page 152). In the absence of insulin, GLUT4 is present in membrane vesicles in the cytoplasm of insulin responsive cells (Figure 15.24). These vesicles fuse with the plasma membrane in response to insulin, a process that is referred to as GLUT4 translocation. The increase in numbers of glucose transporters in the plasma membrane leads to increased glucose uptake (Figure 15.24). GLUT4 translocation depends on activation of PI 3-kinase and PKB. This conclusion is based on experiments showing that inhibitors of PI3K block GLUT4 translocation. In addition, overexpression of PI3K or PKB stimulates GLUT4 translocation. It is well known that many receptors activate PI3K, whereas it is only the insulin receptor that stimulates GLUT4 translocation. This suggests that there is a second pathway downstream of the insulin receptor that is essential for GLUT4 translocation to occur. Detailed understanding of how the two pathways work together to stimulate GLUT4 translocation is still lacking.

Excess glucose that is taken up by muscle and liver cells is stored in the form of glycogen. Glycogen synthesis is carried out by glycogen synthase, an enzyme that is turned off by phosphorylation on serine and threonine residues. Glycogen synthase kinase-3 (GSK-3) has been identified as a negative regulator of glycogen synthase. GSK-3, in turn, is inactivated following phosphorylation by PKB. Thus, activation of the PI 3-kinase-PKB pathway in response to insulin leads to a decrease in GSK-3 kinase activity, resulting in an increase in

glycogen synthase activity (Figure 15.23c). Activation of protein phosphatase 1, an enzyme known to dephosphorylate glycogen synthase, contributes further to glycogen synthase activation.

Diabetes Mellitus   One of the most common human diseases, diabetes mellitus, is caused by defects in insulin signaling. Diabetes occurs in two varieties: type 1, which accounts for 5–10 percent of the cases, and type 2, which accounts for the remaining 90–95 percent. Type 1 diabetes is caused by an inability to produce insulin and is discussed in the Human Perspective of Chapter 17. Type 2 diabetes is a more complex disease whose incidence is increasing around the world at an alarming rate. The rising incidence of the disease is most likely a result of changing lifestyle and eating habits. A high-calorie diet combined with a sedentary lifestyle is thought to lead to a chronic increase in insulin secretion. Elevated levels of insulin overstimulate target cells in the liver and elsewhere in the body, which leads to a condition referred to as *insulin resistance*, in which these target cells stop responding to the presence of the hormone. According to genetic experiments in mice, mutations in the genes for either the insulin receptor or IRS-2, which renders cells unable to respond to insulin, creates a diabetic phenotype. However, mutations in these genes are rare in the human population and the exact basis of insulin resistance remains to be established.

Insulin Signaling and Lifespan   We saw on page 34 that the lifespan of an animal can be significantly increased by reducing its food intake (i.e., calorie restriction). A number of early studies demonstrated that the lifespan of a worm or fruit fly can also be increased by decreasing the level of insulin (or insulin-like growth factors) that circulate in their bloodstream. Studies of humans support this relationship; humans that live exceptionally long lives often exhibit unusually high insulin sensitivity—that is, their tissues respond fully to relatively low circulating insulin levels. Thus, just as increased insulin resistance is associated with poor health, increased insulin sensitivity appears to be associated with good health. It is interesting to note that calorie restriction in laboratory animals leads to decreased insulin levels and increased insulin sensitivity, so that these two paths to increased longevity may be acting by the same mechanism.

Signaling Pathways in Plants

Plants and animals share certain basic signaling mechanisms, including the use of Ca^{2+} and phosphoinositide messengers, but other pathways are unique to each major kingdom. For example, cyclic nucleotides, which may be the most ubiquitous animal cell messengers, appear to play little, if any, role in plant cell signaling. Receptor tyrosine kinases are also lacking in plant cells. On the other hand, plants contain a type of protein kinase that is absent from animal cells.

It has long been known that bacterial cells have a protein kinase that phosphorylates histidine residues and mediates the cell's response to a variety of environmental signals. Until 1993, these enzymes were thought to be restricted to bacterial cells but were then discovered in both yeast and flowering

plants. In both types of eukaryotes, the enzymes are trans-membrane proteins with an extracellular domain that acts as a receptor for external stimuli and a cytoplasmic, histidine kinase domain that transmits the signal to the cytoplasm. One of the best studied of these plant proteins is encoded by the *Etr1* gene. The product of the *Etr1* gene encodes a receptor for the gas ethylene (C_2H_4), a plant hormone that regulates a diverse array of developmental processes, including seed germination, flowering, and fruit ripening. Binding of ethylene to its receptor leads to transmission of signals along a pathway that is very similar to the MAP kinase cascade found in yeast and animal cells. As in other eukaryotes, the downstream targets of the MAP kinase pathway in plants are transcription factors that activate expression of specific genes encoding proteins required for the hormone response. As researchers analyze the massive amount of data obtained from sequencing *Arabidopsis* and other plant genomes, the similarities and differences between plant and animal signaling pathways should become more apparent.

REVIEW ?

1. Describe the steps between the binding of an insulin molecule at the surface of a target cell and the activation of the effector PI3K. How does the action of insulin differ from other ligands that act by means of receptor tyrosine kinases?

2. What is the role of Ras in signaling pathways? How is this affected by the activity of a Ras-GAP? How does Ras differ from a heterotrimeric G protein?

3. What is an SH2 domain, and what role does it play in signaling pathways?

4. How does the MAP kinase cascade alter the transcriptional activity of a cell?

5. What is the relationship between type 2 diabetes and insulin production? How is it that a drug that increases insulin sensitivity might help treat this disease?

15.5 THE ROLE OF CALCIUM AS AN INTRACELLULAR MESSENGER

Calcium ions play a significant role in a remarkable variety of cellular activities, including muscle contraction, cell division, secretion, fertilization, synaptic transmission, metabolism, transcription, cell movement, and cell death. In each of these cases, an extracellular message is received at the cell surface and leads to a dramatic increase in concentration of calcium ions within the cytosol. The concentration of calcium ions in a particular cellular compartment is controlled by the regulated activity of Ca^{2+} pumps, Ca^{2+} exchangers, and/or Ca^{2+} ion channels located within the membranes that surround the compartment (as in Figure 15.26). The concentration of Ca^{2+} ions in the cytosol of a resting cell is maintained at very low levels, typically about 10^{-7} M. In contrast, the concentration of this ion in the

extracellular space or within the lumen of the ER or a plant cell vacuole is typically 10,000 times higher than the cytosol. The cytosolic calcium level is kept very low because (1) Ca^{2+} ion channels in both the plasma and ER membranes are normally kept closed, making these membranes highly impermeable to this ion, and (2) energy-driven Ca^{2+} transport systems of the plasma and ER membranes pump calcium out of the cytosol.[2] Abnormal elevation of cytosolic Ca^{2+} concentration, as can occur in brain cells following a stroke, can lead to massive cell death.

IP$_3$ and Voltage-Gated Ca^{2+} Channels We have described in previous pages two major types of signaling receptors, GPCRs and RTKs. It was noted on page 616 that interaction of an extracellular messenger molecule with a GPCR can lead to the activation of the enzyme phospholipase C-β, which splits the phosphoinositide PIP_2, to release the molecule IP_3, which opens calcium channels in the ER membrane, leading to a rise in cytosolic $[Ca^{2+}]$. Extracellular messengers that signal through RTKs can trigger a similar response. The primary difference is that RTKs activate members of the phospholipase C-γ subfamily, which possess an SH2 domain that allows them to bind to the activated, phosphorylated RTK. There are numerous other PLC isoforms. For example, PLCδ is actived by Ca^{2+} ions, and PLCε is activated by Ras-GTP. All PLC isoforms carry out the same reaction, producing IP_3 and linking a multitude of cell surface receptors to an increase in cytoplasmic Ca^{2+}. There is another major route leading to elevation of cytosolic $[Ca^{2+}]$, which was encountered in our discussion of synaptic transmission on page 163. In this case, a nerve impulse leads to a depolarization of the plasma membrane, which triggers the opening of voltage-gated calcium channels in the plasma membrane, allowing the influx of Ca^{2+} ions from the extracellular medium.

Visualizing Cytoplasmic Ca^{2+} Concentration in Real Time Our understanding of the role of Ca^{2+} ions in cellular responses has been greatly advanced by the development of indicator molecules that emit light in the presence of free calcium. In the mid-1980s, new types of highly sensitive, fluorescent, calcium-binding compounds (e.g., *fura*-2) were developed in the laboratory of Roger Tsien at the University of California, San Diego. These compounds are synthesized in a form that can enter a cell by diffusing across its plasma membrane. Once inside a cell, the compound is modified to a form that is unable to leave the cell. Using these probes, the concentration of free calcium ions in different parts of a living cell can be determined over time by monitoring the light emitted using a fluorescence microscope and computerized imaging techniques. Use of calcium-sensitive, light-emitting molecules has provided dramatic portraits of the complex spatial and temporal changes in free cytosolic calcium concentration that occur in a single cell in response to various types of stimuli. This is one of the advantages of studying calcium-mediated responses compared to responses mediated by other types of messengers whose location in a cell cannot be readily visualized.

---

[2]Mitochondria also play an important role in sequestering and releasing Ca^{2+} ions, but their role was considered in Chapter 5 and will not be discussed here.

Depending on the type of responding cell, a particular stimulus may induce repetitive oscillations in the concentration of free calcium ions, as seen in Figure 15.9; cause a wave of Ca^{2+} release that spreads from one end of the cell to the other (see Figure 15.27); or trigger a localized and transient release of Ca^{2+} in one part of the cell. Figure 15.25 shows a Purkinje cell, a type of neuron in the mammalian cerebellum that maintains synaptic contact with thousands of other cells through an elaborate network of postsynaptic dendrites. The micrograph in Figure 15.25 shows the release of free calcium in a localized region of the dendritic "tree" of the cell following synaptic activation. The burst of calcium release remains restricted to this region of the cell.

IP$_3$ receptors described earlier are one of two main types of Ca^{2+} ion channels present in the ER membrane; the other type are called *ryanodine receptors* (*RyRs*) because they bind the toxic plant alkaloid ryanodine. Ryanodine receptors are found primarily in excitable cells and are best studied in cardiac and skeletal muscle cells, where they mediate the rise in Ca^{2+} levels following the arrival of an action potential. Mutations in the cardiac RyR isoform have been linked to occurrences of sudden death during periods of exercise. Depending on the type of cell in which they are found, RyRs can be opened by a variety of agents, including calcium itself. The influx of a limited amount of calcium through open channels in the plasma membrane induces the opening of ryanodine receptors in the ER, causing the release of Ca^{2+} into the cytosol (Figure 15.26). This phenomenon is called *calcium-induced calcium release* (*CICR*).

FIGURE 15.25 Experimental demonstration of localized release of intracellular Ca^{2+} within a single dendrite of a neuron. The mechanism of IP$_3$-mediated release of Ca^{2+} from intracellular stores was described on page 617. In the micrograph shown here, which pictures an enormously complex Purkinje cell (neuron) of the cerebellum, calcium ions have been released locally within a small portion of the complex "dendritic tree." Calcium release (shown in red) was induced in the dendrite following the local production of IP$_3$, which followed repetitive activation of a nearby synapse. The sites of release of cytosolic Ca^{2+} ions are revealed by fluorescence from a fluorescent calcium indicator that was loaded into the cell prior to stimulation of the cell. (FROM ELIZABETH A. FINCH AND GEORGE J. AUGUSTINE, NATURE, VOL. 396, COVER OF 12/24/98; COPYRIGHT © 1998, BY MACMILLAN MAGAZINES LIMITED.)

FIGURE 15.26 Calcium-induced calcium release, as it occurs in a cardiac muscle cell. A depolarization in membrane voltage causes the opening of voltage-gated calcium channels in the plasma membrane, allowing entry of a small amount of Ca^{2+} into the cytosol (step 1). The calcium ions bind to ryanodine receptors in the SER membrane (step 2), leading to release of stored Ca^{2+} into the cytosol (step 3), which triggers the cell's contraction. The calcium ions are subsequently removed from the cytosol by the action of Ca^{2+} pumps located in the membrane of the SER (step 4) and a Na^+/Ca^{2+} secondary transport system in the plasma membrane (step 5), which leads to relaxation. This cycle is repeated after each heart beat. (REPRINTED WITH PERMISSION AFTER M. J. BERRIDGE, NATURE 361:317, 1993; COPYRIGHT © 1993, BY MACMILLAN MAGAZINES LIMITED.)

15 μm
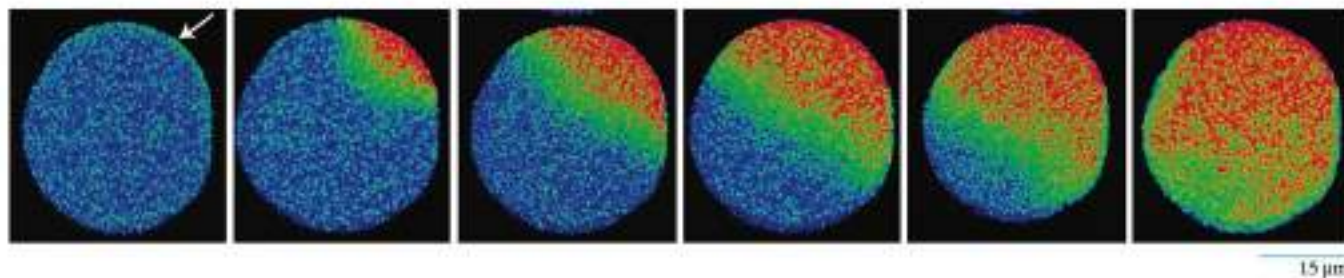
FIGURE 15.27 Calcium wave in a starfish egg induced by a fertilizing sperm. The unfertilized egg was injected with a calcium-sensitive fluorescent dye, fertilized, and photographed at 10-second intervals. The rise in Ca^{2+} concentration is seen to spread from the point of sperm entry (arrow) throughout the entire egg. The blue color indicates low free $[Ca^{2+}]$, whereas the red color indicates high free $[Ca^{2+}]$. A similar Ca^{2+} wave in mammalian eggs is triggered by the formation of IP_3 by a phospholipase C that is brought into the egg by the fertilizing sperm. (COURTESY OF STEPHEN A. STRICKER.)

Extracellular signals that are transmitted by Ca^{2+} ions typically act by opening a small number of Ca^{2+} ion channels at the cell surface at the site of the stimulus. As Ca^{2+} ions rush through these channels and enter the cytosol, they act on nearby Ca^{2+} ion channels in the ER, causing these channels to open and release additional calcium into adjacent regions of the cytosol. In some responses, the elevation of Ca^{2+} levels remains localized to a small region of the cytosol (as in Figure 15.25). In other cases, a propagated wave of calcium release spreads through the entire cytoplasmic compartment. One of the most dramatic Ca^{2+} waves occurs within the first minute or so following fertilization and is induced by the sperm's contact with the plasma membrane of the egg (Figure 15.27). The sudden rise in cytoplasmic calcium concentration following fertilization triggers a number of events, including activation of cyclin-dependent kinases (page 564) that drive the zygote

toward its first mitotic division. Calcium waves are transient because the ions are rapidly pumped out of the cytosol and back into the ER and/or the extracellular space.

Recent research in the field of calcium signaling has focused on a phenomenon known as *store-operated calcium entry* (or *SOCE*), in which the "store" refers to the calcium ions stored in the ER. During periods of repeated cellular responses, the stockpile of intracellular calcium ions can become depleted. During SOCE the depletion of calcium levels in the ER triggers a response leading to the opening of calcium channels in the plasma membrane as depicted in Figure 15.28. Once these channels have opened, Ca^{2+} ions can enter the cytosol from where they can be pumped back into the ER, thereby replenishing the ER's calcium stores. The mechanism responsible for SOCE had been an unsolved mystery for many years until it was discovered recently that these
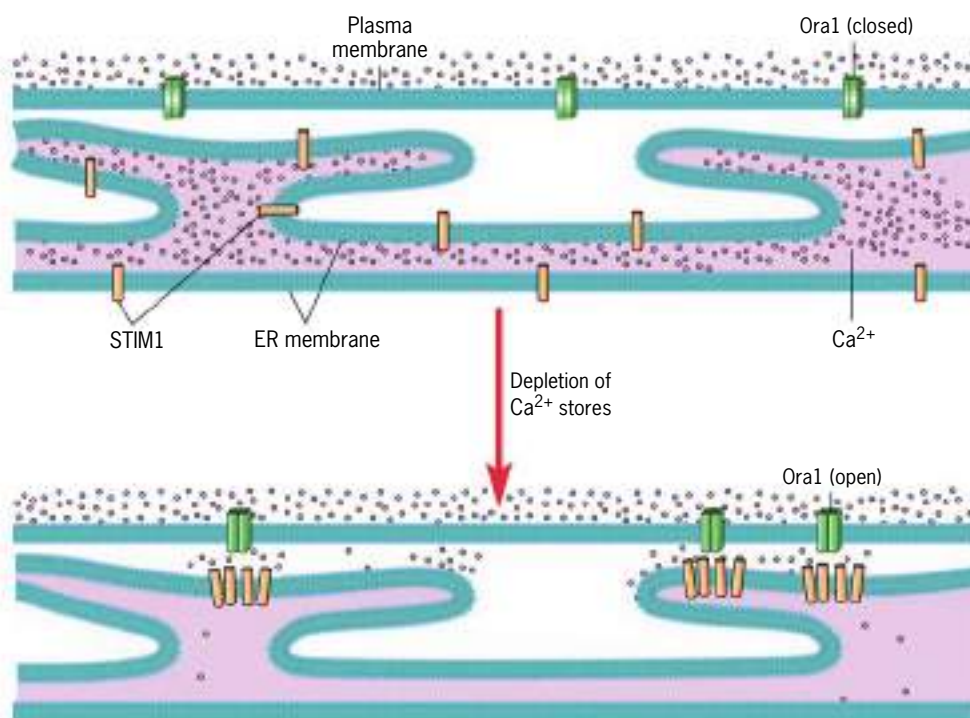
FIGURE 15.28 A model for store-operated calcium entry. When the ER lumen contains abundant Ca^{2+} ions, the STIM1 proteins of the ER membrane and the Ora1 proteins of the plasma membrane are situated diffusely in their respective membranes, and the Ora1 calcium channel is closed. If the ER stores are depleted, a signaling system operates between the two membranes, causing the two proteins to become clustered within their respective membranes in close proximity to one another. Apparent interaction between the two membrane proteins leads to opening of the Ora1 channel and the influx of Ca^{2+} ions into the cytosol from where they can be pumped into the ER lumen.

events are orchestrated by a signaling system operating between the ER and plasma membrane. In this system, the depletion of Ca^{2+} in the ER leads to the clustering within the ER membrane of a Ca^{2+}-sensing protein called STIM1. As the STIM1 proteins are rearranging themselves within the ER membrane, a protein called Ora1 is undergoing a corresponding redistribution within the plasma membrane. Ora1 is a tetrameric Ca^{2+} ion channel that had been identified as being involved in a particular type of inherited human immune deficiency that results from a lack of Ca^{2+} stores in T lymphocytes. The ER and plasma membrane typically come into very close proximity with one another at specific sites, and it is at these sites where the STIM1 and Ora1 clusters, each within their own respective membranes, come into apparent contact with one another (Figure 15.28). Contact between these proteins leads to the opening of the Ora1 channels, the influx of Ca^{2+}, and the refilling of the cell's ER stores.

Ca^{2+}-Binding Proteins   Unlike cAMP, whose action is usually mediated by stimulation of a protein kinase, calcium can affect a number of different types of cellular effectors, including protein kinases (Table 15.5). Depending on the cell type, calcium ions can activate or inhibit various enzyme and transport systems, change the ionic permeability of membranes, induce membrane fusion, or alter cytoskeletal structure and function. Calcium does not bring about these responses by itself but acts in conjunction with a number of **calcium-binding proteins** (examples are discussed on pages 297 and 364). The best-studied calcium-binding protein is **calmodulin**, which participates in many signaling pathways.

Calmodulin is found universally in plants, animals, and eukaryotic microorganisms, and it has virtually the same amino acid sequence from one end of the eukaryotic spectrum to the other. Each molecule of calmodulin (Figure 15.29) contains four binding sites for calcium. Calmodulin does not have sufficient affinity for Ca^{2+} to bind the ion in a nonstimulated cell. If, however, the Ca^{2+} concentration rises in response to a stimulus, the ions bind to calmodulin, changing the conformation of the protein and increasing its affinity for a variety of effectors. Depending on the cell type, the calcium–calmodulin (Ca^{2+}–CaM) complex may bind to a protein kinase, a cyclic nucleotide phosphodiesterase, ion channels, or even to the calcium-transport system of the plasma membrane. In the latter instance, rising levels of calcium activate the system responsible for ridding the cell of excess quantities of the ion, thus constituting a self-regulatory mechanism for maintaining

TABLE 15.5 Examples of Mammalian Proteins Activated by Ca^{2+}

Protein	Protein function
Troponin C	Modulator of muscle contraction
Calmodulin	Ubiquitous modulator of protein kinases and other enzymes (MLCK, CaM kinase II, adenylyl cyclase I)
Calretinin, retinin	Activator of guanylyl cylase
Calcineurin B	Phosphatase
Calpain	Protease
PI-specific PLC	Generator of IP_3 and diacylglycerol
α-Actinin	Actin-bundling protein
Annexin	Implicated in endo- and exocytosis, inhibition of PLA_2
Phospholipase A_2	Producer of arachidonic acid
Protein kinase C	Ubiquitous protein kinase
Gelsolin	Actin-severing protein
IP_3 receptor	Effector of intracellular Ca^{2+} release
Ryanodine receptor	Effector of intracellular Ca^{2+} release
Na^+/Ca^{2+} exchanger	Effector of the exchange of Ca^{2+} for Na^+ across the plasma membrane
Ca^{2+}-ATPase	Pumps Ca^{2+} across membranes
Ca^{2+} antiporters	Exchanger of Ca^{2+} for monovalent ions
Caldesmon	Regulator of muscle contraction
Villin	Actin organizer
Arrestin	Terminator of photoreceptor response
Calsequestrin	Ca^{2+} buffer

Adapted from D. E. Clapham, *Cell* 80:260, 1995, by copyright permission of Cell Press.

FIGURE 15.29 Calmodulin. A ribbon diagram of calmodulin (CaM) with four bound calcium ions (white spheres). Binding of these Ca^{2+} ions changes the conformation of calmodulin, exposing a hydrophobic surface that promotes interaction of Ca^{2+}–CaM with a large number of target proteins. (Courtesy Michael Carson, University of Alabama at Birmingham.)

low intracellular calcium concentrations. The Ca^{2+}–CaM complex can also stimulate gene transcription through activation of various protein kinases (CaMKs) that phosphorylate transcription factors. In the best-studied case, one of these protein kinases phosphorylates CREB on the same serine residue as PKA (Figure 15.12).

(a)

(b)
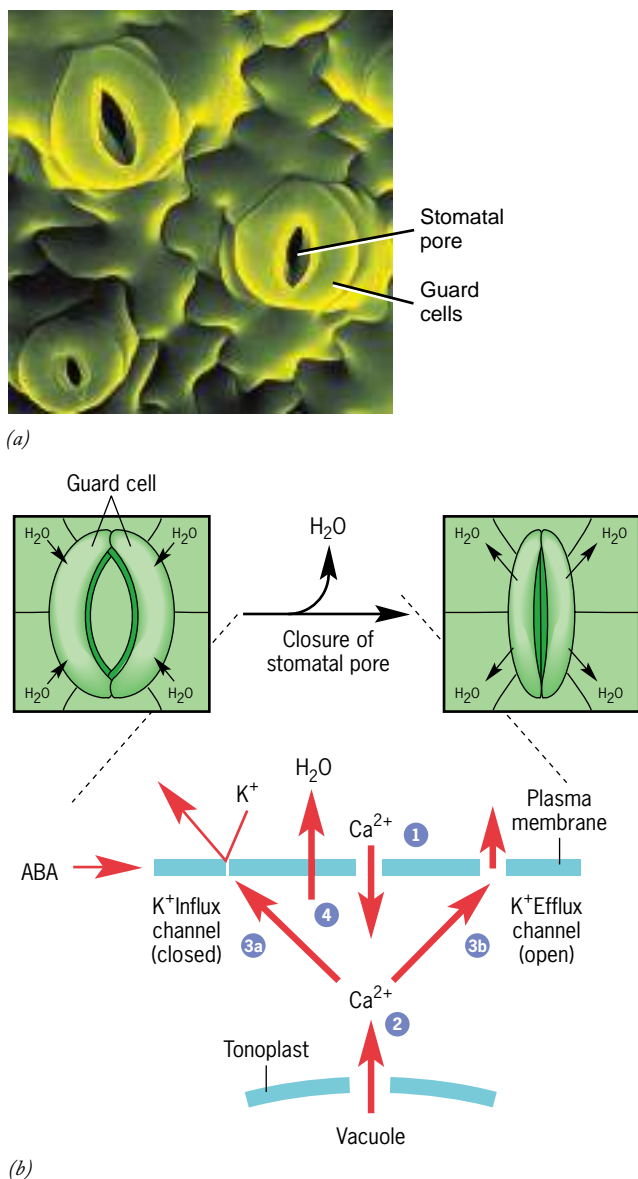
FIGURE 15.30 **A simplified model of the role of Ca^{2+} in guard cell closure.** (*a*) Photograph of stomatal pores, each flanked by a pair of guard cells. The stomata are kept open as turgor pressure is kept high within the guard cells, causing them to bulge outward as seen here. (*b*) One of the factors controlling stomatal pore size is the hormone abscisic acid (ABA). When ABA levels rise, calcium ion channels in the plasma membrane are opened, allowing the influx of Ca^{2+} (step 1), which triggers the release of Ca^{2+} from internal stores (step 2). The subsequent elevation of intracellular $[Ca^{2+}]$ closes K^+ influx channels (step 3a) and opens K^+ efflux channels (step 3b). K^+ efflux is accompanied by Cl^- efflux. These ion movements lead to a drop in internal solute concentration and the osmotic loss of water (step 4). (A: JEREMY BURGESS/PHOTO RESEARCHERS.)

Regulating Calcium Concentrations in Plant Cells

Calcium ions (acting in conjunction with calmodulin) are important intracellular messengers in plant cells. The levels of cytosolic calcium change dramatically within certain plant cells in response to a variety of stimuli, including changes in light, pressure, gravity, and the concentration of plant hormones such as abscisic acid. The concentration of Ca^{2+} in the cytosol of a resting plant cell is kept very low by the action of transport proteins situated in the plasma membrane and vacuolar membrane (tonoplast).

The role of Ca^{2+} in plant cell signaling is illustrated by guard cells that regulate the diameter of the microscopic pores (stomata) of a leaf (Figure 15.30*a*). Stomata are a major site of water loss in plants (page 226), and the diameter of their aperture is tightly controlled, which prevents dehydration. The diameter of the stomatal pore decreases as the fluid (turgor) pressure in the guard cell decreases. The drop in turgor pressure is caused in turn by a decrease in the ionic concentration (osmolarity) of the guard cell. Adverse conditions, such as high temperatures and low humidity, stimulate the release of the plant stress hormone abscisic acid. Recent studies suggest that abscisic acid binds to a GPCR in the plasma membrane of guard cells, triggering the opening of Ca^{2+} ion channels in the same membrane (Figure 15.30*b*). The resulting influx of Ca^{2+} into the cytosol triggers the release of additional Ca^{2+} from intracellular stores. The elevated cytosolic Ca^{2+} concentration leads to closure of K^+ influx channels in the plasma membrane and opening of K^+ efflux channels. These changes produce a net outflow of K^+ ions (and accompanying Cl^- ions) and a decrease in turgor pressure.

REVIEW ?

1. How is the $[Ca^{2+}]$ of the cytosol maintained at such a low level? How does the concentration change in response to stimuli?

2. What is the role of calcium-binding proteins such as calmodulin in eliciting a response?

3. Describe the role of calcium in mediating the diameter of stomata in guard cells.

15.6 CONVERGENCE, DIVERGENCE, AND CROSS-TALK AMONG DIFFERENT SIGNALING PATHWAYS

The signaling pathways described above and illustrated schematically in the various figures depict linear pathways leading directly from a receptor at the cell surface to an end target. In actual fact, signaling pathways in the cell are much more complex. For example:

■ Signals from a variety of unrelated receptors, each binding to its own ligand, can *converge* to activate a common effector, such as Ras or Raf.
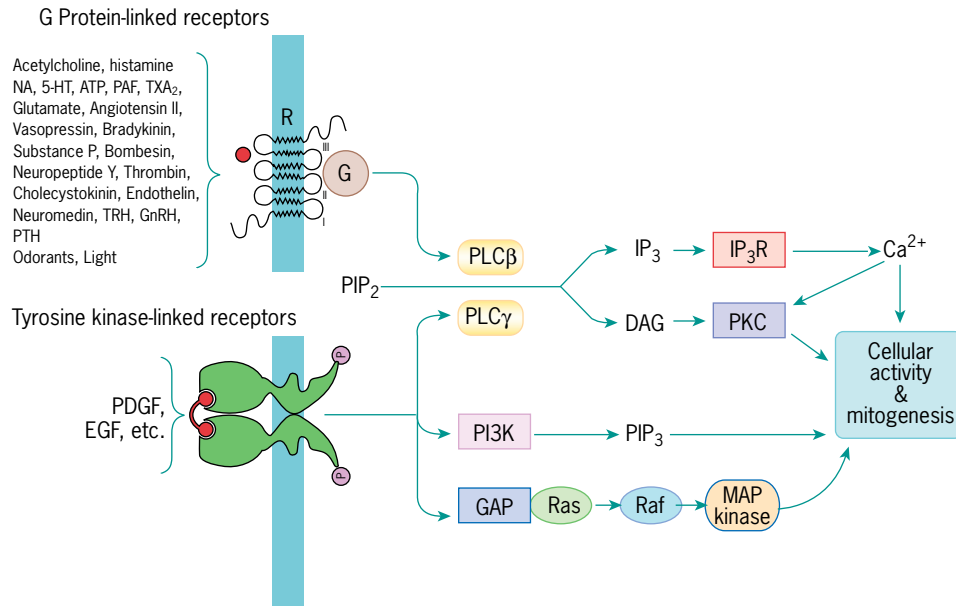
FIGURE 15.31 Examples of convergence, divergence, and cross-talk among various signal-transduction pathways. This drawing shows the outlines of signal-transduction pathways initiated by receptors that act by means of both heterotrimeric G proteins and receptor protein-tyrosine kinases. The two are seen to converge by the activation of different phospholipase C isoforms, both of which lead to the production of the same second messengers (IP$_3$ and DAG). Activation of the RTK by either PDGF or EGF leads to the transmission of signals along three different pathways, an example of divergence. Cross-talk between the two types of pathways is illustrated by calcium ions, which are released from the SER by action of IP$_3$ and can then act on various proteins, including protein kinase C (PKC), whose activity is also stimulated by DAG. (AFTER M. J. BERRIDGE, REPRINTED WITH PERMISSION FROM NATURE VOL. 361, P. 315, 1993; © COPYRIGHT 1993, BY MACMILLAN JOURNALS LIMITED.)

- Signals from the same ligand, such as EGF or insulin, can *diverge* to activate a variety of different effectors, leading to diverse cellular responses.

- Signals can be passed back and forth between different pathways, a phenomenon known as *cross-talk*.

These characteristics of cell-signaling pathways are illustrated schematically in Figure 15.31.

Signaling pathways provide a mechanism for routing information through a cell, not unlike the way the central nervous system routes information to and from the various organs of the body. Just as the central nervous system collects information about the environment from various sense organs, the cell receives information about its environment through the activation of various surface receptors, which act like sensors to detect extracellular stimuli. Like sense organs that are sensitive to specific forms of stimuli (e.g., light, pressure, or sound waves), cell-surface receptors can bind only to specific ligands and are unaffected by the presence of a large variety of unrelated molecules. A single cell may have dozens of different receptors sending signals to the cell interior simultaneously. Once they have been transmitted into the cell, signals from these receptors can be selectively routed along a number of different signaling pathways that may cause a cell to divide, change shape, activate a particular metabolic pathway, or even commit suicide (discussed in a following section). In this way, the cell integrates information arriving from different sources and mounts an appropriate and comprehensive response.

Examples of Convergence, Divergence, and Cross-Talk Among Signaling Pathways

1. *Convergence.* We have discussed two distinct types of cell-surface receptors in this chapter: G protein-coupled receptors and receptor tyrosine kinases. Another type of cell-surface receptor that is capable of signal transduction was discussed in Chapter 7, namely, integrins. Although these three types of receptors may bind to very different ligands, all of them can lead to the formation of phosphotyrosine docking sites for the SH2 domain of the adaptor protein Grb2 in close proximity to the plasma membrane (Figure 15.32). The recruitment of the Grb2-Sos complex results in the activation of Ras and transmission of signals down the MAP kinase pathway. As a result of this convergence, signals from diverse receptors can lead to the
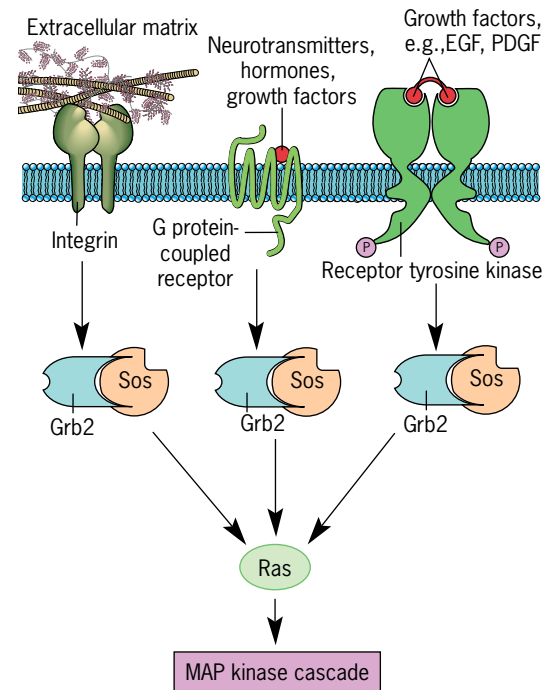
FIGURE 15.32 Signals transmitted from a G protein-coupled receptor, an integrin, and a receptor tyrosine kinase all converge on Ras and are then transmitted along the MAP kinase cascade.

transcription and translation of a similar set of growth-promoting genes in each target cell.

2. *Divergence.* Evidence of signal divergence has been evident in virtually all of the examples of signal transduction that have been described in this chapter. A quick look at Figure 15.13 or 15.23*b,c* illustrates how a single stimulus—a ligand binding to a GPCR or an insulin receptor—sends signals out along a variety of different pathways.

3. *Cross-talk.* In previous sections, we examined a number of signaling pathways as if each operated as an independent, linear chain of events. In fact, the information circuits that operate in cells are more likely to resemble an interconnected web in which components produced in one pathway can participate in events occurring in other pathways. The more that is learned about information signaling in cells, the more cross-talk between signaling pathways that is discovered. Rather than attempting to catalog the ways that information can be passed back and forth within a cell, we will look at an example involving cAMP that illustrates the importance of this type of cross-talk.

Cyclic AMP was depicted earlier as an initiator of a reaction cascade leading to glucose mobilization. However, cAMP can also inhibit the growth of a variety of cells, including fibroblasts and fat cells, by blocking signals transmitted through the MAP kinase cascade. Cyclic AMP is thought to accomplish this by activating PKA, the cAMP-dependent kinase, which can phosphorylate and inhibit Raf, the protein that heads the MAP kinase cascade (Figure 15.33). These two pathways also intersect at another important signaling effector, the transcription factor CREB. CREB was described on page 620 as a terminal effector of cAMP-mediated pathways. It was assumed for years that CREB could only be phosphorylated by the cAMP-dependent kinase, PKA. It has since become apparent that CREB is a substrate of a much wider range of kinases. For example, one of the kinases that phosphorylates CREB is Rsk-2, which is activated as a result of phosphorylation by MAPK (Figure 15.33). In fact, both PKA and Rsk-2 phosphorylate CREB on precisely the same amino acid residue, Ser133, which should endow the transcription factor with the same potential in both pathways.

A major unanswered question is raised by these examples of convergence, divergence, and cross-talk: How are different stimuli able to evoke distinct responses, even though they utilize similar pathways? PI3K, for example, is an enzyme that is activated by a remarkable variety of stimuli, including cell adhesion to the ECM, insulin, and EGF. How is it that activation of PI3K in an insulin-stimulated liver cell promotes GLUT4 translocation and protein synthesis, whereas activation of PI3K in an adherent epithelial cell promotes cell survival? Ultimately, these contrasting cellular responses must be due to differences in the protein composition of different cell types. Part of the answer probably lies in the fact that different cells have different isoforms of these various proteins, including PI3K. Some of these isoforms are encoded by different, but related genes, whereas others are generated by
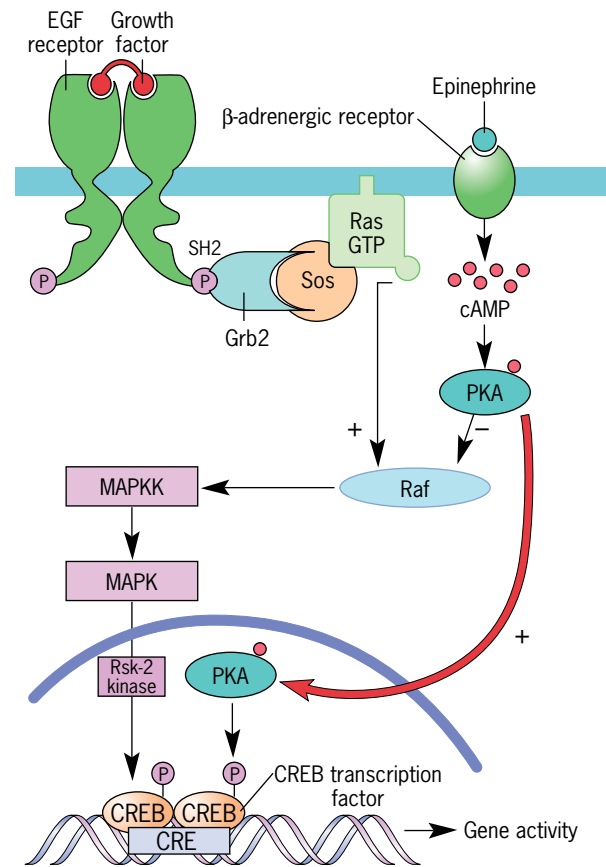
FIGURE 15.33 **An example of cross-talk between two major signaling pathways.** Cyclic AMP acts in some cells, by means of the cAMP-dependent kinase PKA, to block the transmission of signals from Ras to Raf, which inhibits the activation of the MAP kinase cascade. In addition, both PKA and the kinases of the MAP kinase cascade phosphorylate the transcription factor CREB on the same serine residue, activating the transcription factor and allowing it to bind to specific sites on the DNA.

alternative splicing (page 522), or other mechanisms. Different isoforms of PI3K, PKB, or PLC, for example, may bind to different sets of upstream and downstream components, which could allow similar pathways to evoke distinct responses. But it isn't likely that isoform variation can explain the extraordinary diversity of cellular responses any more than differences in the structures of neurons can explain the range of responses evoked by the nervous system. Hopefully, as the signaling pathways of more and more cells are described, we will gain a better understanding of the specificity that can be achieved through the use of similar signaling molecules.

15.7 THE ROLE OF NO AS AN INTERCELLULAR MESSENGER

During the 1980s, a new type of messenger was discovered that was neither an organic compound, such as cAMP, nor an ion, such as Ca^{2+}; it was an inorganic gas—nitric

oxide (NO). NO is unusual because it acts both as an extracellular messenger, mediating intercellular communication, and as a second messenger, acting within the cell in which it is generated. NO is formed from the amino acid L-arginine in a reaction that requires oxygen and NADPH and that is catalyzed by the enzyme nitric oxide synthase (NOS). Since its discovery, it has become evident that NO is involved in a myriad of biological processes including anticoagulation, neurotransmission, smooth muscle relaxation, and visual perception.

As with many other biological phenomena, the discovery that NO functions as a messenger molecule began with an accidental observation. It had been known for many years that acetylcholine acts in the body to relax the smooth muscle cells of blood vessels, but the response could not be duplicated in vitro. When portions of a major blood vessel such as the aorta were incubated in physiologic concentrations of acetylcholine in vitro, the preparation usually showed little or no response. In the late 1970s, Robert Furchgott, a pharmacologist at a New York State medical center, was studying the in vitro response of pieces of rabbit aorta to various agents. In his earlier studies, Furchgott used strips of aorta that had been dissected from the organ. For technical reasons, Furchgott switched from strips of aortic tissue to aortic rings and discovered that the new preparations responded to acetylcholine by undergoing relaxation. Further investigation revealed that the strips had failed to display the relaxation response because the delicate endothelial layer that lines the aorta had been rubbed away during the dissection. This surprising finding suggested that the endothelial cells were somehow involved in the response by the adjacent muscle cells. In subsequent studies, it was found that acetylcholine binds to receptors on the surface of endothelial cells, leading to the production and release of an agent that diffuses through the cell's plasma membrane and causes the muscle cells to relax. The diffusible agent was identified in 1986 as nitric oxide by Louis Ignarro at UCLA and Salvador Moncada at the Wellcome Research Labs in England. The steps in the acetylcholine-induced relaxation response are illustrated in Figure 15.34.

The binding of acetylcholine to the outer surface of an endothelial cell (step 1, Figure 15.34) signals a rise in cytosolic Ca^{2+} concentration (step 2) that activates nitric oxide synthase (step 3). The NO formed in the endothelial cell diffuses across the plasma membrane and into the adjacent smooth muscle cells (step 4), where it binds and stimulates guanylyl cyclase (step 5), the enzyme that synthesizes cyclic GMP (cGMP), which is an important second messenger similar in structure to cAMP. Cyclic GMP binds to a cGMP-dependent protein kinase (a PKG), which phosphorylates specific substrates causing relaxation of the muscle cell (step 6) and dilation of the blood vessel.

NO as an Activator of Guanylyl Cyclase

The discovery that NO acts as an activator of guanylyl cyclase was made in the late 1970s by Ferid Murad and colleagues at the University of Virginia. Murad was working with azide (N_3), a potent inhibitor of electron transport, and chanced to discover that the molecule stimulated cGMP production in cellular ex-
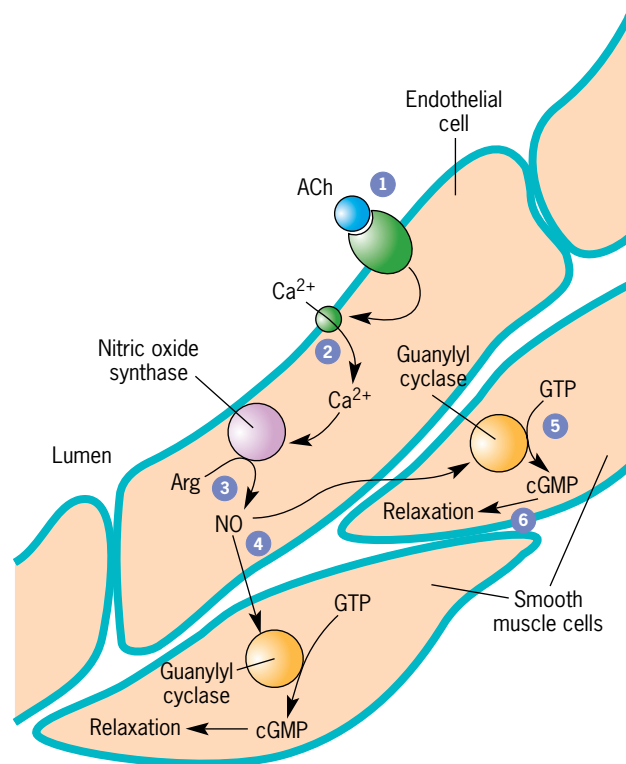
FIGURE 15.34 A signal transduction pathway that operates by means of NO and cyclic GMP that leads to the dilation of blood vessels. The steps illustrated in the figure are described in the text. (FROM R. G. KNOWLES AND S. MONCADA, TRENDS BIOCHEM. SCI. 17:401, 1992.)

tracts. Murad and colleagues ultimately demonstrated that azide was being converted enzymatically into nitric oxide, which served as the actual guanylyl cyclase activator. These studies also explained the action of nitroglycerine, which had been used since the 1860s to treat the pain of angina that results from an inadequate flow of blood to the heart. Nitroglycerine is metabolized to nitric oxide, which stimulates the relaxation of the smooth muscles lining the blood vessels of the heart, increasing blood flow to the organ. The therapeutic benefits of nitroglycerine were discovered through an interesting observation. Persons with heart disease who worked with nitroglycerine in Alfred Nobel's dynamite factory were found to suffer more from the pain of angina on days they weren't at work. It is only fitting that the Nobel Prize, which is funded by a donation from Alfred Nobel's estate, was awarded in 1998 for the discovery of NO as a signaling agent.

Inhibiting Phosphodiesterase

The discovery of NO as a second messenger has also led to the development of Viagra (sildenafil). During sexual arousal, nerve endings in the penis release NO, which causes relaxation of smooth muscle cells in the lining of penile blood vessels and engorgement of the organ with blood. As described above, NO mediates this response in smooth muscle cells by activation of the enzyme guanylyl cyclase and subsequent production of cGMP. Viagra (and related drugs) has no effect on the release of NO or the

activation of guanylyl cyclase, but instead acts as an inhibitor of cGMP phosphodiesterase, the enzyme that destroys cGMP. Inhibition of this enzyme leads to maintained, elevated levels of cGMP, which promotes the development and maintenance of an erection. Viagra is quite specific for one particular isoform of cGMP phosphodiesterase, PDE5, which is the version that acts in the penis. Another isoform of the enzyme, PDE3, plays a key role in the regulation of heart muscle contraction, but fortunately is not inhibited by the drug. Viagra was discovered when a potential angina medication had unexpected side effects.

Recent investigations have revealed that NO has a variety of actions within the body that do not involve production of cGMP. For example, NO is added to the —SH group of certain cysteine residues in a number of proteins, including hemoglobin, Ras, ryanodine channels, and caspases. This posttranslational modification, which is called *S-nitrosylation*, alters the activity, turnover, or interactions of the protein.

REVIEW ?

1. Describe the steps in the signaling pathway by which nitric oxide mediates dilation of blood vessels.

15.8 APOPTOSIS (PROGRAMMED CELL DEATH)

Apoptosis, or programmed cell death, is a normal occurrence in which an orchestrated sequence of events leads to the death of a cell. Death by apoptosis is a neat, orderly process (Figure 15.35) characterized by the overall shrinkage in volume of the cell and its nucleus, the loss of adhesion to neighboring cells, the formation of blebs at the cell surface, the dissection of the chromatin into small fragments, and the rapid engulfment of the "corpse" by phagocytosis. Because it is a safe and orderly process, apoptosis might be compared to the controlled implosion of a building using carefully placed explosives as compared to simply blowing up the structure without concern for what happens to the flying debris.

Why do our bodies have unwanted cells, and where do we find cells that become targeted for elimination? The short answer is: almost anywhere you look. It has been estimated that 10^{10}–10^{11} cells in the human body die every day by apoptosis. For example, apoptosis is involved in the elimination of cells that have sustained irreparable genomic damage. This is important because damage to the genetic blueprint can result in unregulated cell division and the development of cancer. During embryonic development, neurons grow out of the central nervous system to innervate organs that are present in the periphery of the body. Usually many more neurons grow out than are needed for normal innervation. Those neurons that reach their destination receive a signal from the target tissue that allows them to survive. Those neurons that fail to find their way to the target tissue do not receive the survival signal

(a)

(b)

(c)   2 µm
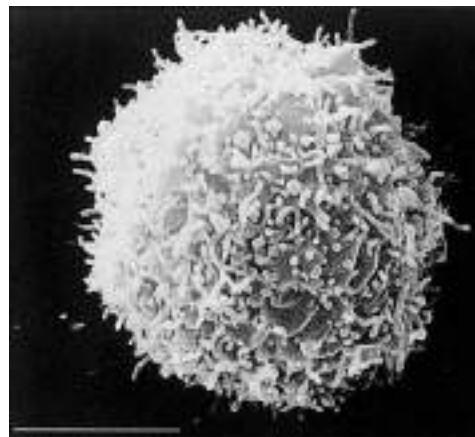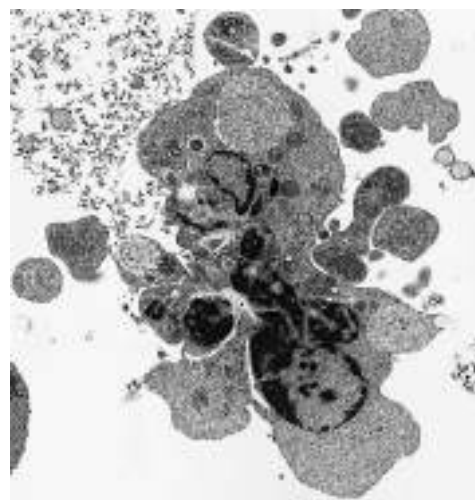
FIGURE 15.35 **A comparison of normal and apoptotic cells.** (*a,b*) Scanning electron micrographs of a normal (*a*) and apoptotic (*b*) T-cell hybridoma. The apoptotic cell exhibits many surface blebs that are budded off in the cell. Bar equals 4 mm. (*c*) Transmission electron micrograph of an apoptotic cell treated with an inhibitor that arrests apoptosis at the membrane blebbing stage. (A,B: FROM Y. SHI AND D. R. GREEN, IN S. J. MARTIN ET AL., TRENDS BIOCHEM. SCI. 19:28, 1994; C: COURTESY OF NICOLA J. MCCARTHY.)

and are ultimately eliminated by apoptosis. T lymphocytes are cells of the immune system that recognize and kill abnormal or pathogen-infected target cells. These target cells are recognized by specific receptors that are present on the surface of T lymphocytes. During embryonic development, T lymphocytes are produced that possess receptors capable of binding tightly to proteins present on the surface of *normal* cells within the body. T lymphocytes that have this dangerous capability are eliminated by apoptosis (see Figure 17.25). Finally, apoptosis appears to be involved in neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, and Huntington's disease. Elimination of essential neurons during disease progression gives rise to loss of memory or decrease in motor coordination. These examples show that apoptosis is important in maintaining homeostasis in multicellular organisms and that failure to regulate apoptosis can result in serious damage to the organism.

The term *apoptosis* was coined in 1972 by John Kerr, Andrew Wyllie, and A. R. Currie of the University of Aberdeen, Scotland, in a landmark paper that described for the first time the coordinated events that occurred during the programmed death of a wide range of cells. Insight into the molecular basis of apoptosis was first revealed in studies on the nematode worm *C. elegans*, whose cells can be followed with absolute precision during embryonic development. Of the 1090 cells produced during the development of this worm, 131 cells are normally destined to die by apoptosis. In 1986, Robert Horvitz and his colleagues at the Massachusetts Institute of Technology discovered that worms carrying a mutation in the *CED-3* gene proceed through development without losing any of their cells to apoptosis. This finding suggested that the product of the *CED-3* gene played a crucial role in the process of apoptosis in this organism. Once a gene has been identified in one organism, such as a nematode, researchers can search for homologous genes in other organisms, such as humans or other mammals. The identification of the *CED-3* gene in nematodes led to the discovery of a homologous family of proteins in mammals, which are now called **caspases**. Caspases are a distinctive group of cysteine proteases (i.e., proteases with a key cysteine residue in their catalytic site) that are activated at an early stage of apoptosis and are responsible for triggering most, if not all, of the changes observed during cell death. Caspases accomplish this feat by cleaving a select group of essential proteins. Among the targets of caspases are the following:

- *More than a dozen protein kinases, including focal adhesion kinase* (*FAK*), *PKB, PKC, and Raf1.* Inactivation of FAK, for example, is presumed to disrupt cell adhesion, leading to detachment of the apoptotic cell from its neighbors.

- *Lamins*, which make up the inner lining of the nuclear envelope. Cleavage of lamins leads to the disassembly of the nuclear lamina and shrinkage of the nucleus.

- *Proteins of the cytoskeleton*, such as those of intermediate filaments, actin, tubulin, and gelsolin. Cleavage and consequent inactivation of these proteins lead to changes in cell shape.

- *An endonuclease called caspase activated DNase* (*CAD*), which is activated following caspase cleavage of an inhibitory protein. Once activated, CAD translocates from the cytoplasm to the nucleus where it attacks DNA, severing it into fragments.

Recent studies have focused on the events that lead to the activation of a cell's suicide program. Apoptosis can be triggered by both internal stimuli, such as abnormalities in the DNA, and external stimuli, such as certain cytokines (proteins secreted by cells of the immune system). For example, epithelial cells of the prostate become apoptotic when deprived of the male sex hormone testosterone. This is the reason why prostate cancer that has spread to other tissues is often treated with drugs that interfere with testosterone production. Studies indicate that external stimuli activate apoptosis by a signaling pathway, called the *extrinsic pathway*, that is distinct from that utilized by internal stimuli, which is called the *intrinsic pathway*. Here we will discuss the extrinsic and intrinsic pathways separately. However, it should be noted that there is cross-talk between these pathways and that extracellular apoptotic signals can cause activation of the intrinsic pathway.

The Extrinsic Pathway of Apoptosis

The steps in the extrinsic pathway are illustrated in Figure 15.36. In the case depicted in the figure, the stimulus for apoptosis is carried by an extracellular messenger protein called tumor necrosis factor (TNF), which was named for its ability to kill tumor cells. TNF is produced by certain cells of the immune system in response to adverse conditions, such as exposure to ionizing radiation, elevated temperature, viral infection, or toxic chemical agents such as those used in cancer chemotherapy. Like other types of first messengers discussed in this chapter, TNF evokes its response by binding to a transmembrane receptor, TNFR1. TNFR1 is a member of a family of related "death receptors" that turns on the apoptotic process. The available evidence suggests that the TNF receptor is present in the plasma membrane as a preassembled trimer. The cytoplasmic domain of each TNF receptor subunit contains a segment of about 70 amino acids called a "death domain" (each green segment in Figure 15.36) that mediates protein–protein interactions. Binding of TNF to the trimeric receptor produces a change in conformation of the receptor's death domain, which leads to the recruitment of a number of proteins, as indicated in Figure 15.36.

The last proteins to join the complex that assembles at the inner surface of the plasma membrane are two procaspase-8 molecules (Figure 15.36). These proteins are called "procaspases" because each is a precursor of a caspase; it contains an extra portion that must be removed by proteolytic processing to activate the enzyme. The synthesis of caspases as proenzymes protects the cell from accidental proteolytic damage. Unlike most proenzymes, procaspases exhibit a low level of proteolytic activity. According to one model, when two or more procaspases are held in close association with one
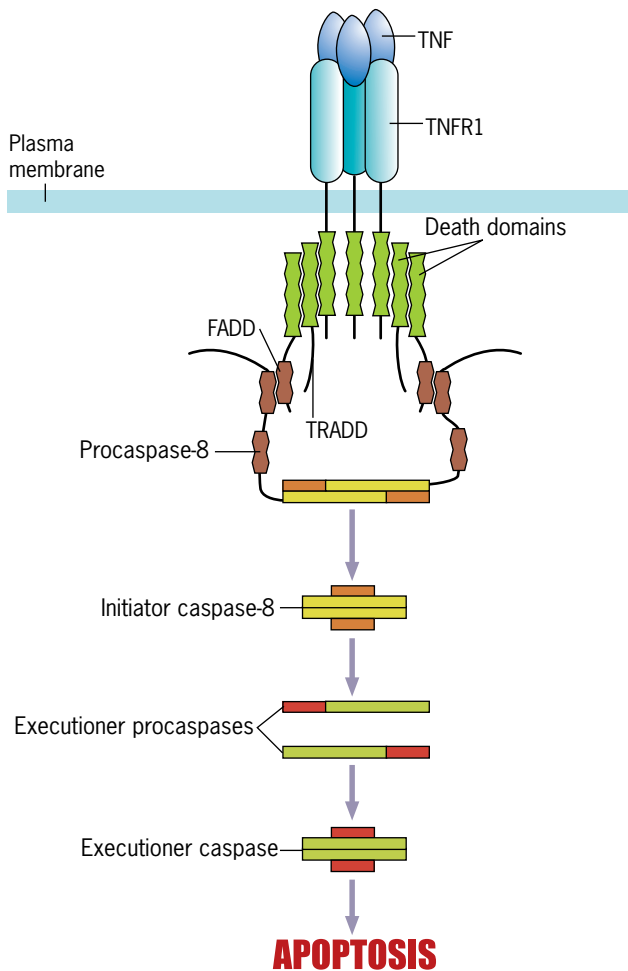
FIGURE 15.36 The extrinsic (receptor-mediated) pathway of apoptosis.
When TNF binds to a TNF receptor (TNFR1), the activated receptor
binds two different cytoplasmic adaptor proteins (TRADD and FADD)
and procaspase-8 to form a multiprotein complex at the inner surface of
the plasma membrane. The cytoplasmic domains of the TNF receptor,
FADD, and TRADD interact with one another by homologous regions
called death domains that are present in each protein (indicated as green
boxes). Procaspase-8 and FADD interact by means of homologous re-
gions called death effector domains (indicated as brown boxes). Once as-
sembled in the complex, the two procaspase molecules cleave one
another to generate an active caspase-8 molecule containing four
polypeptide segments. Caspase-8 is an initiator complex that activates
downstream (executioner) caspases that carry out the death sentence. It
can be noted that the interaction between TNF and TNFR1 also acti-
vates other signaling pathways, one of which leads to cell survival rather
than self-destruction.

another, as they are in Figure 15.36, they are capable of cleaving
one another's polypeptide chain and converting the other
molecule to the fully active caspase. The final mature enzyme
(i.e., caspase-8) contains four polypeptide chains, derived
from two procaspase precursors as illustrated in the figure.

Activation of caspase-8 is similar in principle to the acti-
vation of effectors by a hormone or growth factor. In all of
these signaling pathways, the binding of an extracellular li-
gand causes a change in conformation of a receptor that leads
to the binding and activation of proteins situated downstream

in the pathway. Caspase-8 is described as an *initiator* caspase
because it initiates apoptosis by cleaving and activating down-
stream, or *executioner*, caspases, that carry out the controlled
self-destruction of the cell as described above.

The Intrinsic Pathway of Apoptosis

Internal stimuli, such as irreparable genetic damage, lack of
oxygen (hypoxia), extremely high concentrations of cytosolic
Ca^{2+}, viral infection, or severe oxidative stress (i.e., the pro-
duction of large numbers of destructive free radicals, page 34)
trigger apoptosis by the intrinsic pathway illustrated in Figure
15.37. Activation of the intrinsic pathway is regulated by
members of the Bcl-2 family of proteins, which are charac-
terized by the presence of one or more BH domains. Bcl-2
family members can be subdivided into three groups: (1) pro-
apoptotic members that promote apoptosis (e.g., Bax and
Bak), (2) antiapoptotic members that protect cells from apop-
tosis (e.g., Bcl-x_L, Bcl-w, and Bcl-2),[3] and (3) BH3-only pro-
teins (so-named because they share only one small domain—the
BH3 domain—with other Bcl-2 family members), which
promote apoptosis by an indirect mechanism. According to
the prevailing view, BH3-only proteins (e.g., Bid, Bad, Puma,
and Bim) can exert their proapoptotic effect in two different
ways, depending on the particular proteins involved. In some
cases they apparently promote apoptosis by inhibiting anti-
apoptotic Bcl-2 members, whereas in other cases they appar-
ently promote apoptosis by activating proapoptotic Bcl-2
members. In either case, the BH3-only proteins are the likely
determinants as to whether a cell follows a pathway of survival
or death. In a healthy cell, the BH3-only proteins are either
absent or strongly inhibited, and the antiapoptotic Bcl-2 pro-
teins are able to restrain proapoptotic members. The mecha-
nism by which this occurs is debated. It is only in the face
of certain types of stress that the BH3-only proteins are ex-
pressed or activated, thereby shifting the balance in the direc-
tion of apoptosis. In these circumstances, the restraining
effects of the antiapoptotic Bcl-2 proteins are overridden, and
certain proapoptotic members of the Bcl-2 family, such as
Bax, are free to translocate from the cytosol to the outer mito-
chondrial membrane. Although the mechanism is not entirely
clear, it is thought that Bax (and/or Bak) molecules undergo
a change in conformation that causes them to insert into
the outer mitochondrial membrane and assemble into a multi-
subunit, protein-lined channel. Once formed, this channel
dramatically increases the permeability of the outer mito-
chondrial membrane and promotes the release of certain
mitochondrial proteins, most notably cytochrome *c* (Fig-
ure 15.38), which resides in the intermembrane space (see
Figure 5.17). Mitochondrial membrane permeabilization may
be accelerated by a rise in cytosolic Ca^{2+} levels following re-
lease of the ion from the ER. Nearly all of the cytochrome *c*

---

[3]The first member of the family, Bcl-2 itself, was originally identified in 1985 as
a cancer-causing oncogene in human lymphomas. The gene encoding Bcl-2 was
overexpressed in these malignant cells as the result of a translocation. We now
understand that Bcl-2 acts as an oncogene by promoting survival of potential
cancer cells that would otherwise die by apoptosis.

Internal cellular damage

Activation of proapoptotic Bcl-2
family members (e.g., Bad or Bax)

—Bax

—Cytochrome *c*

+ Cytoplasmic factors
(e.g. Apaf-1)

+ Procaspase-9

Activated initiator
caspase-9 complex
(Apoptosome)

Executioner procaspases

—Executioner caspase
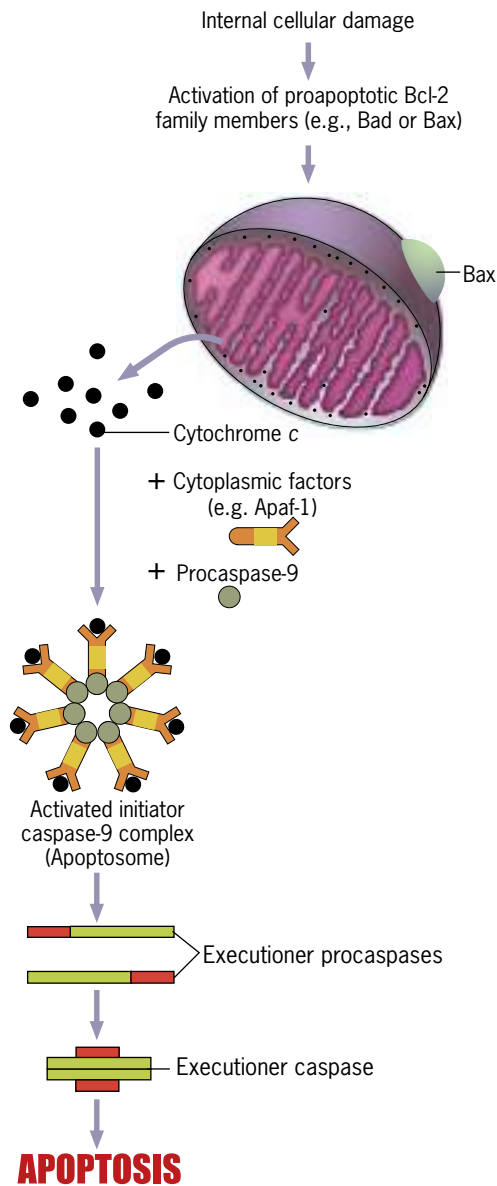
APOPTOSIS

FIGURE 15.37 The intrinsic (mitochondria-mediated) pathway of apoptosis. Various types of cellular stress cause proapoptotic members of the Bcl-2 family of proteins, such as Bax, to become inserted into the outer mitochondrial membrane. Insertion of these proteins leads to the release of cytochrome *c* molecules from the intermembrane space of the mitochondria. Release is thought to be mediated by pores in the mitochondrial membrane that are formed by Bax oligomers. Once in the cytosol, the cytochrome *c* molecules form a multisubunit complex with a cytosolic protein called Apaf-1 and procaspase-9 molecules. Procaspase-9 molecules are apparently activated to their full proteolytic capacity as the result of a conformational change induced by association with Apaf-1. Caspase-9 molecules cleave and activate executioner caspases, which carry out the apoptotic response.

(a)

(b)

FIGURE 15.38 Release of cytochrome *c* and nuclear fragmentation during apoptosis. Fluorescence micrographs of cultured mammalian cell before (*a*) and after (*b*) treatment with a cytotoxic protein-kinase inhibitor that activates the intrinsic apoptotic pathway. In the untreated cell, cytochrome *c* (green) is seen to be localized to the mitochondrial network and the nucleus (blue) is intact. Once apoptosis has been triggered, cytochrome *c* is released from the mitochondria and is present throughout the cell, while the nucleus breaks up into several fragments. (Courtesy of S. E. Wiley, USCD/Walther Cancer Institute.)

molecules present in all of a cell's mitochondria can be released from an apoptotic cell in a period as short as five minutes.

Release of proapoptotic mitochondrial proteins such as cytochrome *c* is apparently the "point of no return," that is, an event that irreversibly commits the cell to apoptosis. Once in the cytosol, cytochrome *c* forms part of a multiprotein

complex called the *apoptosome*, that also includes several molecules of procaspase-9. Procaspase-9 molecules are thought to become activated by simply joining the multiprotein complex and do not require proteolytic cleavage (Figure 15.37). Like caspase-8, which is activated by the receptor-mediated pathway described above, caspase-9 is an initiator caspase that activates downstream executioner caspases, which bring about apoptosis.[4] The extrinsic (receptor-mediated) and intrinsic (mitochondria-mediated) pathways ultimately converge by activating the same executioner caspases, which cleave the same cellular targets.

You might be wondering why cytochrome *c*, a component of the electron transport chain, and the mitochondrion, an organelle that functions as the cell's power plant, would be involved in initiating apoptosis. There is no obvious answer to this question at the present time. The key role of mitochondria in apoptosis is even more perplexing when one considers that these organelles have evolved from prokaryotic endosymbionts and that prokaryotes do not undergo apoptosis.

As cells execute the apoptotic program, they lose contact with their neighbors and start to shrink. Finally, the cell disin-

---

[4]Other intrinsic pathways that are independent of Apaf-1 and caspase-9, and possibly independent of cytochrome *c*, have also been described.
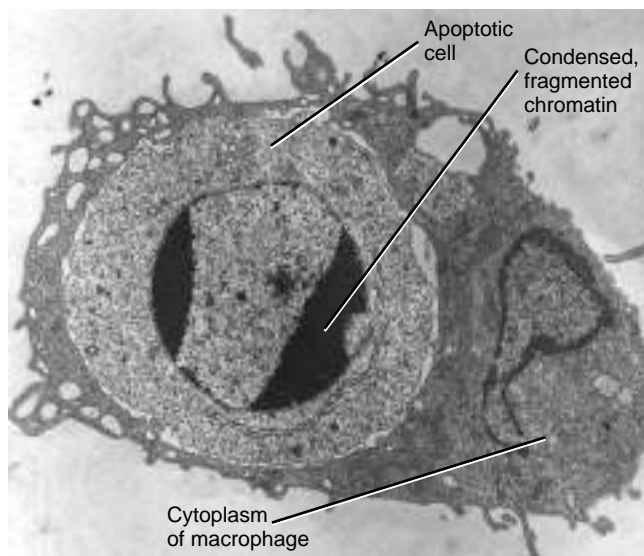
FIGURE 15.39 **Clearance of apoptotic cells is accomplished by phagocytosis.** This electron micrograph shows an apoptotic cell "corpse" within the cytoplasm of a phagocyte. Note the compact nature of the engulfed cell and the dense state of its chromatin. (Reprinted with permission from Peter M. Henson, Donna L. Bratton, and Valerie A. Fadok, Curr. Biol. 11:R796, 2001.)

tegrates into a condensed, membrane-enclosed apoptotic body. This entire apoptotic program can be executed in less than an hour. The apoptotic bodies are recognized by the presence of phosphatidylserine on their surface. Phosphatidylserine is a phospholipid that is normally present only on the inner leaflet of the plasma membrane. During apoptosis, a phospholipid "scramblase" moves phosphatidylserine

molecules to the outer leaflet of the plasma membrane where they are recognized as an "eat me" signal by specialized macrophages. Apoptotic cell death thus occurs without spilling cellular content into the extracellular environment (Figure 15.39). This is important because the release of cellular debris can trigger inflammation, which can cause a significant amount of tissue damage.

Just as there are signals that commit a cell to self-destruction, there are opposing signals that maintain cell survival. In fact, interaction of TNF with a TNF receptor often transmits two distinct and opposing signals into the cell interior: one stimulating apoptosis and another stimulating cell survival. As a result, most cells that possess TNF receptors do not undergo apoptosis when treated with TNF. This was a disappointing finding because it was originally hoped that TNF could be used as an agent to kill tumor cells. Cell survival is typically mediated through the activation of a key transcription factor called NF-κB, which activates the expression of genes encoding cell-survival proteins. It would appear that the fate of a cell—whether survival or death—depends on a delicate balance between proapoptotic and antiapoptotic signals.

REVIEW ?

1. What are some of the functions of apoptosis in vertebrate biology? Describe the steps that occur between (a) the time that a TNF molecule binds to its receptor and the eventual death of the cell and (b) the time a proapoptotic Bcl-2 member binds to the outer mitochondrial membrane and the death of the cell.

2. What is the role of the formation of caspase-containing complexes in the process of apoptosis?

SYNOPSIS

Cell signaling is a phenomenon in which information is relayed across the plasma membrane to the cell interior and often to the cell nucleus. Cell signaling typically includes recognition of the stimulus at the outer surface of the plasma membrane; transfer of the signal across the plasma membrane; and transmission of the signal to the cell interior, triggering a response. Responses may include a change in gene expression, an alteration of the activity of metabolic enzymes, a reconfiguration of the cytoskeleton, a change in ion permeability, the activation of DNA synthesis, or the death of the cell. This process is often called signal transduction. Within the cell, information passes along signaling pathways, which often include protein kinases and protein phosphatases that activate or inhibit their substrates through changes in conformation. Another prominent feature of signaling pathways is the involvement of GTP-binding proteins that serve as switches that turn a pathway on or off. (p. 606)

Many extracellular stimuli (first messengers) initiate responses by interacting with a G protein-coupled receptor (GPCR) on the outer cell surface and stimulating the release of a second messenger within the cell. Many extracellular messenger molecules act by binding to receptors that are integral membrane proteins containing seven membrane-spanning α helices (GPCRs). The signal is trans-

mitted from the receptor to the effector by a heterotrimeric G protein. These proteins are referred to as heterotrimeric because they have three subunits (α, β, and γ) and as G proteins because they bind guanine nucleotides, either GDP or GTP. Each G protein can exist in two states: an active state with a bound GTP or an inactive state with a bound GDP. Hundreds of different G protein-coupled receptors have been identified that respond to a wide variety of stimuli. All of these receptors act through a similar mechanism. The binding of the ligand to its specific receptor causes a change in the conformation of the receptor that increases its affinity for the G protein. As a result, the ligand-bound receptor binds to the G protein, causing the G protein to release its bound GDP and bind a GTP replacement, which switches the G protein into the active state. Exchange of guanine nucleotides changes the conformation of the G_α subunit, causing it to dissociate from the other two subunits, which remain together as a $G_{\beta\gamma}$ complex. Each dissociated G_α subunit with its attached GTP can activate specific effector molecules, such as adenylyl cyclase. The dissociated G_α subunit is also a GTPase and, with the help of an accessory protein, hydrolyzes the bound GTP to form bound GDP, which shuts off the subunit's ability to activate effector molecules. The G_α-GDP then reassociates with the $G_{\beta\gamma}$ subunits to

reform the trimeric complex and return the system to the resting state. Each of the three subunits that make up a heterotrimeric G protein can exist in different isoforms. Different combinations of specific subunits compose G proteins having different properties in their interactions with both receptors and effectors. (p. 609)

Phospholipase C is another important effector on the inner surface of the plasma membrane that can be activated by heterotrimeric G proteins. PI-phospholipase C splits phosphatidylinositol 4,5-bisphosphate (PIP_2) into two different second messengers, inositol 1,4,5-trisphosphate (IP_3) and 1,2-diacylglycerol (DAG). DAG remains in the plasma membrane where it activates the enzyme protein kinase C, which phosphorylates serine and threonine residues on a variety of target proteins. Constitutive activation of protein kinase C leads to loss of growth control. IP_3 is a small, water-soluble molecule that can diffuse into the cytoplasm where it binds to IP_3 receptors located on the surface of the smooth ER. IP_3 receptors are tetrameric calcium ion channels; binding of IP_3 leads to an opening of the ion channels and diffusion of Ca^{2+} into the cytosol. (p. 616)

Utilization of glucose is controlled by a signaling pathway that begins with an activated GPCR. The breakdown of glycogen into glucose is stimulated by the hormones epinephrine and glucagon, which act as first messengers by binding to their respective receptors on the outer surface of target cells. Binding of the hormones activates an effector, adenylyl cyclase, on the membrane's inner surface, leading to the production of the diffusible second messenger cyclic AMP (cAMP). Cyclic AMP evokes its response through a reaction cascade in which a series of enzymes are covalently modified. Cyclic AMP molecules bind to the regulatory subunits of a cAMP-dependent protein kinase called PKA, which phosphorylates phosphorylase kinase and glycogen synthase, leading to the activation of the former enzyme and inhibition of the latter one. The activated phosphorylase kinase molecules add phosphates to glycogen phosphorylase, activating the latter enzyme, leading to the breakdown of glycogen to glucose 1-phosphate, which is converted to glucose. As a result of this reaction cascade, the original message—as delivered by the binding of the hormone at the cell surface—is greatly amplified, and the response time is greatly reduced. Reaction cascades of this type also provide varied sites of regulation. The addition of phosphate groups by kinases is reversed by phosphatases that remove the phosphates. Cyclic AMP is produced in many different cells in response to a wide variety of first messengers. The course of events that occurs in the target cell depends on the specific proteins phosphorylated by the cAMP-dependent kinase. (p. 618)

Many extracellular stimuli initiate a cellular response by binding to the extracellular domain of a receptor protein-tyrosine kinase (RTK), which activates the tyrosine kinase domain located at the inner surface of the plasma membrane. RTKs regulate diverse functions, including cell growth and proliferation, the course of cell differentiation, the uptake of foreign particles, and cell survival. The best-studied growth-stimulating ligands, such as PDGF, EGF, and FGF, activate a signaling pathway called the MAP kinase cascade that includes a small monomeric GTP-binding protein called Ras. Like other G proteins, Ras cycles between an inactive GDP-bound form and an active GTP-bound form. In its active form, Ras stimulates effectors that lie downstream in the signaling pathway. Like other G proteins, Ras has a GTPase activity (stimulated by a GAP) that hydrolyzes the bound GTP to form bound GDP, thus switching itself off. When a ligand binds to the RTK, trans-autophosphorylation of the cytoplasmic domain of the receptor leads to the recruitment of Sos, an activator of Ras, to the inner surface of the membrane. Sos catalyzes the exchange of GDP

for GTP, thus activating Ras. Activated Ras has an increased affinity for another protein called Raf, which is recruited to the plasma membrane where it becomes an active protein kinase that initiates an orderly chain of phosphorylation reactions that are outlined in Figure 15.20. The ultimate targets of the MAP kinase cascade are transcription factors that stimulate the expression of genes whose products play a key role in activating the cell cycle, leading to the initiation of DNA synthesis and cell division. The MAP kinase cascade is found in all eukaryotes from yeast to mammals, though it has adapted through evolution to evoke different responses in different types of cells. (p. 623)

Insulin mediates many of its actions on target cells through interaction with the insulin receptor, which is an RTK. The activated kinase adds phosphate groups to tyrosine residues located on both the receptor and on receptor-associated docking proteins called IRSs. The phosphorylated tyrosine residues in an IRS serve as docking sites for proteins that have SH2 domains, which become activated upon binding the IRS. Several separate signaling pathways may become activated as the result of different signaling proteins binding to the phosphorylated IRS. One pathway may stimulate DNA synthesis and cell division, another may stimulate the movement of glucose transporters to the cell membrane, and yet another may activate transcription factors that turn on the expression of insulin-specific genes. (p. 631)

The rapid elevation of cytosolic Ca^{2+}, whether brought about by the opening of ion channels in cytoplasmic membranes or in the plasma membrane, triggers a wide variety of cellular responses. The concentration of Ca^{2+} ions within the cytosol is normally maintained at about 10^{-7} M by the action of Ca^{2+} pumps located in the plasma membrane and the SER membrane. Many different stimuli—ranging from a fertilizing sperm to a nerve impulse arriving at a muscle cell—cause a sudden rise in cytosolic $[Ca^{2+}]$, which may follow the opening of the plasma membrane Ca^{2+} channels, IP_3 receptors, or ryanodine receptors, which are a different type of calcium channel located in the SER membrane. Depending on the type of cell, ryanodine channels may be opened by an action potential arriving at the cell, or by the influx of a small amount of Ca^{2+} through the plasma membrane. Among the responses, elevated cytosolic $[Ca^{2+}]$ can lead to the activation or inhibition of various enzymes and transport systems, membrane fusion, or alterations in cytoskeletal or contractile functions. Calcium does not act on these various targets in the free ionic state, but instead it binds to one of a small number of calcium-binding proteins, which in turn elicit the response. The most widespread of these proteins is calmodulin, which contains four calcium-binding sites. The calcium ion is also an important intracellular messenger in plant cells, where it mediates responses to a variety of stimuli, including changes in light, pressure, gravity, and the concentration of plant hormones such as abscisic acid. (p. 634)

Different signaling pathways are often interconnected. As a result, signals from a variety of unrelated ligands can converge to activate a common effector, such as Ras; signals from the same ligand can diverge to activate a variety of different effectors; and signals can be passed back and forth between different pathways (cross-talk). (p. 638)

Nitric oxide (NO) acts as an intercellular messenger that diffuses directly through the plasma membrane of the target cell. Included among the activities stimulated by NO is the relaxation of the smooth muscle cells that line blood vessels. NO is produced by the enzyme nitric oxide synthase, which uses arginine as a substrate. NO often functions by activating guanylyl cyclase to produce the second messenger cGMP. (p. 640)

Signaling pathways may end in apoptosis—the programmed death of the cell. Examples of apoptosis include the death of excess nerve cells, the death of T lymphocytes that react with the body's own tissues, and the death of potential cancer cells. Death by apoptosis is characterized by the overall compaction of the cell and its nucleus and the orderly dissection of the chromatin by special endonucleases. Apoptosis is mediated by proteolytic enzymes called caspases that ac-

tivate or deactivate key protein substrates by removing a portion of their polypeptide chain. Two distinct pathways of apoptosis have been identified, one triggered by extracellular stimuli acting through death receptors, such as TNFR1, and the other triggered by internal cellular stress acting through the release of cytochrome c from the intermembrane space of mitochondria and the activation of proapoptotic members of the Bcl-2 protein family. (p. 642)

ANALYTIC QUESTIONS

1. The subject of cell signaling was placed near the end of the book because it ties together so many different topics in cell biology. Now that you have read the chapter, would you agree or disagree with this statement? Support your conclusions by example.

2. Suppose the signaling pathway in Figure 15.3 were to lead to the activation of a gene that inhibits a cyclin-dependent kinase responsible for moving a cell into S phase of the cell cycle. How would a debilitating mutation in protein kinase 3 affect the cell's growth?

3. What might be the effect on liver function of a mutation in a gene that encodes a cAMP phosphodiesterase? of a mutation in a gene encoding a glucagon receptor? of a mutation in a gene encoding phosphorylase kinase? of a mutation that altered the active site of the GTPase of a G_α subunit? (Assume in all cases that the mutation causes a loss of function of the gene product.)

4. Ca^{2+}, IP_3, and cAMP have all been described as second messengers. In what ways are their mechanisms of action similar? In what ways are they different?

5. In the reaction cascade illustrated in Figure 15.20, which steps lead to amplification and which do not?

6. Suppose that epinephrine and norepinephrine could initiate a similar response in a particular target cell. How could you determine whether or not the two compounds act by binding to the same cell-surface receptor?

7. One of the key experiments to show that gap junctions (page 256) allowed the passage of small molecules was carried out by allowing cardiac muscle cells (which respond to norepinephrine by contraction) to form gap junctions with ovarian granulosa cells (which respond to FSH by undergoing various metabolic changes). The researchers then added FSH to the mixed cell culture and observed the contraction of the muscle cells. How could muscle cells respond to FSH, and what does this tell you about the structure and function of gap junctions?

8. How would you expect a GTP analogue that the cell could not hydrolyze (a nonhydrolyzable analogue) to affect signaling events that take place during the stimulation of a liver cell by glucagon? What would be the effect of the same analogue on signal transduction of an epithelial cell after exposure to epidermal growth factor (EGF)? How would this compare to the effects of the cholera toxin (page 614) on these same cells?

9. You suspect that phosphatidylcholine might be serving as a precursor for a second messenger that triggers secretion of a hormone in a type of cultured endocrine cell that you are studying. Furthermore, you suspect that the second messenger released by the plasma membrane in response to a stimulus is choline phosphate. What type of experiment might you perform to verify your hypothesis?

10. Figure 15.25 shows the localized changes in $[Ca^{2+}]$ within the dendritic tree of a Purkinje cell. Calcium ions are small, rapidly

diffusible agents. How is it possible for a cell to maintain different concentrations of this free ion in different regions of its cytosol? What do you suspect would happen if you injected a small volume of a calcium chloride solution into one region of a cell that had been injected previously with a fluorescent calcium probe?

11. Formulate a hypothesis that might explain how contact of the outer surface of an egg by a fertilizing sperm at one site causes a wave of Ca^{2+} release that spreads through the entire egg, as shown in Figure 15.27.

12. Because calmodulin activates many different effectors (e.g., protein kinases, phosphodiesterases, calcium transport proteins), a calmodulin molecule must have many different binding sites on its surface. Would you agree with this statement? Why or why not?

13. Diabetes mellitus is a disease that can result from a number of different defects involving insulin function. Describe three different molecular abnormalities in a liver cell that could cause different patients to exhibit a similar clinical picture, including, for example, high concentrations of glucose in the blood and urine.

14. Would you expect a cell's response to EGF to be more sensitive to the fluidity of the plasma membrane than its response to insulin? Why or why not?

15. Would you expect a mutation in Ras to act dominantly or recessively as a cause of cancer? Why? (A dominant mutation causes its effect when only one of the homologous alleles is mutated, whereas a recessive mutation requires that both alleles of the gene are mutated.)

16. Speculate on a mechanism by which apoptosis might play a crucial role in combating the development of cancer, a topic discussed in the following chapter.

17. You are working with a type of fibroblast that normally responds to epidermal growth factor by increasing its rate of growth and division, and to epinephrine by lowering its rate of growth and division. You have determined that both of these responses require the MAP kinase pathway, and that EGF acts by means of an RTK and epinephrine by means of a G protein-coupled receptor. Suppose you identified a mutant strain of these cells that can still respond to EGF but is no longer inhibited by epinephrine. You suspect that the mutation is affecting the cross-talk between two pathways (shown in Figure 15.33). Which component in this figure might be affected by such a mutation?

18. In what way is the calcium wave that occurs at fertilization similar to a nerve impulse that travels down a neuron?

19. Now that you have read the section on taste perception, why do you suppose it has been difficult to find effective rat poisons?

20. One of the genes of the cowpox virus encodes a protein called CrmA that is a potent inhibitor of caspases. What effect would you expect this inhibitor to have on an infected cell? Why is this advantageous to the infecting virus?

21. Most RTKs act directly on downstream effectors, whereas the insulin RTK acts through an intermediate docking protein, an insulin receptor substrate (IRS). Are there any advantages in signaling that might accrue from the use of these IRS intermediates?

22. Researchers have reported that (1) most of the physiologic effects of insulin on target cells can be blocked by incubation of cells with wortmannin, a compound that specifically inhibits the enzyme PI3K and (2) that causing cells to overexpress a constitutively active form of PKB (i.e., a form of the enzyme that is continually active regardless of circumstances) induces a response in cells that is virtually identical to addition of insulin to these cells. Looking at Figure 15.23, are these observations ones that you might have predicted? Why or why not?

23. Knockout mice that are unable to produce caspase-9 die as a result of a number of defects, most notably a greatly enlarged brain. Why would these mice have this phenotype? How would you expect the phenotype of a cytochrome *c* knockout mice to compare with that of the caspase-9 knockout?

24. Why do you suppose that some people find a compound called PROP to have a bitter taste, whereas others do not report this perception?

16
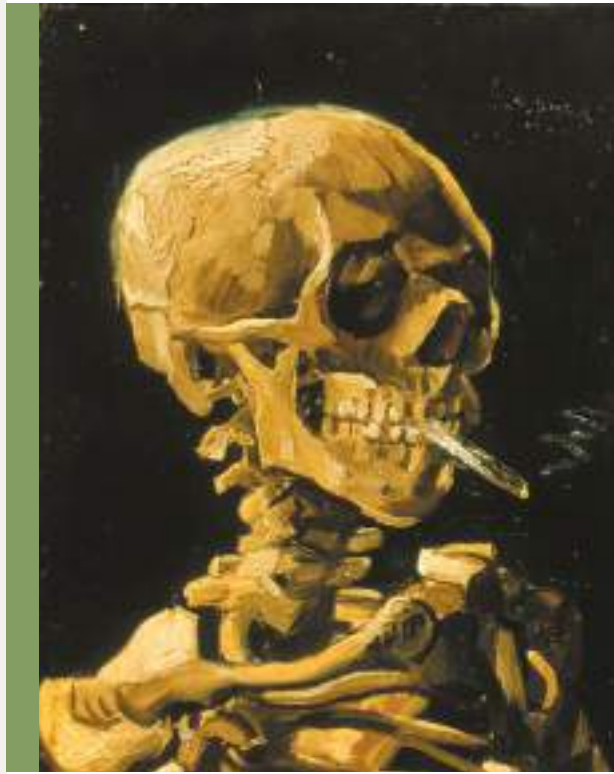
Cancer

Cancer is a genetic disease because it can be traced to alterations within specific genes, but in most cases, it is not an inherited disease. In an inherited disease, the genetic defect is present in the chromosomes of a parent and is transmitted to the zygote. In contrast, the genetic alterations that lead to most cancers arise in the DNA of a somatic cell during the lifetime of the affected individual. Because of these genetic changes, cancer cells proliferate uncontrollably, producing malignant tumors that invade surrounding healthy tissue (Figure 16.1). As long as the growth of the tumor remains localized, the disease can usually be treated and cured by surgical removal of the tumor. But malignant tumors tend to *metastasize*, that is, to spawn cells that break away from the parent mass, enter the lymphatic or vascular circulation, and spread to distant sites in the body where they establish lethal secondary tumors (*metastases*) that are no longer amenable to surgical removal. The subject of metastasis is discussed in the Human Perspective of Chapter 7 on page 248.

Because of its impact on human health and the hope that a cure might be developed, cancer has been the focus of a massive research effort for decades. Though these studies have led to a remarkable breakthrough in our understanding of the cellular and molecular basis of cancer, they have had very little impact on either preventing the occurrence of or increasing the chances of surviving most cancers. The incidence of various types of cancer in the United States and the corresponding mortality rates are shown in Figure 16.2. Current treatments, such as chemotherapy and radiation, lack the specificity needed to kill cancer cells without simultaneously damaging normal cells, as evidenced by the serious side effects that accompany these treatments. As a result, patients cannot usually be subjected to high enough doses

Skull with Cigarette. (BY VINCENT VAN GOGH, 1885, VAN GOGH MUSEUM, AMSTERDAM/© ART RESOURCE, NY.)
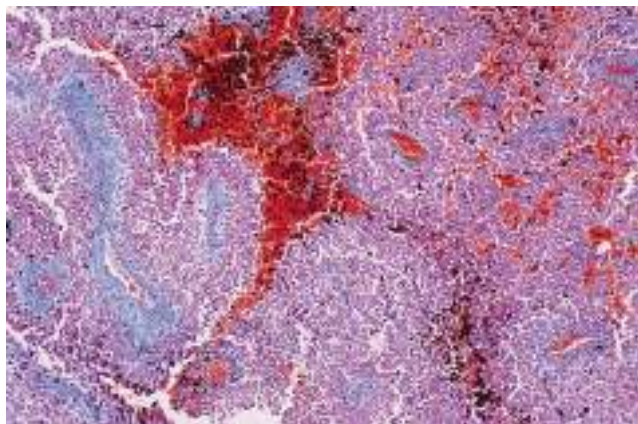
FIGURE 16.1 The invasion of normal tissue by a growing tumor. This light micrograph of a section of human liver shows a metastasized melanosarcoma (in red) that is invading the normal liver tissue. (Micro-graph by Astrid and Hanns-Frieder Michler/Science Photo Library/Photo Researchers, Inc.)

of chemicals or radiation to kill all of the tumor cells in their body. Researchers have been working for many years to develop more effective and less debilitating treatments. Some of these newer strategies in cancer therapy will be discussed at the end of the chapter. ∎

16.1 BASIC PROPERTIES OF A CANCER CELL

The behavior of cancer cells is most easily studied when the cells are growing in culture. Cancer cells can be obtained by removing a malignant tumor, dissociating the tissue into its separate cells, and culturing the cells in vitro. Over the

years, many different lines of cultured cells that were originally derived from human tumors have been collected in cell banks and are available for study. Alternatively, normal cells can be converted to cancer cells by treatment with carcinogenic chemicals, radiation, or tumor viruses. Cells that have been transformed in vitro by chemicals or viruses can generally cause tumors when introduced into a host animal. There are many differences in properties from one type of cancer cell to another. At the same time, there are a number of basic prop-erties that are shared by cancer cells, regardless of their tissue of origin.

At the cellular level, the most important characteristic of a cancer cell—whether residing in the body or on a culture dish—is its loss of growth control. The *capacity* for growth and division is not drastically different between a cancer cell and most normal cells. When normal cells are grown in tissue culture under conditions that promote cell proliferation, they grow and divide at a rate similar to that of their malignant counterparts. However, when the normal cells proliferate to the point where they cover the bottom of the culture dish, their growth rate decreases markedly, and they tend to remain as a single layer (*monolayer*) of cells (Figure 16.3*a*,*b*). Growth rates drop as normal cells respond to inhibitory influences from their environment. Growth-inhibiting influences may arise as the result of depletion of growth factors in the culture medium or from contact with surrounding cells on the dish. In contrast, when malignant cells are cultured under the same conditions, they continue to grow, piling on top of one an-other to form clumps (Figure 16.3*c*,*d*). It is evident that ma-lignant cells are not responsive to the types of signals that cause their normal counterparts to cease growth and division.

Not only do cancer cells ignore inhibitory growth signals, they continue to grow in the absence of stimulatory growth signals that are required by normal cells. Normal cells growing in culture depend on growth factors, such as epidermal growth factor and insulin, that are present in serum (the fluid fraction
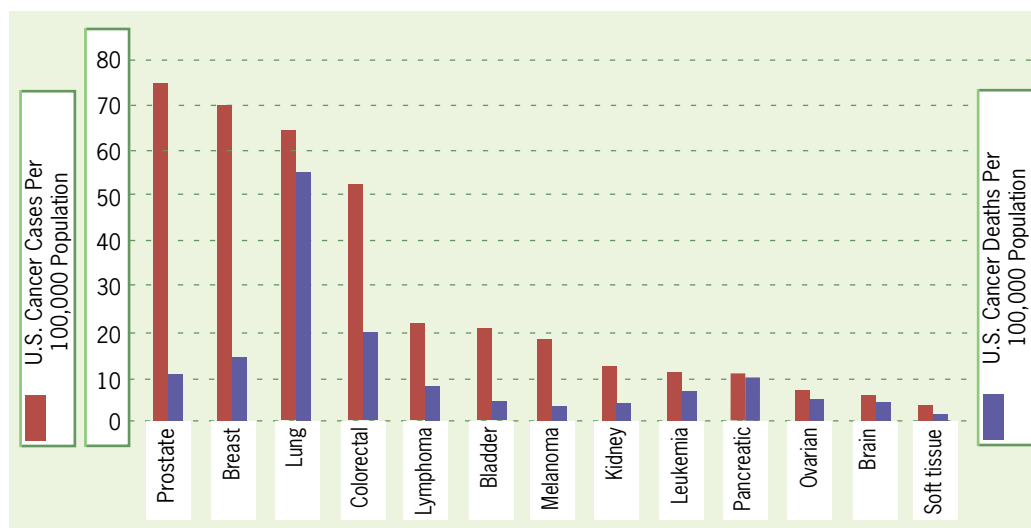
FIGURE 16.2 The incidence of new cancer cases and deaths in the United States (2000–2003).

Normal cells

Normal cells grow in monolayer
(a)

(b)

Cancer cells

Cancer cells grow in clumps (foci)
(c)

(d)
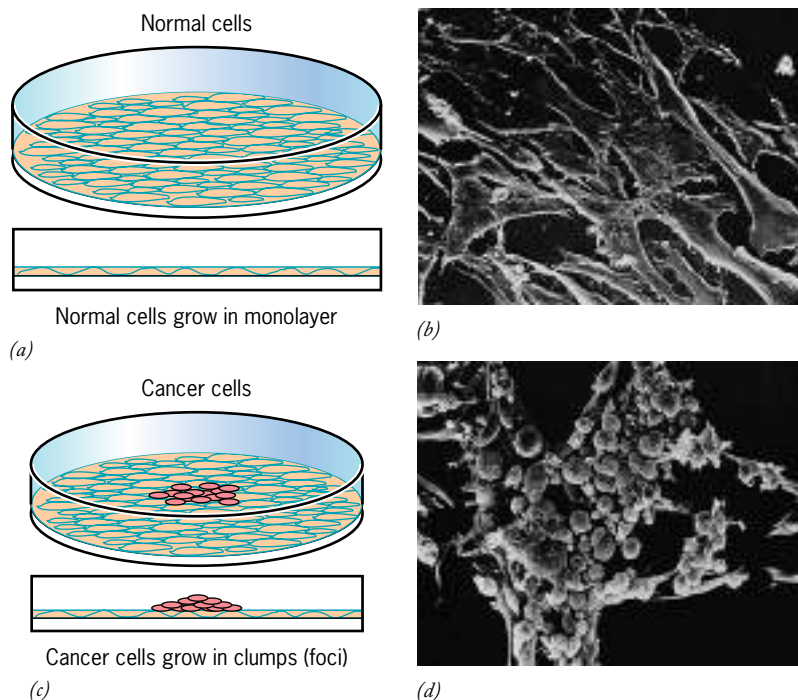
FIGURE 16.3 Growth properties of normal and cancerous cells. Normal cells typically grow in a culture dish until they cover the surface as a monolayer (*a* and *b*). In contrast, cells that have been transformed by viruses or carcinogenic chemicals (or malignant cells that have been cultured from tumors) typically grow in multilayered clumps, or foci (*c* and *d*). (B AND D: COURTESY OF G. STEVEN MARTIN.)

of blood), which is usually added to the growth medium (Figure 16.4). Cancer cells can proliferate in the absence of serum because their cell cycle does not depend on the interaction between growth factors and their receptors, which are located at the cell surface (page 623). As we will see below, this transformation is a result of basic changes in the intracellular pathways that govern cell proliferation and survival.

Normal cells growing in culture exhibit a limited capacity for cell division; after a finite number of mitotic divisions, they undergo an aging process that renders them unfit to continue to grow and divide (page 495). Cancer cells, on the other hand, are seemingly immortal because they continue to divide indefinitely. This difference in growth potential is often attributed to the presence of telomerase in cancer cells and its absence in nor-

mal cells. Recall from page 493 that telomerase is the enzyme that maintains the telomeres at the ends of the chromosomes, thus allowing cells to continue to divide. The absence of telomerase from most types of normal cells is thought to be one of the body's major defenses that protects against tumor growth.

The most striking alterations in the nucleus following transformation occur within the chromosomes. Normal cells maintain their diploid chromosomal complement as they grow and divide, both in vivo and in vitro. In contrast, cancer cells are genetically unstable and often have highly aberrant chromosome complements, a condition termed *aneuploidy* (Figure 16.5), which may occur primarily as a result of defects in the mitotic checkpoint (page 584) or the presence of an abnormal number of centrosomes (see Figure 14.17c).[1] It is evident from Figure 16.5 that the growth of cancer cells is much less dependent on a standard diploid chromosome content than the growth of normal cells. In fact, when the chromosome content of a normal cell becomes disturbed, a signaling pathway is usually activated that leads to the self-destruction (apoptosis) of the cell. In contrast, cancer cells typically fail to elicit the apoptotic response even when their chromosome content becomes highly deranged. Protection from apoptosis is another important hallmark that distinguishes many cancer cells from normal cells. Finally, it can be noted that cancer cells often depend on glycolysis, which is an anaerobic metabolic pathway (Figure 3.24). This property may reflect the high metabolic requirements of cancer cells and an inadequate blood supply within the tumor. Under conditions of hypoxia (reduced O_2), cancer cells activate a transcription factor called

FIGURE 16.4 The effects of serum deprivation on the growth of normal and transformed cells. Whereas the growth of cancer cells continues regardless of the presence or absence of exogenous growth factors, normal cells require these substances in their medium for growth to continue. The growth of normal cells levels off as the growth factors in the medium are depleted.

[1]There is controversy as to whether the development of aneuploidy occurs at an early stage in tumor formation and is a cause of the genetic instability that characterizes cancer cells, or is a late event and is simply a consequence of abnormal cancer growth.

FIGURE 16.5 Karyotype of a cell from a breast cancer line showing a highly abnormal chromosome complement. A normal diploid cell would have 22 pairs of autosomes and two sex chromosomes. The two members of a pair would be identical, and each chromosome would be a single continuous color (as in the karyotype of a normal cell in Figure 12.18*b* that uses a similar spectral visualization technique). The chromosomes of this cell are highly deranged as evidenced by the presence of extra and missing chromosomes and chromosomes of more than one color. These multicolored chromosomes reflect the large numbers of translocations that have occurred in previous cell generations. A cell with normal cell cycle checkpoints and apoptotic pathways could never have attained a chromosome complement approaching that seen here. (COURTESY OF J. DAVIDSON AND PAUL A. W. EDWARDS.)

HIF that induces the formation of new blood vessels and promotes the migratory properties of the cells, which may contribute to the spread of the tumor. However, even when oxygen is plentiful, tumor cells continue to generate much of their ATP by glycolysis. The end product of glycolysis is lactic acid, which is secreted into the tumor's microenvironment, where it may promote tumor growth.

It is these properties, which can be demonstrated in culture, together with their tendency to spread to distant sites within the body, that make cancer cells such a threat to the well-being of the entire organism.

REVIEW ?

1. Describe some of the properties that distinguish cancer cells from normal cells.
2. How do the properties of cancer cells manifest themselves in culture?

16.2 THE CAUSES OF CANCER

In 1775, Percivall Pott, a British surgeon, made the first known correlation between an environmental agent and the development of cancer. Pott concluded that the high incidence of cancer of the nasal cavity and the skin of the scrotum in chimney sweeps was due to their chronic exposure to soot. Within the past several decades, the carcinogenic chemicals in soot have been isolated, along with hundreds of other compounds shown to cause cancer in laboratory animals. In addition to a diverse array of chemicals, a number of other types of agents are also carcinogenic, including ionizing radiation and a variety of DNA- and RNA-containing viruses. All of these agents have one property in common: they alter the genome. Carcinogenic chemicals, such as those present in soot or cigarette smoke, can almost always be shown either to be directly mutagenic or to be converted to mutagenic compounds by cellular enzymes. Similarly, ultraviolet radiation, which is the leading cause of skin cancer, is also strongly mutagenic.

A number of viruses can infect mammalian cells growing in cell culture, transforming them into cancer cells. These viruses are broadly divided into two large groups: **DNA tumor viruses** and **RNA tumor viruses**, depending on the type of nucleic acid found within the mature virus particle. Among the DNA viruses capable of transforming cells are polyoma virus, simian virus 40 (SV40), adenovirus, and herpes-like viruses. RNA tumor viruses, or retroviruses, are similar in structure to HIV (see Figure 1.21*b*) and are the subject of the Experimental Pathways, which can be found at the end of the chapter. Tumor viruses can transform cells because they carry genes whose products interfere with the cell's normal growth-regulating activities. Although tumor viruses were an invaluable tool for researchers in identifying numerous genes

involved in cell transformation, they are associated with only a small number of human cancers. Other types of viruses are, however, linked to as many as 20 percent of cancers worldwide. In most cases, these viruses greatly increase a person's risk of developing the cancer, rather than being the sole determinant responsible for the disease. This relationship between viral infection and cancer is illustrated by human papilloma virus (HPV), which can be transmitted through sexual activity and is increasing in frequency in the population. Although the virus is present in about 90 percent of cervical cancers, indicating its importance in development of the disease, the vast majority of women who have been infected with the virus will never develop this malignancy. HPV is also linked as a causative agent to cancers of the mouth and tongue in both men and women. An effective vaccine against this virus is now available. Other viruses linked to human cancers include hepatitis B virus, which is associated with liver cancer; Epstein-Barr virus, which is associated with Burkitt's lymphoma in areas where malaria is common; and a herpes virus (HHV-8), which is associated with Kaposi's sarcoma.

Certain gastric lymphomas are associated with chronic infection by the stomach-dwelling bacterium *Helicobacter pylori*, which is also responsible for ulcers. Recent evidence suggests that many of these cancers linked to infections are actually caused by the chronic inflammation that is triggered by the presence of the pathogen. Inflammatory bowel disease (IBD), which is also characterized by chronic inflammation, has been associated with an increased risk of colon cancer. These findings have caused researchers to look more closely at the general process of inflammation as a previously unexplored factor in the development of many types of cancers.

Determining the causes of different types of cancer is an endeavor carried out by *epidemiologists*, researchers who study disease patterns in populations. The causes of certain cancers are obvious: smoking causes lung cancer, exposure to ultraviolet radiation causes skin cancer, and inhaling asbestos fibers causes mesothelioma. But despite a large number of studies, we are still uncertain as to the causes of most types of human cancer. Humans live in complex environments and are exposed to many potential carcinogens in a changing pattern over a period of decades. Attempting to determine the causes of cancer from a mountain of statistical data obtained from the answers to questionnaires about individual lifestyles has proven very difficult. The importance of environmental factors (e.g., diet) is seen most clearly in studies of the children of couples that have moved from Asia to the United States or Europe. These individuals no longer exhibit a high rate of gastric cancer, as occurs in Asia, but instead are subject to an elevated risk of colon and breast cancer, which is characteristic of Western countries (Figure 16.6).

There is a general consensus among epidemiologists that some ingredients in the diet, such as animal fat and alcohol, can increase the risk of developing cancer, whereas certain compounds found in fruits, vegetables, and tea can reduce that risk. Several widely prescribed drugs appear also to have a preventive effect. Long-term use of nonsteroidal anti-inflammatory drugs (NSAIDs) such as aspirin and indomethacin has been shown to
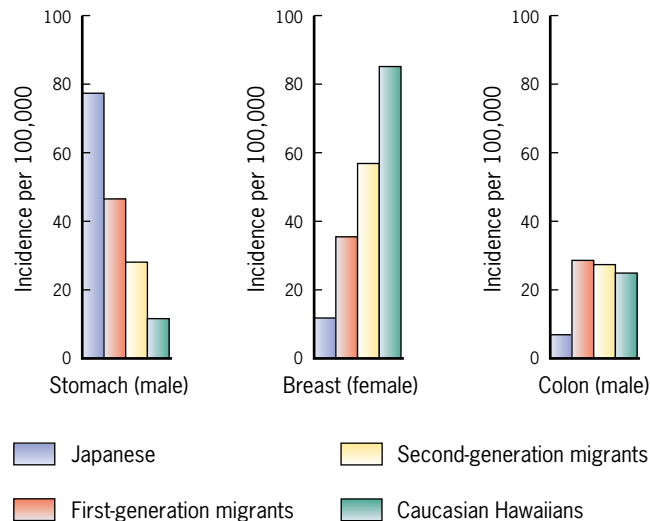
FIGURE 16.6 Changing cancer incidence in persons of Japanese descent following migration to Hawaii. The incidence of stomach cancer declines, whereas that of breast and colon cancer rises. However, of the three types of cancer, only colon cancer has reached rates equivalent to Caucasian Hawaiians by the second generation. (FROM L. N. KOLONEL ET AL, REPRINTED WITH PERMISSION FROM NATURE REVS. CANCER 4:3, 2004; © COPYRIGHT 2004, MACMILLAN MAGAZINES LTD.)

markedly decrease the risk of colon cancer. They are thought to have this effect by inhibiting cyclooxygenase-2, an enzyme that catalyzes the synthesis of hormone-like prostaglandins, which promote the growth of intestinal polyps. The cancer-suppressing action of NSAIDs supports the idea that inflammation plays a major role in the development of various cancers.

16.3 THE GENETICS OF CANCER

Cancer is one of the two leading causes of death in Western countries, afflicting approximately one in every three individuals. Viewed in this way, cancer is a very common disease. But at the cellular level, the development of a cancer is a remarkably rare event. Whenever the cells of a cancerous tumor are genetically scrutinized, they are invariably found to have arisen from a single cell. Thus, unlike other diseases that require modification of a large number of cells, cancer results from the uncontrolled proliferation of a single wayward cell (cancer is said to be *monoclonal*). Consider for a moment that the human body contains trillions of cells, billions of which undergo cell division on any given day. Though almost any one of these dividing cells may have the potential to change in genetic composition and grow into a malignant tumor, this only occurs in about one-third of the human population during an entire lifetime.

One of the primary reasons why a greater number of cells do not give rise to cancerous tumors is that malignant transformation requires more than a single genetic alteration. We can distinguish between two types of genetic alterations that might make us more likely to develop a particular type of

cancer—those that we inherit from our parents (germ-line mutations) and those that occur during our own lifetime (somatic mutations). There are a few types of mutations that we can inherit that make us much more likely to develop cancer. The study of these mutations has taught us a great deal about how malfunctioning genes can lead to the development of cancer; some of these inherited cancer syndromes will be discussed later in this section. However, for the most part, inherited mutations are not a major factor in the occurrence of most cases of the disease. One way to determine an overall estimate of the impact of inheritance in tumor formation is to ascertain the likelihood that two identical twins will develop the same type of cancer by the time the individuals reach a certain age. Studies of this type suggest that the likelihood two 75-year-old identical twins will share a particular cancer, such as breast cancer or prostate cancer, is generally between 10 and 15 percent, depending on the type of cancer. Clearly, the genes that we inherit have a significant influence on our risks of developing cancer, but the greatest impact comes from genes that are altered during our lifetime.

The development of a malignant tumor (*tumorigenesis*) is a multistep process characterized by a progression of permanent genetic alterations in a single line of cells, which may occur over the course of many successive cell divisions and take years to complete. Each genetic change may elicit a particular feature of the malignant state, such as protection from apoptosis, as discussed in Section 16.1. As these genetic changes gradually occur, the cells in the line become increasingly less responsive to the body's normal regulatory machinery and better able to invade normal tissues. According to this concept, tumorigenesis requires that the cell responsible for initiating the cancer be capable of a large number of cell divisions. This requirement has focused a great deal of attention on the types of cells that are present in a tissue that might have the potential to develop into a tumor.

The most common solid tumors—such as those of the breast, colon, prostate, and lung—arise in epithelial tissues that are normally engaged in a relatively high level of cell division. The same is true of leukemias, which develop in rapidly dividing blood-forming tissues. The cells of these tissues can be roughly divided into three groups: (1) stem cells, which possess unlimited proliferation potential, have the capacity to produce more of themselves, and can give rise to all of the cells of the tissue (page 19); (2) progenitor cells, which are derived from stem cells and possess a limited ability to proliferate; and (3) the differentiated end products of the tissue, which generally lack the capability to divide. Examples of these three groups of cells are illustrated in Figure 17.6.

Given the fact that tumor formation requires that a cell be capable of extensive division, two general scenarios have been considered for the origin of tumors. According to one scenario, cancer arises from within the relatively small population of stem cells that inhabit each adult tissue. Given their long life and unlimited division potential, stem cells have the opportunity to accumulate the mutations required for malignant transformation. A number of studies suggest that stem cells are the source of a variety of different types of tumors. According to

another scenario, the committed progenitor cells can give rise to malignant tumors by acquiring certain properties, such as the capacity for unlimited proliferation, as part of the process of tumor progression. These two scenarios are not mutually exclusive, in that some tumors may well arise from stem cells and others from the committed progenitor cell population.

As a cancer grows, the cells in the tumor mass are subjected to a type of natural selection that drives the accumulation of cells with properties that are most favorable for tumor growth. For example, only those tumors containing cells that maintain the length of their telomeres will be capable of unlimited growth (page 495). Any cell that appears within a tumor that happens to express telomerase will have a tremendous growth advantage over other cells that fail to express the enzyme. Over time, the telomere-expressing cells will flourish while the nonexpressing cells will die off, and all of the cells in the tumor will contain telomerase. Expression of telomerase illustrates another important feature of tumor progression; not all of these changes result from genetic mutation. The activation of telomerase expression can be considered an epigenetic change, one that results from the activation of a gene that is normally repressed. As discussed in Chapter 12, this type of activation process likely involves a change in the structure of chromatin in and around the gene and/or a change in the state of DNA methylation. Once the epigenetic change has occurred, it is transmitted to all of the progeny of that cell and, consequently, represents a permanent, inheritable alteration. Even after they have become malignant, cancer cells continue to accumulate mutations and epigenetic changes that make them increasingly abnormal (as is evident in Figure 16.5). This genetic instability makes the disease difficult to treat by conventional chemotherapy because cells often arise within the tumor mass that are resistant to the drug.

The genetic changes that occur during tumor progression are often accompanied by histological changes, that is, changes in the appearance of the cells. The initial changes often produce cells that can be identified as "precancerous," which indicates that they have gained some of the properties of a cancer cell, such as loss of certain growth controls, but lack the capability to invade normal tissues or metastasize to distant sites. The *Pap smear* is a test for detecting precancerous cells in the epithelial lining of the cervix. The development of cervical cancer typically progresses over a period of more than 10 years and is characterized by cells that appear increasingly abnormal (less well differentiated than normal cells, with larger nuclei, as in Figure 16.7). When cells having an abnormal appearance are detected, the precancerous lesion in the cervix can be located and destroyed by laser treatment, freezing, or surgery. Some tissues often generate *benign* tumors, which contain cells that have proliferated to form a mass that poses little threat of becoming malignant. The moles that we all possess are an example of benign tumors. Recent studies indicate that the pigment cells that compose a mole have undergone a response that causes them to enter a permanent state of growth arrest, referred to as *senescence*. Senescence is apparently triggered in these pigment cells after they have undergone certain of the genetic changes that would have other-

(a)

(b)
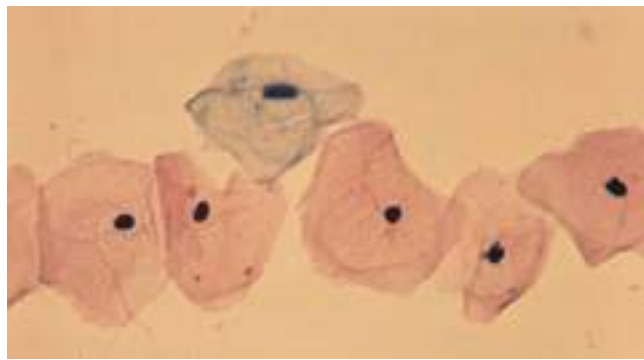
FIGURE 16.7 Detection of abnormal (premalignant) cells in a Pap smear. (*a*) Normal squamous epithelial cells of the cervix. The cells have a uniform shape with a small centrally located nucleus. (*b*) Abnormal cells from a case of carcinoma in situ, which is a preinvasive cancer of the cervix. The cells have heterogeneous shapes and large nuclei. (MICROGRAPHS FROM © Vu/Cabisco/Visuals Unlimited.)
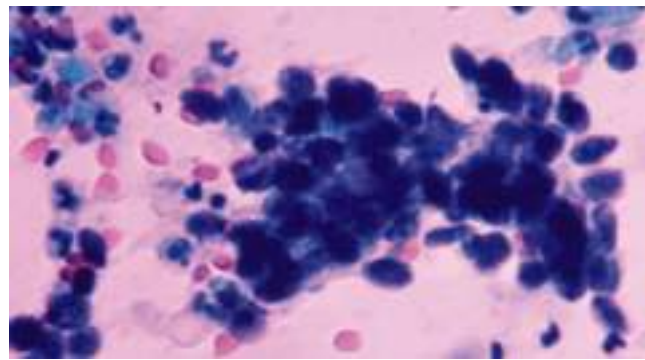
wise set them on a course to becoming a malignant cancer. This concept of "forced senescence" may represent another pathway that has evolved to restrict the development of cancers in higher organisms. The molecular basis of senescence is discussed further on page 663.

Tumor-Suppressor Genes and Oncogenes: Brakes and Accelerators

The genes that have been implicated in carcinogenesis are divided into two broad categories: tumor-suppressor genes and oncogenes. **Tumor-suppressor genes** act as a cell's

brakes; they encode proteins that restrain cell growth and prevent cells from becoming malignant (Figure 16.8*a*). The existence of such genes originally came to light from studies in the late 1960s in which normal and malignant rodent cells were fused to one another. Some of the cell hybrids formed from this type of fusion lost their malignant characteristics, suggesting that a normal cell possesses factors that can suppress the uncontrolled growth of a cancer cell. Further evidence for the existence of tumor-suppressor genes was gathered from observations that specific regions of particular chromosomes are consistently deleted in cells of certain types of cancer. If the absence of such genes is correlated with the development
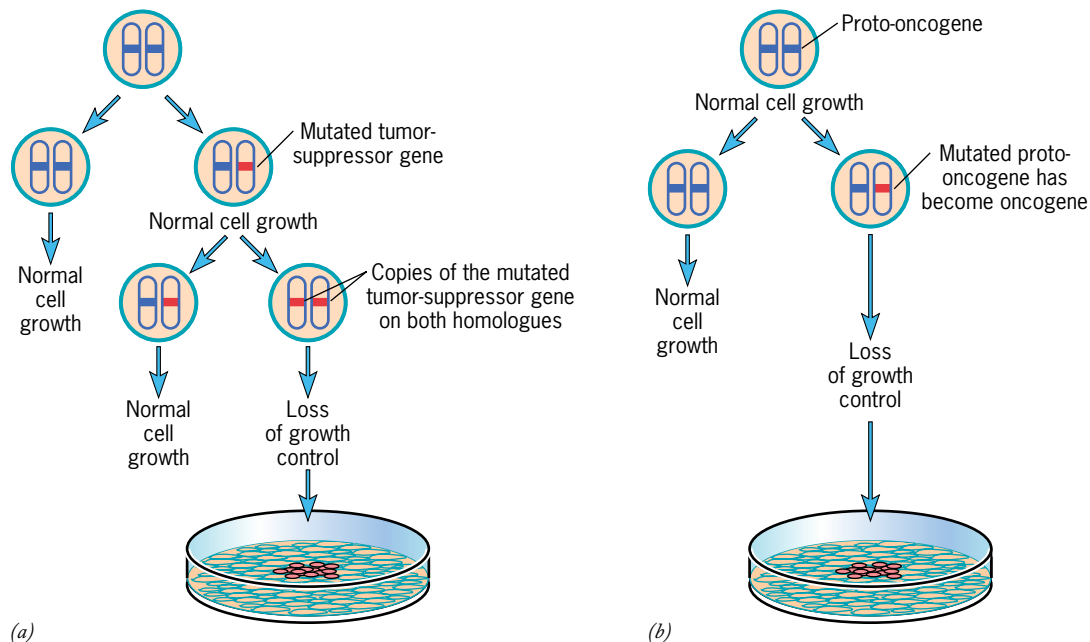
(a)

(b)

FIGURE 16.8 Contrasting effects of mutations in tumor-suppressor genes (*a*) and oncogenes (*b*). Whereas a mutation in one of the two copies (alleles) of an oncogene may be sufficient to cause a cell to lose growth control, both copies of a tumor-suppressor gene must be knocked out to induce the same effect. As discussed shortly,

oncogenes arise from proto-oncogenes as the result of gain-of-function mutations, that is, mutations that cause the gene product to exhibit new functions that lead to malignancy. Tumor-suppressor genes, in contrast, suffer loss-of-function mutations and/or epigenetic inactivation that render them unable to restrain cell growth.

of a tumor, then it follows that the presence of these genes normally suppresses the formation of the tumor.

Oncogenes, on the other hand, encode proteins that promote the loss of growth control and the conversion of a cell to a malignant state (Figure 16.8*b*). Most oncogenes act as accelerators of cell proliferation, but they have other roles as well. Oncogenes may lead to genetic instability, prevent a cell from becoming a victim of apoptosis, or promote metastasis. The existence of oncogenes was discovered through a series of investigations on RNA tumor viruses that is documented in the Experimental Pathways at the end of this chapter. These viruses transform a normal cell into a malignant cell because they carry an oncogene that encodes a protein that interferes with the cell's normal activities. The turning point in these studies came in 1976, when it was discovered that an oncogene called *src*, carried by an RNA tumor virus called avian sarcoma virus, was actually present in the genome of uninfected cells. The oncogene, in fact, was not a viral gene, but a cellular gene that had become incorporated into the viral genome during a previous infection. It soon became evident that cells possess a variety of genes, now referred to as **proto-oncogenes**, that have the potential to subvert the cell's own activities and push the cell toward the malignant state.

As discussed below, proto-oncogenes encode proteins that have various functions in a cell's *normal* activities. Proto-oncogenes can be converted into oncogenes (i.e., *activated*) by several mechanisms (Figure 16.9):

1. The gene can be mutated in a way that alters the properties of the gene product so that it no longer functions normally (Figure 16.9, path a).

2. The gene can become duplicated one or more times, resulting in gene amplification and excess production of the encoded protein (Figure 16.9, path b).

3. A chromosome rearrangement can occur that brings a DNA sequence from a distant site in the genome into close proximity of the gene, which can either alter the expression of the gene or the nature of the gene product (Figure 16.9, path c).

Any of these genetic alterations can cause a cell to become less responsive to normal growth controls, causing it to behave as a malignant cell. Oncogenes act *dominantly*, which is to say that a single copy of an oncogene can cause the cell to express the altered phenotype, regardless of whether or not there is a normal, unactivated copy of the gene on the homologous chromosome (Figure 16.8*b*). Researchers have taken advantage of this property to identify oncogenes by introducing the DNA suspected of containing the gene into cultured cells and monitoring the cells for evidence of altered growth properties.

We saw earlier that the development of a human malignancy requires more than a single genetic alteration. The reason becomes more apparent with the understanding that there are two types of genes responsible for tumor formation. As long as a cell has its full complement of tumor-suppressor
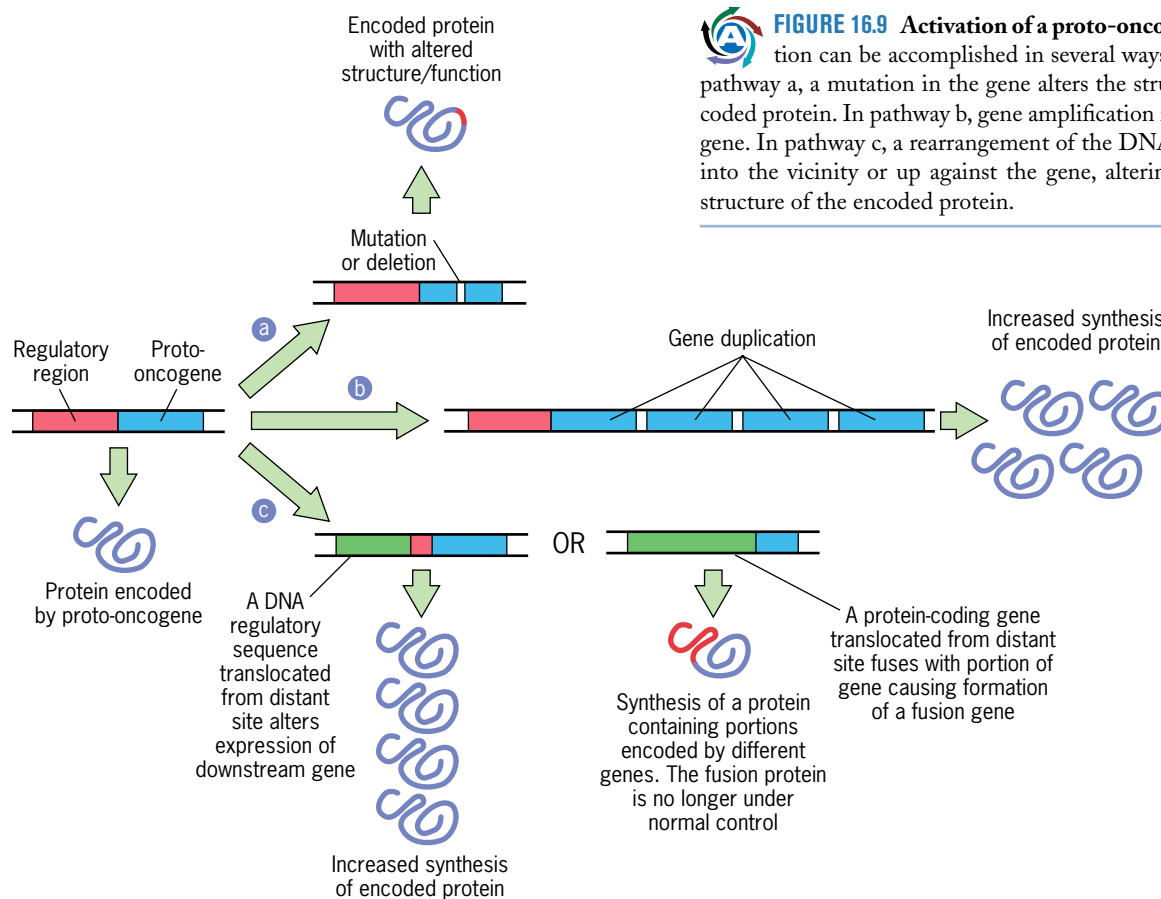
FIGURE 16.9 Activation of a proto-oncogene to an oncogene. Activation can be accomplished in several ways as indicated in this figure. In pathway a, a mutation in the gene alters the structure and function of the encoded protein. In pathway b, gene amplification results in overexpression of the gene. In pathway c, a rearrangement of the DNA brings a new DNA segment into the vicinity or up against the gene, altering either its expression or the structure of the encoded protein.

TABLE 16.1 Tumor-Suppressor Genes

Gene	Primary tumor	Proposed function	Inherited syndrome
APC	Colorectal	Binds β-catenin acting as transcription factor	Familial adenomatous polyposis
ARF	Melanoma	p53 activation (MDM2 antagonist)	Familial melanoma
BRCA1	Breast	DNA repair	Familial breast cancer
MSH2, MLH1	Colorectal	Mismatch repair	HNPCC
E-Cadherin	Breast, colon, etc.	Cell adhesion molecule	Familial gastric cancer
INK4a	Melanoma, pancreatic	p16: Cdk inhibitor $p14^{ARF}$: stabilizes p53	Familial melanoma
NF1	Neurofibromas	Activates GTPase of Ras	Neurofibromatosis type 1
NF2	Meningiomas	Links membrane to cytoskeleton	Neurofibromatosis type 2
TP53	Sarcomas, lymphomas, etc.	Transcription factor (cell cycle and apoptosis)	Li-Fraumeni syndrome
PTEN	Breast, thyroid	PIP_3 phosphatase	Cowden disease
RB	Retinal	Binds E2F (cell cycle transcription regulation)	Retinoblastoma
VHL	Kidney	Protein ubiquitination and degradation	von Hippel-Lindau syndrome
WT1	Wilms tumor of kidney	Transcription factor	Wilms tumor

genes, it is thought to be protected against the effects of an oncogene for reasons that will be evident when the functions of these genes are discussed below. Most tumors contain alterations in both tumor-suppressor genes and oncogenes, suggesting that the loss of a tumor-suppressor function within a cell must be accompanied by the conversion of a proto-oncogene into an oncogene before the cell becomes fully malignant. Even then, the cell may not exhibit all of the properties required to invade surrounding tissues or to form secondary colonies by metastasis. Mutations in additional genes, such as those encoding cell-adhesion molecules or extracellular proteases (discussed on page 248), may be required before these cells acquire the full life-threatening phenotype.

We can now turn to the functions of the products encoded by both tumor-suppressor genes and oncogenes and examine how mutations in these genes can cause a cell to become malignant.

Tumor-Suppressor Genes The transformation of a normal cell to a cancer cell is accompanied by the loss of function of one or more tumor-suppressor genes. At the present time, more than two dozen genes have been implicated as tumor suppressors in humans, some of which are listed in Table 16.1. Included among the genes on the list are those that encode transcription factors (e.g., TP53 and WT1), cell cycle regulators (e.g., RB and p16), components that regulate G proteins, (NF1), a phosphoinositide phosphatase (PTEN), and a protein that regulates protein degradation (VHL).[2] In one way or another, most of the proteins encoded by tumor-suppressor genes act as negative regulators of cell proliferation, which is why their elimination promotes uncontrolled cell growth. The products of tumor-suppressor genes also help maintain genetic stability, which may be a primary reason that tumors contain such an aberrant karyotype (Figure 16.5). Some tumor-suppressor genes are involved in the development of a wide variety of different cancers, whereas others play a role in the formation of one or a few cancer types.

It is common knowledge that members of some families are at high risk of developing certain types of cancers. Although these inherited cancer syndromes are rare, they provide an unprecedented opportunity to identify tumor-suppressor genes that, when missing, contribute to the development of both inherited and sporadic (i.e., noninherited) forms of cancer. The first tumor-suppressor gene to be studied and eventually cloned—and one of the most important—is associated with a rare childhood cancer of the retina of the eye, called *retinoblastoma*. The gene responsible for this disorder is named *RB*. The incidence of retinoblastoma follows two distinct patterns: (1) it occurs at high frequency and at young age in members of certain families, and (2) it occurs sporadically at an older age among members of the population at large. The fact that retinoblastoma runs in certain families suggested that the cancer can be inherited. Examination of cells from children suffering from retinoblastoma revealed that one member of the thirteenth pair of homologous chromosomes was missing a small piece from the interior portion of the chromosome. The deletion was present in all of the children's cells—both the cells of the retinal cancer and cells elsewhere in the body—indicating that the chromosomal aberration had been inherited from one of the parents.

Retinoblastoma is inherited as a dominant genetic trait because members of high-risk families that develop the disease inherit one normal allele and one abnormal allele. But unlike most dominantly inherited conditions, such as Huntington's disease, where an individual who inherits a missing or an altered gene invariably develops the disorder, children who inherit a chromosome missing the retinoblastoma gene inherit a strong disposition toward developing retinoblastoma, rather than the disorder itself. In fact, approximately 10 percent of individuals who inherit a chromosome with an *RB* deletion never develop the retinal cancer. How is it that a small percentage of these predisposed individuals escape the disease?

The genetic basis of retinoblastoma was explained in 1971 by Alfred Knudson of the University of Texas. Knudson proposed that the development of retinoblastoma requires

_____
[2]For the present chapter, which deals primarily with human biology, we will follow a convention that is commonly used: human genes are written in capital letters (e.g., *APC*), mouse genes are written with the first letter capitalized (e.g., *Brca1*), and viral genes are written in lower case (e.g., *src*).

that both copies of the *RB* gene of a retinal cell be either eliminated or mutated before the cell can give rise to a retinoblastoma. In other words, the cancer arises as the result of two independent "hits" in a single cell. In cases of sporadic retinoblastoma, the tumor develops from a retinal cell in which both copies of the *RB* gene have undergone successive spontaneous mutation (Figure 16.10*a*). Because the chance that both alleles of the same gene will be the target of debilitating mutations in the same cell is extremely unlikely, the incidence of the cancer in the general population is extremely low. In contrast, the cells of a person who inherits a chromosome with an *RB* deletion are already halfway along the path to becoming malignant. A mutation in the remaining *RB* allele in any of the cells of the retina produces a cell that lacks a normal *RB* gene and thus cannot produce a functional *RB* gene product (Figure 16.8*b*). This explains why individuals who inherit an abnormal *RB* gene are so highly predisposed to developing the cancer. The second "hit" fails to occur in approximately 10 percent of these individuals, who do not develop the disease. Knudson's hypothesis was subsequently confirmed by examining cells from patients with an inherited disposition to retinoblastoma and finding that, as predicted, both alleles of the gene were mutated in the cancer cells. Individuals with sporadic retinoblastomas had normal cells that lacked *RB* mutations and tumor cells in which both alleles of the gene were mutated.

Although deficiencies in the *RB* gene are first manifested in the development of retinal cancers, this is not the end of the story. People who suffer from the inherited form of retinoblastoma are also at high risk of developing other types of tumors later in life, particularly soft-tissue sarcomas (tumors of mesenchymal rather than epithelial origin). The consequences of *RB* mutations are not confined to persons who inherit a mutant allele. Mutations in *RB* alleles are a common occurrence in sporadic breast, prostate, and lung cancers among individuals who have inherited two normal *RB* alleles. When cells from these tumors are cultured in vitro, the reintroduction of a wild-type *RB* gene back into the cells is sufficient to suppress their cancerous phenotype, indicating that the loss of this gene function contributes significantly to tumorigenesis. Let's look more closely at the role of the *RB* gene.

The Role of pRB in Regulating the Cell Cycle   The importance of the cell cycle in cell growth and proliferation was discussed in Chapters 14 and 15, where it was noted that factors that control the cell cycle can play a pivotal role in the development of cancer. In its best studied role, the protein encoded by the *RB* gene, pRB, helps regulate the passage of cells from the G_1 stage of the cell cycle into S phase, during which DNA synthesis occurs. As discussed on page 564, the transition from G_1 to S is a time of commitment for the cell; once a cell enters S phase, it invariably proceeds through the remainder of the cell cycle and

(a)

(b)
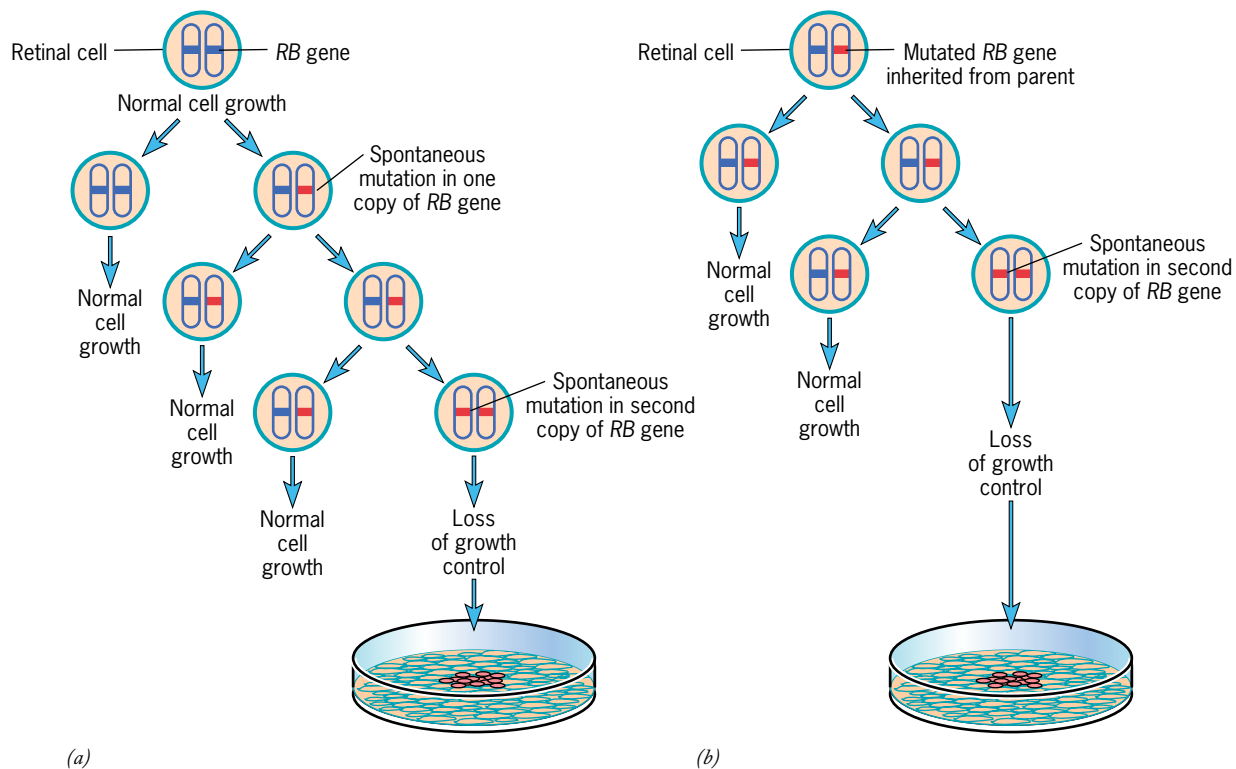
FIGURE 16.10 **Mutations in the *RB* gene that can lead to retinoblastoma.** (*a*) In sporadic (i.e., nonfamilial) cases of the disease, an individual begins life with two normal copies of the *RB* gene in the zygote, and retinoblastoma occurs only in those rare individuals in whom a given retinal cell accumulates independent mutations in both alleles of the gene. (*b*) In familial (i.e., inherited) cases of the disease, an individual begins life with one abnormal allele of the *RB* gene, usually present as a deletion. Thus all the cells of the retina have at least one of their two *RB* genes nonfunctional. If the other *RB* allele in a retinal cell becomes inactivated, usually as the result of a point mutation, that cell gives rise to a retinal tumor.

into mitosis. The transition from G_1 to S is accompanied by the activation of many different genes that encode proteins ranging from DNA polymerases to cyclins and histones. Among the transcription factors involved in activating genes required for S-phase activities are members of the E2F family of transcription factors, which are key targets of pRB. A model depicting the role of pRB in controlling E2F activity is illustrated in Figure 16.11. During G_1, E2F proteins are normally bound to pRB, which prevents them from activating a number of genes encoding proteins required for S-phase activities (e.g., cyclin E and DNA polymerase α). Studies suggest (as indicated in step 1 of Figure 16.11) that the E2F–pRB complex is associated with the DNA but acts as a gene repressor rather than a gene activator. As the end of G_1 approaches, the pRB subunit of the
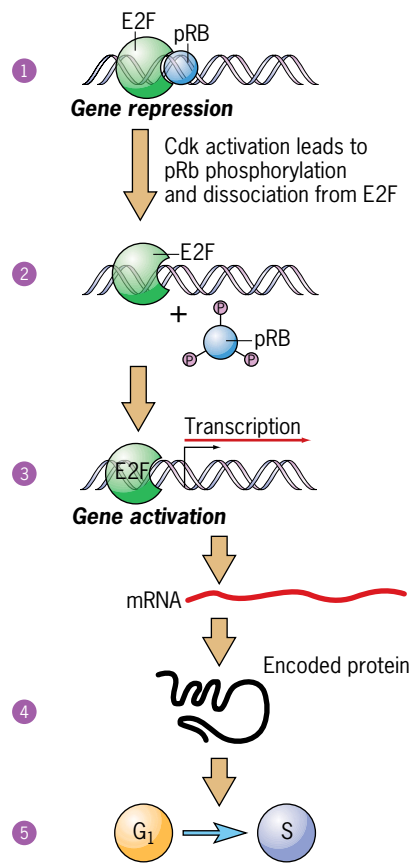
FIGURE 16.11 The role of pRB in controlling transcription of genes required for progression of the cell cycle. During most of G_1, the un-phosphorylated pRB is bound to the E2F protein. The E2F–pRB complex binds to regulatory sites in the promoter regions of numerous genes involved in cell cycle progression, acting as a transcriptional repressor that blocks gene expression. Repression probably involves the methylation of lysine 9 of histone H3 that modulates chromatin architecture (page 488). Activation of the cyclin-dependent kinase (Cdk) leads to the phosphorylation of pRB, which can no longer bind the E2F protein (step 2). In the pathway depicted here, loss of the bound pRB converts the DNA-bound E2F into a transcriptional activator, leading to expression of the genes being regulated (step 3). The mRNA is translated into proteins (step 4) that are required for the progression of cells from G_1 into S phase of the cell cycle (step 5). Other roles of pRB have been identified but are not discussed.

pRB–E2F complex is phosphorylated by the cyclin-dependent kinases that regulate the G_1–S transition. Once phosphorylated, pRB releases its bound E2F, allowing the transcription factor to activate gene expression, which marks the cell's irreversible commitment to enter S phase. A cell that loses pRB activity as the result of *RB* mutation would be expected to lose its ability to inactivate E2F, thereby removing certain restraints over the entry to S phase. E2F is only one of dozens of proteins capable of binding to pRB, suggesting that pRB has numerous other functions. The complexity of pRB interactions is also suggested by the fact that the protein contains at least 16 different serine and threonine residues that can be phosphorylated by cyclin-dependent kinases. It is likely that phosphorylation of different combinations of amino acid residues allows the protein to interact with different downstream targets.

The importance of pRB as a negative regulator of the cell cycle is demonstrated by the fact that DNA tumor viruses (including adenoviruses, human papilloma virus, and SV40) encode a protein that binds to pRB, blocking its ability to bind to E2F. The ability of these viruses to induce cancer in infected cells depends on their ability to block the negative influence that pRB has on progression of a cell through the cell cycle. By using these pRB-blocking proteins, these viruses accomplish the same result as when the *RB* gene is deleted, leading to the development of human tumors.

The Role of p53: Guardian of the Genome   Despite its unassuming name, the *TP53* gene may have more to do with the development of human cancer than any other component of the genome. The gene gets its name from the product it encodes, p53, which is a polypeptide having a molecular mass of 53,000 daltons. In 1990, *TP53* was recognized as the tumor-suppressor gene that, when absent, is responsible for a rare inherited disorder called Li-Fraumeni syndrome. Victims of this disease are afflicted with a very high incidence of various cancers, including breast and brain cancer and leukemia. Like individuals with the inherited form of retinoblastoma, persons with Li-Fraumeni syndrome inherit one normal and one abnormal (or deleted) allele of the *TP53* tumor-suppressor gene and are thus highly susceptible to cancers that result from random mutations in the normal allele.

The importance of p53 as an antitumor weapon is most evident from the fact that *TP53* is the most commonly mutated gene in human cancers; approximately half of all human tumors contain cells with point mutations or deletions in both alleles of the *TP53* gene. Furthermore, tumors composed of cells bearing *TP53* mutations are correlated with a poorer survival rate than those containing a wild-type *TP53* gene. Clearly, the elimination of *TP53* function is an important step in the progression of many cancer cells toward the fully malignant state.

Why is the presence of p53 so important in preventing a cell from becoming malignant? p53 is a transcription factor that activates the expression of a large number of genes involved in cell cycle regulation and apoptosis. The importance of the transcription-regulating role of p53 is evident in Figure 16.12, which shows the location of the six mutations most commonly found to disable p53 in human cancers; all of them map in the region of the protein that interacts with

FIGURE 16.12 p53 function is particularly sensitive to mutations in its DNA-binding domain. p53 functions as a tetramer, each subunit of which consists of several domains with different functions. This image shows a ribbon drawing of the DNA-binding domain. The six amino acid residues most often mutated in p53 molecules that have been debilitated in human cancers are indicated in a single-letter nomenclature (Figure 2.26). These residues occur at or near the protein–DNA interface and either directly impact the binding of the protein to DNA or alter its conformation. (See *Annu. Rev. Biochem.* 77:557, 2008 for a discussion.) (REPRINTED WITH PERMISSION FROM Y. CHO, S. GORINA, P. D. JEFFREY, N. P. PAVLETICH, SCIENCE 265:352, 1994; © COPYRIGHT 1994, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

DNA. One of the best studied genes activated by p53 encodes a protein called p21 that inhibits the cyclin-dependent kinase that normally drives a cell through the G_1 checkpoint. As the level of p53 rises in the damaged G_1 cell, expression of the *p21* gene is activated, and progression through the cell cycle is arrested (see Figure 14.9). This gives the cell time to repair the genetic damage before it initiates DNA replication. When both copies of the *TP53* gene in a cell have been mutated so that their product is no longer functional, the cell can no longer produce the p21 inhibitor or exercise the feedback control that prevents it from entering S phase when it is not prepared to do so. Failure to repair DNA damage leads to the production of abnormal cells that have the potential to become malignant.

Cell cycle arrest is not the only way that p53 protects an organism from developing cancer. Alternatively, p53 can direct a genetically damaged cell along a pathway that leads to death by apoptosis, thereby ridding the body of cells with a malignant potential. Whether p53 moves a cell toward cell cycle arrest or apoptosis apparently depends on the type of posttranslational modifications to which it is subjected. The p53 protein is thought to direct a cell into apoptosis as the result of several events, including the activation of expression of the *BAX* gene, whose encoded product (Bax) initiates apoptosis (page 642). Not all actions of p53 are dependent on its activation of transcription. p53 is also capable of binding directly to several members of the Bcl-2 family proteins (page 644) in a manner that stimulates apoptosis. For example, p53 can bind to Bax proteins at the outer mitochondrial membrane, directly triggering membrane permeabilization and release of apoptotic factors. If both alleles of *TP53* should become inactivated, a cell that is carrying damaged DNA fails to be destroyed, even though it lacks the genetic integrity required for controlled growth (Figure 16.13). Several studies have shown that established tumors in mice will undergo regression when the activity of their p53 genes is restored. This finding suggests that tumor development continues to depend on the absence of a functional *TP53* gene, even after its cells become genetically unstable. For these reasons, the development of drugs (e.g., PRIMA-1) that restore p53 function to mutant p53 proteins has become an active area of research.

The level of p53 in a healthy G_1 cell is very low, which keeps its potentially lethal action under control. However, if a G_1 cell sustains genetic damage, as occurs if the cell is subjected to ultraviolet light or chemical carcinogens, the concentration of p53 rises rapidly. A similar response can be elicited simply by injecting a cell with DNA containing broken strands. The increase in p53 levels is not due to increased expression of the gene but to an
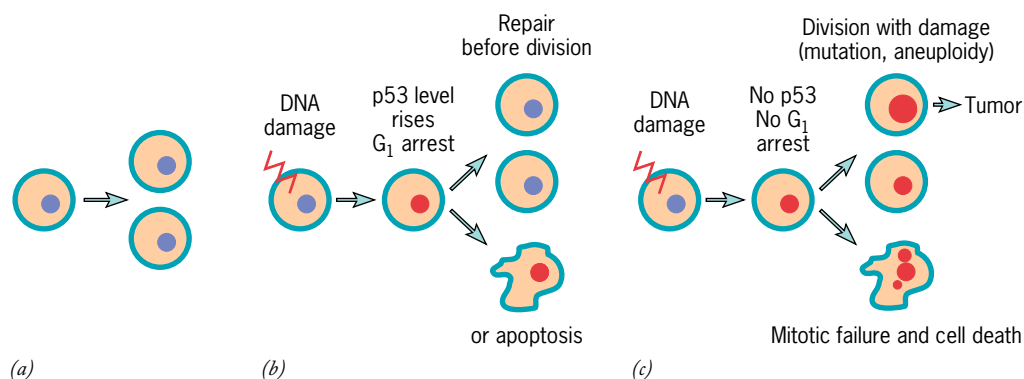
FIGURE 16.13 A model for the function of p53. (*a*) Cell division does not normally require the involvement of p53. (*b*) If, however, the DNA of a cell becomes damaged as the result of exposure to mutagens, the level of p53 rises and acts either to arrest the progression of the cell through G_1 or to direct the cell toward apoptosis. (*c*) If both copies of the *TP53* gene are inactivated, the cell loses the ability to arrest the cell cycle or commit the cell to apoptosis following DNA damage. As a result, the cell either dies from mitotic failure or continues to proliferate with genetic abnormalities that may lead to the formation of a malignant growth. (AFTER D. P. LANE, REPRINTED WITH PERMISSION FROM NATURE 358:15, 1992; © COPYRIGHT 1992, MACMILLAN MAGAZINES LTD.)

increase in the stability of the protein. In unstressed cells, p53 has a half-life of a few minutes. p53 degradation is facilitated by a protein called MDM2, which binds to p53 and escorts it out of the nucleus and into the cytosol. Once in the cytosol, MDM2 adds ubiquitin molecules to the p53 molecule, leading to its destruction by a proteasome (page 530). How does DNA damage lead to stabilization of p53? We saw on page 568 that persons suffering from ataxia telangiectasia lack a protein kinase called ATM and are unable to respond properly to DNA-damaging radiation. ATM is normally activated following DNA damage, and p53 is one of the proteins ATM phosphorylates. The phosphorylated version of the p53 molecule is no longer able to interact with MDM2, which stabilizes existing p53 molecules in the nucleus and allows them to activate the expression of genes such as *p21* and *BAX* (see Figure 16.16).

Some tumor cells have been found that contain a wild-type *TP53* gene but extra copies of *MDM2*. Such cells are thought to produce excessive amounts of MDM2, which prevents p53 from building to required levels to stop the cell cycle or induce apoptosis following DNA damage (or other oncogenic stimuli). A major effort is underway to develop drugs that block the interaction between MDM2 and p53 in an attempt to restore p53 activity in cancer cells that retain this key tumor suppressor. The relationship between MDM2 and p53 has also been demonstrated using gene knockouts. Mice that lack a gene encoding MDM2 die at an early stage of development, presumably because their cells undergo p53-dependent apoptosis. This interpretation is supported by the finding that mice lacking genes that encode *both* MDM2 and p53 (double knockouts) survive to adulthood but are highly prone to cancer. Because these embryos cannot produce p53, they don't require a protein such as MDM2 that facilitates p53 destruction. This observation illustrates an important principle in cancer genetics: even if a "crucial" gene such as *RB* or *TP53* is not mutated or deleted, the *function* of that gene can be affected as the result of alterations in other genes whose products are part of the same pathway as the "crucial" gene. In this case, overexpression of MDM2 can have the same effect as the absence of p53. As long as the tumor-suppressor pathway is blocked, the tumor-suppressor gene itself need not be mutated. Numerous studies indicate that both the p53 and pRB pathways have to be inactivated, one way or another, to allow the progression of most tumor cells.

Because of its ability to trigger apoptosis, p53 plays a pivotal role in treatment of cancer by radiation and chemotherapy. It was assumed for many years that cancer cells are more susceptible than normal cells to drugs and radiation because cancer cells divide more rapidly. But some cancer cells divide more slowly than their normal counterparts, yet they are still more sensitive to drugs and radiation. An alternate theory suggests that normal cells are more resistant to drugs or radiation because, once they sustain genetic damage, they either arrest their cell cycle until the damage is repaired or they undergo apoptosis. In contrast, cancer cells that have sustained DNA damage are more likely to become apoptotic—as long as they possess a functioning *TP53* gene. If cancer cells lose p53 function, they often cannot be directed into apoptosis and they become highly resistant to further treatment (Figure 16.14). This may be the primary reason why tumors that typically lack a functional *TP53* gene (e.g., colon cancer,

| Untreated | 5-Fluorouracil | Etoposide | Adriamycin |

(+/+)

(+/-)

(-/-)
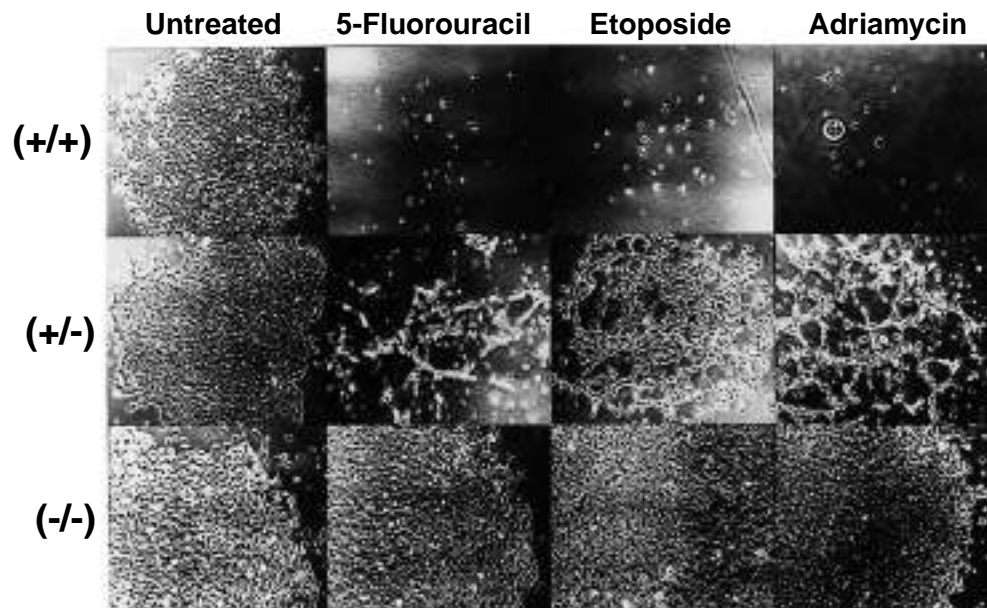
FIGURE 16.14 Experimental demonstration of the role of p53 in the survival of cells treated with chemotherapeutic agents. Cells were cultured from mice that had two functional alleles of the gene encoding p53 (top row), one functional allele of the gene (middle row), or were lacking a functional allele of the gene (bottom row). Cultures of each of these cells were grown either in the absence of a chemotherapeutic agent (first column) or in the presence of one of the three compounds indicated at the top of the other three columns. It is evident that the compounds had a dramatic effect on arresting growth and inducing cell death (apoptosis) in normal cells, whereas the cells lacking p53 continued to proliferate in the presence of these compounds. (FROM SCOTT W. LOWE, H. E. RULEY, T. JACKS, AND D. E. HOUSMAN, CELL 74:959, 1993; BY PERMISSION OF CELL PRESS.)

prostate cancer, and pancreatic cancer) respond much more poorly to radiation and chemotherapy than tumors that possess a wild-type copy of the gene (e.g., testicular cancer and childhood acute lymphoblastic leukemias).

We have seen how p53 can direct a potential cancer cell into either growth arrest or apoptosis. Recent studies indicate that p53 also controls signaling pathways that lead to cellular senescence, another mechanism that has evolved as a barrier that stops wayward cells from developing into malignant tumors. Unlike apoptotic cells, senescent cells remain alive and metabolically active but they are permanently arrested in a nondividing state, as exemplified by the senescent melanocytes found in moles (discussed on page 655). Senescence can be triggered in an otherwise normal cell by the experimental activation of an oncogene, such as Ras, which might occur with some frequency during the day-to-day activities of dividing cells in a normal tissue. Studies suggest that oncogene activation triggers a period of accelerated division after which the senescence program takes effect and slams on the brakes. This is the apparent route that is taken during the formation of benign moles. One of the pathways leading to senescence requires expression of a tumor-suppressor gene called *INK4a,* which is often disabled in human cancers (Table 16.1). *INK4a* encodes two separate tumor-suppressing proteins: p16, which is an inhibitor of cyclin-dependent kinases, and ARF, which stabilizes p53. The precise role of p53 in directing cells towards the senescent state remains unclear.

Other Tumor–Suppressor Genes   Although mutations in *RB* and *TP53* are associated with a wide variety of human malignancies, mutations in a number of other tumor-suppressor genes are detected in only a few types of cancer.

Familial adenomatous polyposis coli (*FAP*) is an inherited disorder in which individuals develop hundreds or even thousands of premalignant polyps (adenomas) from epithelial cells that line the colon wall (Figure 16.15). If not removed, cells within some of these polyps are very likely to progress to a fully malignant stage. The cells of patients with this condition were found to contain a deletion of a small portion of chromosome 5, which was subsequently identified as the site of a tumor-suppressor gene called *APC*. A person inheriting an *APC* deletion is in a similar position to one who inherits an *RB* deletion: if the second allele of the gene is mutated in a given cell, the protective value of the gene function is lost. The loss of the second allele of *APC* causes the cell to lose growth control and proliferate to form a polyp rather than differentiating into normal epithelial cells of the intestinal wall. The conversion of cells in a polyp to the more malignant state, characterized by the ability to metastasize and invade other tissues, is presumably gained by the accumulation of additional mutations, including those in *TP53*. Mutated *APC* genes are found not only in inherited forms of colon cancers, but also in up to 80 percent of sporadic colon tumors, suggesting that the gene plays a major role in the development of this disease. The protein encoded by the *APC* gene binds a number of different proteins, and its mechanism of action is complex. In its best studied role, APC suppresses the Wnt pathway, which activates the transcription of genes (e.g.,
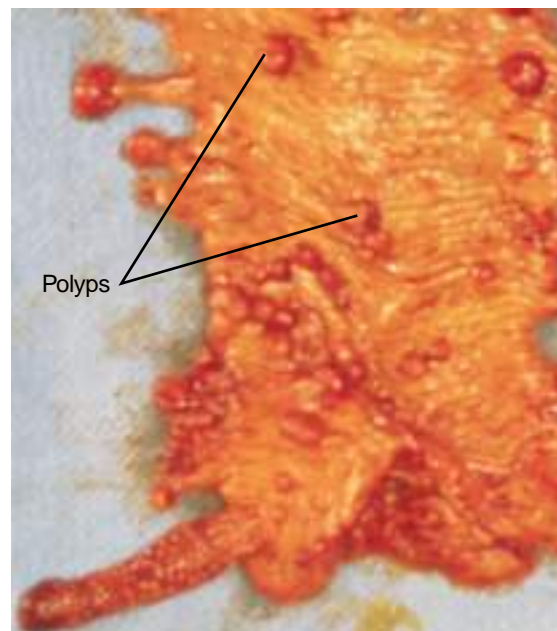
FIGURE 16.15 **Premalignant polyps in the epithelium of the human colon.** Photograph of a portion of a colon removed from a patient with Gardner's syndrome that demonstrates the pattern of polyp formation. Similar patterns are found in the colon of persons with the inherited condition of familial adenomatous polyposis coli (FAP). (COURTESY OF RANDALL W. BURT.)

MYC and *CCND1*) that promote cell proliferation. APC may also play a role in the attachment of microtubules to the kinetochores of mitotic chromosomes. Loss of APC function could therefore lead directly to abnormal chromosome segregation and aneuploidy (page 584). The presence of mutated *APC* DNA has been found in the blood of persons with early-stage colon cancer, which raises the possibility of a diagnostic test for the disease.

It is estimated that breast cancer strikes approximately one in eight women living in the United States, Canada, and Europe. Of these cases, 5–10 percent are due to the inheritance of a gene that predisposes the individual to development of the disease. After an intensive effort by several laboratories, two genes named *BRCA1* and *BRCA2* were identified in the mid-1990s as being responsible for the majority of the inherited cases of breast cancer. *BRCA* mutations also predispose a woman to the development of ovarian cancer, which has an especially high mortality rate.

BRCA1 and BRCA2 proteins have been the focus of a major research effort, but their precise functions remain unclear. It was pointed out on page 567, that cells possess checkpoints that halt progression of the cell cycle following DNA damage. The BRCA proteins are part of one or more large protein complexes that respond to DNA damage and activate DNA repair by means of homologous recombination. Cells with mutant BRCA proteins contain unrepaired DNA and exhibit a highly aneuploid karyotype. In cells with a functional *TP53* gene, failure to repair DNA damage leads to the activation of p53, which causes the cell to either arrest cell cycle
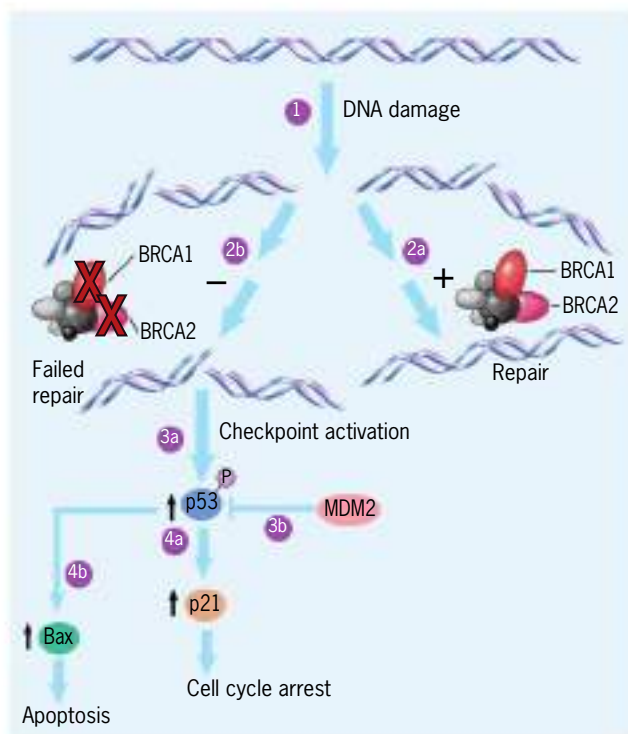
FIGURE 16.16 DNA damage initiates activity of a number of proteins encoded by both tumor-suppressor genes and proto-oncogenes. In this simplified figure, DNA damage is seen to cause double-strand breaks in the DNA (step 1) that are repaired by a proposed multiprotein complex that includes BRCA1 and BRCA2 (step 2a). Mutations in either of the genes that encode these proteins can block the repair process (step 2b). If DNA damage is not repaired, a checkpoint is activated that leads to a rise in the level of p53 activity (step 3a). The p53 protein is normally inhibited by interaction with the protein MDM2 (step 3b). p53 is a transcription factor that activates expression of either (1) the *p21* gene (step 4a), whose product (p21) causes cell cycle arrest, or (2) the *BAX* gene (step 4b), whose product (Bax) causes apoptosis. p53 activation can also promote cellular senescence, but the pathway is unclear. (REPRINTED WITH PERMISSION AFTER J. BRUGAROLAS AND T. JACKS, NATURE MED. 3:721, 1997; COPYRIGHT 1997, MACMILLAN MAGAZINES LTD.)

progress or undergo apoptosis, as illustrated in Figure 16.16. Unlike most tumor-suppressor genes, neither of the *BRCA* genes is commonly mutated in sporadic forms of the cancer, although *BRCA1* is often epigenetically silenced.

We have seen in this chapter that apoptosis is one of the body's primary mechanisms of ridding itself of potential tumor cells. The mechanism of apoptosis was discussed in the last chapter, as were pathways that promoted cell survival rather than cell destruction. The best studied cell-survival pathway involves the activation of a kinase called PKB (AKT) by the phosphoinositide PIP_3. PIP_3, in turn, is formed by the catalytic activity of the lipid kinase PI3K (see Figure 15.23). Activation of the PI3K/PKB pathway leads to an increased likelihood that a cell will survive a stimulus that normally would lead to its destruction. Whether a cell lives or dies following a particular event depends to a large degree on the balance between proapoptotic and antiapoptotic signals. Mutations that affect this balance, such as those that contribute to

the overexpression of PKB or PI3K, can shift this balance in favor of cell survival, which can provide a potential cancer cell with a tremendous advantage. Another protein that can affect the balance between life and death of a cell is the lipid phosphatase, PTEN, which removes the phosphate group from the 3-position of PIP_3, converting the molecule into $PI(4,5)P_2$, which cannot activate PKB. Cells in which both copies of the *PTEN* gene are inactivated tend to have an excessively high level of PIP_3, which leads to an overactive population of PKB molecules. Like the other tumor-suppressor genes listed in Table 16.1, mutations in *PTEN* cause a rare hereditary disease characterized by an increased risk of cancer, and such mutations are also found in a variety of sporadic cancers. Mutation is not the only mechanism by which tumor-suppressor genes can be inactivated. *PTEN* genes are often rendered nonfunctional as the result of epigenetic mechanisms, such as DNA methylation, which silences transcription of the gene (page 520). When a normal *PTEN* gene is introduced into tumor cells that lack a functioning copy of the gene, the cells typically undergo apoptosis, as would be expected.

Oncogenes As described above, oncogenes encode proteins that promote the loss of growth control and the conversion of a cell to a malignant state. Oncogenes are derived from proto-oncogenes (page 657), which are genes that encode proteins having a function in the normal cell. Most known proto-oncogenes play a role in the control of cell growth and division. Numerous oncogenes were initially identified as part of the genomes of RNA tumor viruses, but many more have been identified because of their importance in tumorigenesis as determined in laboratory animals or human tumor samples. Different oncogenes become activated in different types of tumors, which reflects variations in the signaling pathways that operate in diverse cell types. The oncogene mutated most frequently in human tumors is *RAS*, which encodes a GTP-binding protein (Ras) that functions as an on–off switch for a number of key signaling pathways controlling cell proliferation (page 627).[3] Oncogenic *RAS* mutants typically encode a protein whose GTPase activity cannot be stimulated, which leaves the molecule in an active GTP-bound form, sending continuous proliferation signals along the pathway. The functions of a number of oncogenes are summarized in Figure 16.17 and discussed below.[4]

Oncogenes That Encode Growth Factors or Their Receptors The first connection between oncogenes and growth factors was made in 1983, when it was discovered that the cancer-causing simian sarcoma virus contained an oncogene (*sis*) derived from the cellular gene for platelet-derived growth factor (PDGF), a protein present in human blood. Cultured cells that are transformed with this virus secrete large amounts of PDGF into the medium, which causes the cells to proliferate in an uncon-

---

[3]The human genome actually contains three different *RAS* genes and three different *RAF* genes that are active in different tissues. Of these, *KRAS* and *BRAF* are most often implicated in tumor formation.

[4]The reader is referred to the Human Perspective of Chapter 7 (page 248) for a discussion of genes that encode cell-surface molecules and extracellular proteases that play an important role in tissue invasion and metastasis.

Ras) signaling pathway is disturbed. As discussed below, growth factor receptors have become a favored target for therapeutic antibodies, which bind to the receptor's extracellular domain, and for small molecule inhibitors, which bind to the receptor's intracellular tyrosine kinase domain.

Oncogenes That Encode Cytoplasmic Protein Kinases   Overactive protein kinases function as oncogenes by generating signals that lead to inappropriate cell proliferation or survival. Raf, for example, is a serine-threonine protein kinase that heads the MAP kinase cascade, the primary growth-controlling signaling pathway in cells (page 627). It is evident that Raf is well positioned to wreak havoc within a cell should its enzymatic activity become altered as the result of mutation. As with the growth factor receptors and Ras, mutations that turn Raf into an enzyme that remains in the "on" position are most likely to convert the proto-oncogene into an oncogene and contribute to the cell's loss of growth control. Raf is most closely linked to melanoma, where *BRAF* mutations play a causative role in the development of approximately 70 percent of these cancers.

The first oncogene to be discovered, *SRC*, is also a protein kinase, but one that phosphorylates tyrosine residues on protein substrates rather than serine and threonine residues. Transformation of a cell by a *src*-containing tumor virus is accompanied by the phosphorylation of a wide variety of proteins. Included among the apparent substrates of Src are proteins involved in signal transduction, control of the cytoskeleton, and cell adhesion. For an unknown reason, *SRC* mutations appear only rarely among the repertoire of genetic changes in human tumor cells.

Oncogenes That Encode Nuclear Transcription Factors   A number of oncogenes encode proteins that act as transcription factors. The progression of cells through the cell cycle requires the timely activation (and repression) of a large variety of genes whose products contribute in various ways to cell growth and division. It is not surprising, therefore, that alterations in the proteins that control the expression of these genes could seriously disturb a cell's normal growth patterns. Probably the best studied oncogene whose product acts as a transcription factor is *MYC*.

It was noted on page 562 that cells that are not actively growing and dividing tend to withdraw from the cell cycle and enter a stage referred to as G_0, from which they can be retrieved. Myc is normally one of the first proteins to appear when a cell in this quiescent stage has been stimulated by growth factors to reenter the cell cycle and divide. Myc regulates the expression of a huge number of proteins and miRNAs involved in cell growth and proliferation. When *MYC* expression is selectively blocked, the progression of the cell through G_1 is blocked. The *MYC* gene is one of the proto-oncogenes most commonly altered in human cancers, often being amplified within the genome or rearranged as the result of a chromosome translocation. These chromosomal changes are thought to remove the *MYC* gene from its normal regulatory influences and increase its level of expression in the cell, producing an excess of the Myc protein. One of the most

FIGURE 16.17 A schematic diagram summarizing the types of proteins encoded by proto-oncogenes. These include growth factors (1), receptors for growth factors (2), protein kinases and the proteins that activate them (3), proteins that regulate the cell cycle (4), proteins that inhibit apoptosis (5), and transcription factors (6). Proteins involved in mitosis, tissue invasion, and metastasis are not included.

trolled fashion. Overexpression of PDGF has been implicated in the development of brain tumors (gliomas).

Another oncogenic virus, avian erythroblastosis virus, was found to carry an oncogene (*erbB*) that encodes an EGF receptor that is missing part of the extracellular domain of the protein that binds the growth factor. One might expect that the altered receptor would be unable to signal a cell to divide, but just the reverse is true. This altered version of the receptor stimulates the cell constitutively, that is, regardless of whether or not the growth factor is present in the medium. This is the reason why cultured cells that carry the altered gene proliferate in an uncontrolled manner. A number of spontaneous human cancers have been found to contain cells with genetic alterations that affect growth factor receptors, including EGFR. Most commonly, the malignant cells contain a much larger number of the receptors in their plasma membranes than do normal cells. The presence of excess receptors makes the cells sensitive to much lower concentrations of the growth factor, and thus, they are stimulated to divide under conditions that would not affect normal cells. A number of studies suggest that mutations in *EGFR* occur commonly in lung cancers from patients who have never smoked, but are not found in lung cancers from smokers. Mutations in the *KRAS* gene show the opposite distribution. This finding suggests that lung cancers in the two groups of patients have a different course of genetic progression, although the same (EGFR-

common types of cancer among populations in Africa, called Burkitt's lymphoma, results from the translocation of a *MYC* gene to a position adjacent to an antibody gene. The disease occurs primarily in persons who have also been infected with Epstein-Barr virus. This same virus causes only minor infections (e.g., mononucleosis) in people living in Western countries and is not associated with tumorigenesis.

Oncogenes That Encode Products That Affect Apoptosis   Apoptosis is one of the body's key mechanisms to rid itself of tumor cells at an early stage in their progression toward malignancy. Consequently, any alteration that diminishes a cell's ability to self-destruct would be expected to increase the likelihood of that cell giving rise to a tumor. This was evident in the previous discussion of the role of the PI3K/PKB pathway in cell survival and tumorigenesis (page 664). With this in mind, it is not surprising that both PI3K and PKB are encoded by documented oncogenes. The oncogene most closely linked to apoptosis is *BCL-2*, which encodes a membrane-bound protein that inhibits apoptosis (page 644).

The role of *BCL-2* in apoptosis is most clearly revealed in the phenotypes of knockout mice that are lacking a *Bcl-2* gene. Once formed, the lymphoid tissues of these mice undergo dramatic regression as the result of widespread apopto-

sis. Like *MYC*, the product of the *BCL-2* gene becomes oncogenic when it is expressed at higher-than-normal levels, as can occur when the gene is translocated to an abnormal site on the chromosome. Certain human lymphoid cancers (called follicular B-cell lymphomas) are correlated with the translocation of the *BCL-2* gene next to a gene that codes for the heavy chain of antibody molecules. It is suggested that overexpression of the *BCL-2* gene leads to the suppression of apoptosis in lymphoid tissues, allowing abnormal cells to proliferate to form lymphoid tumors. The *BCL-2* gene may also play a role in reducing the effectiveness of chemotherapy by keeping tumor cells alive and proliferating despite damage by the drug treatment. To counteract this property of cancer cells, a number of pharmaceutical companies are developing compounds that make cancer cells more likely to undergo apoptosis.

Over the past few pages, we have discussed a number of the most important tumor suppressors and oncogenes involved in tumorigenesis. Figure 16.18 provides a simplified overview of some of these proteins and the signaling pathways in which they operate. Tumor suppressors and tumor-suppressive pathways are shown in red, oncogenes and tumor-promoting pathways are shown in blue. The basic functions of each of the tumor suppressors and oncogenes depicted in the figure are noted in the legend of Figure 16.18.
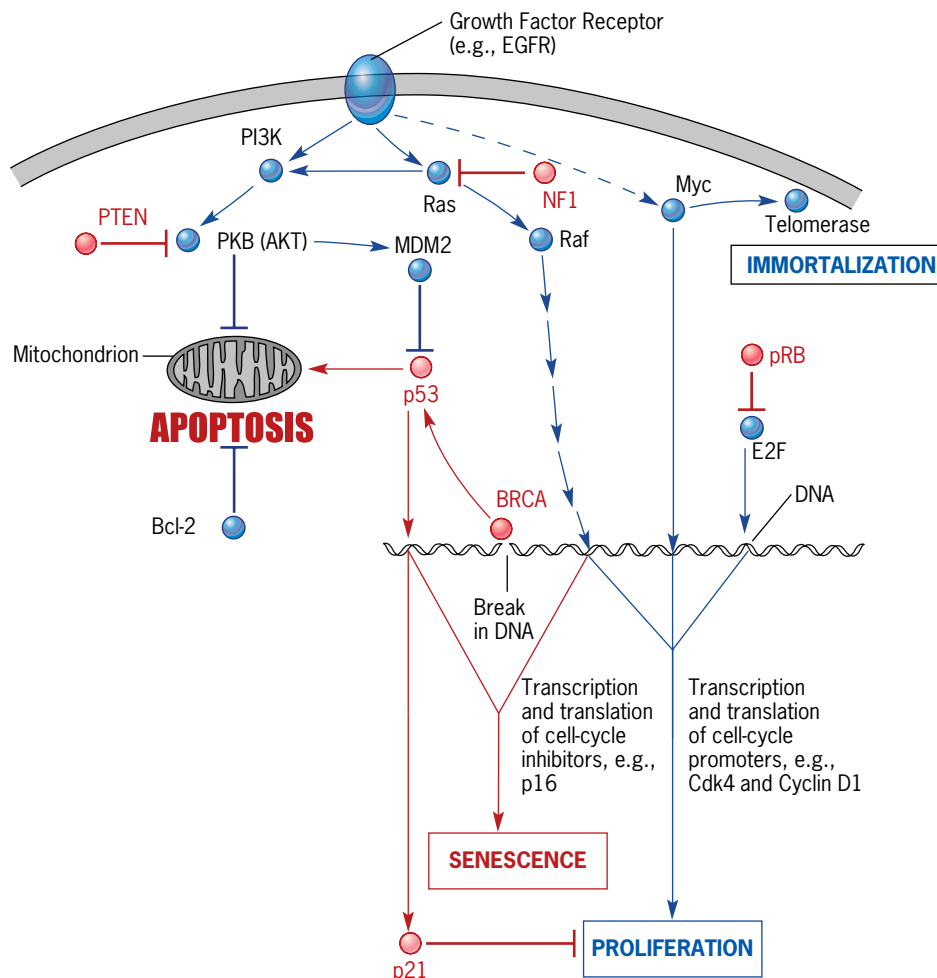
FIGURE 16.18 An overview of several of the signaling pathways involved in tumorigenesis that were discussed in this section. Tumor suppressors and tumor suppression are shown in red, whereas oncogenes and tumor stimulation are shown in blue. Arrows indicate activation, perpendicular lines indicate inhibition. Among the proteins depicted in this figure are transcription factors (p53, Myc, and E2F), a transcriptional coactivator or corepressor (pRB), a lipid kinase (PI3K) and lipid phosphatase (PTEN), a cytoplasmic tyrosine kinase (Raf) and its activator (Ras), a GTPase activating protein for Ras (NF1), a protein kinase that promotes cell survival (PKB/AKT), a protein that senses DNA breaks (BRCA), subunits of a cyclin-dependent kinase (cyclin D1 and Cdk4), a Cdk inhibitor (p21), an antiapoptotic protein (Bcl-2), a ubiquitin ligase (MDM2), an enzyme that elongates DNA (telomerase), and a protein that binds growth factors (e.g., EGFR). The dashed line indicates indirect action by activation of expression of the *MYC* gene. The figure addresses only four of the processes that contribute to tumorigenesis, namely apoptosis, senescence, proliferation, and immortalization.

The remarkable diversity of protein activities that can contribute to tumorigenesis is evident from this list.

The Mutator Phenotype: Mutant Genes Involved in DNA Repair

If one considers cancer as a disease that results from alterations in the DNA of somatic cells, then it follows that any activity that increases the frequency of genetic mutations is likely to increase the risk of developing cancer. As discussed in Chapter 13, nucleotides that become chemically altered or nucleotides that are incorporated incorrectly during replication are selectively removed from the DNA strand by DNA repair. DNA repair processes require the cooperative efforts of a substantial number of proteins, including proteins that recognize the lesion, remove a portion of the strand containing the lesion, and replace the missing segment with complementary nucleotides. If any of these proteins are defective, the affected cell can be expected to display an abnormally high mutation rate, which is described as a "mutator phenotype." Cells with a mutator phenotype are likely to incur secondary mutations in both tumor-suppressor genes and oncogenes, which increases their risk of becoming malignant.

We saw in Chapter 13 that defects in nucleotide excision repair (NER) lead to the development of a cancer syndrome called xeroderma pigmentosum. In 1993, evidence was obtained that defects in a different type of DNA repair might also lead to the initiation of cancer. In this case, studies were performed on the cells of patients with the most common hereditary form of colon cancer, called hereditary nonpolyposis colon cancer (HNPCC) to distinguish it from the polyp-forming type (FAP) described earlier. A defective gene responsible for HNPCC is carried by approximately 0.5 percent of the population and accounts for about 3 percent of all colon cancer cases. It was noted on page 395 that the genome contains large numbers of very short, repetitive DNA sequences called microsatellites. Analysis of the DNA of persons with HNPCC revealed that microsatellite sequences in the tumor cells were often of different length than the corresponding sequences in the DNA of normal cells from the same patient. One would expect such variations in the DNA from different individuals, but not from the DNA of different cells of the same person.

The discovery of microsatellite sequence variation in these hereditary cancers—and in sporadic colon cancers as well—suggested that a deficiency in the mismatch repair system (page 554) is likely at fault. Support for this proposal has come from studies of persons with HNPCC. Whereas extracts from normal cells carry out mismatch repair in vitro, extracts from HNPCC tumor cells display DNA-repair deficiencies. Analysis of the DNA from a large number of people with HNPCC has revealed deletions or debilitating mutations in any one of a number of genes that code for proteins that form the DNA mismatch repair pathway. Cells with mismatch repair deficiencies accumulate secondary mutations throughout the genome.

MicroRNAs: A New Player in the Genetics of Cancer

Recall from Section 11.5 that microRNAs are tiny regulatory RNAs that negatively regulate the expression of target mRNAs. Given that cancers arise as the result of abnormal gene expression, it would not be surprising to discover that miRNAs are somehow involved in tumorigenesis. In 2002, it was reported that the locus that encodes two microRNAs, *miR-15a* and *miR-16*, was either deleted or underexpressed in most cases of chronic lymphocytic leukemia. It was subsequently shown that these two miRNAs act to inhibit expression of the mRNA that encodes the antiapoptotic protein Bcl-2, a known proto-oncogene (page 657). In the absence of the miRNAs, the oncogenic Bcl-2 protein is overexpressed, which promotes development of leukemia. Because these miRNAs act to inhibit tumorigenesis, they can be thought of as tumor suppressors. When the leukemic cells lacking *miR-15a* and *miR-16* were genetically engineered to reexpress these RNAs, they underwent apoptosis as would be expected if a missing tumor suppressor activity was restored. The locus that encodes *miR-15a* and *miR-16* is also deleted in other types of cancers, suggesting it has widespread importance in tumor suppression. The expression of two of the most important human oncogenes, *RAS* and *MYC,* have also been shown to be inhibited by an miRNA, namely, *let-7,* which was the first miRNA to be discovered (page 452).

Some miRNAs act more like oncogenes than tumor suppressors. One specific cluster of miRNA genes, for example, is overexpressed during the formation of certain lymphomas. Overexpression of these miRNAs can occur because the gene cluster encoding them is present in increased number (amplified) in the tumor cells, or it can occur because the gene cluster is excessively transcribed as the result of an overly active transcription factor. When mice are genetically engineered to overexpress these particular miRNAs, the animals develop lymphomas as would be predicted if the genes encoding them were acting as oncogenes. The abnormal expression of miRNAs has also been implicated as a causal factor in tumor cell invasiveness and metastasis, which elevates the interest in these RNAs to an even higher level.

It is not yet clear how important miRNAs are in the overall occurrence of human cancer, but a number of microarray studies that survey large numbers of these tiny regulatory RNAs suggest that most human cancers have a characteristic miRNA expression profile, just as they have a characteristic mRNA expression profile. Recent studies suggest that miRNA expression profiles may serve as sensitive and accurate biomarkers to identify the exact type of tumor a person is suffering from and the best avenue of treatment. These studies also suggest that miRNAs such as *let-7* may ultimately serve as another weapon in the arsenal of anticancer therapeutics.

The Cancer Genome

All cancers arise as the result of genetic alterations. As was evident from the previous discussion, the genes involved in tumorigenesis constitute a specific subset of the genome whose products are involved in such activities as the progression of a cell through the cell cycle, adhesion of a cell to its neighbors, apoptosis, and repair of DNA damage. Taken together, approximately 350 different genes have been identified as "cancer

genes," that is, genes that are thought to have some causal role in the development of at least one type of malignancy. Over the past few years a concerted effort has been made to determine which of these genes are altered—either by point mutation, deletion, or duplication—in various types of tumors. This effort has been aided by recent advances in DNA sequencing that allow researchers to determine the nucleotide sequences of specific regions of the genome much more rapidly and inexpensively than ever before.

It was hoped that most types of cancers would be characterized by alterations in a relatively small number of genes. It has been known for quite a while, for example, that a large percentage of melanomas exhibit mutant *BRAF* genes and a large percentage of colorectal cancers exhibit mutant *APC* genes. In both cases these mutations occur at an early stage in tumor development and are thought to be critically important in turning the cell toward an eventual malignant state. However, the results from the initial studies of cancer genomes suggest that the same types of tumors taken from different patients possess widely divergent combinations of aberrant genes. This observation presumably reflects the many different routes that individual tumors can take to escape the cell's normal antitumor protections. These findings can be displayed as shown in Figure 16.19, where mutated genes identified in a large number of colorectal cancers are shown as peaks within a two-dimensional "mutational landscape." The height of the various peaks reflects the frequency with which that particular gene is mutated in this particular type of cancer. It is evident from this type of display that a small number of genes are mutated in a large proportion of tumors; these can be thought of as "mountains" in the landscape. For the most part, these are the oncogenes that cancer researchers have

been focusing on over the years. However, a surprisingly large number of genes are mutated at a lower frequency (less than 5 percent of cases); these have been described as "hills." There are approximately 50 different genes representing the hills in the mutational landscape of colorectal cancers in Figure 16.19. While we can presume that the genes that are mutated at high frequency (the mountains) are important factors in driving the cells into malignancy, considerable debate is centered on the roles of the genes that are mutated at lower frequency (the hills). Many of the genes represented by the hills almost certainly have a causal role in determining properties of the malignant phenotype, even if they provide the tumor with only a small selective advantage. Other genes that constitute a hill may simply represent "passengers," that is, genes that are subject to mutation but have no effect on the phenotype of the cancer cell. It may be a daunting challenge to determine which of these genes are drivers of cancer and which are simply passengers. In addition to the genes of the mountains and hills, there is a large number of other genes that show up in a mutant state at very low frequency in a population of tumors. Mutations of this class can be seen in Figure 16.19 as the small circles scattered over each landscape that are not at the base of a hill or mountain. Figure 16.19*a,b* show the mutational landscapes of colorectal tumors from two different individuals. On average, an individual tumor carried approximately 80 different mutations. It is apparent that only a very small number of mutated genes in these cancers are shared by the two individuals. Thus, in a sense, each person suffers from his or her own distinct type of disease.

It is evident from this and a large number of other studies on other types of cancers that the mutational landscape of the
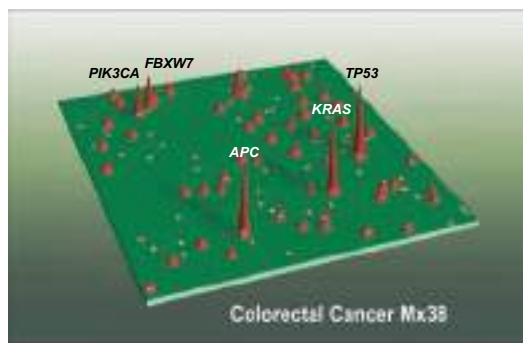
(a)
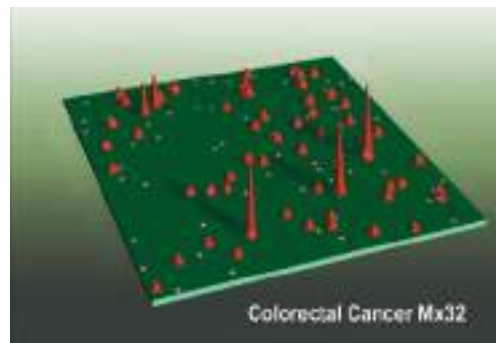
(b)

FIGURE 16.19 The genomic landscape of colorectal cancers. These two-dimensional maps depict the genes most frequently mutated in colorectal tumors. Each reddish projection represents a different gene. The five genes that are mutated in a large proportion of tumors are represented by the tallest projections, referred to as "mountains," and are specifically named. The 50 or so other genes that are mutated at a much lower frequency constitute the smaller "hills" of the genomic landscape. To depict the degree to which colorectal tumors from different patients share commonly mutated genes, the mutational landscapes from two individual tumors (identified as Mx38 and Mx32) are depicted in this illustration. The genes that were found to be somatically mutated in each of these individual tumors are indicated by the white dots. It is evident that very few mutations are shared between the tumors of these two individuals. In the example depicted here, only the *APC* and *TP53* genes are mutated in both cases of the disease. (*Note:* The positions of the genes in this two-dimensional landscape are ordered with loci from one end of chromosome 1 at the bottom left of the landscape, proceeding through each of the autosomes in ascending order, until finally reaching the loci from chromosome X at the right edge of the landscape.) (FROM LAURA D. WOOD ET AL., COURTESY OF BERT VOGELSTEIN, SCIENCE 318:1113, 2007; © COPYRIGHT 2007, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

TABLE 16.2 Core Signaling Pathways and Processes Genetically Altered in Most Pancreatic Cancers

Regulatory process or pathway	Number of genetically altered genes detected	Fraction of tumors with genetic alteration of at least one of the genes
Apoptosis	9	100%
DNA damage control	9	83%
Regulation of G_1/S phase transition	19	100%
Hedgehog signaling	19	100%
Homophilic cell adhesion	30	79%
Integrin signaling	24	67%
c-Jun N-terminal kinase signaling	9	96%
KRAS signaling	5	100%
Regulation of invasion	46	92%
Small GTPase–dependent signaling (other than KRAS)	33	79%
TGF-β signaling	37	100%
Wnt/Notch signaling	29	100%

From S. Jones et al., *Science* 321:1805, 2008; © copyright 2008, American Association for the Advancement of Science.

cancer genome is highly complex. However, a closer analysis of the mutations that make up the mountains and hills suggests that most of these large numbers of genes encode proteins that participate as components of a relatively small number of pathways. In one study of persons with pancreatic cancer, which is one of the most deadly and least treatable types, mutations were found in more than 60 genes, but, for the most part, these mutations affected a core set of 12 cellular pathways or processes (Table 16.2). Most importantly, half of these pathways/processes were disrupted in 100 percent of the tumor samples. In a comparable study of glioblastoma, the most common form of brain cancer, the majority of tumors exhibited mutations that affected three major pathways—p53, pRB, and PI3K. Thus, as discussed elsewhere in this chapter, cancer can be thought of not simply as a disease of aberrant genes but as one of aberrant cellular pathways. Mutations in any one of a number of genes disrupt the same pathway and thus have the same consequence for the cells. A few of the most prominent of these pathways were indicated in Figure 16.18. Looking at cancer as a "pathway" disease rather than a "genetic" disease generates more optimism among drug developers because it suggests that disrupting any one of the key steps in a single essential pathway may be sufficient to derail the malignant cells and lead to tumor regression. If this approach is to be successful, researchers must first identify biomarkers that can inform physicians as to which cellular pathways are disrupted in the cells of a given patient's tumor.

Before leaving the subject of cancer genetics, it should be noted that the view presented here—that cancer is a gradual multistep progression of individual point mutations—is not universally shared. Some researchers argue that the mutation rate in humans is not high enough for cells to accumulate the mutations necessary to become fully malignant during an individual's lifetime. Instead, they have proposed that carcinogenesis is initiated by catastrophic events that lead to widespread genetic instability over a relatively small number of cell divisions. For example, mutations in a gene involved in DNA replication or DNA repair, as occurs in cases of HNPCC, might rapidly spawn cells carrying widespread genetic abnormalities. According to another proposal, cells that have undergone an abnormal cell division and possess aberrant numbers of chromosomes, are likely initiators of cancerous growths. The best way to decide among these possibilities is to analyze the state of the genome in cells at very early stages in tumor formation. Unfortunately, for the interest of both researchers and cancer patients, it is virtually impossible to identify tumors when they are composed of a small number of cells. By the time that tumors are detected, the cells already exhibit a high degree of genetic derangement, making it difficult to determine whether these genetic alterations are a cause or an effect of tumor growth.

Gene-Expression Analysis

Over the past decade or so, a technology for analyzing gene expression has been developed that may, one day, have considerable impact on the way cancer is diagnosed and treated. This technology, which uses DNA microarrays (or DNA chips), was described in detail on page 505. Briefly, a glass slide is prepared that can contain anywhere from a handful to thousands of spots of DNA, each spot containing the DNA corresponding to a single, known gene. Any particular set of genes, such as those thought to be involved in growth and division, or those thought to be involved in the development and differentiation of lymphocytes or some other cell type, can be included within the microarray. Once prepared, the microarray is incubated with fluorescently labeled cDNAs synthesized from the mRNAs of a particular population of cells, such as those from a tumor mass that has been removed during surgery, or from the cancerous blood cells of a patient with leukemia. The fluorescently labeled cDNAs hybridize to spots of complementary DNA immobilized on the slide, and subsequent analysis of the fluorescence pattern tells the investigators which mRNAs are present within the tumor cells and their relative abundance within the mRNA population.

Studies with DNA microarrays have shown that gene-expression profiles can provide invaluable information about the properties of a tumor. It has been found, for example, that (1) progression of a tumor is correlated with a change in expression of particular genes, (2) certain cancers that appear to be similar by conventional criteria can be divided into subtypes that have different clinical features on the basis of their gene-expression profiles, (3) the gene-expression profile of an individual patient's tumor can reveal how aggressive (i.e., how lethal) the cancer is likely to be, (4) the gene-expression profile of an individual patient's tumor can provide clues as to which type of therapeutic strategy will be most likely to induce tumor regression. We can look at a few of these issues in more detail.

Figure 16.20 depicts the levels at which 50 different genes were found to be transcribed in two different types of leukemia, acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML). The genes used in this figure, which are named on the right side, represent those whose transcription showed the greatest difference between these two types of blood-cell cancers. Each column represents the results from a single patient with ALL or AML, so that different columns allow one to compare the similarities in expression of the gene from one patient to the next. Gene-expression levels are indicated from dark blue (lowest level) to dark red (highest level). The top half of the figure identifies genes that are transcribed at a much higher level in ALL cells, whereas the bottom half identifies genes that are transcribed to a much higher level in
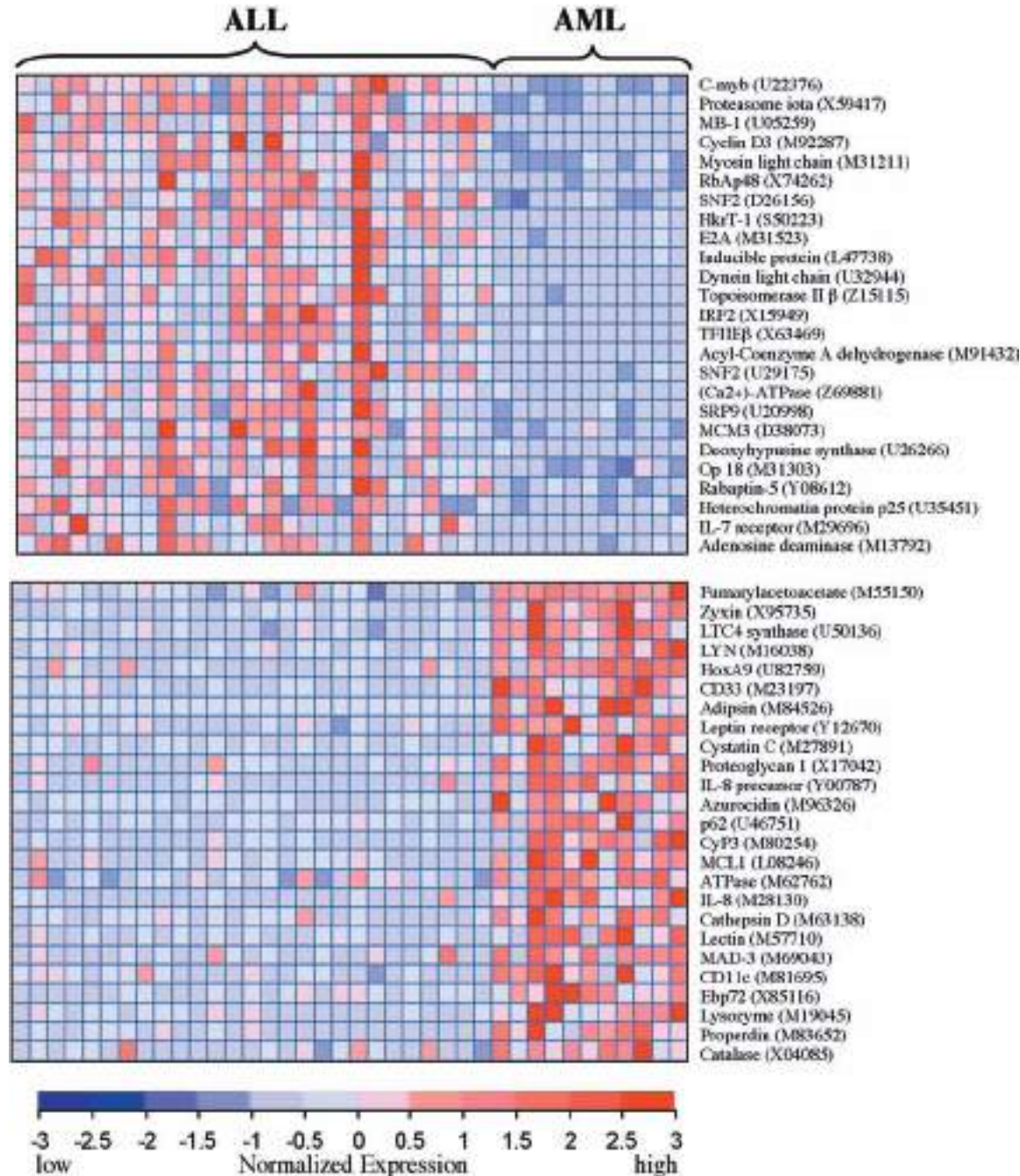
FIGURE 16.20　Gene-expression profiling that distinguishes two types of leukemia. Each row depicts the expression level of a single gene that is named to the right of the row. Altogether the levels of expression of 50 different genes are indicated. The color key is shown at the bottom of the figure, indicating the lowest level of expression is in dark blue and highest in dark red. Each column represents data from a different sample (patient). The columns on the left show the expression profiles from patients with acute lymphoblastic leukemia (ALL), whereas the columns on the right show the profiles from patients with acute myeloid leukemia (AML) (indicated by the brackets at the top). It is evident that the genes in the upper box are expressed at a much greater level in patients with ALL, whereas the genes in the lower box are expressed at a much greater level in patients with AML. (The genes selected for inclusion in this figure were chosen for the illustration because of these differences in expression between the two diseases.) (FROM T. R. GOLUB ET AL., SCIENCE 286:534, 1999; © COPYRIGHT 1999, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

AML cells. These studies make it evident that there are many differences in gene expression between different types of tumors. Some of these differences can be correlated with biological differences between the tumors; one, for example, is derived from a myeloid cell and the other from a lymphoid cell (see Figure 17.6). Most differences, however, cannot be explained. Why, for example, is the gene encoding catalase (the last gene on the list) expressed at a low level in ALL and a high level in AML? Even if these studies can't answer this question, they can provide cancer researchers with a list of genes to look at more closely as potential targets for therapeutic drugs.

The earlier a cancer is discovered, the more likely a person can be cured; this is one of the cardinal principles of cancer treatment. Yet a certain percentage of tumors will prove fatal, even if discovered and removed at an early stage. It is apparent, for example, that some early-stage breast cancers already contain cells capable of seeding the formation of secondary tumors (metastases) at distant locations, whereas others do not. These differences determine the prognosis of the patient. It was found in a landmark study in 2002 that the prognosis of a given breast cancer is likely to be revealed in the level of expression of approximately 70 genes out of the thousands that were studied in DNA microarrays (Figure 16.21). This finding has important clinical applications in guiding the treatment of breast cancer patients. Those patients with early-stage tumors that display a "poor prognosis" profile on the basis of their gene-expression pattern

(Figure 16.21*a*) can be treated aggressively with chemotherapy to maximize the chance that the formation of secondary tumors can be prevented. If gene-expression data is not considered, these individuals are not likely to receive any type of chemotherapy because there is no indication using conventional criteria that their tumor had spread. Conversely, those patients whose tumors exhibit a "good prognosis" profile might be spared the more debilitating chemotherapeutic drugs, even if their tumors appear more advanced (Figure 16.21*b*). In recent years, several companies have introduced laboratory tests (e.g., MammaPrint and Oncotype DX) to analyze gene-expression profiles of individual breast cancers in an attempt to help guide the best course of treatment for these patients. Even though these prognostic tests have been in widespread use for several years, their validity is still being evaluated in large clinical trials. It is hoped that gene-expression profiles can soon be used to improve the diagnosis and optimize treatment for individual patients with all types of cancer.

FIGURE 16.21 **The use of DNA microarray data in determining the choice of treatment.** Each graph shows the survival rate over time in breast cancer patients that had either a good-prognosis signature or a poor-prognosis signature based on the level of expression of 70 selected genes. The patients in *a* showed no visible evidence that the cancer had spread to nearby lymph nodes at the time of surgery. As indicated in the plot: (1) not all of these patients survive and (2) the likelihood of survival can be predicted to a large extent by the gene-expression profiles of their tumors. This allows physicians to treat those patients with a poor signature more aggressively than those with a good signature. The patients in *b* did show visible evidence of spread of cancer cells to nearby lymph nodes. As indicated in the plot, the likelihood of survival in this group can also be predicted by gene-expression data. Normally, all of the patients in this group would be treated very aggressively, which may not be necessary for those with a good signature. (FROM M. J. VAN DE VIJVER ET AL., NEW ENGL. J. MED. 347:2004,2005, 2002. COPYRIGHT © 2002, MASSACHUSETTS MEDICAL SOCIETY.)

REVIEW ?

1. Contrast a benign tumor and malignant tumor; tumor-suppressor genes and oncogenes; dominantly acting and recessively acting mutations; proto-oncogenes and oncogenes.

2. What is meant by the statement that cancer arises as the result of a genetic progression?

3. Why is p53 described as the "guardian of the genome"?

4. Give three mechanisms by which p53 acts to prevent a cell from becoming malignant.

5. How can DNA microarrays be used to determine the type of cancer a patient suffers from? How might they be used to optimize cancer treatment?

6. What types of proteins are encoded by proto-oncogenes, and how do mutations in each type of proto-oncogene cause a cell to become malignant?

16.4 NEW STRATEGIES FOR COMBATING CANCER

It is painfully evident that conventional approaches to combating cancer, namely, surgery, chemotherapy, and radiation, are not usually successful in curing a patient of metastatic cancer, that is, one that has already spread from a primary tumor. Because they kill large numbers of normal cells, along with the cancer cells, chemotherapy and radiation tend to have serious side effects, in addition to having limited curative value for most advanced cancers. There has been hope for decades that these "brute-force" strategies would be replaced by *targeted therapies*, based on our newly formed insights into the molecular basis of malignancy. There are several ways that a therapy can be considered to be "targeted": it can be targeted to attack only cancer cells, leaving normal cells unscathed; it can be targeted to a particular protein whose inactivation leaves the cancer cells unable to grow or survive; and/or it can be targeted to the cancer cells of a particular

patient based on their unique pattern of somatic mutations. Although the cure rate for most types of cancers has not improved significantly over the past 50 years, there is reason to believe that effective targeted therapies will become available to treat many of the common cancers in the foreseeable future. This optimism is based largely on the remarkable success that has been achieved with a small number of targeted treatments, which will be discussed in the following pages. Even though these successes have been scattered among a much larger number of failed prospective treatments, they demonstrate that the concept of targeted therapy is sound, which can be considered as "proof-of-principle." Just as important, they give both researchers and biotechnology companies the incentive to spend time and money to continue the pursuit of better cancer treatments.

The anticancer strategies to be discussed in the following sections can be divided into three groups: (1) those that depend on antibodies or immune cells to attack tumor cells, (2) those that inhibit the activity of cancer-promoting proteins, and (3) those that prevent the growth of blood vessels that nourish the tumor.

Immunotherapy

We have all heard or read about persons with metastatic cancer who were told they had only months to live, yet they defy the prognosis and remain alive and cancer-free years later. The best studied cases of such "spontaneous remissions" were recorded in the late 1800s by a New York physician named William Coley. Coley's interest in the subject began in 1891 when he came across the hospital records of a patient with an inoperable tumor of the neck, who had gone into remission after contracting a streptococcal infection beneath the skin. Coley located the patient and found him to be free of the cancer that had once threatened his life. Coley spent his remaining years trying to develop a bacterial extract that, when injected under the skin, would stimulate a patient's immune system to attack and destroy the malignancy. The work was not without its successes, particularly against certain uncommon soft-tissue sarcomas. Although the use of Coley's toxin, as it was later called, was never widely embraced, Coley's results confirmed anecdotal observations that the body has the capability to destroy a tumor, even after it has become well established. In recent years, two broad treatment strategies involving the immune system have been pursued: passive immunotherapy and active immunotherapy.

Passive immunotherapy is an approach that attempts to treat cancer patients by administering antibodies as therapeutic agents. These antibodies recognize and bind to specific proteins on the surface of the tumor cells being targeted. Once bound, the antibody orchestrates an attack on the cell carried out by other elements of the immune system. As discussed in Section 18.19, the production of monoclonal antibodies capable of binding to particular target antigens was initially developed in the mid-1970s. During the next 20 years or so, attempts to use these proteins as therapeutic agents were foiled for a number of reasons. Most importantly, these antibodies were produced by mouse cells and encoded by mouse genes. As a result, the antibodies were recognized as foreign and cleared from the bloodstream before they had a chance to work. In subsequent efforts, investigators were able to produce "humanized antibodies," which are antibodies that are largely human proteins except for a relatively small part that recognizes the antigen, which remains "mousey." In the past few years, researchers have been able to produce antibodies that have a completely human amino acid sequence. In one approach, mice have been genetically engineered so that their immune system produces fully human antibody molecules.

At the present time 20 or so monoclonal antibodies have been approved to treat cancer and other medical conditions. More than 100 others are being tested. Herceptin is a humanized antibody directed against a cell-surface receptor (Her2) that binds a growth factor that stimulates the proliferation of breast cancer cells. Herceptin is thought to inhibit activation of the receptor by the growth factor and stimulate receptor internalization (page 305). Approximately 25 percent of breast cancers are composed of cells that overexpress the *HER2* gene, which causes these cells to be especially sensitive to growth factor stimulation. Up until the development of Herceptin, patients whose tumors overexpressed HER2 had a very poor prognosis. Their prognosis is now greatly improved. For example, one study of 3000 women with early-stage breast cancer reported that Herceptin reduced the chance of recurrence of the disease by about 50 percent in a four-year period. To date, the most effective humanized antibody is Rituxan, which was approved in 1997 for treatment of non-Hodgkin's B-cell lymphoma. Rituxan binds to a cell-surface protein (CD20) that is present on the malignant B cells in approximately 95 percent of the cases of this disease. Binding of the antibody to the CD20 protein inhibits cell growth and induces the cells to undergo apoptosis. Introduction of this antibody has completely reversed the prospects for people with this particular cancer. People that would have once had a very grim prognosis now stand an excellent chance for complete remission of their disease.

Over the past few years, a number of fully human antibodies have been tested in clinical trials against various types of cancers. One of these, called Vectibix, which is directed against the EGF receptor, has been approved as a single-agent treatment of EGFR-expressing metastatic colon cancer. Because it is a human protein, Vectibix remains in the circulation long enough to be administered once every other week. Another human monoclonal antibody, Arzerra, is likely to be approved for treatment of chronic lymphocytic leukemia. In addition, a new generation of antibodies are being developed that contain a radioactive atom or a toxic compound conjugated to the antibody molecule. According to plan, the antibody targets the complex to the cancer cell, and the associated atom or compound kills the targeted cell. At the time of this writing, two radioactively labeled anti-CD20 antibodies (Zevalin and Bexxar) have been approved for the treatment of non-Hodgkin's B-cell lymphoma, and a toxic drug-linked antibody (Mylotarg) has been approved to treat acute myeloid leukemia.

Active (or *adoptive*) *immunotherapy* is an approach that tries to get a person's own immune system more involved in the fight against malignant cells. The immune system has evolved to recognize and destroy foreign materials, but cancers are derived from a person's own cells. Although many tumor cells do contain proteins that are not expressed by their

normal counterparts, or mutated proteins (e.g., Ras) that are different from those present in normal cells, they are still basically host proteins present in host cells. As a result, the immune system typically fails to recognize these proteins as "inappropriate." Even if a person does possess immune cells that recognize certain tumor-associated antigens, tumors evolve a number of mechanisms that allow them to escape immune destruction. Many different strategies have been formulated to overcome these hurdles and artificially stimulate the immune system to mount a more vigorous response against tumor cells. In most studies, immune cells (typically dendritic cells) are isolated from the patient, stimulated in one way or another in vitro, allowed to proliferate in culture, and then reintroduced into the patient. In some studies, the isolated immune cells are genetically modified before proliferation to enhance their tumor-attacking potential. For many years, clinical trials using these "cancer vaccines" were disappointing, but recent publications have provided reason for cautious optimism. In many of these studies, a significant minority of patients have exhibited a positive response to the treatment, which means that their tumors have at least shrunk in size or extent and their expected time of survival has increased significantly. Because these types of treatments are personalized to the individual patient, they are likely to be both time-consuming to prepare and very costly. The vaccine that has reached the most advanced stage of testing is called Provenge, which is designed to treat patients with advanced prostate cancer that is no longer responsive to hormone treatment. Immune cells are isolated from the blood, exposed to a prostate-specific protein (prostatic acid phosphatase) that is expressed by this cancer, and then reinfused into the body. Provenge was rejected once by the FDA in 2007, a decision that generated great controversy. At the time of this writing, at least two Phase III studies have been completed, but there has been some difficulty reconciling the results of the studies, and the regulatory agency is deliberating the matter. Another personalized cancer vaccine, DCVax, has shown remarkable promise in Phase I and II trials in the treatment of brain cancer and is currently being tested in larger trials. Whether any of these vaccines will ultimately prove to be effective cancer therapies remains to be seen. The ultimate goal of cancer immunotherapy is to vaccinate people with antigens that would prevent them from ever developing life-threatening cancers.

Inhibiting the Activity of Cancer-Promoting Proteins

Cancer cells behave as they do because they contain proteins that are either present at abnormal concentration or display abnormal activity. A number of these proteins were illustrated in Figure 16.17. In many cases, the growth and/or survival of tumor cells is dependent on the continued activity of one or more of these deviant proteins. This dependency is known as "oncogene addiction." If the activity of one of these proteins can be selectively blocked, it should be possible to kill the entire population of malignant cells. With this goal in mind, researchers have synthesized a virtual arsenal of small-molecular-weight compounds that inhibit the activity of cancer-promoting proteins. Some of these drugs were custom-designed to inhibit a particular protein, whereas others were identified by randomly screening large numbers of compounds that have been synthesized by pharmaceutical companies. Once a protein-inhibiting compound has been identified, it is typically tested for effectiveness against a panel of approximately 60 different types of cultured cells, each originally isolated from a different human cancer. Success against cultured cells generally leads to tests of the agent against mice carrying human tumor transplants (xenografts). A list of many of the agents that have been tested in clinical trials is shown in Table 16.3. Although numerous compounds

TABLE 16.3 Examples of Small-Molecule Targeted Therapies That Have Either Been FDA Approved or Are Being Tested

Drug	Target	Mechanism of action
Gleevec	BCR-ABL, KIT, PDGFR	Tyrosine kinase inhibitor
Iressa	EGFR	Tyrosine kinase inhibitor
Tarceva	EGFR	Tyrosine kinase inhibitor
Sutent	VEGFR, PDGFR, KIT	Tyrosine kinase inhibitor
Tykerb	EGFR, HER2	Tyrosine kinase inhibitor
Nexavar	BRAF, EGFR, EGFR	Kinase inhibitor
Velcade	proteasome	Inhibits protein degradation
Zolinza	HDACs	Inhibits histone acetylation (epigenetic effect?)
Torisel	mTOR	Blocks cell-survival pathway
Tamoxifen, Raloxifene	Estrogen receptor	Blocks estrogen action
Arimidex, Aromasin	Aromatase	Inhibits synthesis of estrogen
Genasense	BCL-2	Inhibits synthesis of this proapoptotic protein
ABT-737	BCL-X_L	Inhibits this proapoptotic protein
Geldanamycin	HSP90	Inhibits this molecular chaperone
Nutlins, RITA	p53	Inhibits p53-MDM2 interaction
PRIMA-1	p53	Restores activity of mutant p53
PX-478	HIF-1	Inhibits this transcription factor that is activated by hypoxia
BSI-201	PARP-1	Inhibits this enzyme involved in DNA repair
Decitabine	DNMT	Inhibits DNA methylation

on this list have shown some promise in halting the growth of various types of tumors, one compound has had unparalleled success in clinical trials on patients with chronic myelogenous leukemia (CML).

It was noted earlier that certain types of cancers are caused by specific chromosome translocations. CML is caused by a translocation that brings a proto-oncogene (*ABL*) in contact with another gene (*BCR*) to form a chimeric gene (*BCR–ABL*). Blood-forming cells that carry this translocation express a high level of Abl tyrosine kinase activity, which causes the cells to proliferate uncontrollably and initiate the process of tumorigenesis. As discussed on page 73, a compound called Gleevec was identified that selectively inhibits Abl kinase by binding to the inactive form of the protein and preventing its phosphorylation by another kinase, which is required for Abl activation. Initial clinical trials on Gleevec were remarkably successful, causing nearly all CML patients who received the drug at sufficiently high dose to go into remission. These studies confirmed the notion that elimination of a single required oncogene product could stop the growth of a human cancer. The drug was rapidly approved and has been in use for several years. Patients have to continue to take the drug to remain in remission, and many of them, especially those who begin treatment at an advanced stage, eventually develop drug resistance. Most cases of resistance result from mutations in the *ABL* portion of the fusion gene. This has prompted the development of a second generation of targeted inhibitors that remain active against most of the mutated forms of the ABL kinase (see Figure 2.51*d*). These drugs appear to be effective in treating Gleevec-resistant cases of CML and suggest that the ideal drug regimen may consist of a cocktail of several different inhibitors that target different parts of the same protein, ensuring that drug-resistant mutants will not emerge.

It was hoped that Gleevec would be followed rapidly by numerous other highly effective protein-inhibiting drugs. Although a number of protein-targeting, small-molecule inhibitors have shown modest success in clinical trials and have been approved by the FDA, and hundreds of others are being tested in the clinic, none of them studied to date have been able to fully stop the growth of any of the common solid cancers, such as those of the breast, lung, prostate, or pancreas. The reasons for the difficulty in treating these tumors are not entirely clear. One reason may be that these tumors are more complex genetically and the cells are not as dependent on a single oncogene product and aberrant signaling pathway as are the blood-cell cancers. Another reason may be that only a fraction of the patients with a particular type of tumor are sensitive to a particular drug. This was suggested to be the case for Iressa, an inhibitor of the tyrosine kinase of the EGF receptor (EGFR). Iressa was originally tested on lung cancer patients because these tumors were known to exhibit high levels of EGFR. Initial clinical trials found that approximately 10 percent of patients in the United States and 30 percent of Japanese patients responded positively to the drug, whereas the remaining population was unaffected. Further analysis revealed that responders and nonresponders had mutations that affected different regions of the EGFR protein. This finding verifies the notion that targeted cancer therapies, as well as nontargeted chemotherapies, will ultimately have to be tailored to fit the specific genetic modifications present in the tumors of individual patients.

We have focused in this section upon targeted therapies that inhibit proteins that are either abnormal themselves or are abnormally expressed in cancer cells. But there may also be proteins that have a normal structure and expression, yet for some reason play an important role in the lives of cancer cells. Inhibitors that target such proteins might have considerable potential as therapeutic agents in the treatment of cancer. We have seen in this chapter how cancer cells contain mutations in a wide variety of genes, many of which lead to the overall inhibition of certain metabolic or signaling pathways. While this may help promote the growth and survival of these cancer cells, it also may cause them to become more dependent than normal cells upon other pathways that continue to operate in a normal fashion. Recent announcements hailing the promise of PARP-1 inhibitors in the treatment of several types of cancer provide an example of this type of reasoning. PARP (which is an acronym for poly(ADP-ribose) polymerase) is a poorly understood enzyme that is involved in numerous processes involving DNA metabolism, including DNA repair. PARP-1 inhibitors (e.g., olaparib and BSI-201) have shown particular promise in the treatment of breast and ovarian cancers that exhibit deficiencies in BRCA1 or BRCA2. As discussed on page 663, BRCA proteins are involved in DNA repair and it is reasoned that such cancer cells are more dependent than normal cells upon other DNA repair pathways, including those that require PARP-1. When PARP-1 is inhibited in BRCA-deficient tumor cells, certain types of DNA damage cannot be repaired, causing the cells to die by apoptosis. This type of treatment is based on a strategy of "synthetic lethality," which suggests that mutations or inhibition of only one protein (e.g., BRCA or PARP) has no effect on cell viability but that mutations and/or inhibition of two different proteins leaves the cell unable to carry out one or more essential functions.

Another reason for the failure to develop more effective targeted therapies may be that the agents are not targeting the appropriate cells within the tumor. This possibility requires further explanation but raises an important issue concerning both the basic biology of cancer and its treatment. Throughout this chapter, we have considered a tumor to be a mass of relatively homogeneous cells. When viewed in this way, all of the cells in a tumor are capable of unlimited proliferation and all of the cells have the opportunity to evolve into a more malignant phenotype as the result of ongoing genetic change. In recent years, a new concept has emerged, which suggests that while most of the cells of a tumor may be dividing at a rapid rate, they have relatively limited long-term potential to sustain the primary tumor or initiate a new secondary tumor. Instead, a relatively small number of cells scattered throughout the tumor are responsible for maintaining the tumor and promoting its spread. These "special" cells are known as *cancer stem cells*, and there is considerable experimental evidence for their existence in leukemias, brain tumors, breast tumors, and other cancers. There have also been a number of recent experiments

that argue against the existence of these rare cancer stem cells, and the issue is the focus of current debate.

The concept of the cancer stem cell is raised at this point in the chapter because it has important consequences for cancer therapy. If it is true that only a small subpopulation of cells in a tumor have the capability of continuing the life of that tumor, then developing drugs that rapidly kill off the bulk of a tumor mass but spare the cancer stem cells is doomed ultimately to fail. While there are efforts underway to identify cancer stem cells in various types of tumors and learn more about their properties, these new views are just beginning to have an impact on drug development.

Inhibiting the Formation of New Blood Vessels (Angiogenesis)

As a tumor grows in size, it stimulates the formation of new blood vessels, a process termed *angiogenesis* (Figure 16.22). Blood vessels are required to deliver nutrients and oxygen to the rapidly growing tumor cells and to remove their waste products. Blood vessels also provide the conduits for cancer cells to spread to other sites in the body. In 1971, Judah Folkman of Harvard University suggested that solid tumors might be destroyed by inhibiting their ability to form new blood vessels. After a quarter of a century of relative obscurity, this idea has become an approved anticancer strategy.

Cancer cells promote angiogenesis by secreting growth factors, such as VEGF, that act on the endothelial cells of surrounding blood vessels, stimulating them to proliferate and develop into new vessels. Just as there are stimulants of angiogenesis, there are also inhibitors. Naturally occurring inhibitors, such as endostatin and thrombospondin, have been identified, but most angiogenic inhibitors have been developed by biotechnology companies. These include antibodies and synthetic compounds directed against integrins, growth factors, and growth-factor receptors. Preclinical studies on mice and rats suggested that angiogenic inhibitors might be effective in stopping tumor growth. Most importantly, tumors treated with these inhibitors did not become resistant to repeated drug application. Tumor cells become resistant to the usual chemotherapeutic agents because the cells are genetically unstable and can evolve into resistant forms. Inhibitors of angiogenesis, however, target normal, genetically stable endothelial cells, which continue to respond to the presence of these agents. There are several other reasons that make angiogenesis inhibitors a promising therapy: they should not interfere with normal physiologic activities, because angiogenesis is not a required activity in a mature adult; they act on cells lining the bloodstream, which are directly accessible to blood-borne drugs; and they should be broadly effective against many different types of tumors, which are presumed to employ the same mechanism of angiogenesis.

Inhibiting angiogenesis in human tumors is not as easily accomplished as might have been expected based on studies with mice. To date, the most promising results have been obtained with a humanized antibody (called Avastin) that is directed against VEGF, the endothelial cell growth factor that is overexpressed in most solid tumors. Avastin blocks VEGF from binding to and activating its receptor, VEGFR. FDA approval was based on clinical trials demonstrating that Avastin, in combination with standard chemotherapy, could prolong the lives of patients with metastatic colorectal cancer for several months. Although far from representing a cure, this is a significant achievement in patients with advanced colorectal cancer (and also lung cancer). Two other antiangiogenic drugs—the small-molecular-weight kinase inhibitors Sutent and Nexavar—have since been approved for the treatment of specific types of advanced cancers, but none of these agents have a major impact on the course of the disease and their mechanism of action has been debated. According to one hypothesis, choking off the growth of blood vessels in a tumor might even have a negative impact by creating a more O_2-deficient environment for the tumor cells and driving them to "seek out" other sites in the body.
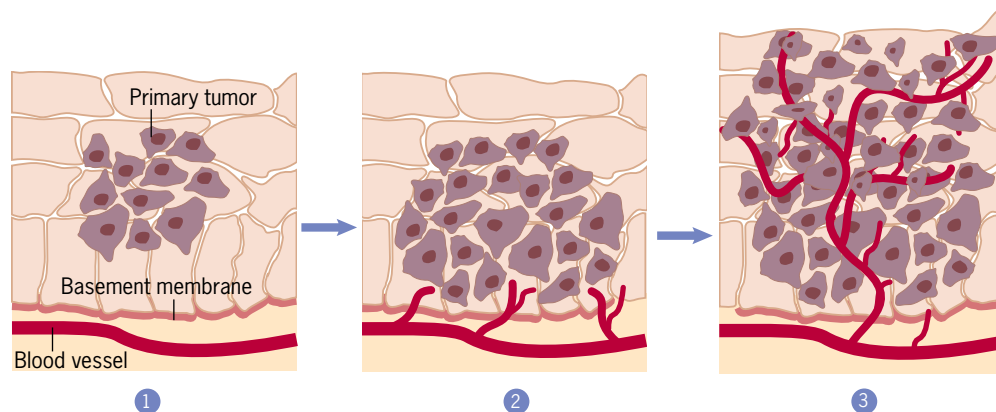
FIGURE 16.22 Angiogenesis and tumor growth. Steps in the vascularization of a primary tumor. In step 1, the tumor proliferates to form a small mass of cells. As long as it is avascular (without blood vessels), the tumor remains very small (1–2 mm). In step 2, the tumor mass has produced angiogenic factors that stimulate the endothelial cells of nearby vessels to grow out toward the tumor cells. In step 3, the tumor has become vascularized and is now capable of unlimited growth. (AFTER B. R. ZETTER, WITH PERMISSION FROM THE ANNUAL REVIEW OF MEDICINE, VOLUME 49; © 1998, BY ANNUAL REVIEWS WWW.ANNUALREVIEWS.ORG.)

For the forseeable future, the best anticancer strategy is early detection. A number of screening procedures are currently in use, including mammography for detecting breast cancer, Pap smears for detecting cervical cancer, PSA determinations for detecting prostate cancer, and colonoscopy for detecting colorectal cancer. It is hoped that advances in proteomics will evenually lead to the development of new screening tests based on the relative levels of various proteins present in the blood. This approach was discussed in some detail on page 70. There are other blood-borne biological indicators (biomarkers) that could reveal the presence of cancer,

including mutant DNA, abnormal carbohydrates, distinctive metabolites, and the presence of cancer cells themselves.

Advances in genomics are also likely to contribute to the screening process by informing each of us of the types of cancer we are most likely to fall victim to. Genomic screening tests are already available for persons whose family history suggests that they might carry mutations in the *BRCA1* gene and, consequently, might be at risk for development of breast cancer. The earlier a cancer is discovered, the greater the chance of survival. Consequently, these screening procedures could have a significant impact in lowering the death rates from cancer.

EXPERIMENTAL PATHWAYS

The Discovery of Oncogenes

In 1911, Peyton Rous of The Rockefeller Institute for Medical Research published a paper that was less than one page in length (it shared the page with a note on the treatment of syphilis) and had virtually no impact on the scientific community. Yet this paper reported one of the most farsighted observations in the field of cell and molecular biology.[1] Rous had been working with a chicken sarcoma that could be propagated from one hen to another by inoculating a host of the same strain with pieces of tumor tissue. In this paper, Rous described a series of experiments that strongly suggested that the tumor could be transmitted from one animal to another by a "filterable virus," which is a term that had been coined a decade or so earlier to describe pathogenic agents that were small enough to pass through filters that were impermeable to bacteria.

In his experiments, Rous removed the tumors from the breasts of hens, ground the cells in a mortar with sterile sand, centrifuged the particulate material into a pellet, removed the supernatant, and forced the supernatant fluid through filters of varying porosity including those small enough to prevent the passage of bacteria. He then injected the filtrate into the breast muscle of a recipient hen and found that a significant percentage of the injected animals developed the tumor.

The virus discovered by Rous in 1911 is an RNA-containing virus. By the end of the 1960s, similar viruses were found to be associated with mammary tumors and leukemias in rodents and cats. Certain strains of mice had been bred that developed specific tumors with very high frequency. RNA-containing viral particles could be seen within the tumor cells and also budding from the cell surface, as shown in the micrograph of Figure 1. It was apparent that the gene(s) causing tumors in these inbred strains are transmitted vertically, that is, through the fertilized egg from mother to offspring, so that the adults of each generation invariably develop the tumor. These studies provided evidence that the viral genome can be inherited through the gametes and subsequently transmitted from cell to cell by means of mitosis without having any obvious effect on the behavior of the cells. The presence of inherited viral genomes is not a peculiarity of inbred laboratory strains, because it was shown that wild (feral) mice treated with chemical carcinogens develop tumors that often contain the same antigens characteristic of RNA tumor viruses and that exhibit virus particles under the electron microscope.

One of the major questions concerning the vertical transmission of RNA tumor viruses was whether the viral genome is passed from parents to progeny as free RNA molecules or is somehow integrated

into the DNA of the host cell. Evidence indicated that infection and transformation by these viruses required the synthesis of DNA. Howard Temin of the University of Wisconsin suggested that the replication of RNA tumor viruses occurs by means of a DNA intermediate—a provirus—which then serves as a template for the synthesis of viral RNA. But this model requires a unique enzyme—an RNA-dependent DNA polymerase—which had never been found in any type of cell. Then in 1970, an enzyme having this activity was discovered independently by David Baltimore of the Massachusetts Institute of Technology and by Temin and Satoshi Mizutani.[2,3]

Baltimore examined the virions (the mature viral particles) from two RNA tumor viruses, Rauscher mouse leukemia virus (R-MLV)

FIGURE 1 Electron micrograph of a Friend mouse leukemia virus budding from the surface of a cultured leukemic cell. (COURTESY OF E. DE HARVEN.)

and Rous sarcoma virus (RSV). A preparation of purified virus was incubated under conditions that would promote the activity of a DNA polymerase, including Mg^{2+} (or Mn^{2+}), NaCl, dithiothreitol (which prevents the —SH groups of the enzyme from becoming oxidized), and all four deoxyribonucleoside triphosphates, one of which (TTP) was radioactively labeled. Under these conditions, the preparation incorporated the labeled DNA precursor into an acid-insoluble product that exhibited the properties of DNA (Table 1). As is characteristic of DNA, the reaction product was rendered acid soluble (indicating that it had been converted to low-molecular-weight products) by treatment with pancreatic deoxyribonuclease or micrococcal nuclease, but it was unaffected by pancreatic ribonuclease or by alkaline hydrolysis (to which RNA is sensitive; Table 1). The DNA-polymerizing enzyme was found to co-sediment with the mature virus particles, suggesting that it was part of the virion itself and not an enzyme donated by the host cell. Although the product was insensitive to treatment with pancreatic ribonuclease, the template was very sensitive to this enzyme (Figure 2), particularly if the virions were pretreated with the ribonuclease prior to addition of the other components of the reaction mixture (Figure 2, curve 4). These results strengthened the suggestion that the viral RNA was providing the template for synthesis of a DNA copy, which presumably served as a template for the synthesis of viral mRNAs required for infection and transformation. Not only did these experiments suggest that cellular transformation by RNA tumor viruses proceeded through a DNA intermediate, they also overturned the long-standing concept originally proposed by Francis Crick and known as the Central Dogma, which stated that information in a cell always flowed from DNA to RNA to protein. The RNA-dependent DNA polymerase became known as reverse transcriptase.

During the 1970s, attention turned to the identification of the genes carried by tumor viruses that were responsible for transformation and the mechanism of action of the gene products. Evidence from genetic analyses indicated that mutant strains of viruses could be isolated that retained the ability to grow in host cells, but were unable to transform the cell into one exhibiting malignant properties.[4] Thus, the capacity to transform a cell resided in a restricted portion of the viral genome.

These findings set the stage for a series of papers by Harold Varmus, J. Michael Bishop, Dominique Stehelin, and their co-workers at the University of California, San Francisco. These researchers began by isolating mutant strains of the avian sarcoma virus (ASV) carrying deletions of 10 to 20 percent of the genome that render the virus unable to induce sarcomas in chickens or to transform fibroblasts in culture. The gene responsible for transformation, which is
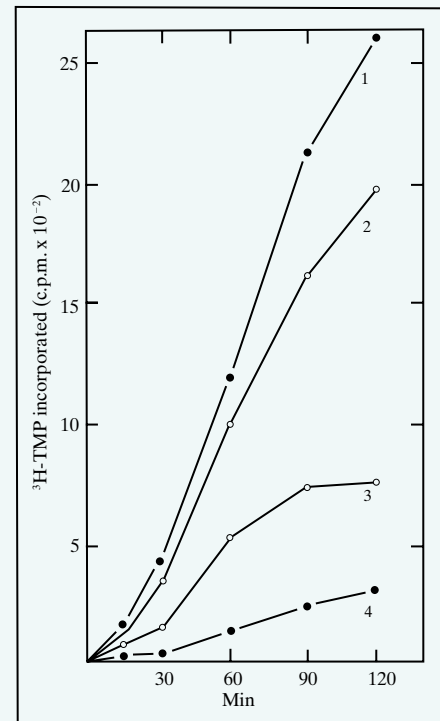
FIGURE 2 Incorporation of radioactivity from [³H]TTP into an acid-insoluble precipitate by the Rauscher murine leukemia virus DNA polymerase in the presence and absence of ribonuclease. (*Note:* the labeled TTP precursor is converted to TMP as it is incorporated into DNA.) Curve 1, no added ribonuclease; curve 2, preincubated without added ribonuclease for 20 minutes before addition of [³H]TTP; curve 3, ribonuclease added to the reaction mixture; curve 4, preincubated with ribonuclease before addition of [³H]TTP. (REPRINTED WITH PERMISSION FROM D. BALTIMORE, NATURE 226:1210, 1970. © COPYRIGHT 1970, MACMILLAN MAGAZINES LTD.)

missing in these mutants, was referred to as *src* (for sarcoma). To isolate the DNA corresponding to the deleted regions of these mutants, which presumably carry the genes required for transformation, the following experimental strategy was adopted.[5] RNA from the genomes of complete (oncogenic) virions was used as a template for the formation of a radioactively labeled, single-stranded, complementary DNA (cDNA) using reverse transcriptase. The labeled cDNA (which is present as fragments) was then hybridized to RNA obtained from one of the deletion mutants. Those DNA fragments that failed to hybridize to the RNA represent portions of the genome that had been deleted from the transformation-defective mutant and thus were presumed to contain the gene required by the virus to cause transformation. DNA fragments that did not hybridize to the RNA were separated from those that were part of DNA–RNA hybrids by column chromatography. Using this basic strategy, a DNA sequence referred to as $cDNA_{sarc}$ was isolated, which corresponded to approximately 16 percent of the viral genome (1600 nucleotides out of a total genomic length of 10,000 nucleotides).

Once isolated, $cDNA_{sarc}$ proved to be a very useful probe. It was first shown that this labeled cDNA hybridizes to DNA extracted from cells of a variety of avian species (chicken, turkey, quail, duck, and emu), indicating that the cellular genomes of these birds contain a DNA sequence that is closely related to *src*.[6]

TABLE 1 Characterization of the Polymerase Product

Expt.	Treatment	Acid-insoluble radioactivity	Percentage undigested product
1	Untreated	1,425	(100)
	20 μg deoxyribonuclease	125	9
	20 μg micrococcal nuclease	69	5
	20 μg ribonuclease	1,361	96
2	Untreated	1,644	(100)
	NaOH hydrolyzed	1,684	100

Source: D. Baltimore, reprinted with permission from *Nature* 226:1210, 1970. © Copyright 1970, Macmillan Magazines, Ltd.

These findings provided the first strong evidence that a gene carried by a tumor virus that causes cell transformation is actually present in the DNA of normal (uninfected) cells and thus is presumably a part of the cells' normal genome. These results indicated that the transforming genes of the viral genome (the oncogenes) are not true viral genes, but rather are cellular genes that were picked up by RNA tumor viruses during a previous infection. Possession of this cell-derived gene apparently endows the virus with the power to transform the very same cells in which this gene is normally found. The fact that the *src* sequence is present in all of the avian species tested suggests that the sequence has been conserved during avian evolution and, thus, is presumed to govern a basic activity of normal cells. In a subsequent study, it was found that cDNA$_{sarc}$ binds to DNA from all vertebrate classes, including mammals, but not to the DNA from sea urchins, fruit flies, or bacteria. Based on these results it was possible to conclude that the *src* gene is not only present in the RNA of the ASV genome and the genome of the chicken cells it can infect, but a homologous gene is also present in the DNA of distantly related vertebrates, suggesting that it plays some critical function in the cells of all vertebrates.[7]

These findings raised numerous questions; foremost among these were (1) what is the function of the *src* gene product, and (2) how does the presence of the viral *src* gene (referred to as v-*src*) alter the behavior of a normal cell that already possesses a copy of the cellular gene (referred to as c-*src*)?

The product of the *src* gene was initially identified by Ray Erikson and co-workers at the University of Colorado by two independent procedures: (1) precipitation of the protein from extracts of transformed cells by antibodies prepared from RSV-infected animals, and (2) synthesis of the protein in a cell-free protein-synthesizing system using the isolated viral gene as a template. Using these procedures, the *src* gene product was found to be a protein of 60,000 daltons, which they named pp60^{src}.[8] When pp60^{src} was incubated with [^{32}P]ATP, radioactive phosphate groups were transferred to the heavy chains of the associated antibody (IgG) molecules used in the immunoprecipitation. This finding suggested that the *src* gene codes for an enzyme that possesses protein kinase activity.[9] When cells infected with ASV were fixed, sectioned, and incubated with ferritin-labeled antibodies against pp60^{src}, the antibodies were found to be localized on the inner surface of the plasma membrane, suggesting a concentration of the *src* gene product in this part of the cell (Figure 3).[10]

These were the first studies to elucidate the function of an oncogene. A protein kinase is the type of gene product that might be expected to have potential transforming activity, because it can regulate the activities of numerous other proteins, each of which might serve a critical function in one or another activity related to cell growth. Further analysis of the role of the *src* gene product turned up an unexpected finding. Unlike all the other protein kinases whose function had been studied, pp60^{src} transferred phosphate groups to tyrosine residues on the substrate protein rather than to serine or threonine residues.[11] The existence of phosphorylated tyrosine residues had escaped previous detection because phosphorylated serine and threonine residues are approximately 3000 times more abundant in cells than phosphotyrosine, and because phosphothreonine and phosphotyrosine residues are difficult to separate from one another by traditional electrophoretic procedures. Not only did the product of the viral *src* gene (v-*src*) code for a tyrosine protein kinase, so too did c-*src*, the cellular version of the gene. However, the number of phosphorylated tyrosine residues in proteins of RSV-transformed cells was approximately eight times higher than that of

0.1 μm

FIGURE 3 Electron micrograph of a section through a pair of adjacent fibroblasts that had been treated with ferritin-labeled antibodies against the pp60^{src} protein. The protein is localized (as revealed by the dense ferritin granules) at the plasma membrane of the cell and is particularly concentrated at the sites of gap junctions. (FROM MARK C. WILLINGHAM, GILBERT JAY, AND IRA PASTAN, CELL 18:128, 1979, BY PERMISSION OF CELL PRESS.)

control cells. This finding suggested that the viral version of the gene may induce transformation because it functions at a higher level of activity than the cellular version.

The results from the study of RSV provided preliminary evidence that an increased activity of an oncogene product could be a key to converting a normal cell into a malignant cell. Evidence soon became available that the malignant phenotype could also be induced by an oncogene that contained an altered nucleotide sequence. A key initial study was conducted by Robert Weinberg and his col-

leagues at the Massachusetts Institute of Technology using the technique of DNA transfection.[12]

Weinberg began the studies by obtaining 15 different malignant cell lines that were derived from mouse cells that had been treated with a carcinogenic chemical. Thus these cells had been made malignant without exposing them to viruses. The DNA from each of these cell lines was extracted and used to transfect a type of nonmalignant mouse fibroblast called an NIH3T3 cell. NIH3T3 cells were selected for these experiments because they take up exogenous DNA with high efficiency and they are readily transformed into malignant cells in culture. After transfection with DNA from the tumor cells, the fibroblasts were grown in vitro, and the cultures were screened for the formation of clumps (foci) that contained cells that had been transformed by the added DNA. Of the 15 cell lines tested, five of them yielded DNA that could transform the recipient NIH3T3 cells. DNA from normal cells lacked this capability. These results demonstrated that carcinogenic chemicals produced alterations in the nucleotide sequences of genes that gave the altered genes the ability to transform other cells. Thus cellular genes could be converted into oncogenes in two different ways: as the result of becoming incorporated into the genome of a virus or by becoming altered by carcinogenic chemicals.

Up to this point, virtually all of the studies on cancer-causing genes had been conducted in mice, chickens, or other organisms whose cells were highly susceptible to transformation. In 1981, attention turned to human cancer when it was shown that DNA isolated from human tumor cells can transform mouse NIH3T3 cells following transfection.[13] Of 26 different human tumors that were tested in this study, two provided DNA that was capable of transforming mouse fibroblasts. In both cases, the DNA had been extracted from cell lines taken from a bladder carcinoma (identified as EJ and J82). Extensive efforts were undertaken to determine if the genes had been derived from a tumor virus, but no evidence of viral DNA was detected in these cells. These results provided the first evidence that some human cancer cells contain an activated oncogene that can be transmitted to other cells, causing their transformation.

The finding that cancer can be transmitted from one cell to another by DNA fragments provided a basis for determining which genes in a cell, when activated by mutation or some other mechanism, are responsible for causing the cell to become malignant. To make this determination, it was necessary to isolate the DNA that was taken up by cells, causing their transformation. Once the foreign DNA responsible for transformation had been isolated, it could be analyzed for the presence of cancer-causing alleles. Within two months of one another in 1982, three different laboratories reported the isolation and cloning of an unidentified gene from human bladder carcinoma cells that can transform mouse NIH3T3 fibroblasts.[14–16]

Once the transforming gene from human bladder cancer cells had been isolated and cloned, the next step was to determine if this gene bears any relationship to the oncogenes carried by RNA tumor viruses. Once again, within two months of one another, three papers appeared from different laboratories reporting similar results.[17–19] All three showed that the oncogene from human bladder carcinomas that transforms NIH3T3 cells is the same oncogene (named *ras*) that is carried by the Harvey sarcoma virus, which is a rat RNA tumor virus. Preliminary comparisons of the two versions of *ras*—the viral version and its cellular homologue—failed to show any differences, indicating that the two genes are either very similar or identical. These findings suggested that cancers that develop spontaneously in the human population are caused by a genetic alter-

ation that is similar to the changes in cells that have been virally transformed in the laboratory. It is important to note that the types of cancers induced by the Harvey sarcoma virus (sarcomas and erythroleukemias) are quite different from the bladder tumors, which have an epithelial origin. This was the first indication that alterations in the same human gene—*RAS*—can cause a wide range of different tumors.

By the end of 1982, three more papers from different laboratories reported on the precise changes in the human *RAS* gene that leads to its activation as an oncogene.[20–22] Once the section of the large DNA fragment that is responsible for causing transformation was pinned down, nucleotide sequence analysis indicated that the DNA from the malignant bladder cells is activated as a result of a single base substitution within the coding region of the gene. Remarkably, cells of both human bladder carcinomas studied (identified as EJ and T24) contain DNA with precisely the same alteration: a guanine-containing nucleotide at a specific site in the DNA of the proto-oncogene had been converted to a thymidine in the activated oncogene. This base substitution results in the replacement of a valine for a glycine as the twelfth amino acid residue of the polypeptide.

Determination of the nucleotide sequence of the v-*ras* gene carried by the Harvey sarcoma virus revealed an alteration in base sequence that affected precisely the same codon found to be modified in the DNA of the human bladder carcinomas. The change in the viral gene substitutes an arginine for the normal glycine. It was apparent that this particular glycine residue plays a critical role in the structure and function of this protein. It is interesting to note that the human *RAS* gene is a proto-oncogene that, like *SRC*, can be activated by linkage to a viral promoter. Thus *RAS* can be activated to induce transformation by two totally different pathways: either by increasing its expression or by altering the amino acid sequence of its encoded polypeptide.

The research described in this Experimental Pathway provided a great leap forward in our understanding of the genetic basis of malignant transformation. Much of the initial research on RNA tumor viruses stemmed from the belief that these agents may be an important causal agent in the development of human cancer. The search for viruses as a cause of cancer led to the discovery of the oncogene, which led to the realization that the oncogene is a cellular sequence that is acquired by the virus, which ultimately led to the discovery that an oncogene can cause cancer without the involvement of a viral genome. Thus tumor viruses, which are not themselves directly involved in most human cancers, have provided the necessary window through which we can view our own genetic inheritance for the presence of information that can lead to our own undoing.

References

1. ROUS, P. 1911. Transmission of a malignant new growth by means of a cell-free filtrate. *J. Am. Med. Assoc.* 56:198.

2. BALTIMORE, D. 1970. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 226:1209–1211.

3. TEMIN, H. & MIZUTANI, S. 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 226:1211–1213.

4. MARTIN, G. S. 1970. Rous sarcoma virus: A function required for the maintenance of the transformed state. *Nature* 227:1021–1023.

5. STEHELIN, D. ET AL. 1976. Purification of DNA complementary to nucleotide sequences required for neoplastic transformation of fibroblasts by avian sarcoma viruses. *J. Mol. Biol.* 101:349–365.

6. STEHELIN, D. ET AL. 1976. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* 260:170–173.

7. SPECTOR, D. H., VARMUS, H. E., & BISHOP, J. M. 1978. Nucleotide sequences related to the transforming gene of avian sarcoma virus are present in DNA of uninfected vertebrates. *Proc. Nat'l. Acad. Sci. U.S.A.* 75:4102–4106.

8. PURCHIO, A. F. ET AL. 1978. Identification of a polypeptide encoded by the avian sarcoma virus src gene. *Proc. Nat'l. Acad. Sci. U.S.A.* 75:1567–1671.

9. COLLETT, M. S. & ERIKSON, R. L. 1978. Protein kinase activity associated with the avian sarcoma virus src gene product. *Proc. Nat'l. Acad. Sci. U.S.A.* 75:2021–2924.

10. WILLINGHAM, M. C., JAY, G., & PASTAN, I. 1979. Localization of the ASV src gene product to the plasma membrane of transformed cells by electron microscopic immunocytochemistry. *Cell* 18:125–134.

11. HUNTER, T. & SEFTON, B. M. 1980. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc. Nat'l. Acad. Sci. U.S.A.* 77:1311–1315.

12. SHIH, C. ET AL. 1979. Passage of phenotypes of chemically transformed cells via transfection of DNA and chromatin. *Proc. Nat'l. Acad. Sci. U.S.A.* 76:5714–5718.

13. KRONTIRIS, T. G. & COOPER, G. M. 1981. Transforming activity of human tumor DNAs. *Proc. Nat'l. Acad. Sci. U.S.A.* 78:1181–1184.

14. GOLDFARB, M. ET AL. 1982. Isolation and preliminary characterization of a human transforming gene from T24 bladder carcinoma cells. *Nature* 296:404–409.

15. SHIH, C. & WEINBERG, R. A. 1982. Isolation of a transforming sequence from a human bladder carcinoma cell line. *Cell* 29:161–169.

16. PULCIANI, S. ET AL. 1982. Oncogenes in human tumor cell lines: Molecular cloning of a transforming gene from human bladder carcinoma cells. *Proc. Nat'l. Acad. Sci. U.S.A.* 79:2845–2849.

17. PARADA, L. F. ET AL. 1982. Human EJ bladder carcinoma oncogene is a homologue of Harvey sarcoma virus ras gene. *Nature* 297:474–478.

18. DER, C. J. ET AL. 1982. Transforming genes of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses. *Proc. Nat'l. Acad. Sci. U.S.A.* 79:3637–3640.

19. SANTOS, E. ET AL. 1982. T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB- and Harvey-MSV transforming genes. *Nature* 298:343–347.

20. TABIN, C. J. ET AL. 1982. Mechanism of activation of a human oncogene. *Nature* 300:143–149.

21. REDDY, E. P. ET AL. 1982. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature* 300:149–152.

22. TAPAROWSKY, E. ET AL. 1982. Activation of the T24 bladder carcinoma transforming gene is linked to a single amino acid change. *Nature* 300:762–765.

SYNOPSIS

Cancer is a disease involving heritable defects in cellular control mechanisms that result in the formation of invasive tumors capable of releasing cells that can spread the disease to distant sites in the body. Many of the characteristics of tumor cells can be observed in culture. Whereas normal cells proliferate until they form a single layer (monolayer) over the bottom of the dish, cancer cells continue to grow in culture, piling on top of one another to form clumps. Other characteristics often revealed by cancer cells include an abnormal number of chromosomes, an ability to continue to divide indefinitely, a dependence on glycolysis, and a lack of responsiveness to neighboring cells. (p. 650)

Normal cells can be converted to cancer cells by treatment with a wide variety of chemicals, ionizing radiation, and a variety of DNA- and RNA-containing viruses; all of these agents act by causing changes in the genome of the transformed cell. Analysis of the cells of a cancerous tumor almost always shows that the cells have arisen by growth of a single cell (the tumor is said to be monoclonal). The development of a malignant tumor is a multistep process characterized by a progression of genetic alterations that make the cells increasingly less responsive to the body's normal regulatory machinery and better able to invade normal tissues. The genes involved in carcinogenesis constitute a specific subset of the genome whose products are involved in such activities as control of the cell cycle, intercellular adhesion, and DNA repair. The level of expression of particular genes in different types of cancers can be determined using DNA microarrays. In addition to genetic alterations, growth of tumor cells is also influenced by both nongenetic and epigenetic influences that allow the cell to express its malignant phenotype. (p. 653)

Genes that have been implicated in carcinogenesis are divided into two broad categories: tumor-suppressor genes and oncogenes. Tumor-suppressor genes encode proteins that restrain cell growth and prevent cells from becoming malignant. Tumor-suppressor genes act recessively since both copies must be deleted or mutated before their protective function is lost. Oncogenes, in contrast, encode proteins that promote the loss of growth control and malignancy. Oncogenes arise from proto-oncogenes—genes that encode proteins having a role in a cell's normal activities. Mutations that alter either the protein or its expression cause the proto-oncogene to act abnormally and promote the formation of a tumor. Oncogenes act dominantly; that is, a single copy will cause the cell to express the altered phenotype. Most tumors contain alterations in both tumor-suppressor genes and oncogenes. As long as a cell retains at least one copy of all of its tumor-suppressor genes, it should be protected against the consequences of oncogene formation. Conversely, the loss of a tumor-suppressor function should not be sufficient, by itself, to cause the cell to become malignant. (p. 656)

The first tumor-suppressor gene to be identified was *RB*, which is responsible for a rare retinal tumor called retinoblastoma, which occurs with high frequency in certain families but can also occur sporadically. Children with the familial form of the disease inherit one mutated copy of the gene. These individuals develop the cancer only after sporadic damage to the second allele in one of the retinal cells. *RB* encodes a protein called pRB, which is involved in regulating passage of a cell from G_1 to S in the cell cycle. The unphosphorylated form of pRB interacts with certain transcription factors, preventing them from activating the genes required for certain S phase activities. Once pRB has been phosphorylated, the protein releases its bound transcription factor, which can then activate gene expression, leading to the initiation of S phase. (p. 658)

The tumor-suppressor gene most often implicated in human cancer is *TP53*, whose product (p53) may be able to suppress cancer formation by several different mechanisms. In one of its actions, p53 acts as a transcription factor that activates the expression of a protein (p21) that inhibits the cyclin-dependent kinase that moves a cell through the cell cycle. Damage to DNA triggers the phosphorylation and stabilization of p53, leading to the arrest of the cell cycle until the damage can be repaired. p53 can also redirect cells that are on the path toward malignancy onto an alternate path leading to

either apoptosis or senescence. *TP53* knockout mice begin to develop tumors several weeks after birth. Other tumor-suppressor genes include *APC* that, when mutated, predisposes an individual to developing colon cancer, and *BRCA1* and *BRCA2* that, when mutated, predispose an individual to developing breast cancer. (p. 660)

Most of the known oncogenes are derived from proto-oncogenes that play a role in the pathways that transmit growth signals from the extracellular environment to the cell interior, particularly the cell nucleus. A number of oncogenes have been identified that encode growth factor receptors, including platelet-derived growth factor (PDGF) and epidermal growth factor (EGF) receptors. Malignant cells may contain a much larger number of one of these growth factor receptors in their plasma membranes when compared to normal cells. The excess receptors make the cells sensitive to lower concentrations of the growth factor and, thus, are stimulated to divide under conditions that would not affect normal cells. A number of cytoplasmic protein kinases, including both serine/threonine and tyrosine kinases, are included on the list of oncogenes. These include *RAF*, which encodes a protein kinase in the MAP kinase cascade. Mutations in *RAS* are among the most common oncogenes found in human cancers. As discussed in Chapter 15, Ras activates the protein kinase activity of Raf. If Raf remains in the activated state, it sends continual signals along the MAP kinase pathway, leading to the continual stimulation of cell proliferation. A number of oncogenes, such as *MYC*, encode proteins that act as transcription factors. Myc is normally one of the first proteins to appear when a cell is stimulated to reenter the cell cycle from the quiescent G_0 phase. Another group of oncogenes, such as *BCL-2*, encode proteins involved in apoptosis. Overexpression of the *BCL-2* gene leads to the suppression of apoptosis in lymphoid tissues, allowing abnormal cells to proliferate to form lymphoid tumors. (p. 664)

Genes that encode proteins involved in DNA repair have also been implicated in carcinogenesis. The genome of patients with the most common hereditary form of colon cancer, called hereditary nonpolyposis colon cancer (HNPCC), contains microsatellite sequences of abnormal nucleotide number. Changes in the length of a microsatellite sequence arise as an error during replication, which is normally recognized by mismatch repair enzymes, suggesting that defects in such repair systems may be responsible for the cancers. This conclusion is supported by findings that extracts from HNPCC tumor cells display DNA repair deficiencies. Cells that contain such deficiencies would be expected to display a greatly elevated level of secondary mutations in tumor-suppressor genes and oncogenes that would lead to a greatly increased risk of malignancy. Certain microRNAs have also been implicated as oncogenes (or tumor suppressors). (p. 667)

Cancer is currently treated by surgery, chemotherapy, and radiation. Several other strategies are being tested; these include immunotherapy, inhibition of proteins encoded by oncogenes, and inhibition of angiogenesis. To date, the greatest success has come with the development of an inhibitor of Abl kinase in patients with chronic myelogenous leukemia. A second success story has been the development of humanized antibodies that bind to a protein on the surface of malignant B cells in cases of non-Hodgkin's lymphoma. Antiangiogenic strategies attempt to prevent a solid tumor from inducing the formation of new blood vessels that are required to supply the tumor cells with nutrients and other materials. A number of agents have been identified that block angiogenesis in mice and have had limited success in clinical trials. (p. 671)

17
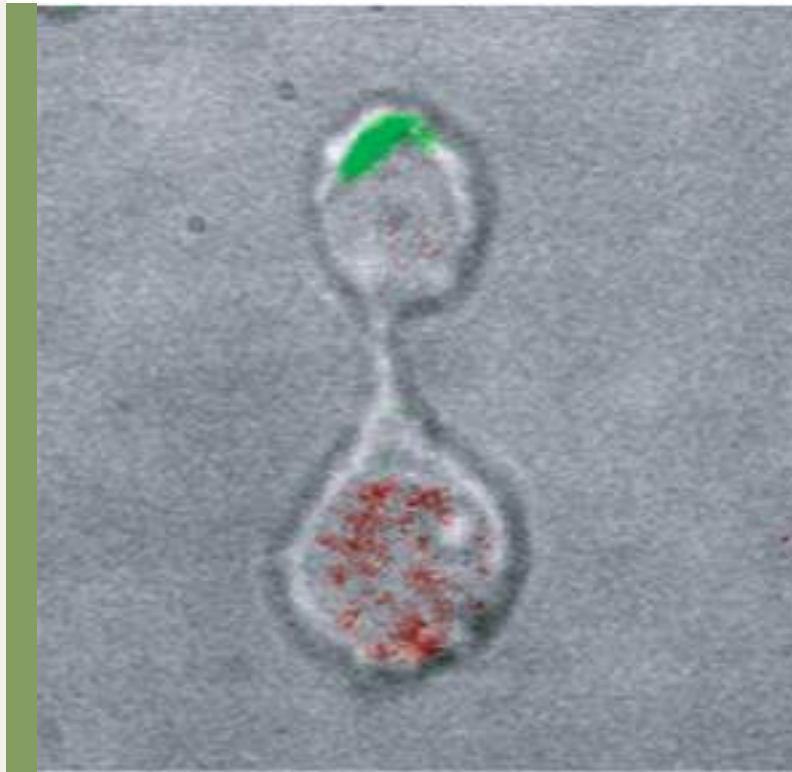
The Immune Response

L iving organisms provide ideal habitats in which other organisms can grow. It is not surprising, therefore, that animals are subject to infection by viruses, bacteria, protists, fungi, and animal parasites. Vertebrates have evolved several mechanisms that allow them to recognize and destroy these infectious agents. As a result, vertebrates are able to develop an **immunity** against invading pathogens. Immunity results from the combined activities of many different cells, some of which patrol the body, whereas others are concentrated in lymphoid organs, such as the bone marrow, thymus, spleen, and lymph nodes (Figure 17.1). Together, these dispersed cells and discrete organs form the body's **immune system**.

The cells of the immune system engage in a type of molecular screening by which they recognize "foreign" macromolecules, that is, ones whose structure is different from those of the body's normal macromolecules. If foreign material is encountered, the immune system mounts a specific and concerted attack against it. The weapons of the immune system include (1) cells that kill or ingest infected or altered cells and (2) soluble proteins that can neutralize, immobilize, agglutinate, or kill pathogens.

T lymphocytes are cells of the immune system that become activated when they contact cells (called APCs) displaying foreign peptides on their surface. This image shows a T cell that has undergone its first division after its association with an APC. The dividing cell has been stained for two different proteins, one appearing red and the other green. It is evident that the two proteins are being distributed into different daughter cells. The evidence from this study suggests that these two cells go on to have different fates, one becoming a short-lived effector cell that carries out a specific T-cell response and the other a memory cell that will remain dormant until some later date when it is reactivated by a subsequent contact with antigen. On a more general note, this image demonstrates that not all cell divisions produce identical daughter cells and that the process of cell division provides a means by which cells with different fates can be generated. (FROM JOHN T. CHANG ET AL., COURTESY OF STEVEN L. REINER, SCIENCE 315:1690, 2007. © COPYRIGHT 2007, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

Pathogens, in turn, are continually evolving countermechanisms to avoid immune destruction. The fact that humans suffer from a number of chronic infective diseases, such as AIDS (caused by a virus), tuberculosis (caused by a bacterium), and malaria (caused by a protozoan) illustrates how our immune systems are not always successful in combating these microscopic pathogens. The immune system is also implicated in the body's fight against cancer, but the degree to which the system can recognize and kill cancer cells remains controversial. In some cases, the immune system may mount an inappropriate response that attacks the body's own tissues. As discussed in the Human Perspective on page 707, these incidents can lead to serious disease.

It is impossible to cover the entire subject of immunity in a single chapter. Instead, we will focus on a number of selected aspects that illustrate principles of cell and molecular biology that were discussed in previous chapters. First, however, it is necessary to examine the basic events in the body's response to the presence of an intruding microbe.[1] ∎

17.1 AN OVERVIEW OF THE IMMUNE RESPONSE

The outer surface of the body and the linings of its internal tracts provide an excellent barrier to prevent penetration by viruses, bacteria, and parasites. If these surface barriers are breached, a series of **immune responses** are initiated that attempt to contain and then kill the invaders. Immune responses can be divided into two general categories: innate responses and adaptive (or acquired) responses. Both types of responses depend on the ability of the body to distinguish between materials that are supposed to be there (i.e., "self") and those that are not (i.e., foreign, or "nonself"). We can also distinguish two categories of pathogens: those that occur primarily inside a host cell (all viruses, some bacteria, and certain protozoan parasites) and those that occur primarily in the extracellular compartments of the host (most bacteria and other cellular pathogens). Different types of immune mechanisms have evolved to combat these two types of infections. An overview of some of these mechanisms is shown in Figure 17.2.

---

[1]Before beginning our study of the immune system, it is worth recalling that many animals are able to defend themselves against viral infections by synthesizing siRNAs that guide the destruction of double-stranded viral RNAs. Although vertebrate cells have the machinery for RNA interference (RNAi) (because they produce miRNAs) and can utilize small dsRNAs provided by researchers to destroy viral genomes (as discussed in the Human Perspective of Chapter 11), most studies suggest that vertebrates do not use siRNAs as a viral defense mechanism. It is generally thought that the evolution of a highly effective immune system in primitive vertebrates led to the abandonment of RNAi as a major weapon in the ongoing war against viruses.
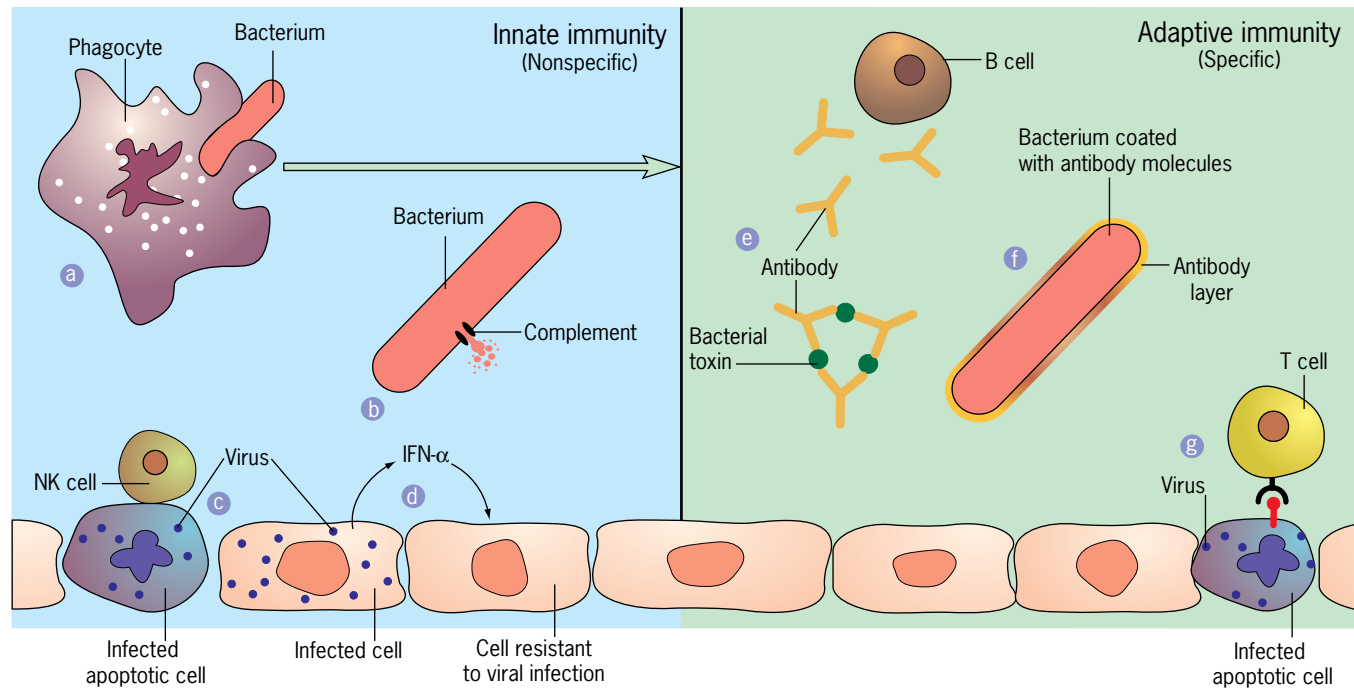
FIGURE 17.2 **An overview of some of the mechanisms by which the immune system rids the body of invading pathogens.** The left panel depicts several types of innate immunity: (*a*), phagocytosis of a bacterial cell; (*b*), bacterial cell killing by complement; (*c*), apoptosis induced in an infected cell by a natural killer (NK) cell; and (*d*), induction of viral resistance by interferon α (IFN-α). The right panel depicts several types of adaptive immunity: (*e*), B cell producing antibodies that neutralize a bacterial toxin; (*f*), bacterial cell coated with antibody (opsonized), which makes it susceptible to phagocytosis or complement-induced death; and (*g*), apoptosis induced in an infected cell by an activated T lymphocyte (T cell). Innate and adaptive immune responses are linked to one another (horizontal green arrow) because cells, such as macrophages, that phagocytize pathogens, use the foreign proteins to stimulate the production of specific antibodies and T cells directed against the pathogen. NK cells also produce substances, for example, IFNγ, that influence a T-cell response.

Innate Immune Responses

Innate immune responses are those the body mounts immediately without requiring previous contact with the microbe. Thus, they provide the body with a first line of defense. An invading pathogen typically makes its initial contact with the innate immune system when it is greeted by a phagocytic cell, such as a macrophage or a dendritic cell (see Figure 17.10), whose function is to recognize foreign objects and sound an appropriate alarm. These phagocytes possess a variety of receptor proteins on their surface that recognize certain types of highly conserved macromolecules that play essential roles in viruses or bacteria but are not produced by cells of the body. Several such receptors (called *pattern recognition receptors* or *PRRs*) have been identified, the most important of which are the **Toll-like receptors (TLRs)**. The discovery of TLRs came about by way of an interesting and unexpected series of events.

The fruit fly *Drosophila melanogaster* is well known for its iconic status in the field of genetics, and also for key contributions to the study of development and neurobiology. But, as an invertebrate, *Drosophila* would not have been considered a likely organism for a major discovery concerning the workings of the human immune system. However, in 1996, a group of researchers in France identified a mutant fruit fly that was highly susceptible to fungal infections, as is evident from the photograph in Figure 17.3. The fly pictured in Figure 17.3 is lacking a protein called Toll, which had been previously iden-

tified as a protein required for the normal development of dorsoventral ("top-bottom") polarity of the fly embryo. In fact, "Toll" is the German word for "weird," which described

FIGURE 17.3 **Scanning electron micrograph of a mutant fruit fly that had died from a fungal infection.** The body is covered with germinating fungal hyphae. This individual was susceptible to infection because it lacked a functional *Toll* gene. (FROM BRUNO LEMAITRE ET AL., CELL 86:978, 1996; BY PERMISSION OF CELL PRESS, COURTESY OF JULES A. HOFFMANN.)

the jumbled appearance of fly embroyos lacking a functional *Toll* gene. It appeared that Toll serves a dual function in these animals, as both a director of embryonic polarity and a factor that promotes immunity from infection. This discovery led researchers to carry out a search for similar proteins, that is, Toll-like receptors (TLRs), in other organisms. As it happens, humans express at least ten functional TLRs, all of which are transmembrane proteins present on the surfaces (or within endosomal/lysosomal membranes) of many different types of cells. Within the human TLR family are receptors that recognize the lipopolysaccharide or peptidoglycan components of the bacterial cell wall, the protein flagellin found in bacterial flagella, double-stranded RNA characteristic of replicating viruses, and unmethylated CpG dinucleotides (which are characteristic of bacterial DNA). A model of a TLR bound to its double-stranded RNA ligand is shown in Figure 17.4.

Activation of a TLR by one of these pathogen-derived molecules initiates a signal cascade within the cell that can lead to a variety of protective immune responses including the activation of cells of the adaptive immune system (represented by the horizontal green arrow in Figure 17.2). For this reason, a number of pharmaceutical companies are working on drugs that stimulate TLRs with the aim of enhancing the body's response against stubborn infections, such as that caused by the hepatitis C virus. Aldara, which was approved in 1997 and is prescribed for a number of skin conditions including genital warts, was later discovered to act by stimulating a TLR.

Innate responses to invading pathogens are typically accompanied by a process of **inflammation** at the site of infection where certain cells and plasma proteins leave the blood vessels and enter the affected tissues (page 247). These events are accompanied by local redness, swelling, and fever. Inflammation provides a means for concentrating the body's defensive agents at the site where they are needed. During inflammation, phagocytic cells migrate toward a site of infection in response to chemicals (chemoattractants) released at the site (page 369). Once there, these cells recognize, engulf, and destroy the pathogen (Figure 17.2*a*). Inflammation is a dual-edged sword. Although it protects the body against an invading pathogen, if inflammation is not terminated in a timely manner it can lead to damage to the body's normal tissues and chronic disease. Regulation of inflammation is a complex and poorly understood process involving a balance between pro- and anti-inflammatory activities.

A number of other mechanisms are also geared to attacking extracellular pathogens. Both epithelial cells and lymphocytes secrete a variety of antimicrobial peptides, called *defensins*, which are able to bind to viruses, bacteria, or fungi and bring about their demise. Blood also contains a group of soluble proteins called **complement** that binds to pathogens, triggering their destruction. In one of the complement pathways, an activated assembly of these proteins perforates the plasma membrane of a bacterial cell, leading to cell lysis and death (Figure 17.2*b*).

Innate responses against intracellular pathogens, such as viruses, are targeted primarily against cells that are already infected. Cells infected with certain viruses are recognized by a type of nonspecific lymphocyte called a **natural killer (NK) cell**. As its name implies, NK cells act by causing the death of the infected cell (Figure 17.2*c*). The NK cell induces the infected cell to undergo apoptosis (page 642). NK cells can also kill certain types of cancer cells in vitro (Figure 17.5) and may

FIGURE 17.4 Model of a Toll-like receptor (TLR3) bound to a double-stranded RNA (dsRNA) molecule. The extracellular ligand-binding domains of TLRs contain a large curved surface composed largely of leucine-rich repeats that provide flexibility in ligand binding. The dsRNA ligand (blue and orange double helix) binds to a TLR dimer whose transmembrane and cytoplasmic domains are brought into close contact as shown in this model. The complex is "ready" to send an intracellular signal alerting the cell of the presence of the foreign nucleic acid molecule. (FROM LIN LIU ET AL., COURTESY OF DAVID R. DAVIES, SCIENCE 320:381, 2008; © COPYRIGHT 2008, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

FIGURE 17.5 Innate immunity. Scanning electron micrograph of a natural killer cell bound to a target cell, in this case, a malignant erythroleukemia cell. NK cells kill their targets by the same mechanism described for CTLs on page 692. (COURTESY OF GIUSEPPE ARANCIA, DEPT. OF ULTRASTRUCTURES, INSTITUTO SUPERIORE DI SANITÁ, ROME, FROM BLOOD CELLS 17:165, 1991.)

provide a mechanism for destroying such cells in vivo before they develop into a tumor. Normal (i.e., noninfected and non-malignant) cells possess surface molecules that protect them from attack by NK cells.

Another type of innate antiviral response is initiated within the infected cell itself. Virus-infected cells produce proteins called type 1 *interferons* (IFN-α and IFN-β) that are secreted into the extracellular space, where they bind to the surface of noninfected cells rendering them resistant to subsequent infection (Figure 17.2*d*). Interferons accomplish this by several means, including activation of a signal transduction pathway that results in phosphorylation and consequent inactivation of the translation factor eIF2 (page 462). Cells that have undergone this response cannot synthesize viral proteins that are required for virus replication. IFN-β may also induce the synthesis of cellular microRNAs that target viral RNA genomes.

The innate and adaptive immune systems do not function independently but work closely together to destroy a foreign invader. Most importantly, the same phagocytic cells and NK cells that carry out an immediate innate response are also responsible for initiating the much slower, more specific adaptive immune response.

Adaptive Immune Responses

Unlike innate responses, **adaptive** (or **acquired**) **immune responses** require a lag period during which the immune system gears up for an attack against a foreign agent. Unlike innate responses, adaptive immune responses are highly specific and can discriminate between two very similar molecules. For example, the blood of a person who has just recovered from measles contains antibodies that react with the virus that causes measles, but not with a related virus, such as that which causes mumps. Unlike the innate system, the adaptive system also has a "memory," which usually means that the person will not suffer again from the same pathogen later in life.

Whereas all animals possess some type of innate immunity against microbes and parasites, only vertebrates are known to mount an adaptive response. There are two broad categories of adaptive immunity:

- **Humoral immunity**, which is carried out by **antibodies** (Figure 17.2*e,f*). Antibodies are globular, blood-borne proteins of the **immunoglobulin superfamily (IgSF)**.
- **Cell-mediated immunity**, which is carried out by cells (Figure 17.2*g*).

Both types of adaptive immunity are mediated by **lymphocytes**, which are nucleated leukocytes (white blood cells) that circulate between the blood and lymphoid organs. Humoral immunity is mediated by **B lymphocytes** (or **B cells**) that, when activated, differentiate into cells that secrete antibodies. Antibodies are directed primarily against foreign materials that are situated outside the body's cells. Such materials include the protein and polysaccharide components of bacterial cell walls, bacterial toxins, and viral coat proteins. In some cases, antibodies can bind to a bacterial toxin or virus particle and directly prevent the agent from entering a host cell (Fig-
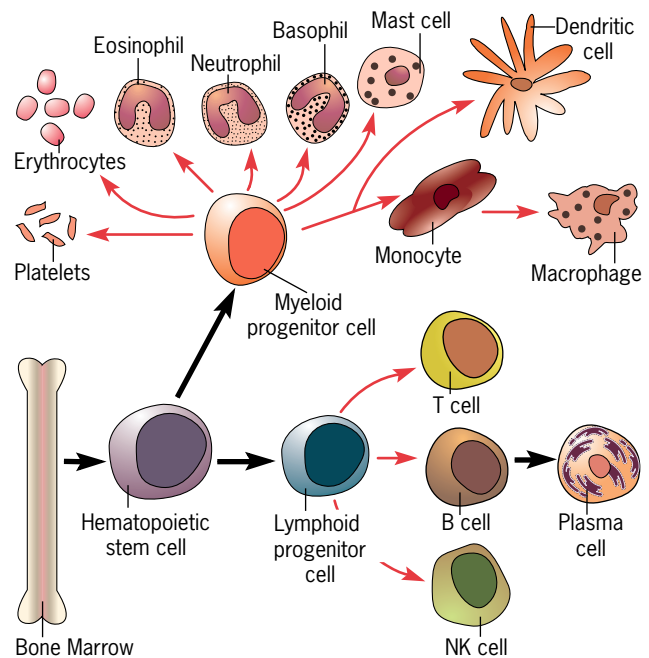
FIGURE 17.6 Pathways of differentiation of a hematopoietic stem cell of the bone marrow. A hematopoietic stem cell can give rise to two different progenitor cells: a myeloid progenitor cell that can differentiate into most of the various blood cells (e.g., erythrocytes, basophils, and neutrophils), macrophages, or dendritic cells; or a lymphoid progenitor cell that can differentiate into any of the various types of lymphocytes (NK cells, T cells, or B cells). T-cell precursors migrate to the thymus where they differentiate into T cells. In contrast, B cells undergo differentiation in the bone marrow. Cells in the various stages of B- and T-cell differentiation can be distinguished by the species of proteins at their cell surface and/or the transcription factors that determine the genes being expressed.

ure 17.2*e*). In other cases, antibodies function as "molecular tags" that bind to an invading pathogen and mark it for destruction. Bacterial cells coated with antibody molecules (Figure 17.2*f*) are rapidly ingested by wandering phagocytes or destroyed by complement molecules carried in the blood. Antibodies are not effective against pathogens that are present inside cells, hence the need for a second type of weapon system. Cell-mediated immunity is carried out by **T lymphocytes** (or **T cells**) that, when activated, can specifically recognize and kill an infected (or foreign) cell (Figure 17.2*g*).

B and T cells arise from the same type of precursor cell (a *hematopoietic stem cell*), but they differentiate along different pathways in different lymphoid organs. A summary of the various pathways of differentiation of the hematopoietic stem cell is shown in Figure 17.6. B lymphocytes differentiate in the fetal liver or adult bone marrow, whereas T lymphocytes differentiate in the thymus gland, an organ located in the chest that reaches its peak size during childhood. Because of these differences, cell-mediated and humoral immunity can be dissociated to a large extent. For instance, humans may suffer from a rare disease called congenital agammaglobulinemia in which humoral antibody is deficient and cell-mediated immunity is normal.

1. Contrast the general properties of innate and adaptive immune responses.

2. List four types of innate immune responses. Which would be most effective against a pathogen inside an infected cell?

3. What is meant by the terms "humoral" and "cell-mediated" immunity?

17.2 THE CLONAL SELECTION THEORY AS IT APPLIES TO B CELLS

If a person is infected with a virus or exposed to foreign material, his or her blood soon contains a high concentration of antibodies capable of reacting with the foreign substance, which is known as an **antigen**. Most antigens consist of proteins or polysaccharides, but lipids and nucleic acids can also serve in this capacity. How can the body produce antibodies that react *specifically* with an antigen to which the body is exposed? In other words, how does an antigen induce an adaptive immune response? For many years it was thought that antigens somehow *instructed* lymphocytes to produce complementary antibodies. It was suggested that an antigen wraps itself around an antibody molecule, molding the antibody into a shape capable of combining with that particular antigen. In this "instructive" model, the lymphocyte only gains its ability to produce a specific antibody after its initial contact with antigen. In 1955, Niels Jerne, a Danish immunologist, proposed a radically different mechanism. Jerne suggested that the body produces small amounts of randomly structured antibodies in the absence of any antigen. As a group, these antibodies would be able to combine with any type of antigen to which a person might someday be exposed. According to Jerne's model, when a person is exposed to an antigen, the antigen combines with a specific antibody, which somehow leads to the subsequent production of that particular antibody molecule. Thus in Jerne's model, the antigen *selects* those preexisting antibodies capable of binding to it. In 1957, the concept of antigen selection of antibodies was expanded into a comprehensive model of antibody formation by the Australian immunologist F. MacFarlane Burnet. Burnet's **clonal selection theory** quickly gained widespread acceptance. An overview of the steps that occur during the clonal selection of B cells is shown in Figure 17.7. More detailed discussion of these events is provided later in the chapter. The clonal selection of T cells is described in the following section. The main features of B-cell clonal selection are the following:

1. *Each B cell becomes committed to produce one species of antibody.* B cells arise from a population of undifferentiated and indistinguishable progenitor cells. As it differentiates, a B cell becomes committed as the result of DNA rearrangements (see Figure 17.18) to producing only one species of antibody molecule (Figure 17.7, step 1). Thousands of different DNA rearrangements are possible, so that different B cells produce different antibody molecules. Thus, even though mature B cells appear identical under the microscope, they can be distinguished by the antibodies they produce.

2. *B cells become committed to antibody formation in the absence of antigen.* The basic repertoire of antibody-producing cells that a person will possess throughout his
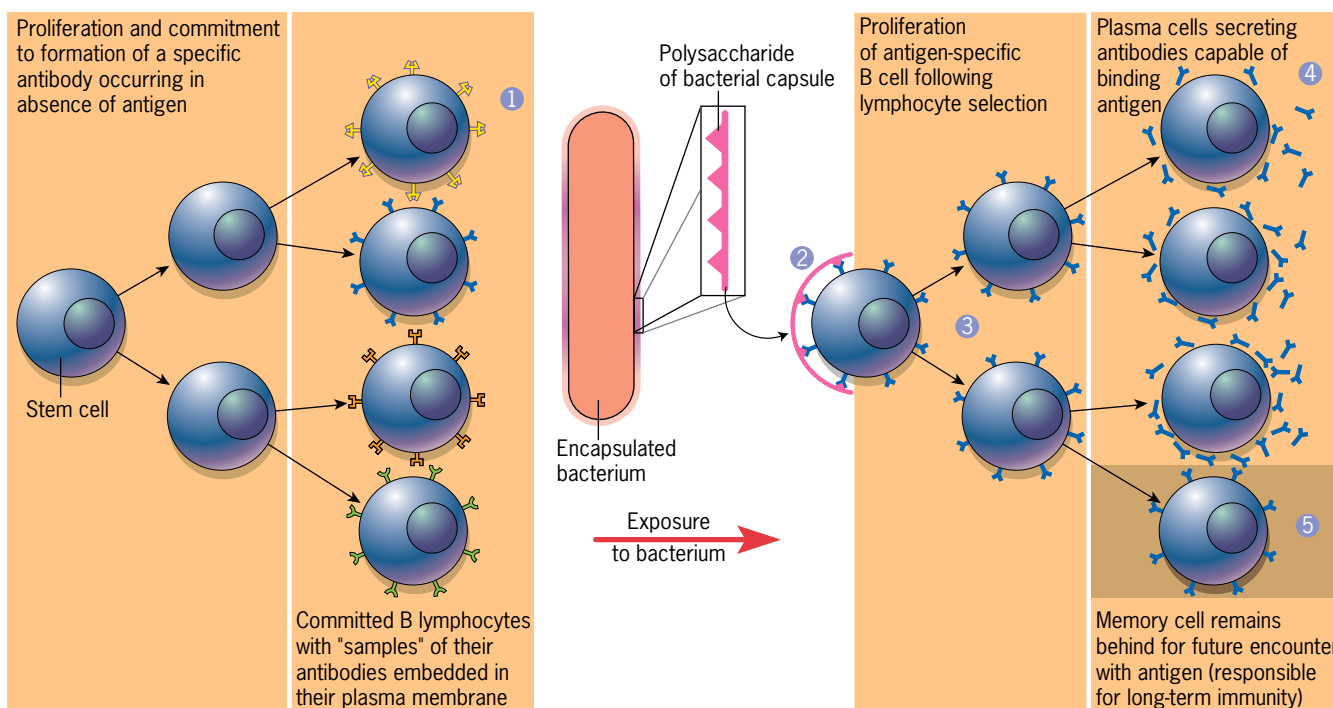
FIGURE 17.7 **The clonal selection of B cells by a thymus-independent antigen.** The steps are described in the figure and also in the text.

or her lifetime is already present within the lymphoid tissues *prior to stimulation by an antigen* and is independent of the presence of foreign materials. Each B cell displays its particular antibody on its surface with the antigen-reactive portion facing outward. As a result, the cell is coated with antigen receptors that can bind specifically with antigens having a complementary structure. Although most lymphoid cells are never required during a person's lifetime, the immune system is primed to respond immediately to any antigen that a person may be exposed to. The presence of cells with different membrane-bound antibodies can be demonstrated experimentally, as shown in Figure 17.8.
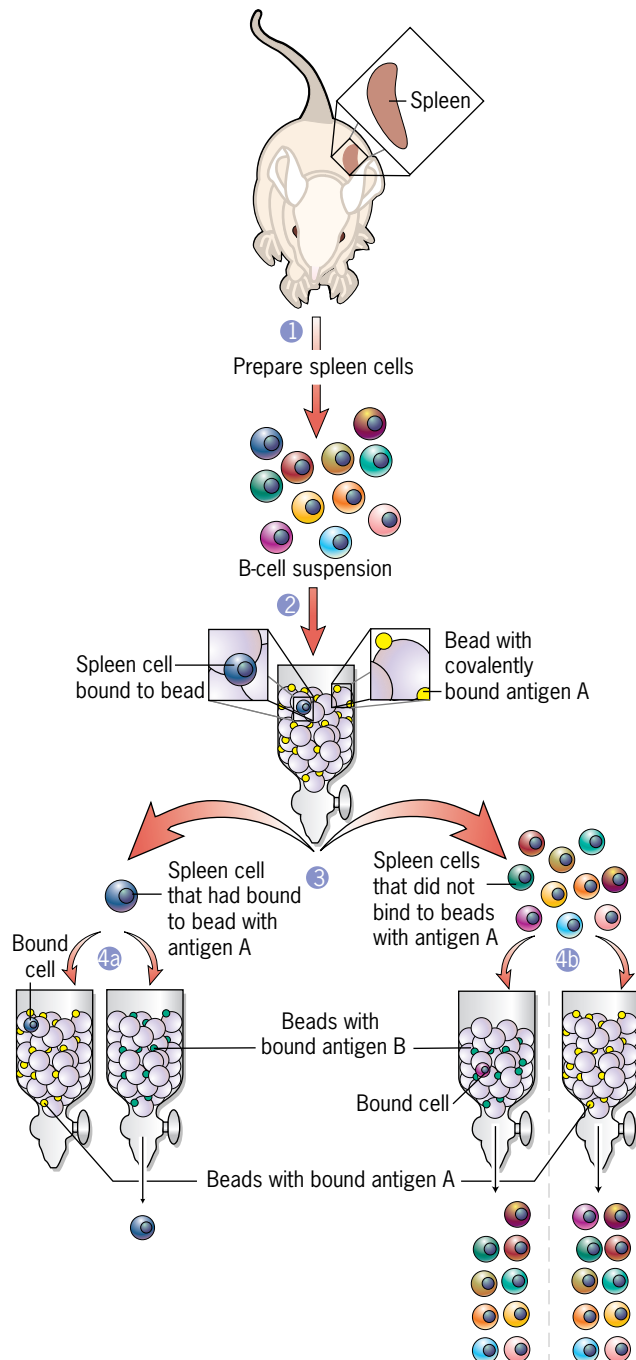
3. ***Antibody production follows selection of B cells by antigen.*** In most cases, the activation of a B cell by antigen requires the involvement of T cells (discussed on pages 692 and 705). A few antigens, however, such as the polysaccharides present in bacterial cell walls, activate B cells by themselves; antigens of this type are described as *thymus-independent antigens*. For simplicity, we will restrict the discussion at this point in the chapter to a thymus-independent antigen. Suppose a person were to be exposed to *Haemophilus influenzae* type B, an encapsulated bacterium that can cause fatal meningitis. The capsule of these bacteria contains a polysaccharide that can bind to a tiny fraction of the body's B cells (Figure 17.7, step 2). The B cells that bind the polysaccharide contain membrane-bound antibodies whose combining site allows them to interact specifically with that antigen. In this way, an antigen *selects* those lymphocytes that produce antibodies capable of interacting with that antigen. Antigen binding activates the B cell, causing it to proliferate (Figure 17.7, step 3) and form a population (or *clone*) of lymphocytes, all of which make the same antibody. Some of these activated cells differentiate into short-lived **plasma cells** that secrete large amounts of antibody molecules (Figure 17.7, step 4). Unlike their B-cell precursors (Figure 17.9*a*), plasma cells possess an extensive rough ER characteristic of cells that are specialized for protein synthesis and secretion (Figure 17.9*b*).

4. ***Immunologic memory provides long-term immunity.*** Not all B lymphocytes that are activated by antigen differentiate into antibody-secreting plasma cells. Some remain in lymphoid tissues as *memory B cells* (Figure 17.7, step 5) that can respond rapidly at a later date if the antigen reappears in the body. Although plasma cells die off following removal of the antigenic stimulus, memory B cells may persist for a person's lifetime. When stimulated by the same antigen, some of the memory B cells rapidly proliferate into plasma cells, generating a *secondary immune response* in a matter of hours rather than the days required for the original response (see Figure 17.13).

FIGURE 17.8 **Experimental demonstration that different B cells contain a different membrane-bound antibody and that these antibodies are produced in the absence of antigen.** In this experiment, B cells are prepared from a mouse spleen (step 1). In step 2, the spleen cells are passed through a column containing beads coated with an antigen (antigen A) to which the mouse had never been exposed. A tiny fraction of the spleen cells bind to the beads, while the vast majority of spleen cells pass directly through the column (shown in step 3). In step 4, spleen cells from the previous experiment are passed through one of two different columns: a column whose beads are coated with antigen A or a column whose beads are coated with an unrelated antigen (antigen B) to which the mouse had never been exposed. In step 4a, the spleen cells tested are those that had bound to the beads in the previous step. These cells are found to rebind to beads coated with antigen A, but do not bind to beads coated with antigen B. In step 4b, the spleen cells tested are those that did not bind to the beads in the previous step. None of these cells bind to beads coated with antigen A, but a tiny fraction binds to beads coated with antigen B.
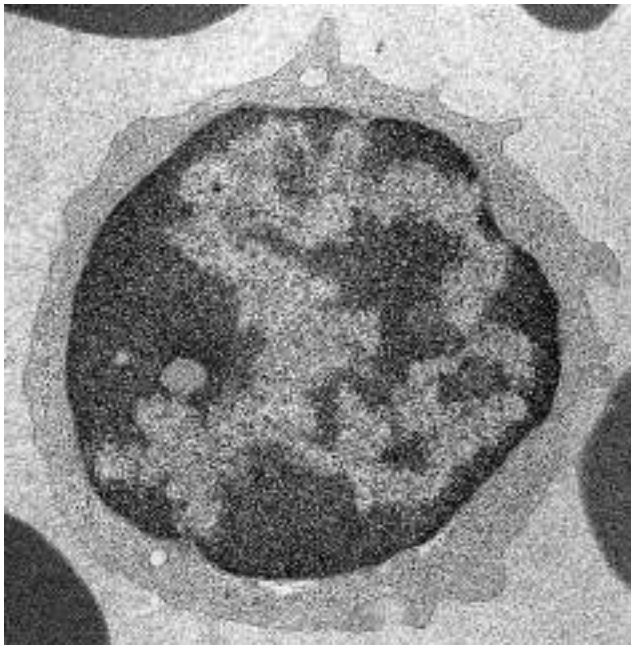
(a)
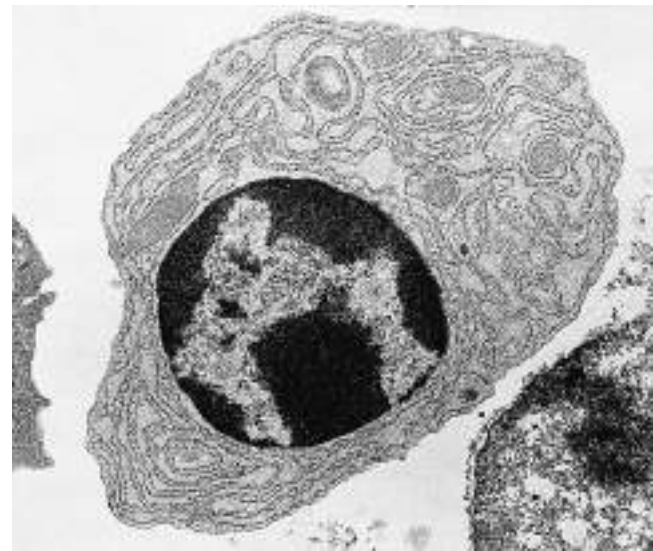
(b)

FIGURE 17.9 Comparison of the structure of a B cell (*a*) and a plasma cell (*b*). The plasma cell has a much larger cytoplasmic compartment than the B cell, with more mitochondria and an extensively developed rough endoplasmic reticulum. These characteristics reflect the synthesis of large numbers of antibody molecules by the plasma cell. (A: © VU/DAVID PHILLIPS/VISUALS UNLIMITED; B: © VU/K. G. MURTI/VISUALS UNLIMITED.)

5. ***Immunologic tolerance prevents the production of antibodies against self.*** As discussed below, genes encoding antibodies are generated by a process in which DNA segments are randomly combined. As a result, genes are invariably formed that encode antibodies that can react with the body's own tissues, which could produce widespread organ destruction and subsequent disease. It is obviously in the best interest of the body to prevent the production of such proteins, which are called **autoantibodies**. As they develop, many of the B cells capable of producing autoantibodies are either destroyed or rendered inactive. As a result, the body develops an *immunologic tolerance* toward itself. As discussed in the Human Perspective on page 707, a breakdown of the tolerant state can lead to the development of debilitating autoimmune diseases.

Several principles of the clonal selection theory can be illustrated by briefly considering the subject of vaccination.

Vaccination

Edward Jenner practiced medicine in the English countryside at a time when smallpox was one of the most prevalent and dreaded diseases. Over the years, he noticed that the maids who tended the cows were typically spared the ravages of the disease. Jenner concluded that milkmaids were somehow "immune" to smallpox because they were infected at an early age with cowpox, a harmless disease they contracted from their cows. Cowpox produces blisters that resemble the pus-filled blisters of smallpox, but the cowpox blisters are localized and disappear, causing nothing more serious than a scar at the site of infection.

In 1796, Jenner performed one of the most famous (and risky) medical experiments of all time. First, he infected an eight-year-old boy with cowpox and gave the boy time to recover. Six weeks later, Jenner intentionally infected the boy with smallpox by injecting pus from a smallpox lesion directly under the boy's skin. The boy showed no signs of the deadly disease. Within a few years, thousands of people had become immune to smallpox by intentionally infecting themselves with cowpox. This procedure was termed *vaccination*, after *vacca*, the Latin word for cow.

Jenner's experiment was successful because the immune response generated against the virus that causes cowpox happens to be effective against the closely related virus that causes smallpox. Most modern vaccines contain *attenuated* pathogens, which are pathogens that are capable of stimulating immunity but have been genetically "crippled" so that they are unable to cause disease. Most vaccines currently in use are B-cell vaccines, such as that employed to fight tetanus. Tetanus results from infection by the anaerobic soil bacterium *Clostridium tetani*, which can enter the body through a puncture wound. As they grow, the bacteria produce a powerful neurotoxin that blocks transmission across inhibitory synapses on motor neurons, leading to sustained muscle contraction and asphyxiation. At two months of age, most infants are *immunized* against tetanus by inoculation with a modified and harmless version of the tetanus toxin (called a *toxoid*). The tetanus toxoid binds to the surfaces of B cells whose membrane-

bound antibody molecules have a complementary binding site. These B cells proliferate to form a clone of cells that produce antibodies capable of binding to the actual tetanus toxin. This initial response soon wanes, but the person is left with memory cells that respond rapidly if the person should happen to develop a *C. tetani* infection at a later date. Unlike most immunizations, immunity to the tetanus toxin does not last a lifetime, which is the reason that people are given a booster shot every ten years or so. The booster shot contains the toxoid protein and stimulates the production of additional memory cells. What if a person receives a wound that has the potential to cause tetanus, and they cannot remember ever receiving a booster shot? In these cases, the person is likely to be given a *passive immunization*, consisting of antibodies that can bind the tetanus toxin. Passive immunization is effective for only a short period of time and does not protect the recipient against a subsequent infection.

Whereas any infected cell can serve as an APC in activating T cells, certain types of "professional" APCs are specialized for this function. Professional APCs include dendritic cells and macrophages (Figure 17.10). We will focus on dendritic cells (DCs), which were first discovered and characterized by Ralph Steinman of Rockefeller University in the early 1970s and are often described as the "sentinels" of the immune system. Dendritic cells have earned this title because they "stand guard" in the body's peripheral tissues where pathogens

---

REVIEW ?

1. Contrast instructive and selective mechanisms of antibody production.

2. What are the basic tenets of the clonal selection theory?

3. What does it mean for a B cell to become committed to antibody formation? How is this process influenced by the presence of antigen? What role does antigen play in antibody production?

4. What is meant by the terms *immunologic memory* and *immunologic tolerance*?

(a)

17.3 T LYMPHOCYTES: ACTIVATION AND MECHANISM OF ACTION

Like B cells, T cells are also subject to a process of clonal selection. T cells possess a cell-surface protein, called a **T-cell receptor**, that allows them to interact specifically with a particular antigen. Like the antibody molecules that serve as B-cell receptors, the proteins that serve as T-cell receptors exist as a large population of molecules that have differently shaped combining sites. Just as each B cell produces only one species of antibody, each T cell has only a single species of T-cell receptor. It is estimated that adult humans possess approximately 10^{12} T cells that, collectively, exhibit more than 10^7 different antigen receptors.

Unlike B cells, which are activated by intact antigens, T cells are activated by fragments of antigens that are displayed on the surfaces of other cells, called **antigen-presenting cells (APCs)**. Consider what would happen if a liver or kidney cell were infected with a virus. The infected cell would display portions of the viral proteins on its surface (see Figure 17.24), enabling the infected cell to bind to a T cell with the appropriate T-cell receptor. As a result of this presentation, the immune system becomes alerted to the entry of a specific pathogen. The process of antigen presentation is discussed at

(b)

FIGURE 17.10 **Professional antigen-presenting cells (APCs).** Colorized scanning electron micrographs of a macrophage (*a*) and a dendritic cell (*b*). Dendritic cells are irregular-shaped cells with long cytoplasmic (or dendritic) processes. (A: FROM A. POLLIACK/PHOTO RESEARCHERS, INC.; B: FROM DAVID SCHARF/PETER ARNOLD, INC.)

are likely to enter (such as the skin and airways). DCs are equipped with a wide variety of receptors capable of recognizing virtually every type of pathogen. Most importantly, DCs are particularly adept at initiating an adaptive immune response.

When present in the body's peripheral tissues, immature DCs recognize and ingest microbes and other foreign materials by phagocytosis. Once a microbe is taken into a dendritic cell, it has to be processed before its components can be presented to another cell. Antigen processing requires that the ingested material is fragmented enzymatically in the cytoplasm and the fragments moved to the cell surface (see Figure 17.23). DCs that have processed antigen migrate to nearby lymph nodes where they differentiate into mature antigen-presenting cells. Once in a lymph node, DCs come into contact with a large pool of T cells, including a minute percentage whose T-cell receptors can bind specifically with the processed foreign antigen, which activates the T cell. This dynamic process of DC-T cell interactions has been visualized recently in living lymph-node tissues by microscopic imaging of fluorescently labeled cells (Figure 17.11). In the absence of antigen, a given DC may interact transiently with as many as 500 to 5000 different T cells, remaining in contact with each cell for only a few minutes (as in Figure 17.11). In contrast, when the DC displays an antigen that is specifically recognized by the TCR of the T cell, the interaction between the cells is seen to last a period of hours, leading to the activation of the T cell, as evidenced by a transient increase in the cytosolic Ca^{2+} concentration.

Once activated, a T cell proliferates to form a clone of T cells having the same T-cell receptor. It is estimated that a single activated T cell can divide three to four times per day for several days, generating a tremendous population of T cells capable of interacting with the foreign antigen. The massive proliferation of specific T lymphocytes in response to an infection is often reflected in the enlargement of local lymph nodes. Once the foreign antigen has been cleared, the vast majority of the expanded T-cell population dies by apoptosis, leaving behind a relatively small population of memory T cells capable of responding rapidly in the event of future contact with the same pathogen.

Unlike B cells, which secrete antibodies, T cells carry out their assigned function through direct interactions with other cells, including B cells, other T cells, or target cells located throughout the body. This cell–cell interaction may lead to the activation, inactivation, or death of the other cell. In addition to direct cell contact, many T-cell interactions are mediated by highly active chemical messengers, called **cytokines**, that work at very low concentrations. Cytokines are small proteins produced by a wide variety of cells and include interferons (IFNs), interleukins (ILs), and tumor necrosis factors (TNFs). Cytokines bind to specific receptors on the surface of

(a)

(b)

(c)
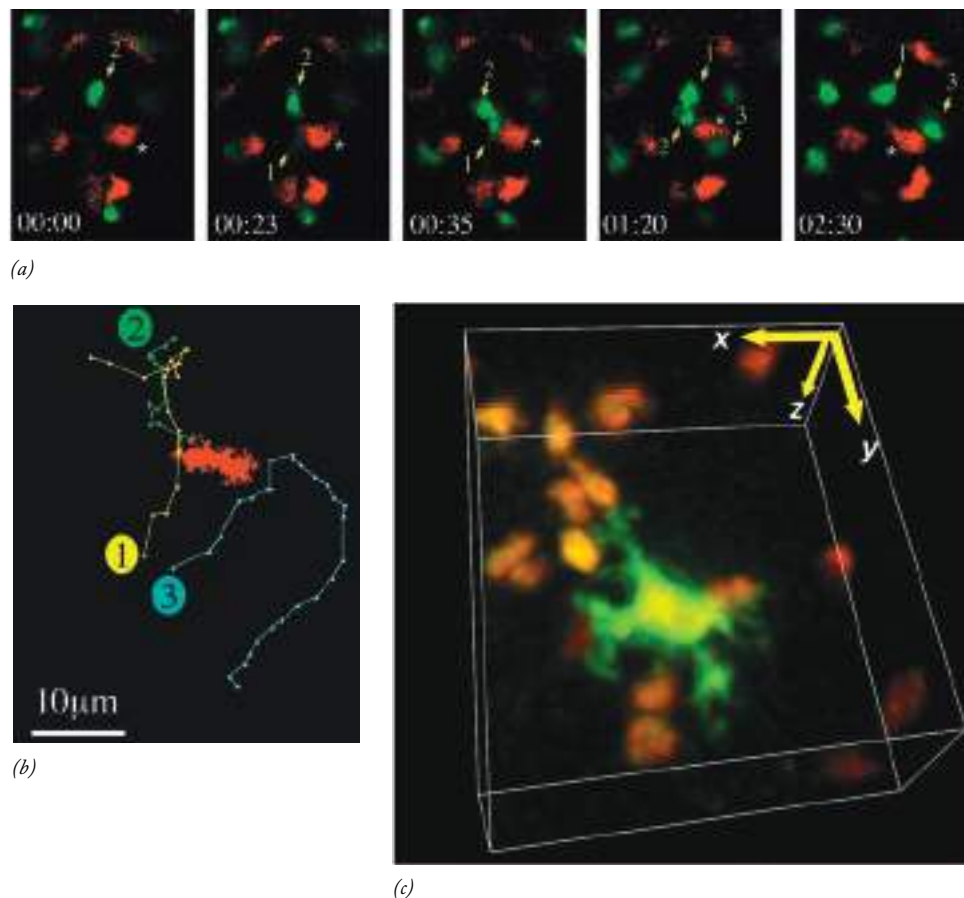
FIGURE 17.11 Live cell imaging of DCs and T cells, and their interactions, within a lymph node. T cells travel from lymph node to lymph node within the body. When they enter a lymph node, they migrate within the tissue as their surface is scanned by individual DCs with which they come into contact. (*a*) Several fluorescently labeled T cells (stained green and indicated with labeled numbers) are seen moving around within a lymph node and coming in contact with an individual DC (stained red and indicated by the asterisk) over a period of 2.5 minutes. (*b*) The trajectory taken by each of the three labeled T cells and the asterisk-marked DC shown in part *a*. (*c*) Contacts within a lymph node between a single DC (green) and several T cells (red) rendered in three dimensions. Contacts between the cells are dynamic, changing rapidly in size over a period of tens of seconds. (A–B: FROM PHILIPPE BOUSSO AND ELLEN ROBEY, NATURE IMMUNOL. 4:581, 2003; C: FROM MARK J. MILLER ET AL., COURTESY OF MICHAEL D. CAHALAN, PROC. NAT'L. ACAD. SCI. U.S.A. 101:1002, 2004.)

TABLE 17.1   Selected Cytokines

Cytokine	Source	Major functions
IL-1	Diverse	Induces inflammation, stimulates T_H-cell proliferation
IL-2	T_H cells	Stimulates T-cell and B-cell proliferation
IL-4	T cells	Induces IgM to IgG class switching in B cells, supresses inflammatory cytokine action
IL-5	T_H cells	Stimulates B-cell differentiation
IL-10	T cells, macrophages	Inhibits macrophage function, suppresses inflammatory cytokine action
IFN-γ	T_H cells, CTLs	Induces MHC expression in APCs, activates NK cells
TNF-α	Diverse	Induces inflammation, activates NO production in macrophages
GM-CSF	T_H cells, CTLs	Stimulates growth and proliferation of granulocytes and macrophages

a responding cell, generating an internal signal that alters the activity of the cell. In responding to a cytokine, a cell may prepare to divide, undergo differentiation, or secrete its own cytokines. One family of small cytokines, called *chemokines*, act primarily as chemoattractants that stimulate the migration of lymphocytes into inflamed tissue. Different types of lymphocytes and phagocytes possess receptors for different chemokines, so that their migration patterns can be separately controlled. A list of some of the best studied cytokines is given in Table 17.1.

Three major subclasses of T cells can be distinguished by the proteins on their surfaces and their biological functions.

1. ***Cytotoxic T lymphocytes* (CTLs)** screen the cells of the body for abnormalities. Under normal circumstances, healthy cells are not harmed by CTLs, but aged or infected cells, and possibly some malignant cells, are attacked and killed. CTLs kill target cells by inducing them to undergo apoptosis. Two distinct cell-killing pathways have been described. In one pathway, the CTL releases perforins and granzymes into a tightly enclosed space between the cells. *Perforins* are proteins that assemble within the membrane of the target cell to form transmembrane channels. *Granzymes* are proteolytic enzymes that enter the perforin channels and activate caspases, which are the proteolytic enzymes that initiate the apoptotic response (page 643). In the alternate pathway, the CTL binds to a receptor on the target cell surface, activating a suicide pathway in the target cell similar to that of Figure 15.36. By killing infected cells, CTLs eliminate viruses, bacteria, yeast, protozoa, and parasites after they have found their way into host cells and are no longer accessible to circulating antibodies. CTLs possess a surface

protein called CD8 (cluster designation 8) and are referred to as $CD8^+$ cells.

2. ***Helper T lymphocytes*** (T_H cells) activate other immune cells in an orchestrated attack against a specific pathogen. They are distinguished from CTLs by the presence of the protein CD4 at their surface rather than CD8.[2] T_H cells are activated by professional antigen-presenting cells, such as dendritic cells and macrophages as shown in Figure 17.12. This is one of the first and most important steps in initiating an adaptive immune response. Nearly all B cells require the help of T_H cells before they can mature and differentiate into antibody-secreting plasma cells. B cells are activated by direct interaction with a T_H cell that is specific for the same antigen, as shown in Figure 17.12 (and in more detail in Figure 17.26). Thus, the formation of antibodies requires the activation of both T cells and B cells capable of interacting specifically with the same antigen. The importance of T_H cells becomes apparent when one considers the devastating effects wrought by HIV, the virus that causes AIDS. T_H cells are the primary targets of HIV. Most HIV-infected people remain free of symptoms as long as their T_H-cell count remains relatively high—above 500 cells/μl (the normal count is over 1000 cells/μl). Once the count drops below approximately 200 cells/μl, a person develops full-blown AIDS and becomes prone to attack by viral and cellular pathogens.

3. ***Regulatory T lymphocytes*** (T_{Reg} cells) are primarily inhibitory cells that suppress the proliferation and activities of other types of immune cells. T_{Reg} cells are characterized by possession of $CD4^+CD25^+$ surface markers and are thought to play an important role in limiting inflammation and maintaining immunologic self-tolerance. T_{Reg} cells carry out this latter activity by suppressing T cells that carry self-reactive receptors from attacking the body's own cells. On the other hand, these same cells may prove detrimental to our health by preventing the immune system from ridding the body of tumor cells. The differentiation of T_{Reg} cells requires the presence of a key transcription factor, FOXP3. Mutations in the *FOXP3* gene result in a fatal disease (IPEX), which is characterized by severe autoimmunity in newborn infants. Studies of T_{Reg} cells have provided direct demonstration that homeostasis in the immune system requires a tight balance between stimulatory and inhibitory influences.

---

[2]There are three major classes of helper T cells, T_H1, T_H2, and T_H17 cells, which can be distinguished by the cytokines they secrete. All of these helper T cells stimulate B cells to produce antibodies, but each also has a unique function. T_H1 cells produce IFN-γ, which protects the body by activating macrophages to kill intracellular pathogens they might harbor (page 308). T_H2 cells produce IL-4, which mobilizes mast cells, basophils, and eosinophils to protect against extracellular pathogens, especially parasitic worms. T_H17 cells produce IL-17, which is thought to stimulate epithelial cells to recruit phagocytes and thereby prevent the entry of extracellular bacteria and fungi into the body. The three types of T_H cells differentiate from a common precursor following stimulation by different cytokines and are defined by the presence of different "master" transcription factors.
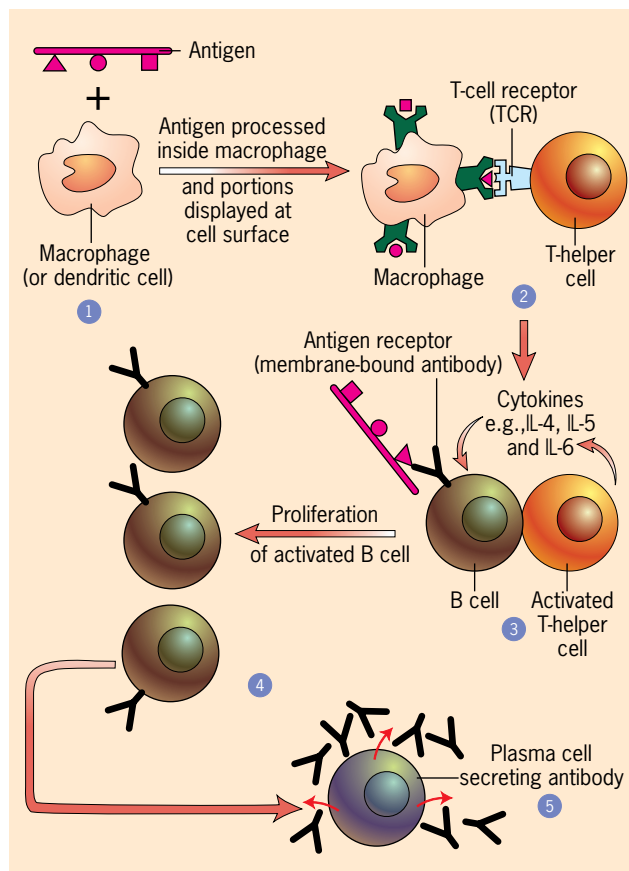
FIGURE 17.12 Highly simplified, schematic drawing showing the role of T$_H$ cells in antibody formation. In step 1, the macrophage interacts with the complex antigen. The antigen is taken into the macrophage and cleaved into fragments, which are displayed at the cell surface. In step 2, the macrophage interacts with a T$_H$ cell whose TCR is bound to one of the displayed antigen fragments (the green membrane protein is an MHC molecule, page 700). This interaction activates the T cell. In step 3, the activated T$_H$ cell interacts with a B cell whose antigen receptor is bound to an intact, soluble antigen. B-cell activation is stimulated by cytokines (e.g., IL-4, IL-5, and IL-6) released by the T$_H$ cell into the space separating it from the adjacent B cell. Interaction with the T$_H$ cell activates the B cell, causing it to proliferate (step 4). The progeny of the activated B cell differentiate into plasma cells that synthesize antibodies that can bind the antigen (step 5).

REVIEW **?**

1. How does an infected cell in the body reveal its condition to a T cell? What is the T cell's response?
2. What is an APC? What types of cells can act as APCs?
3. Compare and contrast the properties and functions of a T$_H$ cell and a CTL. A T$_H$ cell and a T$_{Reg}$ cell.

17.4 SELECTED TOPICS ON THE CELLULAR AND MOLECULAR BASIS OF IMMUNITY

The Modular Structure of Antibodies

Antibodies are proteins produced by B cells and their descendants (plasma cells). B cells incorporate antibody molecules into their plasma membrane, where they serve as antigen receptors, whereas plasma cells secrete these proteins into the blood or other bodily fluids, where they serve as a molecular arsenal in the body's war against invading pathogens. Interaction between blood-borne antibodies and antigens on the surface of a virus or bacterial cell can neutralize the pathogen's ability to infect a host cell and facilitate the pathogen's ingestion and destruction by wandering phagocytes. The immune system produces millions of different antibody molecules that, taken together, can bind any type of foreign substance to which the body may be exposed. Though the immune system exhibits great diversity through the antibodies it produces, a single antibody molecule can interact with only one or a few closely related antigenic structures.

Antibodies are globular proteins called *immunoglobulins*. Immunoglobulins are built of two types of polypeptide chains, larger **heavy chains** (molecular mass of 50,000 to 70,000 daltons) and smaller **light chains** (molecular mass of 23,000 daltons). The two types of chains are linked to one another in pairs by disulfide bonds. Five different classes of immunoglobulin (*IgA, IgD, IgE, IgG,* and *IgM*) have been identified. The different immunoglobulins appear at different times after exposure to a foreign substance and have different biological functions (Table 17.2). IgM molecules are the first antibodies secreted by B cells following stimulation by an antigen,

TABLE 17.2 Classes of Human Immunoglobulins

Class	Heavy chain	Light chain	Molecular mass (kDa)	Properties
IgA	α	κ or λ	360–720	Present in tears, nasal mucus, breast milk, intestinal secretions
IgD	δ	κ or λ	160	Present in B-cell plasma membranes; function uncertain
IgE	ε	κ or λ	190	Binds to mast cells, releasing histamine responsible for allergic reactions
IgG	γ	κ or λ	150	Primary blood-borne soluble antibodies; crosses placenta
IgM	μ	κ or λ	950	Present in B-cell plasma membranes; mediates initial immune response; activates bacteria-killing complement
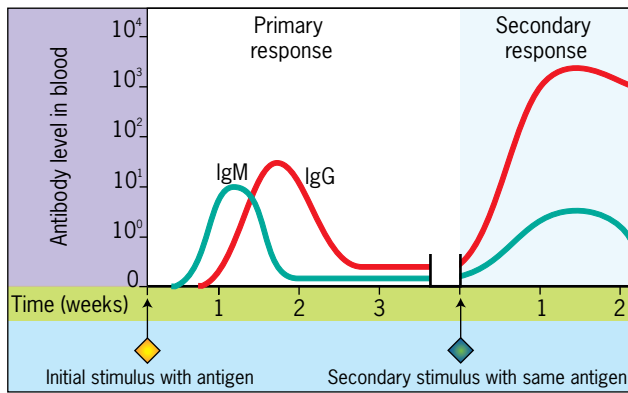
FIGURE 17.13 **Primary and secondary antibody responses.** A primary response, which is elicited by an initial exposure to an antigen, leads first to the production of soluble IgM antibody molecules, followed by production of soluble IgG antibody molecules. When the antigen is reintroduced at a later time, a secondary response is initiated. In contrast to the primary response, the secondary response begins with production of IgG (as well as IgM) molecules, leads to a much higher antibody level in the blood, and occurs with almost no delay.

appearing in the blood after a lag of a few days (Figure 17.13). IgM molecules have a relatively short half-life (about 5 days), and their appearance is followed by secretion of longer-lived IgG and/or IgE molecules. IgG molecules are the predominant antibodies found in the blood and lymph during a secondary response to most antigens (Figure 17.13). IgE molecules are produced at high levels in response to many parasitic infections. IgE molecules are also bound with high affinity to the surface of mast cells, triggering histamine release, which causes inflammation and symptoms of allergy. IgA is the predominant antibody in secretions of the respiratory, digestive, and urogenital tracts. The function of IgD is unclear.

There are two types of light chains, kappa (κ) chains and lambda (λ) chains, both of which are present in the immunoglobulins of all five classes. In contrast, each immunoglobulin class has a unique heavy chain that defines that class (Table 17.2).[3] We will focus primarily on the structure of IgGs. An IgG molecule is composed of two identical light chains and two identical heavy chains arranged to form a Y-shaped molecule, as shown in Figure 17.14a and described below.

To determine the basis of antibody specificity, it was first necessary to determine the amino acid sequence of a number of specific antibodies. Normally, the first step in amino acid sequencing is to purify the particular protein to be studied. Under normal conditions, however, it is impossible to obtain a purified preparation of a specific antibody from the blood because each person produces a large number of different antibody molecules that are too similar in structure to be separated from one another. The problem was solved when it was
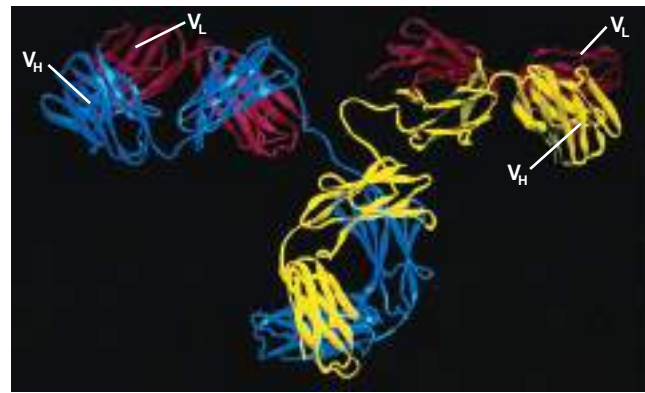
(a)
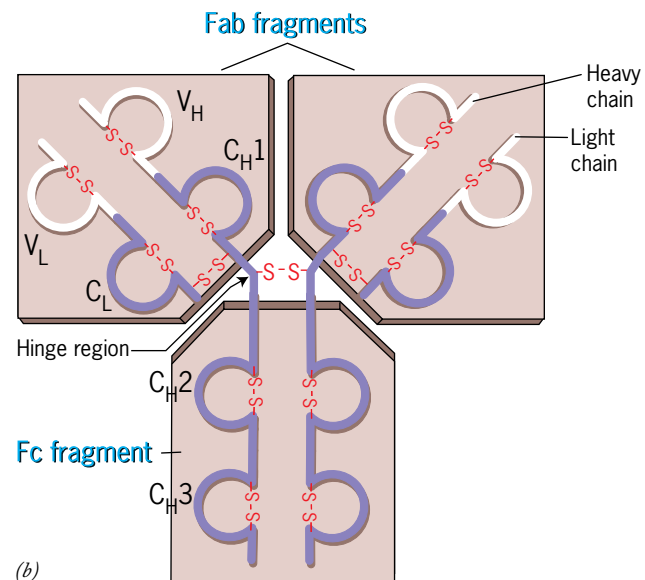
(b)

FIGURE 17.14 **Antibody structure.** (*a*) Ribbon model of an IgG molecule. The molecule contains four polypeptide chains: two identical light chains and two identical heavy chains. One of the heavy chains is shown in blue, the other in yellow, while both light chains are shown in red. The domains of each chain (two per light chain and four per heavy chain) are evident. (*b*) Schematic model showing the domain structure of an IgG molecule. The tertiary structure of each Ig domain is maintained by a disulfide bond. Domains comprising a constant region of the polypeptide chain are indicated by the letter C; domains comprising a variable region are indicated by the letter V. Each heavy chain contains three C_H regions (C_H1, C_H2, C_H3) and one V_H region at the N-terminus of the polypeptide. Each light chain contains one C_L and one V_L region at its N-terminus. The variable regions of each light and heavy chain form an antigen-combining site. Each Y-shaped IgG molecule contains two antigen-combining sites. Each IgG molecule can be fragmented by mild proteolytic treatment into two Fab fragments that contain the antigen-combining sites and an Fc fragment, as indicated. (A: COURTESY OF ALEXANDER MCPHERSON.)

discovered that the blood of patients suffering from a type of lymphoid cancer called multiple myeloma contained large quantities of a single antibody molecule.

As described in Chapter 16, cancer is a monoclonal disease; that is, the cells of a tumor arise from the proliferation of

a single wayward cell. Because a single lymphocyte *normally* produces only a single species of antibody, a patient with multiple myeloma produces large amounts of the specific antibody that is synthesized by the particular cell that became malignant. One patient produces one highly abundant antibody species, and other patients produce different antibodies. As a result, investigators were able to purify substantial quantities of several antibodies from a number of patients and compare their amino acid sequences. An important pattern was soon revealed. It was found that half of each kappa light chain (110 amino acids at the amino end of the polypeptide) has a constant amino acid sequence among all kappa chains, whereas the other half varies from patient to patient. A similar comparison of amino acid sequences of several lambda chains from different patients revealed that they too consist of a section of constant sequence and a section whose sequence varies from one immunoglobulin to the next. The heavy chains of the purified IgGs also contain a variable (V) and a constant (C) portion. A schematic structure of one of these IgG molecules is shown in Figure 17.14*b*.

It was further found that, whereas approximately half of each light chain consists of a variable region (V_L), only one-quarter of each heavy chain is variable (V_H) among different patients; the remaining three-quarters of the heavy chain (C_H) is constant for all IgGs. The constant portion of the heavy chain can be divided into three sections of approximately equal length that are clearly homologous to one another. These homologous Ig units are designated C_H1, C_H2, and C_H3 in Figure 17.14*b*. It appears that the three sections of the C part of the IgG heavy chain (as well as those of heavy chains of the other Ig classes and the C portions of both kappa and lambda light chains) arose during evolution by the duplication of an ancestral gene that coded for an Ig unit of approximately 110 amino acids. The variable regions (V_H or V_L) are also thought to have arisen by evolution from the same ancestral Ig unit. Structural analysis indicates that each of the homologous Ig units of a light or heavy chain folds independently to form a compact domain that is held together by a disulfide bond (Figure 17.15). In an intact IgG molecule, each light chain domain associates with a heavy chain domain as shown in Figure 17.14*a,b*. Genetic analysis indicates that each domain is encoded by its own exon.

The specificity of an antibody is determined by the amino acids of the antigen-combining sites at the ends of each arm of the Y-shaped antibody molecule (Figure 17.14). The two combining sites of a single IgG molecule are identical, and each is formed by the association of the variable portion of a light chain with the variable portion of a heavy chain (Figure 17.14). Assembly of antibodies from different combinations of light and heavy chains allows a person to produce a tremendous variety of antibodies from a modest number of different polypeptides (page 698).

A closer look at the polypeptides of immunoglobulins reveals that the variable portions of both heavy and light chains contain subregions that are especially variable, or *hypervariable*, from one antibody to another (labeled Hv in Figure 17.15). Light and heavy chains both contain three
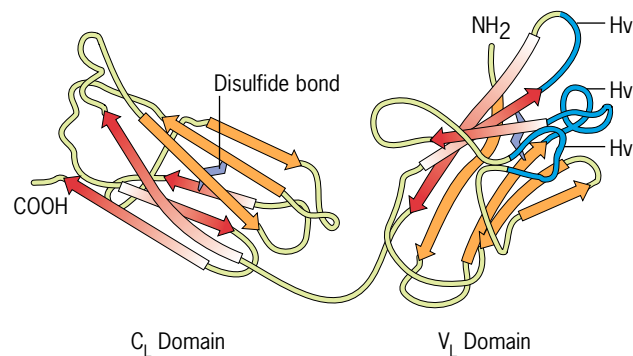
FIGURE 17.15 Antibody domains. Schematic drawing of a human lambda light chain synthesized by cells from a patient with multiple myeloma. The polypeptide undergoes folding so that the constant and variable portions are present in separate domains. Thick arrows represent β strands, which are assembled into β sheets. Each domain has two β sheets, which are distinguished by the red and orange colors. The three hypervariable (Hv) segments of the chain are present as loops at one end of the variable domain, which forms part of the antigen-combining site of the antibody. (AFTER M. SCHIFFER ET AL., BIOCHEMISTRY 12:4628, 1973; © COPYRIGHT 1973, AMERICAN CHEMICAL SOCIETY.)

hypervariable stretches that are clustered at the ends of each arm on the antibody molecule. As might be expected, the hypervariable regions play a prominent role in forming the structure of the antigen-combining site, which can range from a deep cleft to a narrow groove or a relatively flat pocket. Variations in the amino acid sequence of hypervariable regions account for the great diversity of antibody specificity, allowing these molecules to bind to antigens of every conceivable shape.

The combining site of an antibody has a complementary stereochemical structure to a particular portion of the antigen, which is called the **epitope** (or *antigenic determinant*). Because of their close fit, antibodies and antigens form stable complexes, even though they are joined only by noncovalent forces that individually are quite weak. The precise interaction between a particular antigen and antibody, as determined by X-ray crystallography, is shown in Figure 17.16. The two hinge regions within the molecule (Figure 17.14) provide the flexibility necessary for the antibody to bind to two separate antigen molecules or to a single molecule with two identical epitopes.

Whereas the hypervariable portions of light and heavy chains determine an antibody's combining site specificity, the remaining portions of the variable domains provide a scaffold that maintains the overall structure of the combining site. The constant portions of antibody molecules are also important. Different classes of antibodies (IgA, IgD, IgE, IgG, and IgM) have different heavy chains, whose constant regions differ considerably in length and sequence. These differences enable antibodies of various classes to carry out different biological (effector) functions. For example, heavy chains of an IgM molecule bind and activate one of the proteins of the complement system, which leads to the lysis of bacterial cells to which the IgM molecules are bound. Heavy chains of IgE
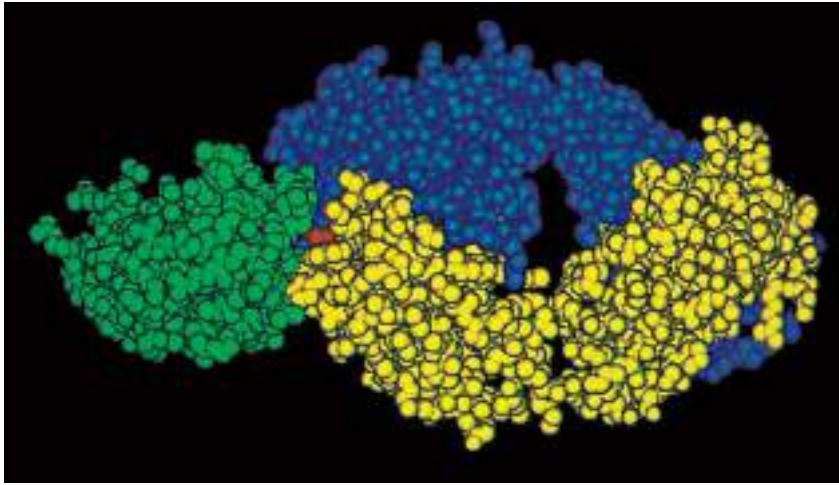
FIGURE 17.16 **Antigen–antibody interaction.** Space-filling model based on X-ray crystallography of a complex between lysozyme (green) and the Fab portion of an antibody molecule (see Figure 17.14). The heavy chain of the antibody is blue; the light chain is yellow. A glutamine residue of lysozyme is red. (REPRINTED WITH PERMISSION FROM A. G. AMIT, R. A. MARIUZZA, S. E. V. PHILLIPS, AND R. J. POLJAK, SCIENCE 233:749, 1986; © COPYRIGHT 1986, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

molecules play an important role in allergic reactions by binding to specific receptors on the surfaces of mast cells, triggering the release of histamine. In contrast, heavy chains of an IgG molecule bind specifically to the surface receptors of macrophages and neutrophils, inducing these phagocytic cells to ingest the particle to which the antibodies are bound. The heavy chains of IgG molecules are also important in allowing this class of antibodies to pass from the blood vessels of a mother to those of her fetus during pregnancy. While this provides the fetus and newborn with passive immunity to infectious organisms, it may cause a life-threatening condition called erythroblastosis fetalis. For this condition to occur, an Rh^- mother must have given birth to a child with an Rh^+ phenotype (Rh^+/Rh^- genotype) during a previous pregnancy. The mother is usually exposed to the Rh^+ fetal antigen during delivery of the first child, who is not affected. If, however, the mother should have a second Rh^+ pregnancy, antibodies present in her bloodstream can enter the fetal circulation and destroy the red blood cells of the fetus. Babies with this condition are given a blood transfusion, either before or after birth, which cleanses the blood of maternal antibodies.

DNA Rearrangement of Genes Encoding B- and T-Cell Antigen Receptors

As discussed above, each IgG molecule is composed of two light (L) chains and two heavy (H) chains. Both types of polypeptides consist of two recognizable parts—a variable (V) portion, whose amino acid sequence varies from one antibody species to another, and a constant (C) portion, whose amino acid sequence is identical among all H or L chains of the same class. What is the genetic basis for the synthesis of polypeptides having a combination of shared and unique amino acid sequences?

In 1965, William Dreyer of the California Institute of Technology and J. Claude Bennett of the University of Alabama put forward the "two gene–one polypeptide" hypothesis to account for antibody structure. In essence, Dreyer and Bennett proposed that each antibody chain is encoded by two separate genes—a C gene and a V gene—that somehow combine to form one continuous "gene" that codes for a single light or heavy chain. In 1976, Susumu Tonegawa, working at a research institute in Basel, Switzerland, provided clear evidence in favor of the DNA rearrangement hypothesis. The basic outline of the experiment is shown in Figure 17.17. In this experiment, Tonegawa and his colleagues compared the length of DNA between the nucleotide sequences coding for the C and V portions of a specific antibody chain in two different types of mouse cells: early embryonic cells and malignant, antibody-producing myeloma cells. The DNA segments encoding the C and V portions of the antibody were widely separated in embryonic DNA but were very close to each other in DNA obtained from antibody-producing myeloma cells (Figure 17.17). These findings strongly suggested that DNA segments encoding the parts of antibody molecules became rearranged during the formation of antibody-producing cells.

Subsequent research revealed the precise arrangement of the DNA sequences that give rise to antibody genes. To simplify the discussion, we will consider only those DNA sequences involved in the formation of human kappa light chains, which are located on chromosome 2. The organization of sequences in germ-line DNA (i.e., the DNA of a sperm or egg) that are involved in the formation of human kappa light chains is shown in the top line (step 1) of Figure 17.18. In this case, a variety of different V_κ genes are located in a linear array and separated from a single C_κ gene by some distance. Nucleotide-sequence analysis of these V genes indicated that they are shorter than required to encode the V region of the kappa light chain. The reason became clear when other segments in the region were sequenced. The stretch of nucleotides that encodes the 13 amino acids at the carboxyl end of a V region is located at some distance from the remainder of the V_κ gene sequence. This small portion that encodes the carboxyl end of the V region is termed the *J segment*. As shown in Figure 17.18 there are five distinct J_κ segments of related nucleotide sequence arranged in tandem. The cluster of J_κ segments is separated from the C_κ gene by an additional stretch of over 2000 nucleotides. A complete kappa V gene is formed, as shown in Figure 17.18 (steps 2–3), as a specific V_κ gene is joined to one of the J_κ segments with the intervening

DNA excised. This process is catalyzed by a protein complex called *V(D)J recombinase*. As indicated in Figure 17.18, the V_κ gene sequence generated by this DNA rearrangement is still separated from the C_κ gene by more than 2000 nucleotides. No further DNA rearrangement occurs in kappa gene assembly prior to transcription; the entire genetic region is transcribed into a large primary transcript (step 4) from which introns are excised by RNA splicing (step 5).

Rearrangement begins as double-stranded cuts are made in the DNA between a V gene and a J gene. The cuts are catalyzed by a pair of proteins called RAG1 and RAG2 that are part of the V(D)J recombinase. The four free ends that are generated are then joined in such a way that the V and J coding

FIGURE 17.17 Experimental demonstration that genes encoding antibody light chains are formed by DNA rearrangement. DNA is first extracted from either embryonic cells or antibody-producing cancer cells and fragmented by a restriction endonuclease (step 1) that cleaves both strands of the DNA at a specific sequence. The DNA fragments from the two preparations are fractionated separately by gel electrophoresis; two identical gels are prepared from each DNA sample (step 2). Following electrophoresis, each of the gels is incubated with a labeled probe containing either the variable (V) or constant (C) gene sequence (step 3). The location of the bound, labeled DNA in the gel is revealed by autoradiography and shown at the bottom of the figure. Whereas the V and C gene sequences are located on separate fragments in DNA from embryonic cells, the two sequences are located on the same small fragment in DNA from the antibody-producing cells. The V and C gene sequences are brought together during B-cell development by a process of DNA rearrangement.

FIGURE 17.18 DNA rearrangements that lead to the formation of a functional gene that encodes an immunoglobulin κ chain. The organization of the variable (V), joining (J), and constant (C) DNA sequences within the genome is shown in step 1. Steps leading to the synthesis of the mature mRNA that codes for the κ chain polypeptide are described in the text. The random union of a V and J segment (steps 2 and 3) determines the amino acid sequence of the polypeptide. The space between the "chosen" J segment and the C segment (which may contain one or more J segments, as shown in the figure) remains as an intron in the gene. The portion of the primary transcript (step 4) that corresponds to this intron is removed during RNA processing (step 5). DNA rearrangement and subsequent transcription of the "stitched" gene occurs on only one allele in each cell, which ensures that the cell will only express a single κ chain. The other allele on the homologous chromosome typically remains unaltered and is not transcribed.

segments are linked to form an exon that encodes the variable region of the polypeptide chain, while the two ends of the intervening DNA are linked to form a small circular piece of DNA that is displaced from the chromosome (Figure 17.18, step 3). Joining of the broken DNA ends is accomplished by the same basic process used to repair DNA strand breaks that was depicted in Figure 13.29.

The rearrangement of Ig DNA sequences has important consequences for a lymphocyte. Once a specific V_κ sequence is joined to a J_κ sequence, no other species of kappa chain can be synthesized by that cell. It is estimated that the DNA of human germ cells contains approximately 40 functional V_κ genes. Thus, if we assume that any V sequence may join to any J sequence, we expect that a person can synthesize approximately 200 different kappa chains (5 J_κ segments $\times$ 40 V_κ genes). But this is not the only source of diversity among these polypeptides. The site at which a J sequence is joined to a V sequence can vary somewhat from one rearrangement to another so that the same V_κ and J_κ genes can be joined in two different cells to produce kappa light chains having different amino acid sequences. Additional variability is achieved by the enzyme deoxynucleotidyl transferase, which inserts nucleotides at the sites of strand breakage. These sources of additional variability increase the diversity of kappa chains an additional tenfold, bringing the number to at least 2000 species. The site at which V and J sequences are joined is part of one of the hypervariable regions of each antibody polypeptide (Figure 17.15). Thus, slight differences at a joining site can have important effects on antibody–antigen interaction.

We have restricted our discussion to kappa light chains for the sake of simplicity. Similar types of DNA rearrangement occur during the commitment of a cell to the synthesis of a particular lambda light chain and to a particular heavy chain. Whereas the variable regions of light chains are formed from two distinct segments (V and J segments), the variable regions of heavy chains are formed from three distinct segments (V, D, and J segments) by similar types of rearrangement. The human genome contains 51 functional V_H segments, 25 D_H segments, and 6 J_H segments. Given the additional diversity stemming from the variability in V_H–D_H and D_H–J_H joining, a person is able to synthesize at least 100,000 different heavy chains. The antigen receptors of T cells (TCRs) also consist of a type of heavy and light chain, whose variable regions are formed by a similar process of DNA rearrangement.

The formation of antibody genes by DNA rearrangement illustrates the potential of the genome to engage in dynamic activities. Because of this rearrangement mechanism, a handful of DNA sequences that are present in the germ line can give rise to a remarkable diversity of gene products. As discussed above, a person synthesizes roughly 2000 different species of kappa light chains and 100,000 different species of heavy chains. If any kappa light chain can combine with any heavy chain, a person can theoretically produce more than 200 million different antibody species from a few hundred genetic elements present in the germ line.[4]

We have seen how antibody diversity arises from (1) the presence of multiple V exons, J exons, and D exons in the DNA of the germ line, (2) variability in V–J and V–D–J joining, and (3) the enzymatic insertion of nucleotides. An additional mechanism for generating antibody diversity, referred to as *somatic hypermutation*, occurs long after DNA rearrangement is complete. When a specific antigen is reintroduced into an animal following a period of time, antibodies produced during the secondary response have a much greater affinity for the antigen than those produced during the primary response. Increased affinity is due to small changes in amino acid sequence of variable regions of the heavy and light antibody chains. These sequence changes result from mutations in the genes that encode these polypeptides. It is estimated that rearranged DNA elements encoding antibody V regions have a mutation rate 10^5 times greater than that of other genetic loci in the same cell. The mechanism responsible for this increased level of V-region mutation has been the focus of a number of interesting studies in recent years. Included in the mechanism is (1) an enzyme—known as activation-induced cytosine deaminase (AID)—that converts cytosine residues in DNA into uracil residues and (2) one or more translesion DNA polymerases (page 557) that tend to make errors when DNA containing uracils is copied or repaired. Persons who carry mutations in AID and are unable to generate somatic hypermutation are plagued by infections and often die at an early age.

Somatic hypermutation generates random changes in the V regions of Ig genes. Those B cells whose genes produce Ig molecules with greater antigen affinity are preferentially selected following antigen reexposure. Selected cells proliferate to form clones that undergo additional rounds of somatic mutation and selection, whereas nonselected cells that express low affinity Igs undergo apoptosis. In this way, the antibody response to recurrent or chronic infections improves markedly over time.

Once a B cell is committed to form a specific antibody, it can switch the class of Ig it produces (e.g., from IgM to IgG) by changing the heavy chain produced in the cell. This process, known as *class switching*, occurs without changing the combining site of the antibodies synthesized. Recall that there are five different types of heavy chains distinguished by their constant regions. The genes that encode the constant regions of heavy chains (C_H portions) are clustered together in a complex, as shown in Figure 17.19. Class switching is accomplished by moving a different C_H gene next to the VDJ gene that was formed previously by DNA rearrangement. Class switching is under the direction of cytokines secreted by helper T cells during their interaction with the B cell produc-
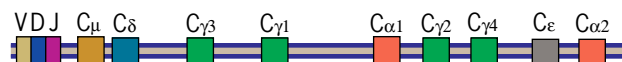
FIGURE 17.19 Arrangement of the C genes for the various human heavy chains. In humans, the heavy chains of IgM, IgD, and IgE are encoded by a single gene, whereas those of IgG are encoded by four different genes and IgA by two different genes (see footnote, page 694).

---

[4] A roughly comparable number of antibodies containing lambda light chains can also be generated.

ing the antibody molecule. For example, a helper T cell that secretes IFN-γ, induces a switch in the adjacent B cell from its initial synthesis of IgM to synthesis of one of the IgG classes (Figure 17.13). Class switching allows a lineage of B cells to continue to produce antibodies having the same specificity but different effector functions (page 695).

Membrane-Bound Antigen Receptor Complexes

The recognition of antigen by both B and T lymphocytes occurs at the cell surface. An antigen receptor on a B cell (a B-cell receptor, or BCR) consists of a membrane-bound immunoglobulin that binds selectively to a portion of an intact antigen (i.e., the epitope) (Figure 17.20a). In contrast, the antigen receptor on a T cell (a T-cell receptor, or TCR, Figure 17.20b) recognizes and binds to a small fragment of an antigen, typically a peptide about 7 to 25 amino acids in length, that is held at the surface of another cell (described below). Both types of antigen receptors are part of large membrane-bound protein complexes that include invariant proteins as depicted in Figure 17.20. The invariant polypeptides associated with BCRs and TCRs play a key role in transmitting signals to the interior that lead to changes in activity of the B cell or T cell.

Each subunit of a TCR contains two Ig-like domains, indicating that they share a common ancestry with BCRs. Like the heavy and light chains of the immunoglobulins, one of the Ig-like domains of each subunit of a TCR has a variable amino acid sequence; the other domain has a constant amino acid sequence (Figure 17.20). X-ray crystallographic studies have shown that the two types of antigen receptors also share a similar three-dimensional shape.

The Major Histocompatibility Complex

During the first part of the twentieth century, clinical researchers discovered that blood could be transfused from one person to another, as long as the two individuals were compatible for the ABO blood group system. The success of blood transfusion led to the proposal that skin might also be grafted between individuals. This idea was tested during World War II when skin grafts were attempted on pilots and other military personnel who had received serious burns. The grafts were rapidly and completely rejected. After the war, researchers set out to determine the basis for tissue rejection. It was discovered that skin could be grafted successfully between mice of the same inbred strain, but that grafts between mice of different strains were rapidly rejected. Mice of the same inbred strain are like identical twins; they are genetically identical. Subsequent studies revealed that the genes that governed tissue graft rejection were clustered in a region of the genome that was named the **major histocompatibility complex (MHC)**. Approximately 20 different MHC genes have been characterized, most of which are highly polymorphic: over 2000 different alleles of MHC genes have been identified (Table 17.3), far more than any other loci in the human genome. It is very unlikely, therefore, that two individuals in a population have the same combination of MHC alleles. This
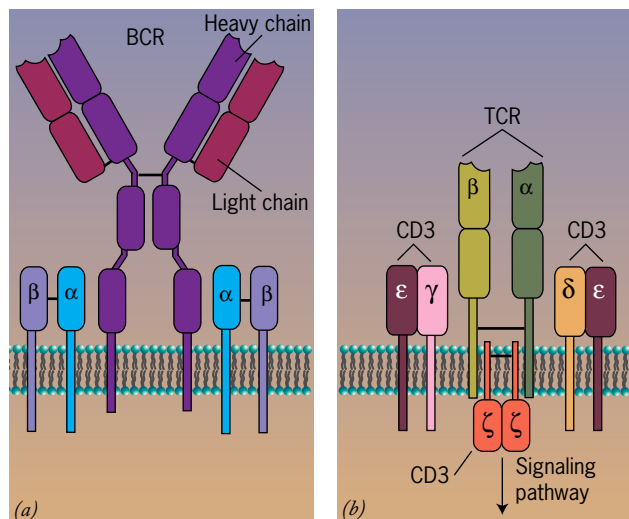
FIGURE 17.20 Structure of the antigen receptors of a B cell and a T cell. (a) The BCR of a B cell is a membrane-bound version of an immunoglobulin associated with a pair of invariant α chains and a pair of invariant β chains. The α and β chains are also members of the Ig superfamily. (b) The TCR of a T cell consists of an α and β polypeptide chain linked to one another by a disulfide bridge. Each polypeptide contains a variable domain that forms the antigen-binding site and a constant domain. The TCR is associated with six other invariant polypeptides of the CD3 protein as indicated in the illustration. (A small fraction of T cells contain a different type of TCR consisting of a γ and δ subunit. These cells are not restricted to recognition of MHC–peptide complexes, and their function is not well understood.)

TABLE 17.3

MHC Class II Alleles	
Locus	Number of alleles
HLA-DRA	3
HLA-DRB	542
HLA-DQA	34
HLA-DQB	73
HLA-DPA	23
HLA-DPB	125
HLA-DMA	4
HLA-DMB	7
HLA-DOA	12
HLA-DOB	9

MHC Class I Alleles	
Locus	Number of alleles
HLA-A	479
HLA-B	805
HLA-C	257
HLA-E	9
HLA-F	20
HLA-G	7

Note: Several other class I alleles are not listed.

is the reason that transplanted organs are so likely to be rejected and why transplant patients are given drugs, such as cyclosporin A, to suppress the immune system following surgery. Cyclosporin A is a cyclic peptide produced by a soil fungus. Cyclosporin A inhibits a particular phosphatase in the signaling pathway leading to the production of cytokines required for T-cell activation. Although these drugs help prevent graft rejection, they make patients susceptible to opportunistic infections similar to those that strike people with immunodeficiency diseases such as AIDS.

It is obvious that proteins encoded by the MHC did not evolve to prevent indiscriminate organ transplantation, which raises the question of their normal role. Long after their discovery as transplantation antigens, MHC proteins were shown to be involved in antigen presentation. Some of the key experiments that led to our current understanding of antigen presentation are discussed in the Experimental Pathways at the end of the chapter.

It was noted earlier that T cells are activated by an antigen that has been dissected into small peptides and displayed on the surface of an antigen-presenting cell (an APC). These small fragments of antigen are held at the surface of the APC in the grip of MHC proteins. Each species of MHC molecule can bind a large number of different peptides that share certain structural features that allow them to fit into its binding site (see Figure 17.24). For example, all of the peptides that are capable of binding to a protein encoded by a particular MHC allele, such as HLA-B8, may contain a specific amino acid at a certain position, which allows it to fit into the peptide-binding groove.

Given the fact that each individual expresses a number of different MHC proteins (as in Figure 17.21a), and each MHC variant may be able to bind large numbers of different peptides (as in Figure 17.21b), a dendritic cell or macrophage should be able to display a vast array of peptides. At the same time, not every person is capable of presenting every possible peptide in an effective manner, which is thought to be a major factor in determining differences in susceptibility in a population to different infectious diseases, including AIDS. For example, the HLA-B*35 allele is associated with rapid progression to full-blown AIDS and the HLA-DRB1*1302 allele is correlated with resistance to a certain type of malaria and hepatitis B infection. The MHC alleles present in a given population have been shaped by natural selection. Those persons who possess MHC alleles that are best able to present the peptides of a particular infectious agent will be the most likely to survive an infection by that agent. Conversely, persons who lack these alleles are more likely to die without passing on their alleles to offspring. As a result, populations tend to be more resistant to diseases to which their ancestors have been routinely exposed. This could explain why Native American populations have been devastated by certain diseases, such as measles, that produce only mild symptoms in persons of European ancestry.

The entire process of T-cell-mediated immunity rests on the basis that small peptides derived from the proteins of a pathogen differ in structure from those derived from the pro-

teins of the host. Consequently, one or more peptides held at the surface of an APC serve as a small representation of the pathogen, providing the cells of the immune system with a "glimpse" of the type of pathogen that is hidden within the cytoplasm of the infected cell. Nearly any cell in the body can function as an APC. Most cells present antigen as an incidental activity that alerts the immune system to the presence of a pathogen, but some professional APCs (e.g., dendritic cells, macrophages, and B cells) are specialized for this function, as discussed later in the chapter.

When a T cell interacts with an APC, it does so by having its TCRs dock onto the MHC molecules projecting from the APC's surface (Figure 17.22). This interaction brings a TCR of a T cell into an orientation that allows it to recognize the specific peptide displayed within a groove of an MHC molecule. The interaction between MHC proteins and TCRs is strengthened by additional contacts that form between cell-surface components, such as occur between CD4 or CD8 molecules on a T cell and MHC proteins on an APC (Figure 17.22). This specialized region that develops between a T cell and an APC is referred to as an *immunologic synapse*.

(a)

(b)
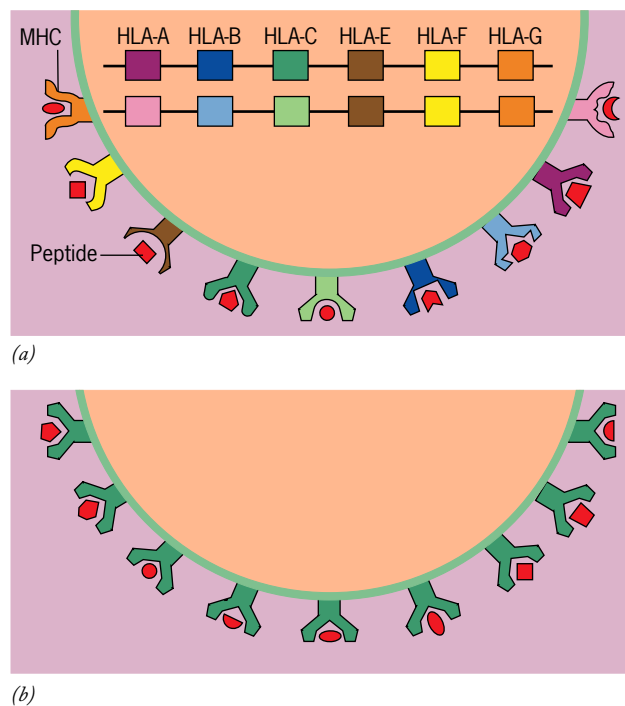
FIGURE 17.21 Human APCs can present large numbers of peptides.
(*a*) Schematic model of the variety of class I MHC molecules an individual might possess. As indicated in Table 17.3, this class of MHC protein is encoded by a number of genes, three of which are represented by a large number of alleles. This particular individual is heterozygous at the HLA-A, -B, and -C loci and homozygous at the HLA-E, -F, and -G loci, which gives them a total of nine different MHC class I molecules. (The difference between class I and II MHC is discussed on page 701.) (*b*) Schematic model illustrating the variety of peptides that can be presented by the protein encoded by a single MHC allele. (The term *HLA* is an acronym for human leukocyte antigen reflecting the discovery of these proteins on the surface of leukocytes.)

(a)

(b)

(c)
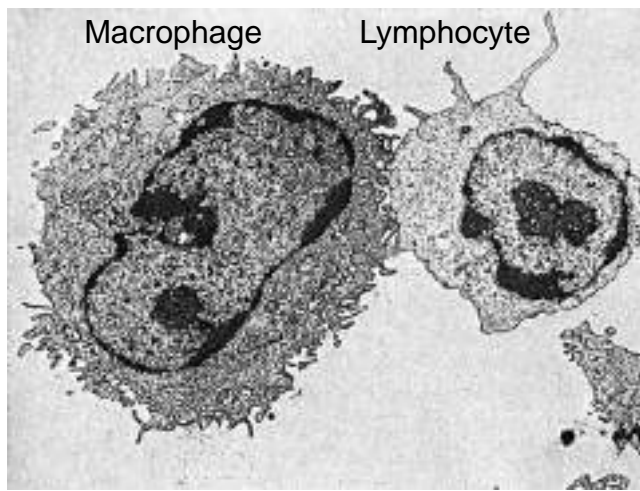
FIGURE 17.22 Interaction between an antigen-presenting cell and a T cell during antigen presentation. (*a*) Electron micrograph of the two types of cells as they interact. (*b*) Schematic model showing some of the proteins present in the immunologic synapse that forms in the region of interaction between an APC and a cytotoxic T lymphocyte (CTL) or helper T (T$_H$) cell. Antigen recognition occurs as the TCR of the T cell recognizes a peptide fragment of the antigen bound to a groove in an MHC molecule of the APC. As discussed in the text, CTLs recognize antigen in combination with an MHC class I molecule, whereas T$_H$ cells recognize antigen in combination with an MHC class II molecule. CD8 and CD4 are integral membrane proteins expressed by the two types of T cells that bind MHC class I and class II molecules, respectively. CD8 and CD4 are described as coreceptors. (*c*) Fluorescence micrograph showing the immunologic synapse between an APC and a T cell. The TCRs of the T cell appear green, the MHC II molecules of the APC appear red. The colocalization of the TCR and MHC molecules generates the yellow fluoresence in the immunologic synapse. (A: ALAN S. ROSENTHAL, FROM REGULATION OF THE IMMUNE RESPONSE—ROLE OF THE MACROPHAGE, NEW ENGLAND JOURNAL OF MEDICINE, NOVEMBER
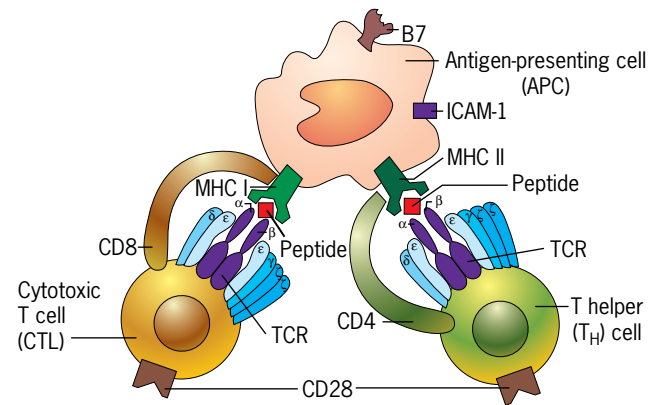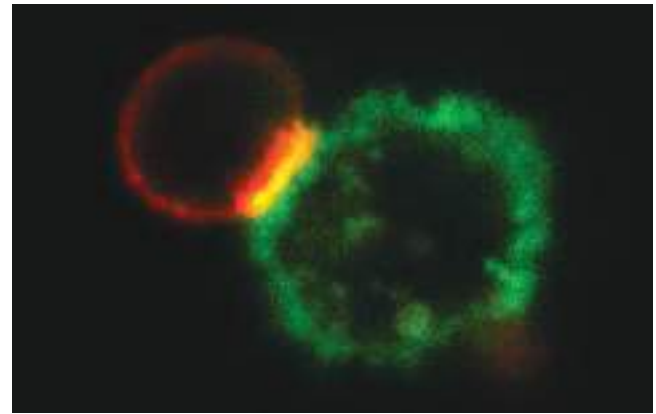
1980, VOL. 303, #20, P. 1154 © 1980 MASSACHUSETTS MEDICAL SOCIETY; B: AFTER L. CHATENOUD, MOL. MED. TODAY 4:26, 1998, WITH PERMISSION FROM ELSEVIER SCIENCE; C: FROM BRIAN A. COBB ET AL., CELL 117:683, 2004, COURTESY OF DENNIS L. KASPER; BY PERMISSION OF CELL PRESS.)

MHC proteins can be subdivided into two major groups, **MHC class I** and **MHC class II** molecules. MHC class I molecules consist of one polypeptide chain encoded by an MHC allele (known as the heavy chain) associated noncovalently with a non-MHC polypeptide known as β_2-microglobulin (see Figure 1, page 711). Differences in the amino acid sequence of the heavy chain are responsible for dramatic changes in the shape of the molecule's peptide-binding groove. MHC class II molecules also consist of a heterodimer, but both subunits are encoded by MHC alleles. Both classes of MHC molecule as well as β_2-microglobulin contain Ig-like domains and are thus members of the immunoglobulin superfamily. Whereas most cells of the body express MHC class I molecules on their surface, MHC class II molecules are expressed primarily by professional APCs.

The two classes of MHC molecules display antigens that originate from different sites within a cell, though some overlap can occur. MHC class I molecules are predominantly responsible for displaying antigens that originate within the cytosol of a cell, that is, *endogenous* proteins. In contrast,

MHC class II molecules primarily display fragments of exogenous antigens that are taken into a cell by phagocytosis. The proposed pathways by which these two classes of MHC molecules pick up their antigen fragments and display them on the plasma membrane are described here and shown in Figure 17.23.

■ *Processing of class I MHC–peptide complexes* (Figure 17.23*a*). Antigens located in the cytosol of an APC are degraded into short peptides by proteases that are part of the cell's proteasomes (page 529). These proteases cleave cytosolic proteins into fragments approximately 8 to 10 residues long that are suitable for binding within a groove of an MHC class I molecule (Figure 17.24*a*). The peptides are then transported across the membrane of the rough endoplasmic reticulum and into its lumen by a dimeric protein called TAP (Figure 17.23*a*). Once in the ER lumen, the peptide can bind to a newly synthesized MHC class I molecule, which is an integral protein of the ER membrane. The MHC–peptide complex moves through the
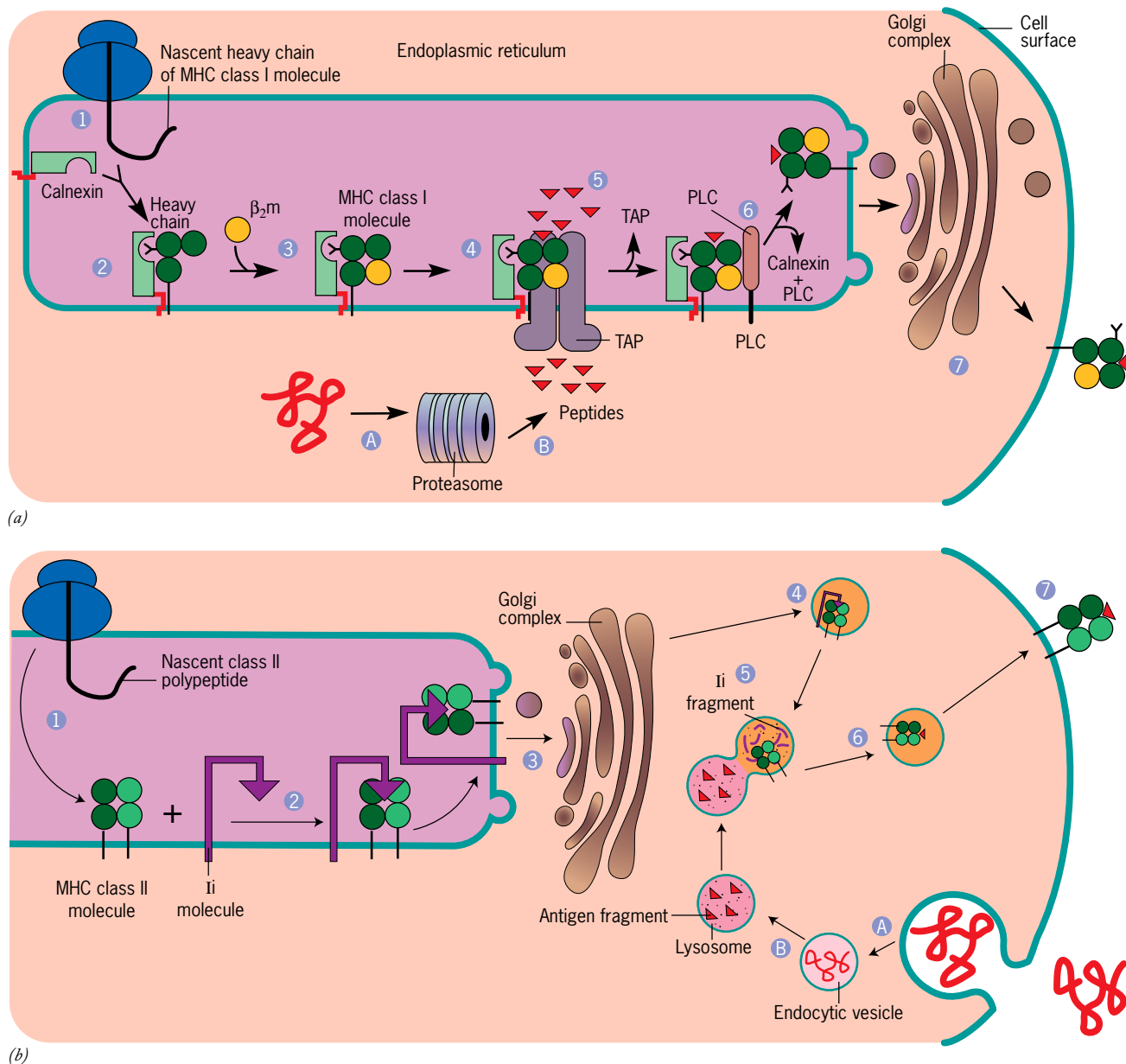
FIGURE 17.23 Classical processing pathways for antigens that become associated with MHC class I and class II molecules. (*a*) A proposed pathway for the assembly of an MHC class I–peptide complex. This pathway occurs in almost all cell types. In step 1, the heavy chain of the MHC protein is synthesized by a membrane-bound ribosome and translocated into the ER membrane. The MHC heavy chain becomes associated with calnexin (step 2), a chaperone in the ER membrane, and the dimeric complex binds the invariant β₂m chain (step 3). The MHC complex then becomes associated with another ER membrane protein, TAP (step 4). Meanwhile, cytosolic antigens are taken into proteasomes (step A) and degraded into small peptides (step B). The peptides are transported into the ER lumen by the TAP protein, where they become bound within the groove of the MHC molecule (step 5) with the help of a large complex of chaperones that is labeled PLC in the figure. PLC and calnexin dissociate from the MHC complex (step 6), which is transported along the biosynthetic/secretory pathway through the Golgi complex (step 7) to the plasma membrane, where it is ready to interact with the TCR of a CTL. (*b*) A proposed pathway for the assembly of an MHC class II–peptide complex. This pathway occurs in dendritic cells and other professional APCs. In step 1, the MHC pro-

tein is synthesized by a membrane-bound ribosome and translocated into the ER membrane, where it becomes associated with Ii (step 2), a trimeric protein that blocks the MHC–peptide-binding site. The MHC–Ii complex passes through the Golgi complex (step 3) and into a transport vesicle (step 4). Meanwhile, an extracellular protein antigen is taken into the APC by endocytosis (step A) and delivered to a lysosome (step B), where the antigen is fragmented into peptides. The lysosome containing antigenic fragments fuses with the transport vesicle containing the MHC–Ii complex (step 5), leading to the degradation of the Ii protein and association between the antigenic peptide fragment and the MHC class II molecule (step 6). The MHC–peptide complex is transported to the plasma membrane (step 7) where it is ready to interact with the TCR of a T_H cell. (*Note*: Not all exogenous antigens follow the classical MHC class II pathway shown in part *b*. A pathway also exists by which exogenous antigens can be taken into APCs by endocytosis and degraded into peptides, which are then bound and displayed by MHC class I molecules. This cross-presentation pathway, as it is called, allows CTLs to become activated by exogenous antigen that would otherwise go "unseen.") (A: AFTER D. B. WILLIAMS ET AL., TRENDS CELL BIOL. 6:271, 1996, WITH PERMISSION FROM ELSEVIER SCIENCE.)
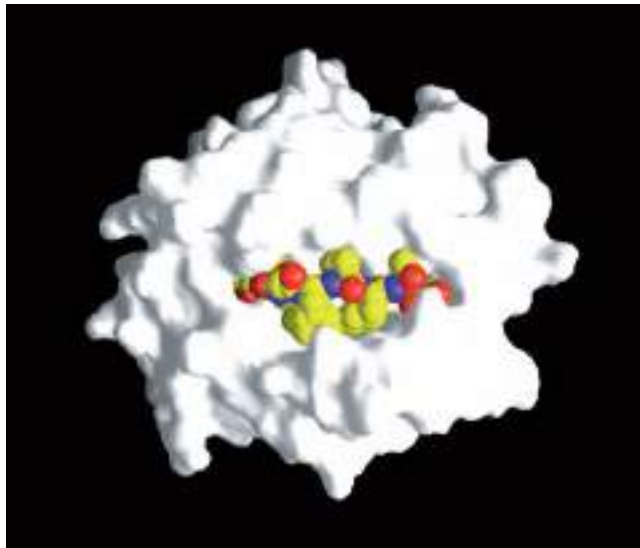
(a)
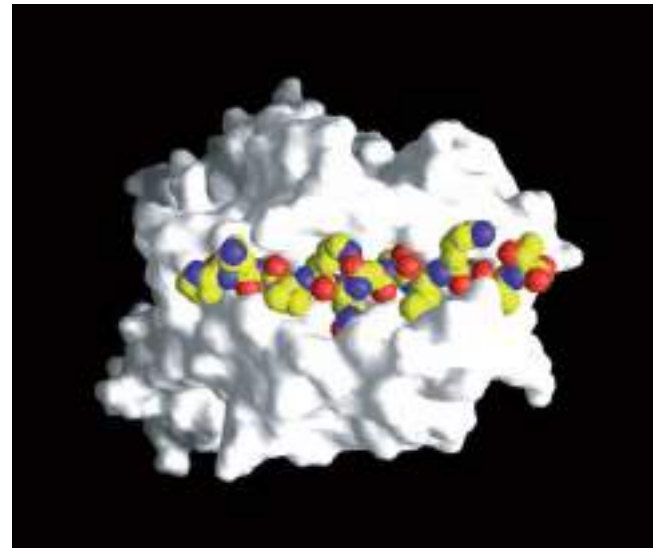
(b)

FIGURE 17.24 Peptides produced by antigen processing bind within a groove of the MHC protein molecule. These models illustrate the binding of peptides to an MHC class I (*a*) and MHC class II (*b*) molecule. The molecular surfaces of the MHC molecules are shown in white and the peptide in the peptide-binding site in color. The peptide in *a* is derived from the influenza virus matrix protein, and the peptide in *b* is derived from influenza virus hemagglutinin protein. The N-terminus of each peptide is on the left. (A,B: COURTESY OF T. JARDETZKY; FROM TRENDS BIOCHEM. SCI. 22:378, 1997, WITH PERMISSION FROM ELSEVIER SCIENCE.)

biosynthetic pathway (Figure 8.2*b*) until it reaches the plasma membrane where the peptide is displayed.

■ *Processing of class II MHC–peptide complexes* (Figure 17.23*b*). MHC class II molecules are also synthesized as membrane proteins of the RER, but they are joined noncovalently to a protein called Ii, which blocks the peptide-binding site of the MHC molecule (Figure 17.23*b*). Following its synthesis, the MHC class II–Ii complex moves out of the ER along the biosynthetic pathway, directed by targeting sequences located within the cytoplasmic domain of Ii. MHC class I and II molecules are thought to separate from one another in the *trans* Golgi network (TGN), which is the primary sorting compartment along the biosynthetic pathway (page 292). An MHC class I–peptide complex is directed toward the cell surface, whereas an MHC class II–Ii complex is directed into an endosome or lysosome, where the Ii protein is digested by acid proteases. An MHC class II molecule is then free to bind to peptides digested from antigens that were taken into the cell and directed along the endocytic pathway (Figure 17.23*b*).[5] The MHC class II–peptide complex is then moved to the plasma membrane where it is displayed, as shown in Figure 17.24*b*.

Once on the surface of an APC, MHC molecules direct the cell's interaction with different types of T cells (Figure 17.22). Cytotoxic T lymphocytes (CTLs) recognize their antigen in association with MHC class I molecules; they are said to be *MHC class I restricted*. Under normal circumstances, cells of the body that come into contact with CTLs display fragments of their own normal proteins in association with their MHC class I molecules. Normal cells displaying normal protein fragments are ignored by the body's T cells because T cells capable of binding with high affinity to peptides derived from normal cellular proteins are eliminated as they develop in the thymus. In contrast, when a cell is infected, it displays fragments of viral proteins in association with its MHC class I molecules. These cells are recognized by CTLs bearing TCRs whose binding sites are complementary to the viral peptides, and the infected cell is destroyed. The presentation of a single foreign peptide on the surface of a cell is probably sufficient to invite an attack by a CTL. Because virtually all cells of the body express MHC class I molecules on their surface, CTLs can combat an infection regardless of the type of cell affected. CTLs may also recognize and destroy cells that display abnormal (mutated) proteins on their surfaces, which could play a role in the elimination of potentially life-threatening tumor cells.

In contrast to CTLs, helper T cells recognize their antigen in association with MHC class II molecules; they are said to be *MHC class II restricted*. As a result, helper T cells are activated primarily by exogenous (i.e., extracellular) antigens (Figure 17.23*b*), such as those that occur as part of bacterial cell walls or bacterial toxins. MHC class II molecules are found predominantly on B cells, dendritic cells, and macrophages. These are the lymphoid cells that ingest foreign, extracellular materials and present the fragments to helper T cells. Helper T cells that are activated in this way can then stimulate B cells to produce soluble antibodies that bind to the exogenous antigen wherever it is located in the body.

---

[5]Peptides generated in lysosomes and attached to MHC class II molecules tend to be longer (10 to 25 residues) than those generated in proteasomes and attached to MHC class I molecules (typically 8 to 10 residues).

Distinguishing Self from Nonself

T cells gain their identity in the thymus. When a stem cell migrates from the bone marrow to the thymus, it lacks the cell-surface proteins that mediate T-cell function, most notably its TCRs. The stem cells proliferate in the thymus to generate a population of T-cell progenitors. Each of these cells then undergoes the DNA rearrangements that enable it to produce a specific TCR. These cells are then subjected to a complex screening process in the thymus that selects for cells having potentially useful T-cell receptors (Figure 17.25). Studies suggest that epithelial cells of the thymus produce small quantities of a great variety of proteins normally restricted to other tissues throughout the body. Production of these proteins may be under the control of a special transcriptional regulator (called AIRE) that is present only in the thymus. According to this model, the thymus re-creates an environment in which developing T cells can sample proteins containing a vast array of the body's own unique epitopes. T cells whose TCRs have a high affinity for peptides derived from the body's own proteins are destroyed (Figure 17.25a). This process of *negative selection* greatly reduces the likelihood that the immune system will attack its own tissues. Humans lacking a functional *AIRE* gene suffer from a severe autoimmune disease (called APECED), in which numerous organs come under immunologic attack.

The generation of T cells requires more than negative selection. When a TCR interacts with a foreign peptide on the surface of an APC, it must recognize both the peptide and the MHC molecule holding that peptide (discussed on page 712). Consequently, T cells whose TCRs do not recognize self-MHC molecules are of little value. The immune system screens out such cells by requiring that T cells recognize self-MHC–self-peptide complexes with low affinity. T cells whose TCRs are unable to recognize self-MHC complexes die within the thymus due to a lack of growth signals—a process referred to as "death by neglect" (Figure 17.25b). In contrast, T cells whose TCRs exhibit weak (low affinity) recognition toward self-MHC complexes are stimulated to remain alive but are not activated (Figure 17.25c). This process of selective survival is referred to as *positive selection*. It is estimated that less than 5 percent of thymic T cells survive these negative and positive screening events.[6]

Those cells that recognize class I MHC molecules are thought to develop into cytotoxic ($CD4^-CD8^+$) T lymphocytes, whereas those that recognize class II MHC molecules are thought to differentiate into helper ($CD4^+CD8^-$) T lymphocytes. Both types of T cells leave the thymus and circulate for extended periods through the blood and lymph. T cells at this stage are described as *naïve* T cells because they have not yet encountered the specific antigen to which their TCR can bind. As they pass through the lymphoid tissues, naïve T cells come into contact with various cells that either maintain their survival in a resting state or trigger their activation.

As they percolate through the lymph nodes and other peripheral lymphoid tissues, T cells scan the surfaces of cells for the presence of an inappropriate peptide bound to a self-MHC molecule. $CD4^+$ T cells are activated by a foreign peptide bound to a class II self-MHC molecule, whereas $CD8^+$ T cells are activated by foreign peptides bound to a class I self-MHC molecule. $CD8^+$ T cells also respond strongly to cells bearing nonself-MHC molecules, such as the cells of a transplanted organ from a mismatched donor. In this latter case, they initiate a widespread attack against the graft cells that can result in organ rejection. Under normal physiologic conditions, autoreactive lymphocytes (i.e., lymphocytes capable of reacting to the body's own tissues) are prevented from becoming activated by a number of poorly understood mechanisms that operate outside of the thymus in the body's periphery. As discussed in the Human Perspective, a breakdown in these mechanisms leads to the production of autoantibodies and autoreactive T cells that can cause chronic tissue damage.

Lymphocytes Are Activated by Cell-Surface Signals

Lymphocytes communicate with other cells through an array of cell-surface proteins. As discussed above, T-cell activation requires an interaction between the TCR of the T cell and an MHC–peptide complex on the surface of another cell. This interaction provides specificity, which ensures that only T cells that bind the antigen are activated. T-cell activation also requires a second signal, called the *costimulatory signal*, which is delivered through a second type of receptor on the surface of

FIGURE 17.25   **Determining the fate of a newly formed T cell in the thymus.** A screening process takes place in the thymus that selects for T cells with appropriate TCRs. (*a*) Those T cells whose TCR exhibits strong affinity for MHC molecules bearing self-peptides are eliminated by apoptosis (negative selection). (*b*) Those T cells whose TCR fails to recognize MHC molecules bearing self-peptides also die by apoptosis (death by neglect). (*c*) In contrast, those T cells whose TCR exhibits weak affinity for MHC molecules bearing self-peptides survive (positive selection) and ultimately leave the thymus to constitute the peripheral T-cell population of the body.

Figure labels: Thymic cell · MHC · T cell · Strong affinity · APOPTOSIS (negative selection) (a) · No affinity · APOPTOSIS (death by neglect) (b) · Weak affinity · SURVIVAL (positive selection) (c)

---

[6]B cells are also subjected to selective processes that lead to the death or inactivation of cells capable of producing autoreactive antibodies. In some cases, the light chains of autoreactive antibodies can be replaced by a new light chain encoded by an Ig gene that has been secondarily rearranged in a process called receptor editing.

a T cell. This receptor is distinct and spatially separate from the TCR. Unlike the TCR, the receptor that delivers the co-stimulatory signal is not specific for a particular antigen and does not require an MHC molecule to bind. The best studied of these interactions occur between helper T cells and pro-fessional antigen-presenting cells (e.g., dendritic cells and macrophages).

Activation of Helper T Cells by Professional APCs T_H cells recognize antigen fragments on the surface of dendritic cells and macrophages that are lodged in the binding cleft of MHC class II molecules. A costimulatory signal is delivered to a T_H cell as the result of an interaction between a protein known as CD28 on the surface of the T_H cell and a member of the B7 family of proteins on the surface of the APC (Fig-ure 17.26a). The B7 protein appears on the surface of the APC after the phagocyte ingests foreign antigen. If the T_H cell does not receive this second signal from the APC, rather than becoming activated, the T_H cell either becomes nonre-sponsive (anergized) or is stimulated to undergo apoptosis (deleted). Because professional APCs are the only cells capa-ble of delivering the costimulatory signal, they are the only cells that can initiate a T_H response. As a result, normal cells of the body that bear proteins capable of combining with the TCRs of a T cell cannot activate T_H cells. Thus, the require-ment by a T_H cell for two activation signals protects normal body cells from autoimmune attack involving T_H cells.

Prior to its interaction with an APC, a T_H cell can be de-scribed as a resting cell, that is, one that has withdrawn from the cell cycle (a G_0 cell, page 562). Once it receives the dual ac-tivation signals, a T_H cell is stimulated to reenter the G_1 phase of the cell cycle and eventually to progress through S phase into mitosis. Thus, interaction of T cells with a specific antigen leads to the proliferation (*clonal expansion*) of those cells capa-ble of responding to that antigen. In addition to triggering cell division, activation of a T_H cell causes it to synthesize and se-crete cytokines (most notably IL-2). Cytokines produced by activated T_H cells act on other cells of the immune system (in-cluding B cells and macrophages) and also back on the T_H cells that secreted the molecules. The source and function of various cytokines were indicated in Table 17.1.

We have seen in this chapter how immune responses are stimulated by ligands that activate receptor signaling path-ways. But many of these events are also influenced by in-hibitory stimuli, so that the ultimate response by the cell is determined by a balance of positive and negative influences. For example, interaction between CD28 and a B7 protein delivers a positive signal to the T cell leading to its activation. Once the T cell has been activated, it produces another cell-surface protein called CTLA4 that is similar to CD28 in

(a)   *(b)*
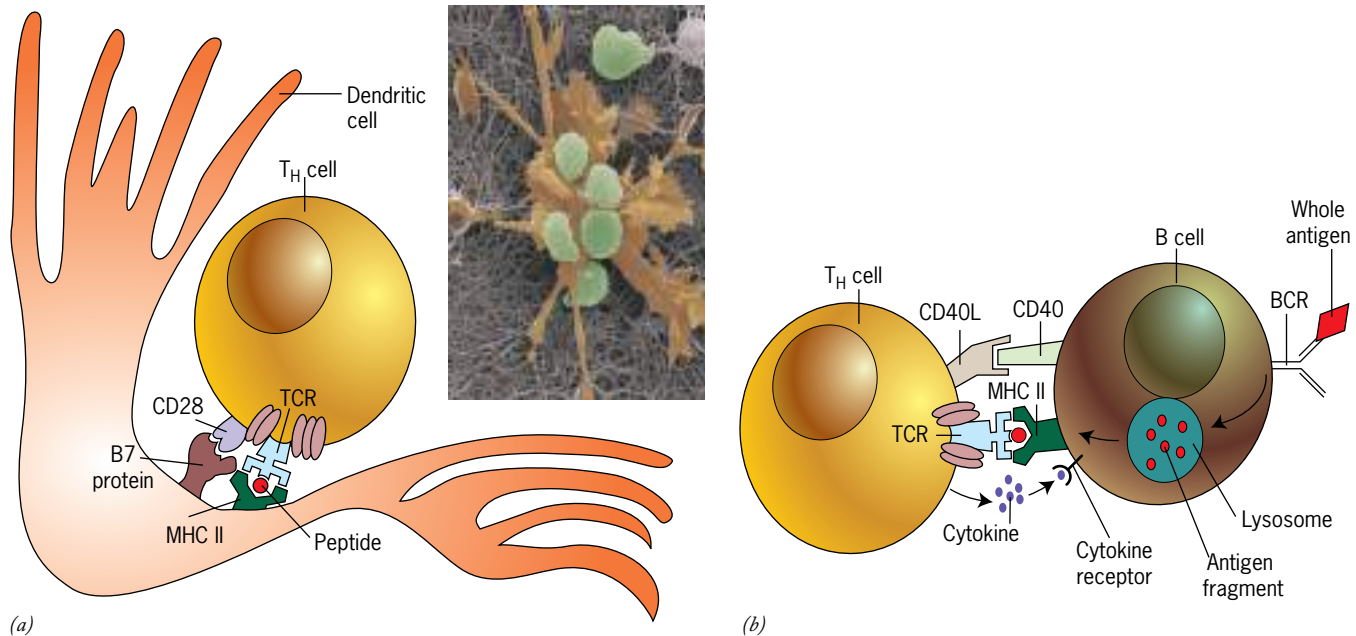
FIGURE 17.26 **Lymphocyte activation.** (*a*) Schematic drawing depicting the interaction between a professional APC, in this case a mature den-dritic cell, and a T_H cell. Specificity in this cell–cell interaction derives from recognition by the TCR on the T_H cell of the MHC class II–peptide complex displayed on the surface of the dendritic cell. Interac-tion between CD28 of the T cell and a B7 protein of the dendritic cell provides a nonspecific costimulatory signal that is required for T-cell activation. (Inset) Scanning electron micrograph of a mature dendritic cell (orange) presenting antigen to a number of T cells (green). (*b*) Schematic drawing of the interaction between an activated T_H cell and a B cell. Specificity in this cell–cell interaction derives from recognition by the TCR on the T_H cell of the MHC class II–peptide complex displayed on the surface of the B cell. The peptide displayed by the B cell is derived from protein molecules that had initially bound to the BCRs at the cell surface. These bound antigens are taken up by endocytosis, fragmented in lysosomes, and bound to MHC class II molecules, as in Figure 17.23b. (AFTER E. LINDHOUT ET AL., IMMUNOL. TODAY 18:574, 1997, WITH PERMIS-SION FROM ELSEVIER SCIENCE. INSET PHOTO; REPRINTED WITH PERMIS-SION FROM P. FRIEDL, A. T. DEN BOER, AND M. GUNZER, NATURE REVS. IMMUNOLOGY 5:533, 2005; © COPYRIGHT 2005, BY MACMILLAN MAGAZINES LIMITED.)

structure and also interacts with B7 proteins of the APC. Unlike the CD28–B7 interaction, however, contact between CTLA4 and B7 leads to inhibition of the T cell's response, rather than activation. The need for balance between activation and inhibition is most evident in mice that have been genetically engineered to lack the gene encoding CTLA4. These mice die as a result of massive overproliferation of T cells. As noted on page 708, insights into CTLA4 function have recently been put to clinical use.

Activation of B Cells by T_H Cells T_H cells bind to B cells whose receptors recognize the same antigen. The antigen initially binds to the immunoglobulin (BCR) at the B-cell surface. Antigens taken up by B cells can be soluble proteins from the extracellular medium or proteins bound to the plasma membrane of other cells. In the latter case, the B cell obtains the antigen by spreading itself over the outer surface of the target cell and collecting the BCR-antigen complexes into a central cluster. The bound antigen is then taken into the B cell, where it is processed enzymatically, and its fragments are displayed in combination with MHC class II molecules (Figure 17.26*b*). Recognition of the peptide fragment by the TCR leads to activation of the T_H cell, which responds by activating the B cell. Activation of a B cell follows transmission of several signals from the T_H cell to the B cell. Some of these signals are transmitted directly from one cell surface to the other through an interaction between complementary proteins, such as CD40 and CD40 ligand (CD40L) (Figure 17.26*b*). Binding between CD40 and CD40L generates signals that help move the B cell from the G_0 resting state back into the cell cycle. Other signals are transmitted by cytokines released by the T cell into the immunologic synapse separating it from the nearby B cell. This process is not unlike the way neurotransmitters act across a neural synapse (page 163). Cytokines released by the T cells into the immunologic synapse include IL-4, IL-5, IL-6, and IL-10. Interleukin 4 is thought to stimulate the B cell to switch from producing the IgM class to producing the IgG or IgE class. Other cytokines induce the proliferation, differentiation, and secretory activities of B cells.

Signal Transduction Pathways in Lymphocyte Activation

We saw in Chapter 15 how hormones, growth factors, and other chemical messengers bind to receptors on the outer surfaces of target cells to initiate a process of signal transduction, which transmits information into a cell's internal compartments. We saw in that chapter how a large variety of extracellular messenger molecules transmit information along a small number of shared signal transduction pathways. The stimulation of lymphocytes occurs by a similar mechanism and utilizes many of the same components employed by hormones and growth factors that act on other cell types.

When a T cell is activated by a dendritic cell, or a B cell is activated by a T_H cell, signals are transmitted from plasma membrane to cytoplasm by tyrosine kinases, similar to the

signals described in Chapter 15 for insulin and growth factors. Unlike the receptors for insulin and growth factors (page 623), lymphocyte antigen receptors lack an inherent tyrosine kinase activity. Instead, ligand binding to antigen receptors leads to the recruitment of cytoplasmic tyrosine kinase molecules to the inner surface of the plasma membrane. This process may be facilitated by the movement of activated receptors into lipid rafts (page 135). Several different tyrosine kinases have been implicated in signal transduction during lymphocyte activation, including members of the Src and Tec families. Src was the first tyrosine kinase to be identified and is the product of the first cancer-causing oncogene to be discovered (Section 16.3).

Activation of these tyrosine kinases leads to a cascade of events and the activation of numerous signal transduction pathways, including:

1. Activation of phospholipase C, which leads to the formation of IP_3 and DAG. As discussed on page 617, IP_3 causes a marked elevation in the levels of cytosolic Ca^{2+}, whereas DAG stimulates protein kinase C activity.

2. Activation of Ras, which leads to the activation of the MAP kinase cascade (page 627).

3. Activation of PI3K, which catalyzes the formation of membrane-bound lipid messengers having diverse functions in cells (page 632).

Transmission of signals along these various pathways, and others, leads to the activation of a number of transcription factors (e.g., NF-κB and NFAT) and the resulting transcription of dozens of genes that are not expressed in the resting T cell or B cell.

As noted above, one of the most important responses by an activated lymphocyte is the production and secretion of cytokines. Some of these cytokines may act as part of an autocrine circuit by binding to receptors on the cell that released them. Like other extracellular signals, cytokines bind to receptors on the outer surface of target cells, generating cytoplasmic signals that act on various intracellular targets. Cytokines utilize a novel signal transduction pathway referred to as the *JAK–STAT pathway*, which operates without the involvement of second messengers. The "JAK" portion of the name is an acronym for Janus kinases, a family of cytoplasmic tyrosine kinases whose members become activated following the binding of a cytokine to a cell-surface receptor. (Janus is a two-faced Roman god who protected entrances and doorways.) "STAT" is an acronym for "signal transducers and activators of transcription," a family of transcription factors that become activated when one of their tyrosine residues is phosphorylated by a JAK (see Figure 15.17*c*). Once phosphorylated, STAT molecules interact to form dimers that translocate from the cytoplasm to the nucleus where they bind to specific DNA sequences, such as an interferon-stimulated response element (ISRE). ISREs are found in the regulatory regions of a dozen or so genes that are activated when a cell is exposed to the cytokine interferon α (IFN-α).

As with the hormones and growth factors discussed in Chapter 15, the specific response of a cell depends on the

particular cytokine receptor engaged and the particular JAKs and STATs that are present in that cell. For example, it was noted earlier that IL-4 acts to induce Ig class switching in B cells. This response follows the IL-4-induced phosphorylation of the transcription factor STAT6, which is present in the cytoplasm of activated B cells. Resistance to viral infection induced by interferons (page 686) is mediated through the phosphorylation of STAT1. Phosphorylation of other STATs may lead to the progression of the target cell through the cell cycle.

REVIEW ?

1. Draw the basic structure of an IgG molecule bound to an epitope of an antigen. Label the heavy and light chains; the variable and constant regions of each chain; the regions that would contain hypervariable sequences.

2. Draw the basic arrangement of the genes of the germ line that are involved in encoding the light and heavy chains of an IgG molecule. How does this differ from their arrangement in the genome of an antibody-producing cell? What steps occur to bring about this DNA rearrangement?

3. Give three different mechanisms that contribute to variability in the V regions of antibody chains.

4. Compare and contrast the structure of antigen receptors on B and T cells.

5. Describe the steps in processing a cytosolic antigen in an APC. What is the role of MHC proteins in this process?

6. Compare and contrast the roles of an MHC class I and class II protein molecule. What types of APCs utilize each class of MHC molecule, and what types of cells recognize them?

7. Describe the steps between the stage when a bacterium is ingested by a macrophage and the stage when plasma cells are producing antibodies that bind to the bacterium and neutralize its infectivity.

THE HUMAN PERSPECTIVE

Autoimmune Diseases

The immune system requires complex and highly specific interactions between many different types of cells and molecules. Numerous events must take place before a humoral or cell-mediated immune response can be initiated, which makes these processes vulnerable to disruption at various stages by numerous factors. Included among the various types of immune dysfunction are **autoimmune diseases**, which result when the body mounts an immune response against part of itself. More than 80 different autoimmune diseases have been characterized, affecting approximately 5 percent of the population.

Because the specificity of the antigen receptors of both T and B cells is determined by a process of random gene rearrangement, it is inevitable that some members of these cell populations possess receptors directed against the body's own proteins (*self-antigens*). Lymphocytes that bind self-antigens with high affinity tend to be removed from the lymphocyte population, making the immune system *tolerant* toward self. However, some of the self-reactive lymphocytes generated in the thymus and bone marrow escape the body's negative selection processes, giving them the *potential* to attack normal body tissues. The presence of B and T lymphocytes capable of reacting against the body's tissues is readily demonstrated in healthy individuals. For example, when T cells are isolated from the blood and treated in vitro with a normal self-protein, together with the cytokine IL-2, a small number of cells in the population are likely to proliferate to form a clone of cells that can react to that self-antigen. Similarly, if laboratory animals are injected with a purified self-protein together with an *adjuvant*, which is a nonspecific substance that enhances the response to the injected antigen, they mount an immune response against the tissues in which that protein is normally found. Under normal circumstances, B and T cells capable of reacting to self-antigens are inhibited by antigen-specific T_{Reg} cells

(page 692) or by other suppressive mechanisms. When these mechanisms fail, a person may suffer from an autoimmune disease, including those described below.

1. **Multiple sclerosis (MS)** is an inflammatory disease that typically strikes young adults, causing severe and often progressive neurological damage. MS results from an attack by immune cells and/or antibodies against the myelin sheath that surrounds the axons of nerve cells (page 162). These sheaths form the white matter of the central nervous system. The demyelination of nerves that results from this immunologic attack interferes with the conduction of nerve impulses along axons, leaving the person with diminished eyesight, problems with motor coordination, and disturbances in sensory perception. A disease similar to MS, called experimental allergic encephalomyelitis, can be induced in laboratory animals by injection of myelin basic protein, a major component of the myelin plasma membrane. Moreover, the disease can be transferred to healthy mice of the same strain by injection of T cells from a diseased individual.

2. **Insulin-dependent diabetes mellitus (IDDM)**, which is often called juvenile-onset or type 1 diabetes because it tends to arise in children, results from the autoimmune destruction of the insulin-secreting β cells of the pancreas. Destruction of these cells is mediated by self-reactive T cells. At present, patients with IDDM are administered daily doses of insulin. While the hormone allows them to survive, these individuals may still be subject to degenerative kidney, vascular, and retinal disease.

3. **Graves' disease and thyroiditis** are autoimmune diseases of the thyroid that produce very different symptoms. In Graves' disease, the target of immune attack is the TSH receptor on the surface of thyroid cells that normally binds the pituitary hormone thyroid-

stimulating hormone (TSH). In patients with this disease, autoantibodies bind to the TSH receptor, causing the prolonged stimulation of thyroid cells, leading to hyperthyroidism (i.e., elevated blood levels of thyroid hormone). Thyroiditis (or Hashimoto's thyroiditis) develops from an immune attack against one or more common proteins of thyroid cells, including thyroglobulin. The resulting destruction of the thyroid gland leads to hypothyroidism (i.e., decreased blood levels of thyroid hormone).

4. **Rheumatoid arthritis** affects approximately one percent of the population and is characterized by the progressive destruction of the body's joints due to a cascade of inflammatory responses. In a normal joint, the synovial membrane that lines the synovial cavity is only a single cell layer thick. In persons with rheumatoid arthritis, this membrane becomes inflamed and thickened due to the infiltration of autoreactive immune cells and/or autoantibodies into the joint. Over time, the cartilage is replaced by fibrous tissue, which causes immobilization of the joint.

5. **Systemic lupus erythematosus (SLE)** gets its name "red wolf" from a reddish rash that develops on the cheeks during the early stages of the disease (Figure 1). Unlike the other autoimmune diseases discussed above, SLE is seldom confined to a particular organ but often attacks tissues throughout the body, including the central nervous system, kidneys, and heart. The serum of patients with SLE contains antibodies directed against a number of components that are found in the nuclei of cells, including small nuclear ribonucleoproteins (snRNPs); proteins of the centromeres of chromosomes; and, most notably, double-stranded DNA. Recent studies suggest that autoimmunity occurs when TLRs that normally recognize microbial DNA and RNA (page 685) mistakenly bind to the body's own informational macromolecules. The incidence of SLE is particularly high in women of child-bearing age, suggesting a role for female hormones in triggering the disease.

Not everyone in the population is equally susceptible to developing one of these autoimmune diseases. Most of these disorders appear much more frequently in certain families than in the general population, indicating a strong genetic component to their development. Though many different genes have been shown to increase susceptibility to autoimmune diseases, those that encode MHC class II polypeptides are most strongly linked. For example, people who inherit certain alleles of the MHC locus are particularly susceptible to developing type 1 diabetes (IDDM). It is thought that cells bearing MHC molecules encoded by the susceptible alleles can bind particular peptides that stimulate the formation of autoantibodies against insulin-secreting β cells of the pancreas.

Possession of high-risk alleles may be necessary for an individual to develop certain autoimmune diseases, but it is not the only contributing factor. Studies of identical twins indicate that if one twin develops an autoimmune disease, the likelihood that the other twin will also develop the disease ranges from about 25 to 75 percent, not 100 percent as expected if genetics were the only contributing factor.

The great progress that has been made over the past two decades in our understanding of the cellular and molecular basis of immunity has led to new treatments for a number of autoimmune diseases that involve manipulation of the body's immune system. These treatments have been tested in animal models (i.e., animals that can be made to develop diseases similar to those of humans) and their safety and efficacy determined in human clinical trials. Among the approaches taken are:

- Treatment with immunosuppressive drugs, such as cyclosporin A or CellCept, that block the autoimmune response. Because these drugs are nonspecific, they inhibit all types of immune responses and, therefore, render the patient susceptible to dangerous infections.

- Restoring immunological tolerance to self-antigens so that the body stops producing autoantibodies and autoreactive T cells. Of all of the approaches discussed in this section, this is the only one that promises the potential for antigen-specific therapy; all of the others exert a nonspecific influence on the immune system. One way to induce tolerance to specific antigens is to administer peptides (called APLs) that resemble the peptides that would be generated from the self-antigens responsible for causing the disease. It is hoped that such APLs would bind to TCRs in a suboptimal manner, blocking T-cell activation and reducing the secretion of inflammatory cytokines (e.g., TNF-α and IFN-α). One drug of this type (Copaxone) consists of a mixture of synthetic peptides whose structure resembles that of myelin basic protein. Copaxone may reduce the frequency of relapses in patients with multiple sclerosis but does not stop disease progression and may elicit severe allergic side effects. Another way to induce immunological tolerance is to administer a modified version of the protein CTLA4, which binds to the B7 costimulatory protein on the surface of APCs and inhibits these cells' ability to activate autoreactive T cells. Orencia, which acts in this manner, has been approved for patients with rheumatoid arthritis. Many researchers believe that the best long-term, antigen-specific approach to reestablishing tolerance is to treat a person with his or her own T_{Reg} cells. In this approach, the desired T_{Reg} cells would be isolated from the patient's blood, allowed to proliferate extensively in vitro, and then reinjected back into the patient in an attempt to suppress the specific self-reactive immune cells that are waging the autoimmune attack. Several clinical trials involving the transfer of T_{Reg} cells have begun in patients who are experiencing immunological complications following organ transplantation.

- Blocking the effect of pro-inflammatory cytokines, which are among the agents that wreak the greatest tissue destruction in many autoimmune diseases. The best studied examples of this approach are the monoclonal antibodies (e.g., Remicade and Humira) that have been developed against the pro-inflammatory cytokine TNF-α. These drugs have been approved for the treatment of rheumatoid arthritis and can have dramatic curative effects in many patients. IL-1 is another key pro-inflammatory cytokine, and a number of IL-1 inhibitors are presently in the drug pipeline.

- Treatment with cytokines. At present, the most widely prescribed treatment for multiple sclerosis is one of several approved IFN-β cytokines (Avonex, Rebif, and Betaseron), which reduce disease progression by an average of 35 percent. IFN-β has many activities, and the precise mechanism of action is debated.

- Treatment with agents that destroy B cells. Although autoimmune diseases are generally thought to result from the dysfunction of T cells, particularly helper T cells, there is considerable evidence that self-reactive B cells are also to blame, at least partly. This conclusion has been confirmed in clinical studies with a monoclonal antibody (called Rituxan) that binds to and destroys B cells. Rituxan had been used as a safe and successful treatment for lymphoma (page 672) before it was tested in patients with rheumatoid arthritis and multiple sclerosis, where it has proved surprisingly effective. It is not clear whether this effect can be traced to the role of B cells as antigen-presenting APCs or to their role in antibody production. Regardless, it is remarkable that even though Rituxan virtually depletes the blood of circulating B cells for a period of several months, it does not limit a patient's ability to mount immune responses against infectious agents.

- Disrupting the movement of self-reactive immune cells to areas of inflammation. The most effective treatment yet developed for multiple sclerosis is an antibody (Tysabri) that is directed against the integrin subunit α_4 present on the surface of activated T cells. The drug is aimed at preventing these T cells from crossing the blood-brain barrier (page 256) and attacking the myelin sheaths of the central nervous system of MS patients. There is always a concern with any type of immune therapy that the treatment will interfere with the body's ability to fight infection. This concern was realized when Tysabri was temporarily removed from the market in 2005 after three patients developed a serious viral brain infection.

Administration of antibodies that interfere with immune regulatory functions must be approached with great care. This point was made abundantly clear in 2006 during a disastrous phase I trial of a monoclonal antibody called TGN1412 directed against a cell-surface protein called CD28. The antibody was designed to activate regulatory T cells in an attempt to dampen the body's immune responses (page 692). Six healthy volunteers were injected with the antibody, and all six experienced rapid and serious multiple organ dysfunction resulting from a systemic release of pro-inflammatory cytokines. The drug had been tested at much higher doses in laboratory monkeys without noticeable side effects.

- Transplantation of hematopoietic stem cells (page 19) from either the patient themselves (an auto-transplant) or a closely matched donor (an allo-transplant). Because this procedure has the potential to cause life-threatening complications, it represents a treatment for only those patients with severe and debilitating autoimmune disease. However, unlike the drugs described above, transplant recipients begin the rest of their lives with a largely "new" immune system and thus the possibility of being completely cured of their disease.

EXPERIMENTAL PATHWAYS

The Role of the Major Histocompatibility Complex in Antigen Presentation

In 1973, Hugh McDevitt and his colleagues at the Scripps Foundation in La Jolla, California, and at Stanford University, demonstrated that the susceptibility of mice to a particular pathogen depends on the allele present at one of the MHC loci.[1] They found that lymphocytic choriomeningitis virus (LCMV) causes a lethal brain infection in mice that are either homozygous or heterozygous for the H-2^g allele but does not cause infections in mice that are homozygous for the H-2^k allele at this locus.

These findings led Rolf Zinkernagel and Peter Doherty of the Australian National University to examine the role of cytotoxic T lymphocytes (CTLs) in the development of this disease. Zinkernagel and Doherty planned experiments to correlate the level of CTL activity with the severity of the disease in mice having different MHC genotypes (or haplotypes, as they are called). Cytotoxic T lymphocyte activity was monitored using the following experimental protocol. Monolayers of fibroblasts (L cells) from one mouse were grown in culture and subsequently infected with the LCM virus. The infected fibroblasts were then overlaid by a preparation of spleen cells from a mouse that had been infected with the LCM virus seven days earlier. Waiting for seven days gives the animal's immune system time to generate CTLs against virus-infected cells. The CTLs become concentrated in the infected animal's spleen. To monitor the effectiveness of the attack by the CTLs on the cultured L cells, the L cells were first labeled with a radioisotope of chromium (^{51}Cr). ^{51}Cr is used as a marker for cell viability: as long as a cell remains alive, the radioisotope remains inside the cell. If a cell should be lysed during the experiment by a CTL, the ^{51}Cr is released into the medium.

Zinkernagel and Doherty found that the level of CTL activity against the cultured fibroblasts—as measured by the release of ^{51}Cr—depended on the relative genotypes of the fibroblasts and spleen cells (Table 1).[2] All of the fibroblasts used to obtain the data shown in Table 1 were taken from an inbred strain of mice that were homozygous for the allele H-2^k at the H-2 locus. When spleen cells were prepared from mice having an H-2^k allele (e.g., CBA/H, AKR, and C3H/HeJ strains of mice), the L cells were effectively lysed. However, spleen cells taken from mice bearing H-2^b or H-2^d alleles at this locus were unable to lyse the infected fibroblasts. (The ^{51}Cr released is approximately the same when noninfected fibroblasts are used in the assay, as shown in the right column of Table 1.)

It was essential to show that the results were not peculiar to mice bearing H-2^k alleles. To make this determination, Zinkernagel and Doherty tested LCMV-activated spleen cells from H-2^b mice against various types of infected cells. Once again, the CTLs would only lyse infected cells having the same H-2 genotype, in this case

TABLE 1 Cytotoxic Activity of Spleen Cells from Various Strains of Mice Injected 7 Days Previously with LCM Virus for Monolayers of LCM-Infected or Normal C$_3$H (H-2^k) Mouse L Cells

Exp	Mouse strain	H-2 type	% ^{51}Cr release Infected	% ^{51}Cr release Normal
1	CBA/H	k	65.1 ± 3.3	17.2 ± 0.7
	Balb/C	d	17.9 ± 0.9	17.2 ± 0.6
	CB57B1	b	22.7 ± 1.4	19.8 ± 0.9
	CBA/H × C57B1	k / b	56.1 ± 0.5	16.7 ± 0.3
	CB57Bl × Balb/C	b / d	24.8 ± 2.4	19.8 ± 0.9
	nu/+ or +/+		42.8 ± 2.0	21.9 ± 0.7
	nu/nu		23.3 ± 0.6	20.0 ± 1.4
2	CBA/H	k	85.5 ± 3.1	20.9 ± 1.2
	AKR	k	71.2 ± 1.6	18.6 ± 1.2
	DBA/2	d	24.5 ± 1.2	21.7 ± 1.7
3	CBA/H	k	77.9 ± 2.7	25.7 ± 1.3
	C3H/HeJ	k	77.8 ± 0.8	24.5 ± 1.5

Reprinted with permission from R. M. Zinkernagel and P. C. Doherty, *Nature* 248:701, 1974. © Copyright 1974, Macmillan Magazine Ltd.

H-2^b. These studies provided the first evidence that the MHC molecules on the surface of an infected cell restricts its interactions with T cells. T-cell function is said to be *MHC restricted*.

These and other experiments during the 1970s raised questions about the role of MHC proteins in immune cell function. Meanwhile, another line of investigation was focusing on the mechanism by which T cells were stimulated by particular antigens. Studies had indicated that T cells respond to antigen that is bound to the surface of other cells. It was presumed that the antigen being displayed had simply bound to the surface of the antigen-presenting cell (APC) from the extracellular medium. During the mid-1970s and early 1980s, studies by Alan Rosenthal at NIH, Emil Unanue at Harvard University, and others demonstrated that antigen had to be internalized by the APC and subjected to some type of processing before it could stimulate T-cell proliferation. Most of these studies were carried out in cell culture utilizing T cells activated by macrophages that had been previously exposed to bacteria, viruses, or other foreign material.

One of the ways to distinguish between antigen that is simply bound to the surface of an APC and antigen that has been processed by metabolic activities is to compare events that can occur at low temperatures (e.g., 4°C) where metabolic processes are blocked with

those that occur at normal body temperatures. In one of these earlier experiments, macrophages were incubated with antigen for one hour at either 4°C or 37°C and then tested for their ability to stimulate T cells prepared from lymph nodes.[3] At lower antigen concentrations, macrophages were nearly 10 times more effective at stimulating T cells at 37°C than at 4°C, suggesting that antigen processing requires metabolic activities. Treatment of cells with sodium azide, a cytochrome oxidase inhibitor, also inhibited the appearance of antigen on the surface of T cells, indicating that antigen presentation requires metabolic energy.[4]

Subsequent experiments by Kirk Ziegler and Emil Unanue provided evidence that processing occurred as extracellular antigens were taken into the macrophage by endocytosis and delivered to the cell's lysosomal compartment.[5] One approach to determining whether lysosomes are involved in a particular process is to treat the cells with substances, such as ammonium chloride or chloroquine, that disrupt lysosomal enzyme activity. Both of these agents raise the pH of the lysosomal compartment, which inactivates the acid hydrolases (page 297). Table 2 shows the effects of these treatments on processing and presentation of antigen derived from the bacteria *Listeria monocytogenes*. It can be seen from the table that neither of these substances affected the uptake (endocytosis) of antigen, but both substances markedly inhibited the processing of the antigen and its ability to stimulate binding of T cells to the macrophage. These data were among the first to suggest that fragmentation of extracellular antigens by lysosomal proteases may be an essential step in preparation of extracellular antigens prior to presentation.

Other studies continued to implicate MHC molecules in the interaction between APCs and T cells. In one series of experiments, Ziegler and Unanue treated macrophages with antibodies directed against MHC proteins encoded by the H-2 locus. They found that these antibodies had no effect on the uptake or catabolism of antigen,[6] but markedly inhibited the macrophages from interacting with T cells.[7] Inhibition of binding of T cells to macrophages was obtained even when macrophages were exposed to the antibodies before addition of antigen.

Evidence from these and many other studies indicated that the interaction between a T cell and a macrophage depended on the recognition of two components on the surface of the antigen-presenting cell: the antigen fragment being displayed and an MHC molecule. But there was no clear-cut picture as to how the antigen fragment and MHC molecule were related. Two models of antigen recognition were considered likely possibilities. According to one model, T cells possess two distinct receptors, one for the antigen and another for the MHC protein. According to the other model, a sin-

TABLE 2 Inhibition of Antigen Presentation with NH$_4$Cl and Chloroquine

Assay	Control (%)	10 mM NH$_4$Cl Observed (%)	10 mM NH$_4$Cl Inhibition (%)	0.1 mM Chloroquine Observed (%)	0.1 mM Chloroquine Inhibition (%)
Antigen uptake	15 ± 1	13 ± 2	13	15 ± 2	0
Antigen ingestion	66 ± 2	63 ± 2	5	67 ± 6	−2
Antigen catabolism	29 ± 4	13 ± 3	55	14 ± 6	52
T cell-macrophage binding					
before antigen handling	70 ± 7	26 ± 8	63	30 ± 8	57
after antigen handling	84 ± 8	70 ± 11	17	60 ± 10	24

Source: H. K. Ziegler and E. R. Unanue, *Proc. Nat'l. Acad. Sci. U.S.A.* 79:176, 1982.

gle T-cell receptor recognizes both the MHC protein and the antigen peptide on the APC surface simultaneously. The balance of opinion began to shift in favor of the one-receptor model as evidence began to point to a physical association between MHC proteins and displayed antigens. In one study, for example, it was shown that antigen that had been processed by T cells could be isolated as a complex with MHC proteins.[8] In this experiment, cultured T cells taken from H-2^k mice were incubated with radioactively labeled antigen for 40 minutes. After the incubation period, processed antigen was prepared from the cells and passed through a column containing beads coated with antibodies directed against MHC proteins. When the beads were coated with antibodies against H-2^k protein, an MHC molecule present in the T cells, large amounts of radioactive antigen adhered to the beads, indicating the association of the processed antigen with the MHC protein. If the beads were instead coated with antibodies against H-2^b protein, an MHC protein that was not present in the T cells, relatively little radioactive antigen remained in the column.

Following these early experiments, investigators turned their attention to the atomic structure of the molecules involved in T-cell interactions. Rather than utilizing MHC class II molecules on the surfaces of macrophages, structural studies have examined MHC

class I molecules of the type found on the surfaces of virally infected cells. The first three-dimensional portrait of an MHC molecule was published in 1987 and was based on X-ray crystallographic studies by Don Wiley and colleagues at Harvard University.[9] The events leading up to this discovery are discussed in Reference 10. MHC class I molecules consist of (1) a heavy chain containing three extracellular domains (α_1, α_2, and α_3) and a single membrane-spanning segment and (2) an invariant β_2m polypeptide (see Figure 17.23). Wiley and colleagues examined the structure of the extracellular (soluble) portion of the MHC molecule (α_1, α_2, α_3, and β_2m) after removing the transmembrane anchor. A ribbon model of the observed structure is shown in Figure 1a, with the outer (antigen-bearing) portion of the protein constructed from the α_1 and α_2 domains. It can be seen from the top portion of this ribbon model that the inner surfaces of these domains form the walls of a deep groove approximately 25 Å long and 10 Å wide. It is this groove that acts as the binding site for peptides produced by antigen processing in the cytoplasm. As shown in Figure 1b, the sides of the antigen-binding pocket are lined by α helices from the α_1 and α_2 domains, and the bottom of the pocket is lined by β sheet that extends from these same domains across the midline. The helices are thought to form relatively flexible side walls enabling peptides of different sequence to bind within the groove.

Subsequent X-ray crystallographic studies described the manner in which peptides are positioned within the MHC antigen-binding pocket. In one of these studies, the spatial arrangement of several naturally processed peptides situated within the antigen-binding pocket of a single MHC class I molecule (HLA-B27) was determined.[11] The backbones of all the peptides bound to HLA-B27 share a single, extended conformation running the length of the
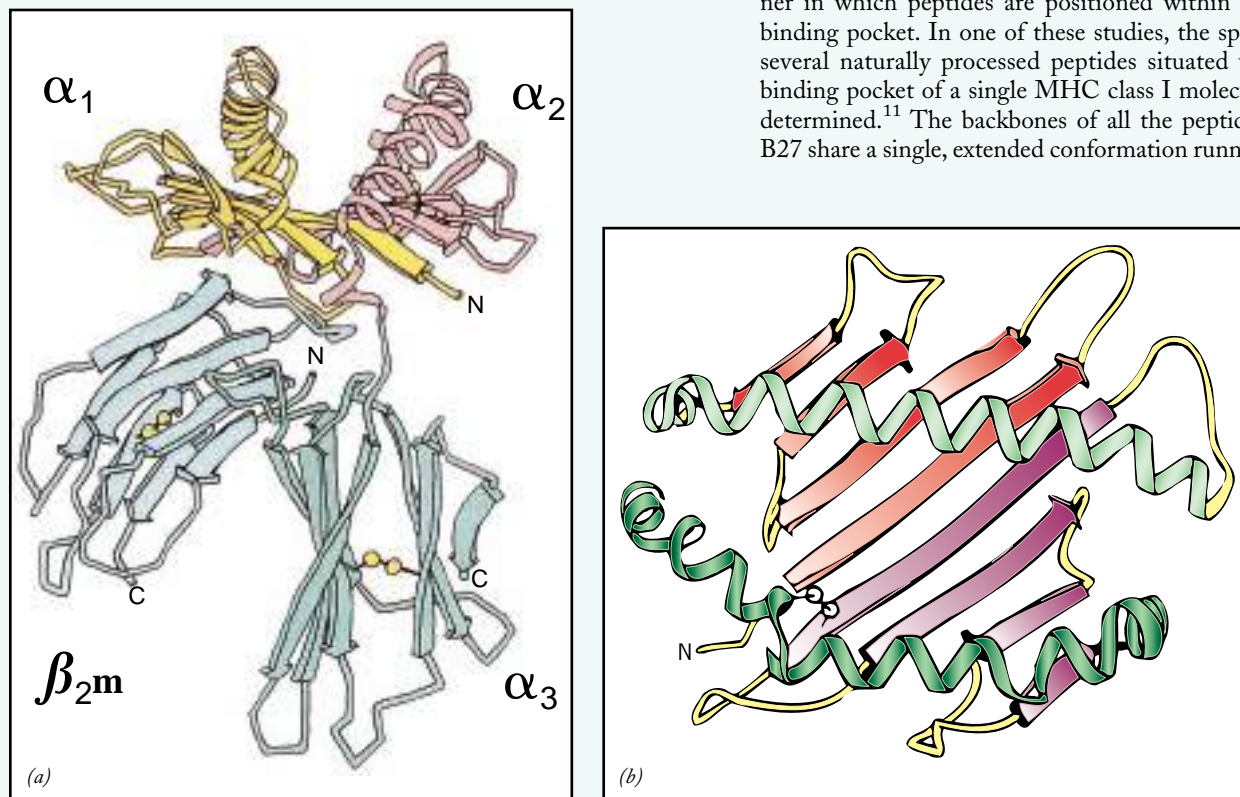
FIGURE 1 (a) Schematic representation of an MHC class I molecule, in this case the human protein HLA-A2. The molecule consists of two subunits: a heavy chain made up of three domains (α_1, α_2, and α_3) and a β_2m chain. The membrane-spanning portion of the heavy chain would connect with the polypeptide at the site marked C (for C-terminus of the remaining portion). Disulfide bonds are indicated as two connected spheres. The peptide-binding groove is shown at the top of the drawing, between the α helical segments of the α_1 and α_2 domains of the heavy chain. (b) Schematic representation of the peptide-binding pocket of the MHC protein as viewed from the top of the molecule. The bottom of the pocket is lined by β sheet (orange-purple arrows) and the walls by α helices (green). The α_1 domain is shown in orange and light green; the α_2 domain is shown in purple and dark green. (REPRINTED WITH PERMISSION FROM PAMELA J. BJORKMAN ET AL., NATURE 329:508, 509, 1987; © COPYRIGHT 1987, BY MACMILLAN MAGAZINES LTD.)

binding cleft. The N- and C-termini of the peptides are precisely positioned by numerous hydrogen bonds at both ends of the cleft. The hydrogen bonds link the peptide to a number of conserved residues in the MHC molecule that are part of both the sides and bottom of the binding groove.

In another key study, Ian Wilson and colleagues at the Scripps Research Institute in La Jolla, California, reported on the X-ray crystallographic structure of a mouse MHC class I protein complexed with two peptides of different length.[12,13] The overall structure of the mouse MHC protein is similar to that of the human MHC protein shown in Figure 1a. In both cases, the peptides are bound in an extended conformation deep within the binding groove of the MHC molecule (Figure 2). This extended conformation allows for numerous interactions between the side chains of the MHC molecule and the backbone of the bound peptide. Because the MHC interacts primarily with the backbone of the peptide rather than its side chains, there are very few restrictions on the particular amino acid residues that can be present at various sites of the binding pocket. As a result, each MHC molecule can bind a diverse array of antigenic peptides.

An MHC–peptide complex projecting from the surface of an infected cell is only half the story of immunologic recognition; the other half is represented by the T-cell receptor (TCR) projecting from the surface of the cytotoxic T cell. It has been evident for more than a decade that a TCR can somehow recognize both an MHC and its contained peptide, but the manner by which this occurs eluded researchers because of difficulties in preparing protein crystals of the TCR that were suitable for X-ray crystallography. These difficulties were eventually overcome, and in 1996, reports were published by both the Wiley and Wilson laboratories that provided a three-dimensional portrait of the interaction between an MHC–peptide and TCR.[14,15] The overall structure of the complex

formed between the two proteins is shown in Figure 3, where the backbones of the polypeptides are portrayed as tubes.

The structure shown in Figure 3 depicts the portions of the proteins that project between a CTL and virally infected host cell. The
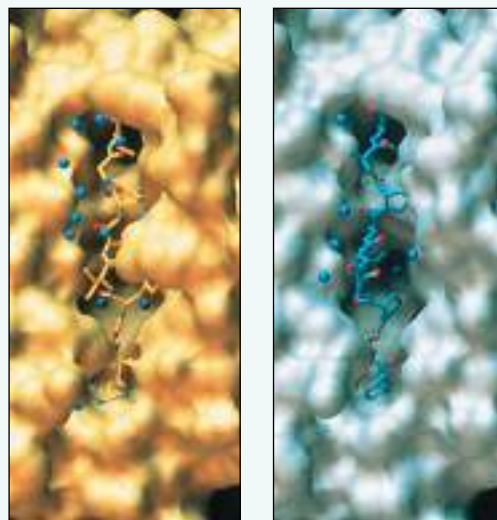
(a)   (b)

FIGURE 2 Three-dimensional models of a peptide (shown in ball-and-stick structure) bound within the antigen-binding pocket of a mouse MHC class I protein (H-2K^b). The peptide in a consists of eight amino acid residues; the peptide in b consists of nine residues. The peptides are seen to be buried deep within the MHC binding groove. (REPRINTED WITH PERMISSION FROM MASAZUMI MATSUMURA, DAVED H. FREMONT, PER A. PETERSON, AND IAN A. WILSON, SCIENCE 257:932, 1992. © COPYRIGHT 1992, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

FIGURE 3 Representation of the interaction between an MHC–peptide complex (at the bottom) and a TCR (at the top). The hypervariable regions (CDRs) of the TCR are shown as colored loops that form the interface between the two proteins. The bound peptide (P1-P8) is shown in yellow-green as it is situated within the binding groove of the MHC class I molecule. The peptide backbones are represented as tubes. (REPRINTED WITH PERMISSION FROM K. CHRISTOPHER GARCIA ET AL., COURTESY OF IAN WILSON, SCIENCE 274:217, 1996; © COPYRIGHT 1996, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

infected cell can activate a CTL, only professional APCs, such as macrophages and dendritic cells, that function in antigen ingestion and processing can activate T$_H$ cells. Cytotoxic T cells kill target cells by secreting proteins that permeabilize the membrane of the target cell and induce it to undergo apoptosis. (p. 690)

Antibodies are globular proteins called immunoglobulins (Igs) that are built of two types of polypeptide chains: heavy and light chains. Antibodies are divided into several classes (IgA, IgG, IgD, IgE, and IgM) depending on their heavy chain. Heavy and light chains both consist of (1) constant (C) regions, that is, regions whose amino acid sequence is virtually identical among all heavy or light chains of a given class, and (2) variable (V) regions, that is, regions whose sequence varies from one antibody species to another. Different classes of antibody appear at different times after exposure to an antigen and have different biological functions. IgG is the predominant form of blood-borne antibody. Each Y-shaped IgG molecule contains two identical heavy chains and two identical light chains. The antigen-combining sites are formed by association of the V region of a light chain with the V region of a heavy chain. Included within the V regions are hypervariable regions that form the walls of the antigen-combining site. (p. 693)

Antigen receptors on both B and T cells are encoded by genes produced by DNA rearrangement. Each variable region of a light Ig chain is composed of two DNA segments (a V and J segment) that are joined together, whereas each variable region of a heavy Ig chain is composed of three joined segments (a V, J, and D segment). Variability arises as the result of the joining of different V genes with different J genes in different antibody-producing cells. Additional variability is introduced by the enzymatic insertion of nucleotides

and variation in the V–J joining site. It is estimated that an individual can synthesize more than 2000 different species of kappa light chains and 100,000 species of heavy chains, which can combine randomly to form more than 200 million antibodies. Additional diversity arises in antibody-producing cells by somatic hypermutation in which the rearranged V genes experience a mutation rate much greater than the remainder of the genome. (p. 696)

Antigens are taken up by APCs, fragmented into short peptides, and combined with proteins encoded by the major histocompatibility complex (MHC) for presentation to T cells. MHC proteins can be subdivided into two classes: class I molecules and class II molecules. MHC class I molecules primarily display peptides derived from endogenous cytosolic proteins, including those derived from an infecting virus. Nearly any cell of the body can display peptides through MHC class I molecules to CTLs. If the TCR of the CTL recognizes a foreign peptide displayed by a cell, the CTL becomes activated to kill the target cell. Professional APCs, including macrophages and dendritic cells, display peptides in combination with MHC class II peptides. The MHC–peptide complexes on the APC are recognized by TCRs on the surface of T$_H$ cells. Helper T cells that become activated by antigens displayed on an APC subsequently associate with and activate B cells whose antigen receptors (BCRs consisting of Igs) recognize the same antigen. Helper T cells stimulate B cells by direct interaction and by secretion of cytokines. Activation of a T cell by a TCR leads to the stimulation of protein tyrosine kinases within the T cell, which in turn leads to the activation of a number of signaling pathways, including the activation of Ras and the MAP kinase cascade and the activation of phospholipase C. (p. 699)

18

Techniques in Cell and Molecular Biology

B ecause of the very small size of the subject matter, cell and molecular biology is more dependent on the development of new instruments and technologies than any other branch of biology. Consequently, it is difficult to learn about cell and molecular biology without also learning about the technology that is required to collect data. In this chapter, we will survey the methods used most commonly in the field without becoming immersed in the details or the many variations that are employed. These are the goals of the present chapter: to describe the ways that selected techniques are used and to provide examples of the types of information that can be learned by using these techniques. We will begin with the instrument that enabled biologists to discover the very existence of cells, providing the starting point for all of the information that has been presented in this text. ■

The use of dual-label fluorescence to follow dynamic events within cellular organelles. The Golgi cisternae of a budding yeast cell are not organized into distinct stacks as in most eukaryotic cells but are dispersed within the cytoplasm. Each of the brightly colored oval structures is an individual cisterna whose color is due to the localization of fluorescently labeled protein molecules. A cisterna that appears green contains a GFP-labeled protein (Vrg4) that is involved in early Golgi activities, whereas a cisterna that appears red contains a DsRed-labeled protein (Sec7) that is involved in late Golgi activities. This series of micrographs reveals the protein composition of individual cisternae over a period of about 13 minutes (elapsed time is indicated in the lower left corner). The white arrowhead and arrow point out two of these cisternae over time. These two cisternae are shown imaged by themselves in the lower rows of micrographs. It can be seen from this series that the protein composition of an individual cisterna changes over time from one containing "early" Golgi proteins (green) to one containing "late" Golgi proteins (red). These findings provide direct visual support for the cisternal maturation model discussed on page 287.

715

18.1 THE LIGHT MICROSCOPE

Microscopes are instruments that produce an enlarged image of an object. Figure 18.1 shows the most important components of a compound light microscope. A light source, which may be external to the microscope or built into its base, illuminates the specimen. The substage *condenser lens* gathers the diffuse rays from the light source and illuminates the specimen with a small cone of bright light that allows very small parts of the specimen to be seen after magnification. The light rays focused on the specimen by the condenser lens are then collected by the microscope's **objective lens**. From this point, we need to consider two sets of light rays that enter the objective lens: those that the specimen has altered and those that it hasn't (Figure 18.2). The latter group consists of light from the condenser that passes directly into the objective lens, forming the background light of the visual field. The former group of light rays emanates from the many parts of the specimen and forms the image of the specimen. These light rays are brought to focus by the objective lens to form a real, enlarged image of the object within the column of the microscope (Figure 18.1). The image formed by the objective lens is used as an object by a second lens system, the *ocular lens*, to form an enlarged and virtual image. A third lens system located in the front part of the eye uses the virtual image produced by the ocular lens as an object to produce a real image on the retina. When the focusing knob of the light microscope is turned, the relative distance between the specimen and the objective lens changes, allowing the final image to become focused precisely on the plane of the retina. The total magnification attained by the microscope is the product of the magnifications produced by the objective lens and the ocular lens.

Resolution

Up to this point we have considered only the magnification of an object without paying any attention to the quality of the image produced, that is, the extent to which the detail of the specimen is retained in the image. Suppose you are looking at a structure in the microscope using a relatively high-power objective lens (for example, 63×) and an ocular lens that magnifies the image of the objective lens another fivefold (a 5× ocular). Suppose the field is composed of chromosomes and it is important to determine the number present, but some of them are very close together and cannot be distinguished as separate structures (Figure 18.3*a*). One solution to the problem might be to change oculars to increase the size of the object being viewed. If you were to switch from the 5× to a 10× ocular, you would most likely increase your ability to determine the number of chromosomes present (Figure 18.3*b*) because you have now spread the image of the chromosomes produced by the objective lens over a greater part of your retina. The more photoreceptors that provide information about the image, the more detail that can be seen (Figure 18.4). If, however, you switch to a 20× ocular, you are not likely to see additional detail, although the image is larger (Figure 18.3*c*), that is, occupies more retinal surface. The
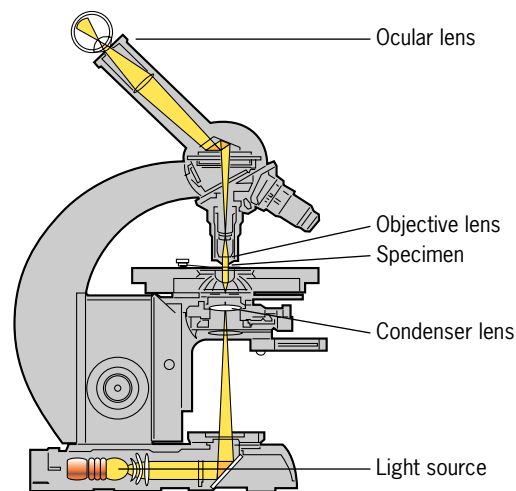
FIGURE 18.1 Sectional diagram through a compound light microscope; that is, a microscope that has both an objective and an ocular lens.

(•••) Light rays that form the image
(—) Background light of the field
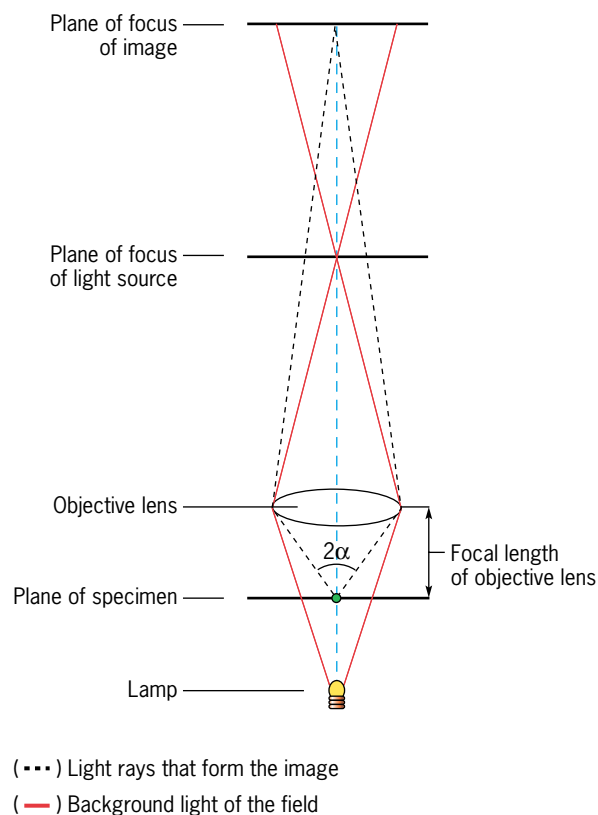
FIGURE 18.2 The paths taken by light rays that form the image of the specimen and those that form the background light of the field. Light rays from the specimen are brought to focus on the retina, whereas background rays are out of focus, producing a diffuse bright field. As discussed later in the text, the resolving power of an objective lens is proportional to the sine of the angle α. Lenses with greater resolving power have shorter focal lengths, which means that the specimen is situated closer to the objective lens when it is brought into focus.
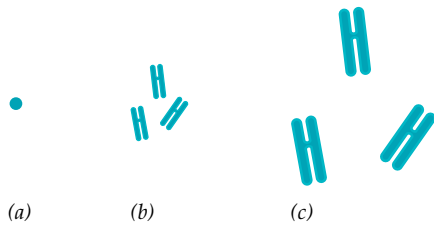
FIGURE 18.3 Magnification versus resolution. The transition from (*a*) to (*b*) provides the observer with increased magnification and resolution, whereas the transition from (*b*) to (*c*) provides only increased magnification (empty magnification). In fact, the quality of the image actually deteriorates as empty magnification increases.

change of ocular fails to provide additional information because the image produced by the objective lens does not possess any further detail to be enhanced by the increased power of the ocular lens. The second switch in oculars provides only *empty magnification* (as in Figure 18.3*c*).

The optical quality of an objective lens is measured by the extent to which the fine detail present in a specimen can be discriminated, or resolved. The **resolution** attained by a microscope is limited by diffraction. Because of diffraction, light emanating from a point in the specimen can never be seen as a point in the image, but only as a small disk. If the disks produced by two nearby points overlap, the points cannot be distinguished in the image (as in Figure 18.4). Thus the resolving power of a microscope can be defined in terms of the ability to see two neighboring points in the visual field as two distinct entities. The resolving power of a microscope is limited by the wavelength of the illumination according to the equation

$$d = \frac{0.61\,\lambda}{n \sin \alpha}$$

where *d* is the minimum distance that two points in the specimen must be separated to be resolved, λ is the wavelength of light (527 nm is used for white light), and *n* is the refractive index of the medium present between the specimen and objective lens. Alpha (α) is equal to half the angle of the cone of light entering the objective lens as shown in Figure 18.2.
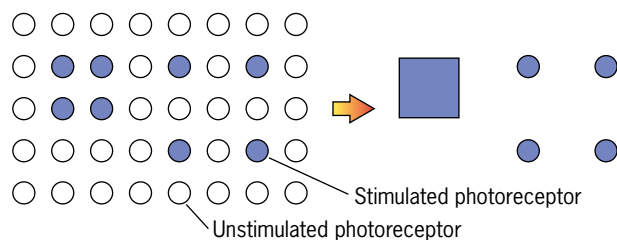
FIGURE 18.4 The resolving power of the eye. A highly schematic illustration of the relationship between the stimulation of individual photoreceptors (left) and the resulting scene one would perceive (right). The diagram illustrates the value of having the image fall over a sufficiently large area of the retina.

Alpha is a measure of the light-gathering ability of the lens and is directly related to its aperture.

The denominator of the equation in column 1 is called the *numerical aperture* (*N.A.*). The numerical aperture is a constant for each lens, a measure of its light-gathering qualities. For an objective that is designed for use in air, the maximum possible N.A. is 1.0 because the sine of the maximum possible angle of α, 90°, is 1, and the refractive index of air is 1.0. For an objective designed to be immersed in oil, the maximum possible N.A. is approximately 1.5. As a common rule of thumb, the maximum useful magnification of a light microscope ranges between 500 and 1000 times the numerical aperture of the objective lens being used. Attempts to enlarge the image beyond this point result in empty magnification, and the quality of the image deteriorates. High numerical aperture is achieved by using lenses with a short focal length, which requires the lens to be placed very close to the specimen.

If we substitute the minimum possible wavelength of illumination and the greatest possible numerical aperture in the preceding equation, we can determine the **limit of resolution** of the standard light microscope. When these substitutions are made, a value of slightly less than 0.2 μm (or 200 nm) is obtained, which is sufficient to resolve larger cellular organelles, such as nuclei and mitochondria. In contrast, the limit of resolution of the naked eye, which has a numerical aperture of about 0.004, is approximately 0.1 mm.

In addition to these theoretical factors, resolving power is also affected by optical flaws, or *aberrations*. There are seven important aberrations, and they are the handicaps that lensmakers must overcome to produce objective lenses whose actual resolving power approaches that of the theoretical limits. Objective lenses are made of a complex series of lenses, rather than a single convergent lens, to eliminate these aberrations. One lens unit typically affords the required magnification, while the others compensate for errors in the first lens to provide a corrected overall image.

Visibility

On the more practical side of microscopy from that of limits of resolution is the subject of *visibility*, which is concerned with factors that allow an object actually to be observed. This may seem like a trivial matter; if an object is there, it should be seen. Consider the case of a clear glass bead. Under most conditions, against most backgrounds, the bead is clearly visible. If, however, a glass bead is dropped into a beaker of immersion oil having the same refractive index as glass, the bead disappears from view because it no longer affects light in any obvious manner that is different from the background fluid. Anyone who has spent any time searching for an amoeba can appreciate the problem of visibility when using the light microscope.

What we see through a window or through a microscope are those objects that affect the light differently from their background. Another term for visibility in this sense of the word is **contrast**, or the difference in appearance between adjacent parts of an object or between an object and its back-

ground. The need for contrast can be appreciated by considering the stars. Whereas a clear night sky can be filled with stars, the same sky during the day appears devoid of celestial bodies. The stars have disappeared from view, but they haven't disappeared from the sky. They are no longer visible against the much brighter background.

In the macroscopic world, we examine objects by having the light fall on them and then observing the light that is reflected back to our eyes. In contrast, when we use a microscope, we place the specimen between the light source and our eyes and view the light that is transmitted through the object (or more properly, diffracted by the object). If you take an object and go into a room with one light source and hold the object between the light source and your eye, you can appreciate part of the difficulty in such illumination; it requires that the object being examined be nearly transparent, that is, translucent. Therein lies another aspect of the problem: objects that are "nearly transparent" can be difficult to see.

One of the best ways to make a thin, translucent specimen visible under the microscope is to stain it with a dye, which absorbs only certain wavelengths within the visible spectrum. Those wavelengths that are not absorbed are transmitted to the eye, causing the stained object to appear colored. Different dyes bind to different types of biological molecules, and therefore, not only do these procedures heighten visibility of the specimen, they can also indicate where in the cells or tissues different types of substances are found. A good example is the Feulgen stain, which is specific for DNA, causing the chromosomes to appear colored under the microscope (Figure 18.5). A problem with stains is that they generally cannot be used with living cells; they are usually toxic, or the staining conditions are toxic, or the stains do not penetrate the plasma membrane. The Feulgen stain, for example, requires that the tissue be hydrolyzed in acid before the stain is applied.

Different types of light microscopes use different types of illumination. In a **bright-field microscope**, the cone of light that illuminates the specimen is seen as a bright background against which the image of the specimen must be contrasted. Bright-field microscopy is ideally suited for specimens of high contrast, such as stained sections of tissues, but it may not provide optimal visibility for other specimens. In the following sections, we will consider various means of making specimens more visible in a light microscope.

Preparation of Specimens for Bright-Field Light Microscopy

Specimens to be observed with the light microscope are broadly divided into two categories: whole mounts and sections. A **whole mount** is an intact object, either living or dead, and can consist of an entire microscopic organism such as a protozoan or a small part of a larger organism. Most tissues of plants and animals are much too opaque for microscopic analysis unless examined as a very thin slice, or **section**. To prepare a section, the cells are first killed by immersing the tissue in a chemical solution, called a **fixative**. A good fixative rapidly penetrates the cell membrane and immobilizes all of its macromolecular material so that the structure of the cell is maintained as close as possible to that of the living state. The most common fixatives for the light microscope are solutions of formaldehyde, alcohol, or acetic acid.

After fixation, the tissue is dehydrated by transfer through a series of alcohols and then usually embedded in paraffin (wax), which provides mechanical support during sectioning. Paraffin is used as an embedding medium because it is readily dissolved by organic solvents. Slides containing adherent paraffin sections are immersed in toluene, which dissolves the wax, leaving the thin slice of tissue attached to the slide, where it can be stained or treated with antibodies or other agents. After staining, a coverslip is mounted over the tissue using a mounting medium that has the same refractive index as the glass slide and coverslip.

FIGURE 18.5   **The Feulgen stain.** This staining procedure is specific for DNA as indicated by the localization of the dye to the chromosomes of this onion root tip cell that was in metaphase of mitosis at the time it was fixed.   (By Ed Reschke/Peter Arnold, Inc.)

Phase-Contrast Microscopy

Small, unstained specimens such as a living cell can be very difficult to see with a bright-field microscope (Figure 18.6*a*). The **phase-contrast microscope** makes highly transparent objects more visible (Figure 18.6*b*). We can distinguish different parts of an object because they affect light differently from one another. One basis for such differences is the refractive index. Cell organelles are made up of different proportions of various molecules: DNA, RNA, protein, lipid, carbohydrate, salts, and water. Regions of different composition are likely to have different refractive indexes. Normally such differences cannot be detected by our eyes. However, the phase-contrast microscope converts differences in refractive index into differences in intensity (relative brightness and darkness), which are
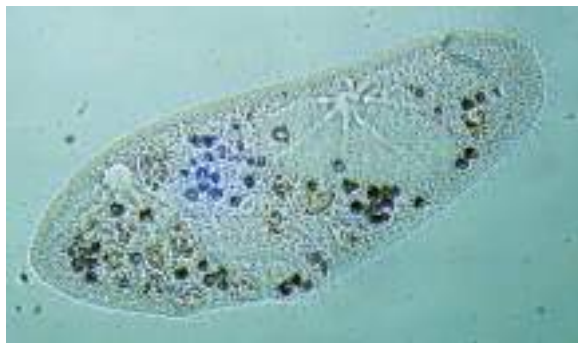
(a)

(b)

(c)

FIGURE 18.6 **A comparison of cells seen with different types of light microscopes.** Light micrographs of a ciliated protist as observed under bright-field (*a*), phase-contrast (*b*), and differential interference contrast (DIC) (or Nomarski) optics (*c*). The organism is barely visible under bright-field illumination but clearly seen under phase-contrast and DIC microscopy. (Micrographs by M. I. Walker/Photo Researchers, Inc.)

cles and vacuoles of cells are bumped around in a random manner in a living cell conveys an excitement about life that is unattainable by observing stained, dead cells. The greatest benefit derived from the invention of the phase-contrast microscope has not been in the discovery of new structures, but in its everyday use in research and teaching labs for observing cells in a more revealing way.

The phase-contrast microscope has optical handicaps that result in loss of resolution, and the image suffers from interfering halos and shading produced where sharp changes in refractive index occur. The phase-contrast microscope is a type of *interference microscope*. Other types of interference microscopes minimize these optical artifacts by achieving a complete separation of direct and diffracted beams using complex light paths and prisms. Another type of interference system, termed *differential interference contrast* (*DIC*), or sometimes Nomarski interference after its developer, delivers an image that has an apparent three-dimensional quality (Figure 18.6*c*). Contrast in DIC microscopy depends on the rate of change of refractive index across a specimen. As a consequence, the edges of structures, where the refractive index varies markedly over a relatively small distance, are seen with especially good contrast.

Fluorescence Microscopy (and Related Fluorescence-Based Techniques)

Over the past couple of decades, the light microscope has been transformed from an instrument designed primarily to examine sections of fixed tissues to one capable of observing the dynamic events occurring at the molecular level in living cells. These advances in live-cell imaging have been made possible to a large extent by innovations in *fluorescence microscopy*. The fluorescence microscope allows viewers to observe the location of certain compounds (called *fluorochromes* or *fluorophores*). Fluorochromes absorb invisible, ultraviolet radiation and release a portion of the energy in the longer, visible wavelengths, a phenomenon called *fluorescence*. The light source in a fluorescence microscope produces a beam of ultraviolet light that travels through a filter, which blocks all wavelengths except that which is capable of exciting the fluorochrome. The beam of monochromatic light is focused on the specimen containing the fluorochrome, which becomes excited and emits light of a visible wavelength that is focused by the objective lens into an image that can be seen by the viewer. Because the light source produces only ultraviolet (black) light, objects stained with a fluorochrome appear brightly colored against a black background, providing very high contrast.

There are many different ways that fluorescent compounds can be used in cell and molecular biology. In one of its most common applications, a fluorochrome (such as rhodamine or fluorescein) is covalently linked (*conjugated*) to an antibody to produce a fluorescent antibody that can be used to determine the location of a specific protein within the cell. This technique is called *immunofluorescence* and is illustrated by the micrograph in Figure 9.29*a*. Immunofluorescence is

visible to the eye. Phase-contrast microscopes accomplish this result by (1) separating the direct light that enters the objective lens from the diffracted light emanating from the specimen and (2) causing light rays from these two sources to *interfere* with one another. The relative brightness or darkness of each part of the image reflects the way in which the light from that part of the specimen interferes with the direct light.

Phase-contrast microscopes are most useful for examining intracellular components of living cells at relatively high resolution. For example, the dynamic motility of mitochondria, mitotic chromosomes, and vacuoles can be followed and filmed with these optics. Simply watching the way tiny parti-

discussed further on page 765. Fluorochromes can also be used to locate DNA or RNA molecules that contain specific nucleotide sequences as described on page 397 and illustrated in Figure 10.19. In other examples, fluorochromes have been used to study the sizes of molecules that can pass between cells (see Figure 7.33), as indicators of transmembrane potentials (see Figure 5.20), or as probes to determine the free Ca^{2+} concentration in the cytosol (see Figure 15.27). The use of calcium-sensitive fluorochromes is discussed on page 634.

Fluorescently labeled proteins can also be used to study dynamic processes as they occur in a living cell. For example, a specific fluorochrome can be linked to a cellular protein, such as actin or tubulin, and the fluorescently labeled protein injected into a living cell (as in Figures 9.4 and 9.73*b*). A noninvasive approach that has been widely employed utilizes a fluorescent protein (called green fluorescent protein, or GFP) from the jellyfish *Aequorea victoria* as illustrated in the chapter-opening photos on page 715. Unlike most other fluorescent proteins, GFP does not require an additional cofactor to absorb and emit light; instead, the light-absorbing/emitting chromophore is formed by self-modification (i.e., by an autocatalytic reaction) of three of the amino acids that make up the primary structure of the GFP polypeptide. In most studies employing GFP, a recombinant DNA is constructed in which the coding region of the GFP gene is joined to the coding region of the gene for the protein under study. This recombinant DNA is used to transfect cells, which then synthesize a chimeric protein containing fluorescent GFP fused to the protein under study. Use of GFP to study membrane dynamics is discussed on page 267. In these various experimental protocols, the labeled proteins participate in the normal activities of the cell, and their location can be followed microscopically to reveal the dynamic activities in which the protein participates (see Figures 8.4 and 9.2).

Studies can often be made more informative by the simultaneous use of GFP variants that exhibit different spectral properties. Variants of GFP that fluoresce in shades of blue (BFP), yellow (YFP), and cyan (CFP) have been generated through directed mutagenesis of the *GFP* gene. In addition, a distantly related red fluorescent tetrameric protein (DsRed) has been isolated from a sea anemone. Monomeric variants of DsRed (e.g., mBanana, mTangerine, and mOrange), which fluoresce in a variety of distinguishable colors, have also been generated by mutagenesis experiments. The type of information that can be obtained using this "palette" of GFP variants is illustrated by the study depicted in Figure 18.7, in which researchers generated strains of mice whose neurons contained differently colored fluorescent proteins. When a muscle of one of these mice was exposed surgically, the investigators could observe the dynamic interactions between the variously colored neurons and the neuromuscular junctions being innervated (see Figure 4.55 for a drawing of this type of junction). They watched, for example, as branches from a CFP-colored neuron competed with branches from a YFP-colored neuron for synaptic contact with the muscle tissue. In each case they found that, when two neurons compete for innervation of different muscle fibers, all of the "winning" branches belong to one of the neurons, while

(a)   *(b)*
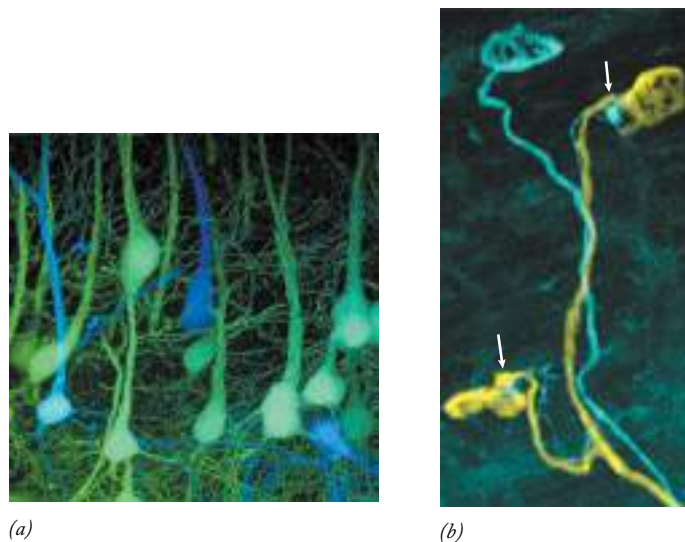
FIGURE 18.7 **Use of GFP variants to follow the dynamic interactions between neurons and their target cells in vivo.** (*a*) A portion of the brain of a mouse with two differently colored, fluorescent neurons. These mice are generated by mating transgenic animals whose neurons are labeled with one or the other flurosecent protein. (*b*) Fluorescence micrograph of portions of two neurons, one labeled with YFP and the other with CFP. The arrows indicate two different neuromuscular junctions on two different muscle fibers in which the YFP-labeled axonal branch has outcompeted the CFP-labeled branch. The third junction is innervated by the CFP-labeled axon in the absence of competition. (A: Courtesy N. Kasthuri & J. W. Lichtman, Washington University School of Medicine; B: Reprinted with permission from Narayanan Kasthuri and Jeff W. Lichtman, Nature 424:429, 2003; © copyright 2003, Macmillan Magazines Ltd.)

all of the "losing" branches belong to the other neuron (Figure 18.7*b*). The chapter-opening micrographs provide another example of how much can be learned by labeling proteins with two different fluorochromes. In this case, the dual-label strategy has allowed investigators to follow the simultaneous movements of two different proteins in real time as they occur within the boundaries of a single cellular organelle.

GFP variants have also been useful in a technique, called *fluorescence resonance energy transfer* (*FRET*), which can measure distances between fluorochromes in the nanoscale range. FRET is typically employed to measure changes in distance between two parts of a protein (or between two separate proteins within a larger structure). FRET can be used to study such changes as they occur in vitro or within a living cell. FRET is based on the fact that excitation energy can be transferred from one fluorescent group (a donor) to another fluorescent group (an acceptor), as long as the two groups are in very close proximity (1 to 10 nm). This transfer of energy reduces the fluorescence intensity of the donor and increases the fluorescence intensity of the acceptor. The efficiency of transfer between two fluorescent groups that are bound to sites on a protein decreases sharply as the distance between the two groups increases. As a result, determination of changes in fluorescence of the donor and acceptor groups that occur during a process provides a
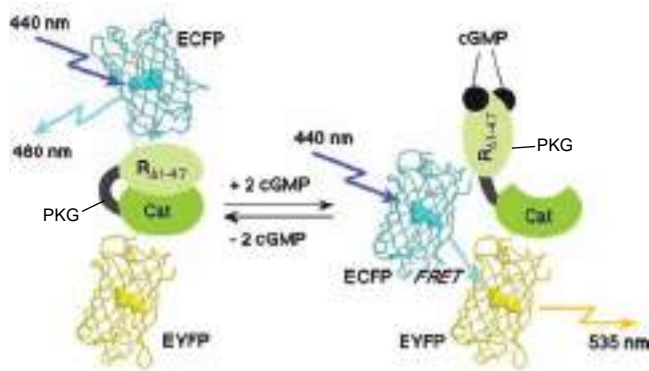
FIGURE 18.8 Fluorescence resonance energy transfer (FRET). This schematic diagram shows an example of the use of FRET technology to follow the change in conformation of a protein (PKG) following cGMP binding. The two small, barrel-shaped fluorescent proteins—enhanced CFP (ECFP) and enhanced YFP (EYFP)—are shown in their fluorescent color. In the absence of cGMP, excitation of ECFP with light of 440 nm leads to emission of light of 480 nm from the fluorescent protein. Following cGMP binding, a conformational change in PKG brings the two fluorescent proteins into close enough proximity for FRET to occur. As a result, excitation of the ECFP donor with light of 440 nm leads to energy transfer to the EYFP acceptor and emission of light of 535 nm from the acceptor. The enhanced versions of these proteins have greater fluorescence intensity and tend to be more stable than the original protein molecule. (FROM MORITOSHI SATO ET AL., COURTESY OF YOSHIO UMEZAWA, ANAL. CHEM. 72:5924, 2000.)

measure of changes in the distance between them at various stages in the process. This technique is illustrated in Figure 18.8, where two different GFP variants (labeled ECFP and EYFP) have been covalently linked to two different parts of a cGMP-dependent protein kinase (PKG). In the absence of bound cGMP, the two fluorochromes are too far apart for energy transfer to occur. Binding of cGMP induces a conformational change in the protein that brings the two fluorochromes into close enough proximity for FRET to occur. FRET can also be used to follow many other processes including protein folding or the association and dissociation of components within a membrane. The separation of the cytoplasmic tails of integrin subunits following activation by talin (Figure 7.14) is an example of an event that has been studied by FRET.

Video Microscopy and Image Processing

Just as a microscopic field can be seen with the eyes or filmed by a camera, it can also be viewed electronically and taped using a video camera. Video cameras offer a number of advantages for viewing specimens. Special types of video cameras (called charge-coupled device, or CCD cameras) are constructed to be very sensitive to light, which allows them to image specimens at very low illumination. This is particularly useful when observing live specimens, which are easily damaged by the heat from a light source, and fluorescently stained specimens, which fade rapidly on exposure to light. In addition, video cameras can detect and amplify very small differences in contrast within

a specimen so that very small objects become visible. The photographs in Figure 9.22*a*, for example, show images of individual microtubules (diameter of 0.025 μm) that are far below the limit of resolution of a standard light microscope (0.2 μm). As an added advantage, images produced by video cameras can be converted to digital electronic images and subjected to computer processing to greatly increase their information content.[1] In one technique, the distracting out-of-focus background in a visual field is stored by the computer and then subtracted from the image containing the specimen. This greatly increases the clarity of the image. Similarly, differences in brightness in an image can be converted to differences in color, which makes them much more apparent to the eye.

Laser Scanning Confocal Microscopy

The use of video cameras, electronic images, and computer processing has led to a renaissance in light microscopy over the past couple of decades. So too has the development of a new type of light microscope. When a whole cell or a section of an organ is examined under a standard light microscope, the observer views the specimen at different depths by changing the position of the objective lens by rotating the focusing knob. As this is done, different parts of the specimen go in and out of focus. But the fact that the specimen contains different levels of focus reduces the ability to form a crisp image because those parts of the specimen above and below the plane of focus interfere with the light rays from that part that is in the plane of focus. In the late 1950s, Marvin Minsky of the Massachusetts Institute of Technology invented a revolutionary new instrument called a **laser scanning confocal microscope** that produces an image of a thin plane situated within a much thicker specimen. A schematic diagram of the optical components and light paths within a modern version of a laser scanning confocal microscope is shown in Figure 18.9*a*. In this type of microscope, the specimen is illuminated by a finely focused laser beam that rapidly scans across the specimen at a single depth, thus illuminating only a thin plane (or "optical section") within the specimen. As indicated in Figure 18.9*a*, confocal microscopes are typically used with fluorescence optics. As described earlier, short-wavelength incident light is absorbed by the fluorochromes in a specimen and reemitted at longer wavelength. Light emitted from the specimen is brought to focus at a site within the microscope that contains a pinhole aperture. Thus, the aperture and the illuminated plane in the specimen are confocal. Light rays emitted from the illuminated plane of the specimen can pass through the aperture, whereas any light rays that might emanate from above or below this plane are prevented from participating in image formation. As a result, out-of-focus points in the specimen become invisible.

---

[1]The conversion of an analog electronic image, such as that obtained by a video camera, to a digital electronic image, such as that obtained by a digital camera and stored on a compact disc, is called digitization. Digital images consist of a discrete number of picture elements (pixels), each of which has an assigned color and brightness value corresponding to that site in the original image.

(a)

(b)
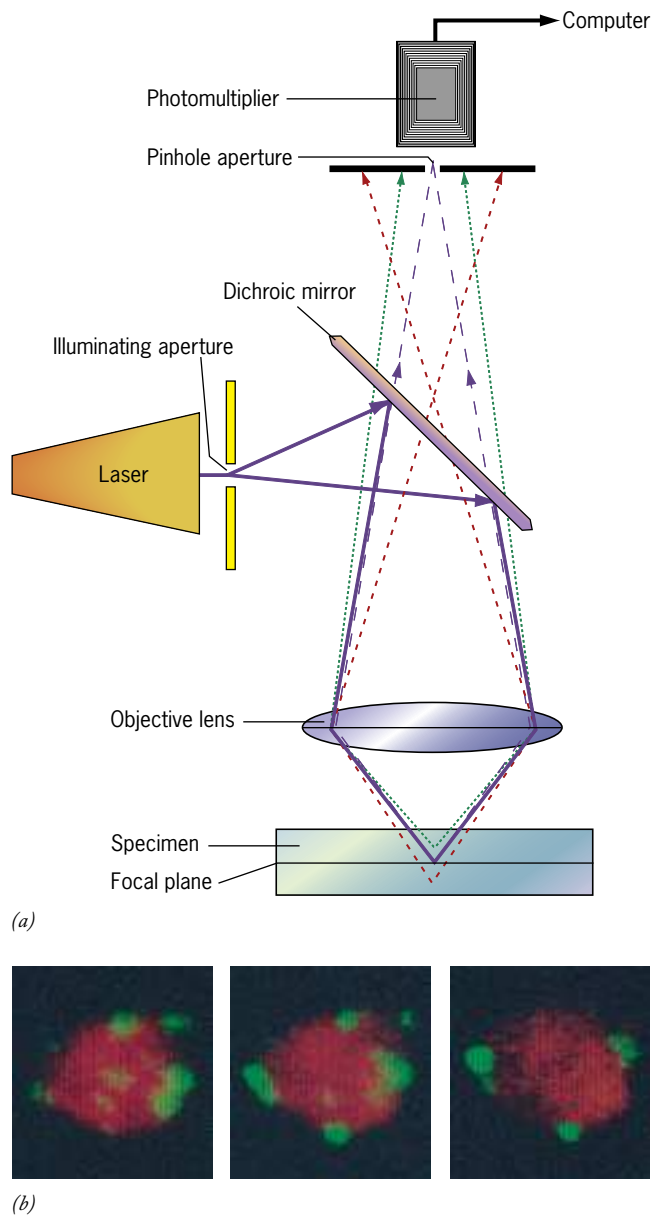
FIGURE 18.9 Laser scanning confocal fluorescence microscopy. (*a*) The light paths in a confocal fluorescence microscope. Light of short (blue) wavelength is emitted by a laser source, passes through a tiny aperture, and is reflected by a dichroic mirror (a type of mirror that reflects certain wavelengths and transmits others) into an objective lens and focused onto a spot in the plane of the specimen. Fluorochromes in the specimen absorb the incident light and emit light of longer wavelength, which is able to pass through the dichroic mirror and come to focus in a plane that contains a pinhole aperture. The light then passes into a photomultiplier tube that amplifies the signal and is transmitted to a computer which forms a processed, digitized image. Any light rays that are emitted from above or below the optical plane in the specimen are prevented from passing through the pinhole aperture and thus do not contribute to formation of the image. This diagram shows the illumination of a single spot in the specimen. Different sites within this specimen plane are illuminated by means of a laser scanning process. The diameter of the pinhole aperture is adjustable. The smaller the aperture, the thinner the optical section and the greater the resolution, but the less intense the signal. (*b*) Confocal fluorescence micrographs of three separate optical sections, each 0.3 μm thick, of a yeast nucleus stained with two different fluorescently labeled antibodies. The red fluorescent antibody has stained the DNA within the nucleus, and the green fluorescent antibody has stained a telomere-binding protein that is localized at the periphery of the nucleus. (FROM THIERRY LAROCHE AND SUSAN M. GASSER, CELL 75:543, 1993, BY PERMISSION OF CELL PRESS.)

croscopy), which allows investigators to localize a single fluorescent molecule within a resolution of less than 20 nm. In this technique, the specimen is illuminated with light of different wavelengths, which has the effect of switching the fluorescence activity of the labeled molecules on and off. During each cycle of illumination, most of the labeled molecules remain dark, but a small fraction of them are randomly activated. As a result, these individual, activated fluorescent molecules can be imaged, and their position determined with very high accuracy. This process is then repeated for a number of imaging cycles, resulting in the construction of a super-resolution image of the field of view. The marked increase in resolution that can be attained with the super-resolution STORM technique versus that of a conventional fluorescence technique is evident by comparing the image of Figure 18.10*a* with those of Figures 18.10 *b* and *c*.

Figure 18.9*b* shows micrographs of a single cell nucleus taken at three different planes within the specimen. It is evident that objects out of the plane of focus have little effect on the quality of the image of each section. If desired, the images from separate optical sections can be stored in a computer and merged with one another to reconstruct a three-dimensional model of the entire object.

Super-Resolution Fluorescence Microscopy

The most recent innovations in fluorescence microscopy fall under the umbrella of "super-resolution" technologies. It was pointed out on page 717 that the limit of resolution of the light microscope is approximately 200 nm owing to the diffraction properties of light. Over the past few years, a number of complex optical techniques have been developed that allow researchers to localize fluorescently labeled proteins at resolutions in the tens of nanometers. We will briefly consider only one of these techniques, called STORM (stochastic optical reconstruction mi-

18.2 TRANSMISSION ELECTRON MICROSCOPY

The electron micrographs shown in this text were taken by two different types of electron microscopes. *Transmission electron microscopes* (*TEMs*) form images using electrons that are transmitted through a specimen, whereas *scanning electron microscopes* (*SEMs*) utilize electrons that have bounced off the surface of the specimen. All of the comments in this section of the chapter concern the use of the TEM; the SEM is discussed separately on page 729.

The transmission electron microscope can provide vastly greater resolution than the light microscope. This is readily illustrated by comparing the two photos in Figure 18.11, which show images of adjacent sections of the same tissue at the same magnification using a light or an electron microscope. Although the photo in Figure 18.11*a* is near the limit of resolution of the light microscope, the photo in Figure 18.11*b* is a

FIGURE 18.10 **Breaking the light microscope's limit of resolution.** (*a*) A conventional fluorescence micrograph of a portion of a cultured mammalian cell with microtubules labeled in green and clathrin-coated pits in red. At this high level of magnification, the image appears pixelated. Moreover, the overlap between the two fluorescent labels produces an orange color, which suggests that the two structures are interconnected. (*b*) A STORM "super-resolution" micrograph of a similar field showing the microtubules and clathrin-coated pits clearly resolved and spatially separated from one another. (*c*) A magnified view of a portion of part *b*. (Discussion of this and other super-resolution techniques can be found in *Nature Revs. Mol. Cell Biol.* 9:929, 2008.)
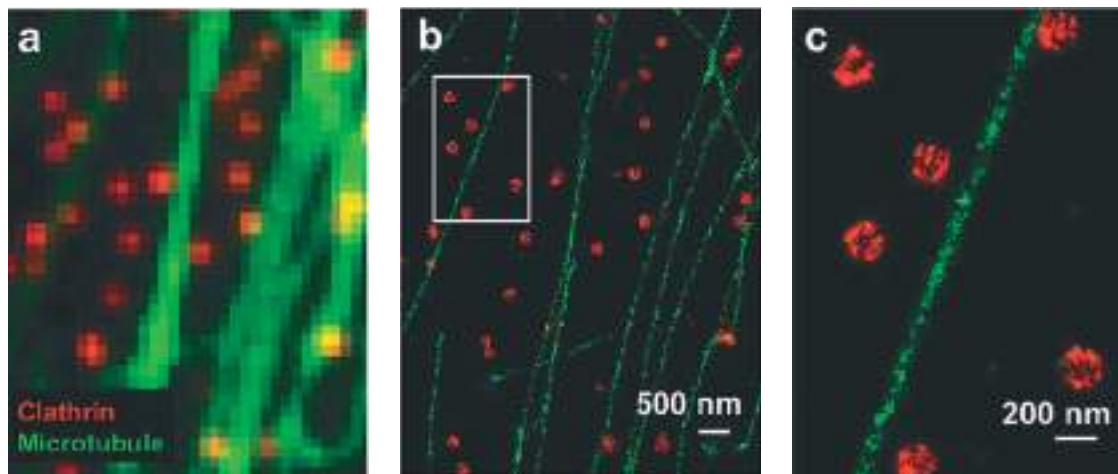
(FROM MARK BATES ET AL., COURTESY OF XIAOWEI ZHUANG, SCIENCE 317:1752, 2007; © COPYRIGHT 2007, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)

very low-power electron micrograph. The great resolving power of the electron microscope derives from the wave properties of electrons. As indicated in the equation on page 717, the limit of resolution of a microscope is directly proportional to the wavelength of the illuminating light: the longer the wavelength, the poorer the resolution. Unlike a photon of light, which has a constant wavelength, the wavelength of an electron varies with the speed at which the particle is traveling, which in turn depends on the accelerating voltage applied in the microscope. This relationship is defined by the equation

$$\lambda = \sqrt{150/V}$$

where λ is the wavelength in angstroms and V is the accelerating voltage in volts. Standard TEMs operate with a voltage range from 10,000 to 100,000 V. At 60,000 V, the wavelength of an electron is approximately 0.05 Å. If this wavelength and

the numerical aperture attainable with the light microscope were substituted into the equation on page 717, the limit of resolution would be about 0.03 Å. In actual fact, the resolution attainable with a standard transmission electron microscope is about two orders of magnitude less than its theoretical limit. This is due to serious spherical aberration of electron-focusing lenses, which requires that the numerical aperture of the lens be made very small (generally between 0.01 and 0.001). The practical limit of resolution of standard TEMs is in the range of 3 to 5 Å. The actual limit when observing cellular structure is more typically in the range of 10 to 15 Å.

Electron microscopes consist largely of a tall, hollow cylindrical column through which the electron beam passes and a console containing a panel of dials that electronically control the operation in the column. The top of the column contains the cathode, a tungsten wire filament that is heated to provide a
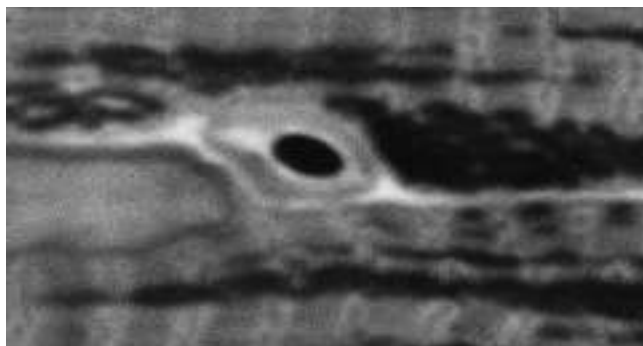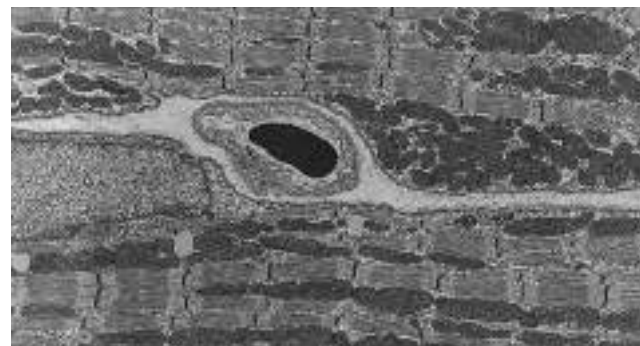
FIGURE 18.11 **A comparison between the information contained in images taken by a light and electron microscope at a comparable magnification of 4500 times actual size.** (*a*) A photo of skeletal muscle tissue that had been embedded in plastic, sectioned at 1 μm, and photographed with a light microscope under an oil immersion objective lens. (*b*) An adjacent section to that used for part *a* that was cut at 0.025 μm and examined under the electron microscope at comparable magnification to that in *a*. The resulting image displays a one- to two-hundred-fold

increase in resolution. Note the difference in the details of the muscle myofibrils, mitochondria, and the capillary containing a red blood cell. Whereas the light microscope cannot provide any additional information to that of *a*, the electron microscope can provide much more information, producing images, for example, of the structure of the individual membranes within a small portion of one of the mitochondria (as in Figure 5.21). (COURTESY OF DOUGLAS E. KELLY AND M. A. CAHILL.)

source of electrons. Electrons are drawn from the hot filament and accelerated as a fine beam by the high voltage applied between the cathode and anode. Air is pumped out of the column prior to operation, producing a vacuum through which the electrons travel. If the air were not removed, electrons would be prematurely scattered by collision with gas molecules.

Just as a beam of light rays can be focused by a ground glass lens, a beam of negatively charged electrons can be focused by electromagnetic lenses, which are located in the wall of the column. The strength of the magnets is controlled by the current provided them, which is determined by the positions of the various dials of the console. A comparison of the lens systems of a light and electron microscope is shown in Figure 18.12. The condenser lenses of an electron microscope are placed between the electron source and the specimen, and they focus the electron beam on the specimen. The specimen is supported on a small, thin metal grid (3 mm diameter) that is inserted with tweezers into a grid holder, which in turn is inserted into the column of the microscope.

Because the focal lengths of the lenses of an electron microscope vary depending on the current supplied to them, one objective lens provides the entire range of magnification delivered by the instrument. As in the light microscope, the image from the objective lens serves as the object for an additional lens system. The image provided by the objective lens of an electron microscope is only magnified about 100 times, but unlike the light microscope, there is sufficient detail present in this image to magnify it an additional 10,000 times. By altering the current applied to the various lenses of the microscope, magnifications
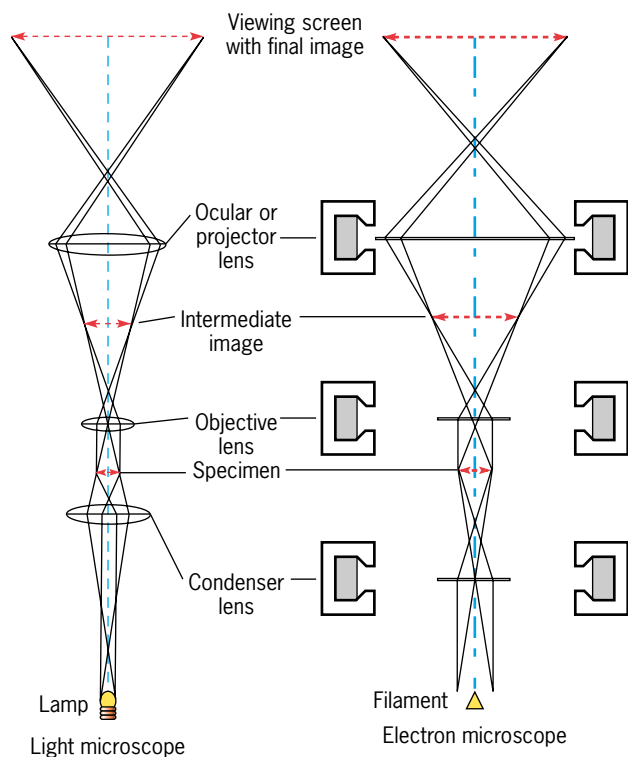
can vary from about 1000 times to 250,000 times. Electrons that have passed through the specimen are brought to focus on a phosphorescent screen situated at the bottom of the column. Electrons striking the screen excite a coating of fluorescent crystals, which emit their own visible light that is perceived by the eye as an image of the specimen.

Image formation in the electron microscope depends on differential scattering of electrons by parts of the specimen. Consider a beam of electrons emitted by the filament and focused on the screen. If no specimen were present in the column, the screen would be evenly illuminated by the beam of electrons, producing an image that is uniformly bright. By contrast, if a specimen is placed in the path of the beam, some of the electrons strike atoms in the specimen and are scattered away. Electrons that bounce off the specimen cannot pass through the very small aperture at the back focal plane of the objective lens and are, therefore, lost as participants in the formation of the image.

The scattering of electrons by a part of the specimen is proportional to the size of the nuclei of the atoms that make up the specimen. Because the insoluble material of cells consists of atoms of relatively low atomic number—carbon, oxygen, nitrogen, and hydrogen—biological material possesses very little intrinsic capability to scatter electrons. To increase electron scattering and obtain required contrast, tissues are fixed and stained with solutions of heavy metals (described below). These metals penetrate into the structure of the cells and become selectively complexed with different parts of the organelles. Those parts of a cell that bind the greatest number of metal atoms allow passage of the least number of electrons. The fewer electrons that are focused on the screen at a given spot, the darker that spot. Photographs of the image are made by lifting the viewing screen out of the way and allowing the electrons to strike a photographic plate in position beneath the screen. Because photographic emulsions are directly sensitive to electrons, much as they are to light, an image of the specimen can be recorded directly on film. Alternatively, the image can be captured on a CCD video camera (page 721) used to monitor the field of view. Although a video camera provides an instant image without the need for chemical development, the image lacks the exceptionally high resolution that is available with film.

Specimen Preparation for Electron Microscopy

As with the light microscope, tissues to be examined in the electron microscope must be fixed, embedded, and sectioned. Fixation of tissue for electron microscopy (Figure 18.13) is much more critical than for light microscopy because the sections are subjected to much greater scrutiny. A fixative must stop the life of the cell without significantly altering the structure of that cell. At the level of resolution of the electron microscope, relatively minor damage, such as swollen mitochondria or ruptured endoplasmic reticulum, becomes very apparent. To obtain the most rapid fixation and the least cellular damage, very small pieces of tissue (less than 1.0 mm^3) are fixed and embedded. Fixatives are chemicals that denature and precipitate cellular macromolecules. Chemicals having such action may cause the coagulation or precipitation of materials that had no structure in the living cell, leading to the formation of an **artifact**. The best argument that

FIGURE 18.12   A comparison of the lens systems of a light and electron microscope. (FROM W. AGAR, PRINCIPLES AND PRACTICE OF ELECTRON MICROSCOPE OPERATION, ELSEVIER/NORTH-HOLLAND, 1974.)

a particular structure is not an artifact is the demonstration of its existence in cells fixed in a variety of different ways or, even better, not fixed at all. To view cells that have not been fixed, the tissue is rapidly frozen, and special techniques are utilized to reveal its structure (see cryofixation and freeze-fracture replication, described later). The most common fixatives for electron microscopy are glutaraldehyde and osmium tetroxide. Glutaraldehyde is a 5-carbon compound with an aldehyde group at each end of the molecule. The aldehyde groups react with amino groups in proteins and cross-link the proteins into an insoluble network. Osmium is a heavy metal that reacts primarily with fatty acids leading to the preservation of cellular membranes.

Once the tissue has been fixed, the water is removed by dehydration in alcohol, and the tissue spaces are filled with a material that supports tissue sectioning. The demands of electron microscopy require the sections to be very thin. The wax sec-

tions cut for light microscopy are rarely thinner than about 5 μm, whereas sections for conventional electron microscopy are best when cut at less than 0.1 μm (equivalent in thickness to about four ribosomes).

Tissues to be sectioned for electron microscopy are usually embedded in epoxy resins, such as Epon or Araldite. Sections are cut by slowly bringing the plastic block down across an extremely sharp cutting edge (Figure 18.13) made of cut glass or a finely polished diamond face. The sections coming off the knife edge float onto the surface of a trough of water that is contained just behind the knife edge. The sections are then picked up with the metal specimen grid and dried onto its surface. The tissue is stained by floating the grids on drops of heavy metal solutions, primarily uranyl acetate and lead citrate. These heavy metal atoms bind to macromolecules and provide the atomic density required to scatter the electron beam. In ad-
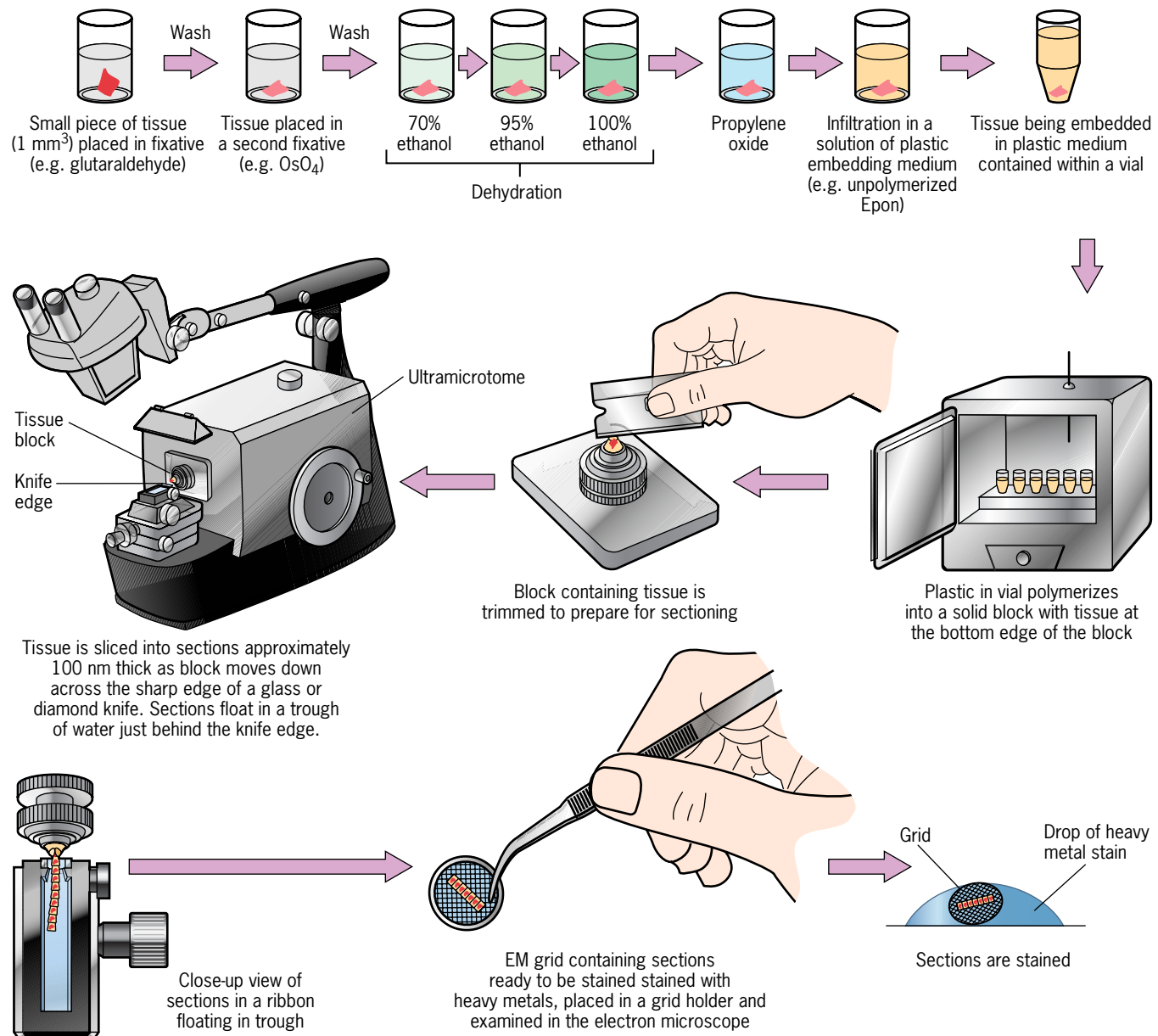
FIGURE 18.13 **Preparation of a specimen for observation in the electron microscope.**

dition to the standard stains, tissue sections can be treated with metal-tagged antibodies or other materials that react with specific molecules in the tissue section. Studies with antibodies are usually carried out on tissues embedded in acrylic resins, which are more permeable to large molecules than epoxy resins.

Cryofixation and the Use of Frozen Specimens   Cells and tissues do not have to be fixed with chemicals and embedded in plastic resins in order to be observed with the electron microscope. Alternatively, cells and tissues can be rapidly frozen. Just as a chemical fixative stops metabolic processes and preserves biological structure, so too does rapid freezing, which is called *cryofixation*. Because cryofixation accomplishes these goals without altering the cell's macromolecules, it is less likely to lead to the formation of artifacts. The major difficulty with cryofixation is the formation of ice crystals, which grow outward from sites where nucleation occurs. As an ice crystal grows, it destroys the fragile contents of the cell in which it develops. The best way to avoid ice crystal formation is to freeze the specimen so rapidly that crystals don't have time to grow. It is as if the water is frozen yet remains in its liquid state. Water in this state is said to be "vitrified." There are several techniques employed to achieve such ultrarapid freezing rates. Smaller specimens are typically plunged into a liquid of very low temperature (such as liquid propane, boiling point of $-42°C$). Larger specimens are best treated by high-pressure freezing. In this technique, the specimen is subjected to high hydrostatic pressure and sprayed with jets of liquid nitrogen. High pressure lowers the freezing point of water, reducing the rate of ice crystal growth.

You might not think that a frozen block of tissue would be of much use to a microscopist, but a surprising number of approaches can be taken to visualize frozen cellular structure in the light or electron microscope. For example, after suitable preparation, a frozen block of tissue can be sectioned with a special microtome in a manner similar to that of a paraffin or plastic block of tissue. Frozen sections (cryosections) can be prepared for examination under either a light or an electron microscope. Frozen sections are particularly useful for studies on enzymes, whose activities tend to be denatured by chemical fixatives. Because frozen sections can be prepared much more rapidly than paraffin or plastic sections, they are often employed by pathologists to examine the light microscopic structure of tissues removed during surgery. As a result, the determination as to whether or not a tumor is malignant can be made while the patient is still on the operating table.

Frozen cells do not have to be sectioned to reveal internal structure. Figure 1.11 shows an image of the thin, peripheral region of an intact cell that had been crawling over the surface of an electron microscope grid an instant before it was rapidly frozen. Unlike the standard electron micrograph, the image in Figure 1.11 has a three-dimensional quality because it was generated by a computer rather than directly by a camera. To obtain the image, the computer aligns a large number of two-dimensional digital images of the cell that are captured as the specimen is tilted at defined angles relative to the path of the electron beam. The three-dimensional, computerized reconstruction is called a *tomogram,* and the technique is called *cryoelectron tomography* (*cryo-ET*).[2] Cryo-ET, which was developed by Wolf-

gang Baumeister of the Max-Planck Institute in Germany, has revolutionized the way in which nanosized intracellular structures can be studied in unfixed, fully hydrated, flash-frozen cells. Cryo-ET can also be utilized to examine the three-dimensional organization of structures present in vitro, as exemplified by the reconstruction of an isolated polysome in the act of translation that is shown in Figure 11.50*b*. Cryo-ET, which can deliver a resolution of a few nms, provides an important bridge between the cellular and molecular worlds. Two other approaches that require the electron microscopic analysis of frozen specimens—freeze fracture replication and single particle analysis—are discussed on pages 728 and 741.

Negative Staining   The electron microscope is also well suited for examining very small particulate materials, such as viruses, ribosomes, multisubunit enzymes, cytoskeletal elements, and protein complexes. The shapes of individual proteins and nucleic acids can also be resolved as long as they are made to have sufficient contrast from their surroundings. One of the best ways to make such substances visible is to employ **negative staining** procedures in which heavy-metal deposits are collected everywhere on a specimen grid except where the particles are present. As a result, the specimen stands out by its relative brightness on the viewing screen. An example of a negatively stained specimen is shown in Figure 18.14*a*.

Shadow Casting   Another technique to visualize isolated particles is to have the objects cast shadows. The technique is described in Figure 18.15. The grids containing the specimens are placed in a sealed chamber, which is then evacuated by vacuum pump. The chamber contains a filament composed of a heavy metal (usually platinum) together with carbon. The filament is heated to high temperature, causing it to evaporate and deposit a metallic coat over accessible surfaces within the chamber. As a result, the metal is deposited on those surfaces facing the filament, while the opposite surfaces of the particles and the grid space in their shadow remain uncoated. When the grid is viewed in the electron microscope, the areas in shadow appear bright on the viewing screen, whereas the metal-coated regions appear dark. This relationship is reversed on the photographic plate, which is a negative of the image. The convention for illustrating shadowed specimens is to print a negative image in which the particle appears illuminated by a bright, white light (corresponding to the coated surface) with a dark shadow cast by the particle (Figure 18.14*b*). The technique provides excellent contrast and produces a three-dimensional effect.

Freeze-Fracture Replication and Freeze Etching   As noted above, a number of electron microscopic techniques have been adapted to work with frozen tissues. The ultrastructure of frozen cells is often viewed using the technique of **freeze-fracture replication**, which is illustrated in Figure 18.16. Small pieces of tissue are placed on a small metal disk and rap-

---

[2]This technique is similar in principle to computerized axial tomography (CAT scans), which uses a multitude of X-ray images taken at different angles to the body to generate a three-dimensional image. Fortunately, the machinery used in radiological tomography allows the X-ray source and detector to rotate while the patient can remain stationary.
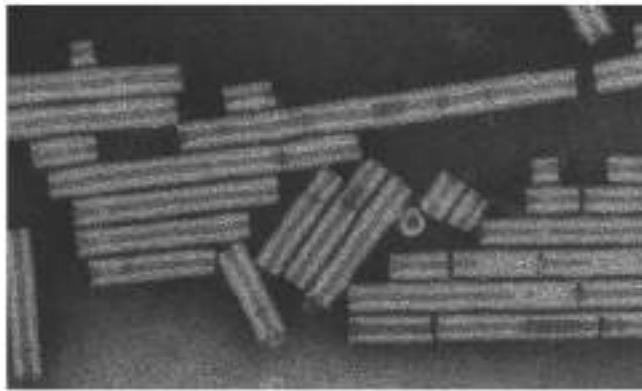
(a)

(b)

FIGURE 18.14 Examples of negatively stained and metal-shadowed specimens. Electron micrographs of a tobacco rattle virus after negative staining with potassium phosphotungstate (*a*) or shadow casting with chromium (*b*). (COURTESY OF M. K. CORBETT.)
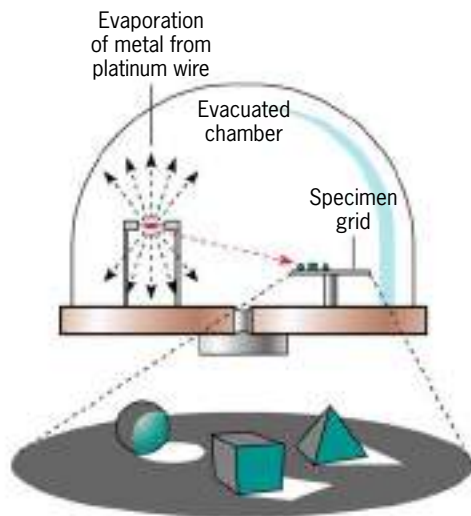
FIGURE 18.15 The procedure used for shadow casting as a means to provide contrast in the electron microscope. This procedure is often used to visualize small particles, such as the viruses shown in the previous figure. DNA and RNA molecules are often made visible by a modification of this procedure known as rotary shadowing in which the metal is evaporated at a very low angle while the specimen is rotated.
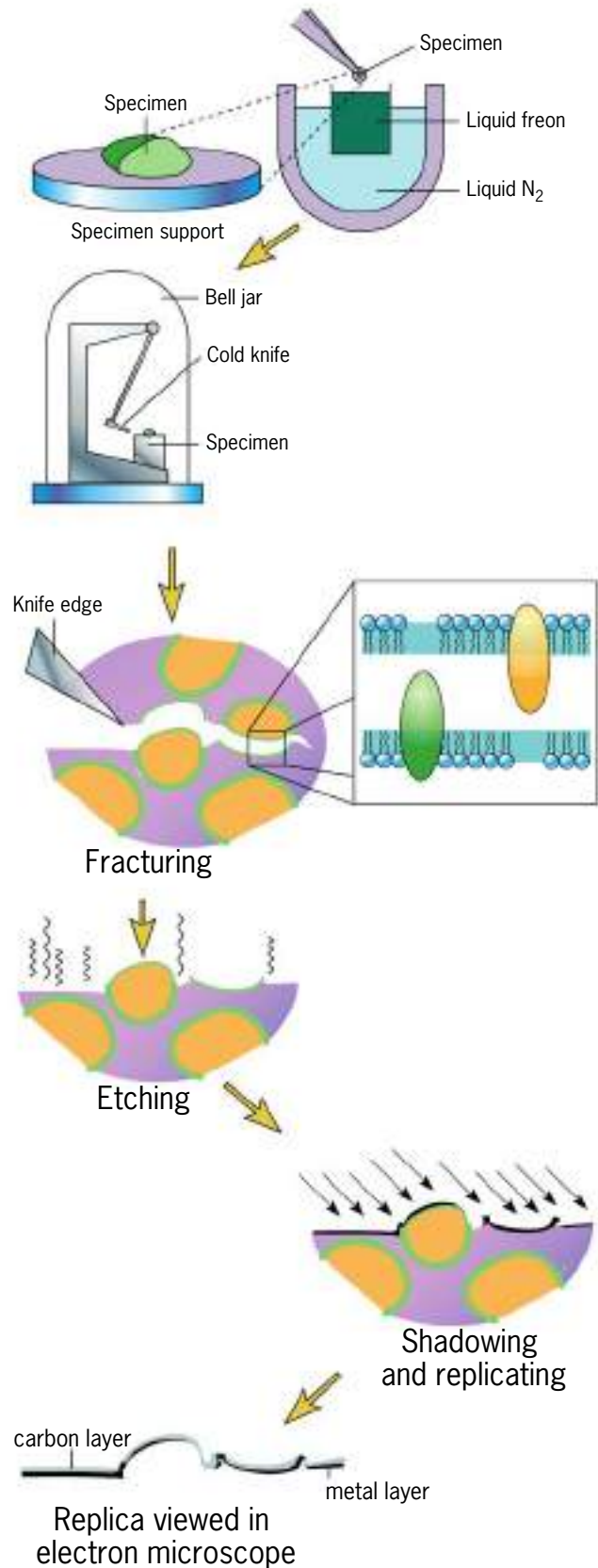
FIGURE 18.16 Procedure for the formation of freeze-fracture replicas as described in the text. Freeze etching is an optional step in which a thin layer of covering ice is evaporated to reveal additional information about the structure of the fractured specimen.

idly frozen. The disk is then mounted on a cooled stage within a vacuum chamber, and the frozen tissue block is struck by a knife edge. The resulting fracture plane spreads out from the point of contact, splitting the tissue into two pieces, not unlike the way that an axe blade splits a piece of wood in two.

Consider what might happen as a fracture plane spreads through a cell containing a variety of organelles of different composition. These structures tend to cause deviations in the fracture plane, either upward or downward, giving the fracture face elevations, depressions, and ridges that reflect the contours of the protoplasm traversed. Consequently, the surfaces exposed by the fracture contain information about the contents of the cell. The goal is to make this information visible. The *replication* process accomplishes this by using the fractured surface as a template on which a heavy-metal layer is deposited. The heavy metal is deposited onto the newly exposed surface of the frozen tissue in the same chamber where the tissue was fractured. The metal is deposited at an angle to provide shadows that accentuate local topography (Figure 18.17), as described in the previous section on shadow casting.

A carbon layer is then deposited on top of the metal layer to cement the patches of metal into a solid surface. Once this cast has been made, the tissue that provided the template can be thawed, removed, and discarded; it is the metal–carbon **replica** that is placed on the specimen grid and viewed in the electron microscope. Variations in thickness of the metal in different parts of the replica cause variations in the numbers of penetrating electrons to reach the viewing screen, producing the necessary contrast in the image. As discussed in Chapter 4, fracture planes take the path of least resistance through the frozen block, which often carries them through the center of cellular membranes. As a result, this technique is particularly well suited for examining the distribution of integral membrane proteins as they span the lipid bilayer (Figure 4.15). Such studies carried out by Daniel Branton and others played an important role in the formulation of the fluid mosaic structure of cellular membranes in the early 1970s (page 121).
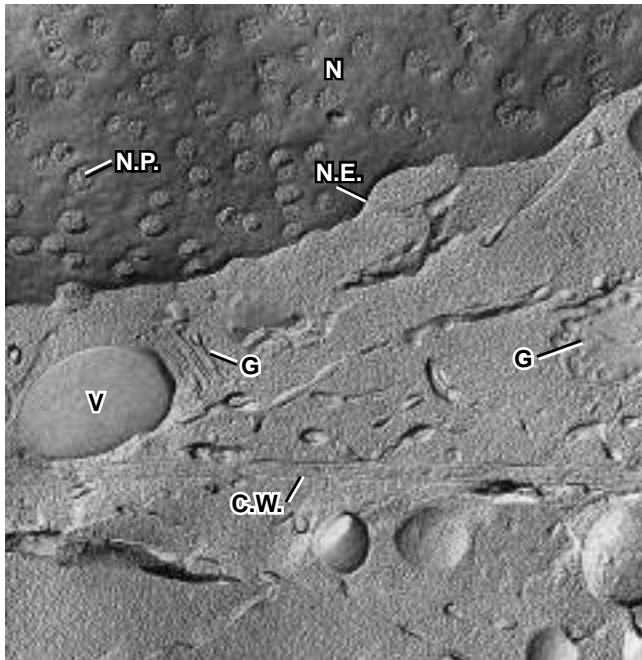
Freeze-fracture replication by itself is an extremely valuable technique, but it can be made even more informative by including a step called **freeze etching** (Figure 18.16). In this step, the frozen, fractured specimen, while still in place within the cold chamber, is exposed to a vacuum at an elevated temperature for one to a few minutes, during which a layer of ice can evaporate (sublime) from the exposed surface. Once some of the ice has been removed, the surface of the structure can be coated with heavy metal and carbon to create a metallic replica that reveals both the external surface and internal structure of cellular membranes. The development by John Heuser of *deep-etching* techniques, in which greater amounts of surface ice are removed, led to a fascinating look at cellular organelles. Examples of specimens prepared by this technique are shown in Figures 18.18, 8.38, and 9.45, where it can be seen that the individual parts of the cell stand out in deep relief against the background. The technique delivers very high resolution and can be used to reveal the structure and the distribution of macromolecular complexes, such as those of the cytoskeleton, as they are presumed to exist within the living cell.

FIGURE 18.17 **Replica of a freeze-fractured onion root cell** showing the nuclear envelope (N.E.) with its pores (N.P.), the Golgi complex (G), a cytoplasmic vacuole (V), and the cell wall (C.W.). (Courtesy of Daniel Branton.)
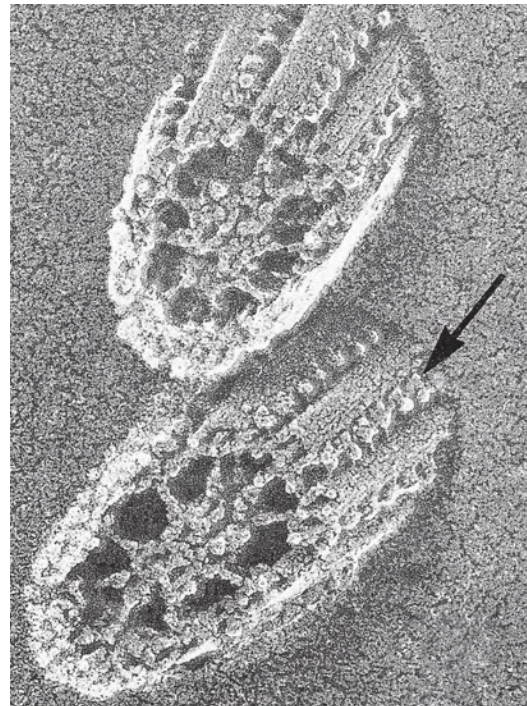
FIGURE 18.18 **Deep etching.** Electron micrograph of ciliary axonemes from the protozoan *Tetrahymena*. The axonemes were fixed, frozen, and fractured, and the frozen water at the surface of the fractured block was evaporated away, leaving a portion of the axonemes standing out in relief, as visualized in this metal replica. The arrow indicates a distinct row of outer dynein arms. (From Ursula W. Goodenough and John E. Heuser, J. Cell Biol. 95:800, 1982; by copyright permission of The Rockefeller University Press.)

18.3 SCANNING ELECTRON AND ATOMIC FORCE MICROSCOPY

The TEM has been exploited most widely in the examination of the internal structure of cells. In contrast, the scanning electron microscope (SEM) is utilized primarily to examine the surfaces of objects ranging in size from a virus to an animal head (Figure 18.19). The construction and operation of the SEM are very different from that of the TEM. The goal of specimen preparation for the SEM is to produce an object that has the same shape and surface properties as the living state, but is devoid of fluid, as required for observing the specimen under vacuum. Because water constitutes such a high percentage of the weight of living cells and is present in association with virtually every macromolecule, its removal can have a very destructive effect on cell structure. When cells are simply air dried, destruction results largely from surface tension at air–water interfaces. Specimens to be examined in the SEM are fixed, passed through a series of alcohols, and then dried by a process of *critical-point drying*. Critical-point drying takes advantage of the fact that a critical temperature and pressure exist for each solvent at which the density of the vapor is equal to the density of the liquid. At this point, there is no surface tension between the gas and the liquid. The solvent of the cells is replaced with a liquid transitional fluid (generally carbon dioxide), which is vaporized under pressure so that the cells are not exposed to any surface tension that might distort their three-dimensional configuration.

Once the specimen is dried, it is coated with a thin layer of metal, which makes it suitable as a target for an electron beam. In the TEM, the electron beam is focused by the condenser lenses to simultaneously illuminate the entire viewing field. In the SEM, electrons are accelerated as a fine beam that scans the specimen. In the TEM, electrons pass through the specimen to form the image. In the SEM, the image is formed by electrons that are reflected back from the specimen (backscattered) or by secondary electrons given off by the specimen after being struck by the primary electron beam. These electrons strike a detector that is located near the surface of the specimen.

Image formation in the SEM is indirect. In addition to the beam that scans the surface of the specimen, another electron beam synchronously scans the face of a cathode-ray tube, producing an image similar to that seen on a television screen. The electrons that bounce off the specimen and reach the detector control the strength of the beam in the cathode-ray tube. The more electrons collected from the specimen at a given spot, the stronger the signal to the tube and the greater the intensity of the beam on the screen at the corresponding spot. The result is an image on the screen that reflects the surface topology of the specimen because it is this topology (the crevices, hills, and pits) that determines the number of electrons collected from the various parts of the surface.

As evident in the micrographs of Figure 18.19, an SEM can provide a great range of magnification (from about 15 to 150,000 times for a standard instrument). Resolving power of

(a)

(b)

FIGURE 18.19 Scanning electron microscopy. Scanning electron micrographs of (*a*) a T4 bacteriophage (×275,000) and (*b*) the head of an insect (×40). (A: Reprinted with permission from A. N. Broers, B. J. Panessa, and J. F. Gennaro, Science 189:635, 1975; © copyright 1975, American Association for the Advancement of Science; b: courtesy of H. F. Howden and L. E. C. Ling.)

an SEM is related to the diameter of the electron beam. Newer models are capable of delivering resolutions of less than 5 nm, which can be used to localize gold-labeled antibodies bound to a cell's surface. The SEM also provides remarkable depth of focus, approximately 500 times that of the light microscope at a corresponding magnification. This property gives SEM images their three-dimensional quality. At the cellular level, the SEM allows the observer to appreciate the structure of the outer cell surface and all of the various processes, extensions, and extracellular materials that interact with the environment.

Atomic Force Microscopy

Although it is not an electron microscope, the **atomic force microscope (AFM)** is a high-resolution scanning instrument that is becoming increasingly important in nanotechnology and molecular biology. Several AFM-derived images can be found in the text: Figures 4.24*a*, 5.24, 7.32*c*, and the Chapter 12 opener. The AFM operates by scanning a sharp, nanosized tip (probe) over the surface of the specimen. In one type of AFM, the probe is attached to a tiny oscillating beam (or cantilever), whose frequency of oscillations changes as the tip encounters variations in the topography of the specimen. These changes in oscillation of the beam can be converted into a three-dimensional topographic image of the surface of the specimen. Unlike other techniques for the determination of molecular structure, such as X-ray crystallography and cryo-EM, which average the structure of many individual molecules, AFM provides an image of each individual molecule as it is oriented in the field (see Figure 5.24). The probe of an AFM can be used as more than a monitoring device; it can also be employed as a "nanomanipulator" to push or pull on the specimen in an attempt to measure various mechanical properties. This capability is illustrated in Figure 9.7 where it is shown that a single intermediate filament can be stretched to several times its normal length. In a different protocol, the AFM tip can be coated with ligands for a particular receptor, and measurements can be made of the affinity of that receptor for the ligand in question. Because of its many potential uses, the AFM has been referred to as "a lab on a tip."

18.4 THE USE OF RADIOISOTOPES

A tracer is a substance that reveals its presence in one way or another and thus can be localized or monitored during the course of an experiment. Depending on the substance and the type of experiment, a tracer might be fluorescently labeled, spin labeled, density labeled, or radioactively labeled. In each case, a labeled group enables a molecule to be detected without affecting the specificity of its interactions. Radioactive molecules, for example, participate in the same reactions as nonradioactive species, but their location can be followed and the amount present can be measured.

The identity of an atom (whether it be an iron atom, a chlorine atom, or some other type), and thus its chemical properties, is determined by the number of positively charged protons in its nucleus. All hydrogen atoms have a single proton, all helium atoms have two protons, all lithium atoms have three protons, and so forth. Not all hydrogen, helium, or lithium atoms, however, have the same number of neutrons. Atoms having the same number of protons and different numbers of neutrons are said to be *isotopes* of one another. Even hydrogen, the simplest element, can exist as three different isotopes, depending on whether the atom has 0, 1, or 2 neutrons in its nucleus. Of these three isotopes of hydrogen, only the one containing two neutrons is radioactive; it is tritium (^3H).

Isotopes are radioactive when they contain an unstable combination of protons and neutrons. Atoms that are unstable tend to break apart, or disintegrate, thus achieving a more stable configuration. When an atom disintegrates, it releases particles or electromagnetic radiation that can be monitored by appropriate instruments. Radioactive isotopes occur throughout the periodic table of elements, and they can be produced from nonradioactive elements in the laboratory. Many biological molecules can be purchased in a radioactive state, that is, containing one or more radioactive atoms as part of their structure.

The **half-life ($t_{1/2}$)** of a radioisotope is a measure of its instability. The more unstable a particular isotope, the greater the likelihood that a given atom will disintegrate in a given amount of time. If one starts with one Curie[3] of tritium, half that amount of radioactive material will be left after approximately 12 years (which is the half-life of this radioisotope). In the early years of research into photosynthesis and other metabolic pathways, the only available radioisotope of carbon was ^{11}C, which has a half-life of approximately 20 minutes. Experiments with ^{11}C were literally carried out on the run so that the amount of incorporated isotope could be measured before the substance disappeared. The availability in the 1950s of ^{14}C, a radioisotope having a half-life of 5700 years, was greeted with great celebration. The radioisotopes of greatest importance in cell biological research are listed in Table 18.1, along with information on their half-lives and the nature of their radiation.

Three main forms of radiation can be released by atoms during their disintegration. The atom may release an *alpha particle*, which consists of two protons and two neutrons and is equivalent to the nucleus of a helium atom; a *beta particle*, which is equivalent to an electron; and/or *gamma radiation*, which consists of electromagnetic radiation or photons. The most commonly used isotopes are beta emitters, which are monitored by either of two different methodologies: liquid scintillation spectrometry or autoradiography.

Liquid scintillation spectrometry is used to measure the amount of radioactivity in a given sample. The technique is based on the property of certain molecules, termed *phosphors* or *scintillants*, to absorb some of the energy of an emitted particle and release that energy in the form of light. In preparing a sample for liquid scintillation counting, one mixes the sam-

---

[3]A Curie is that amount of radioactivity required to yield 3.7×10^{10} disintegrations per second.

TABLE 18.1 Properties of a Variety of Radioisotopes Used in Biological Research

Symbol and atomic weight	Half-life	Type of particle(s) emitted
^{3}H	12.3 yr	Beta
^{11}C	20 min	Beta
^{14}C	5700 yr	Beta
^{24}Na	15.1 hr	Beta, Gamma
^{32}P	14.3 d	Beta
^{35}S	87.1 d	Beta
^{42}K	12.4 hr	Beta, Gamma
^{45}Ca	152 d	Beta
^{59}Fe	45 d	Beta, Gamma
^{60}Co	5.3 yr	Beta, Gamma
^{64}Cu	12.8 hr	Beta, Gamma
^{65}Zn	250 d	Beta, Gamma
^{131}I	8.0 d	Beta, Gamma
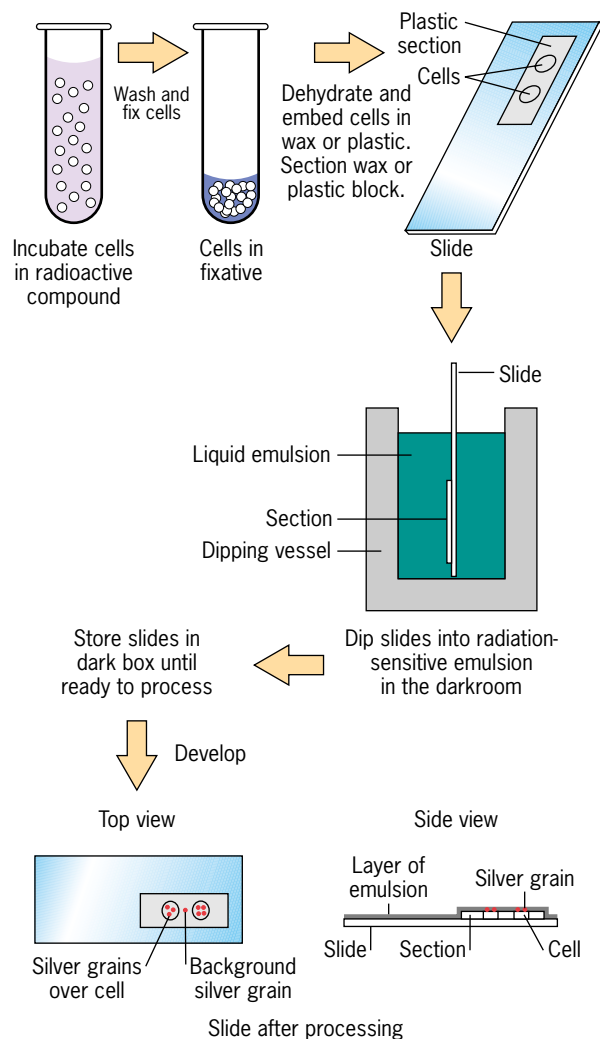
FIGURE 18.20 **Preparation of a light microscopic autoradiograph.**

ple with a solution of the phosphor in a glass or plastic scintillation vial. This brings the phosphor and radioactive isotope into very close contact so that radiation from even the weakest beta emitters can be efficiently measured. Once mixed, the vial is placed into the counting instrument, where it is lowered into a well whose walls contain an extremely sensitive photodetector. As radioactive atoms within the vial disintegrate, the emitted particles activate the scintillants, which emit flashes of light. The light is detected by a photocell, and the signal is amplified by a photomultiplier tube within the counter. After correcting for background noise, the amount of radioactivity present in each vial is displayed on a printout.

Autoradiography is a broad-based technique used to determine where a particular isotope is located, whether in a cell, in a polyacrylamide gel, or on a nitrocellulose filter. The importance of autoradiography in early discoveries on the synthetic activities of cells was described in the pulse-chase experiments on page 267. Autoradiography takes advantage of the ability of a particle emitted from a radioactive atom to activate a photographic emulsion, much like light or X-rays activate the emulsion that coats a piece of film. If the photographic emulsion is brought into close contact with a radioactive source, the particles emitted by the source leave tiny, black silver grains in the emulsion after photographic development. Autoradiography is used to localize radioisotopes within tissue sections that have been immobilized on a slide or TEM grid. The steps involved in the preparation of a light microscopic autoradiograph are shown in Figure 18.20. The emulsion is applied to the sections on the slide or grid as a very thin overlying layer, and the specimen is put into a lightproof container to allow the emulsion to be exposed by the emissions. The longer the specimen is left before development, the greater the number of silver grains that are formed. When the developed slide or grid is examined in the microscope, the location of the silver grains in the layer of emulsion just above the tissue indicates the location of the radioactivity in the underlying cells (Figure 18.21).

18.5 CELL CULTURE

Throughout this book we have stressed the approach in cell biology that attempts to understand particular processes by their analysis in a simplified, controlled, in vitro system. The same approach can be applied to the study of cells themselves because they too can be removed from the influences they are normally subject to within a complex multicellular organism. Learning to grow cells outside the organism, that is, in **cell culture**, has proved to be one of the most valuable technical achievements in the entire study of biology. A quick glance through any journal in cell biology reveals that the majority of the articles describe research carried out on cultured cells. The reasons for this are numerous: cultured cells can be obtained in large quantity; most cultures contain only a single type of cell; many different cellular activities, including endocytosis, cell movement, cell division, membrane trafficking, and macromolecular synthesis, can be studied in cell culture; cells can

(a)　　　　　　　　　　　　　　(b)
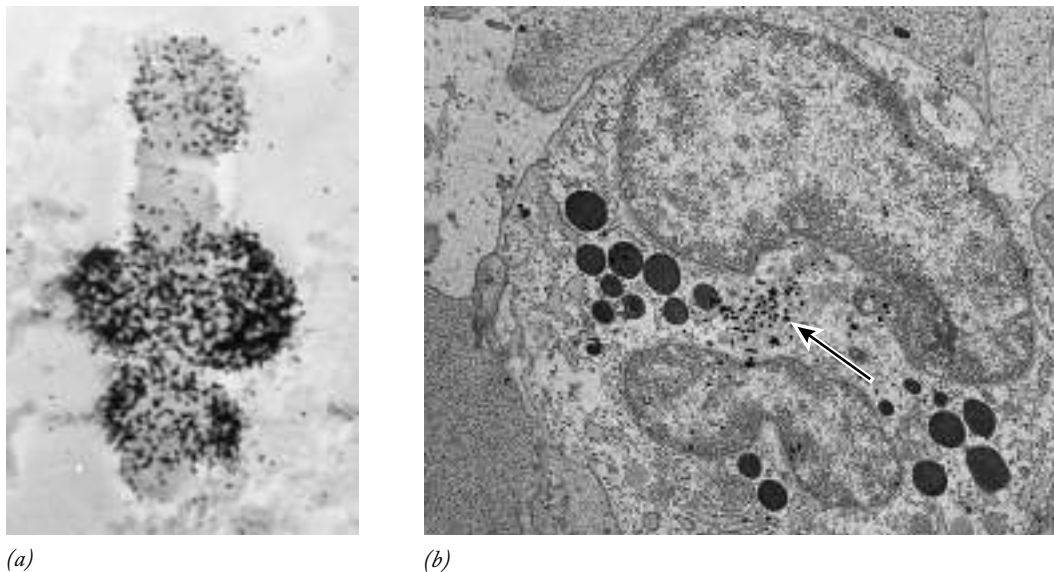
FIGURE 18.21 Examples of a light microscopic and electron microscopic autoradiograph. (*a*) Light microscopic autoradiograph of a polytene chromosome from the insect *Chironomus*, showing extensive incorporation of [³H]uridine into RNA in the puffed regions of the chromosome. Autoradiographs of this type confirmed that the chromosome puffs are sites of transcription. This micrograph can be compared with that of Figure 10.8*b*, which shows the same chromosome as seen in the scanning electron microscope. (*b*) Electron microscopic autoradiograph of a bone marrow cell incubated in [³⁵S]O₄ for 5 minutes and immediately fixed. Sulfate incorporation—as revealed by the small black silver grains—is localized in the Golgi complex (arrow). (A: FROM C. PELLING, CHROMOSOMA 15:98, 1964; B: FROM R. W. YOUNG, J. CELL BIOL. 57:177, 1973; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

differentiate in culture; and cultured cells respond to treatment with drugs, hormones, growth factors, and other active substances.

The early tissue culturists employed media containing a great variety of unknown substances. Cell growth was accomplished by adding fluids obtained from living systems, such as lymph, blood serum, or embryo homogenates. It was found that cells required a considerable variety of nutrients, hormones, growth factors, and cofactors to remain healthy and proliferate. Even today, most culture media contain large amounts of serum. The importance of serum (or the growth factors it contains) on the proliferation of cultured cells is shown in the cell-growth curves in Figure 16.4.

One of the primary goals of cell culturists has been to develop defined, serum-free media that supports the growth of cells. Using a pragmatic approach in which combinations of various ingredients are tested for their ability to support cell growth and proliferation, a growing number of cell types have been successfully cultured in "artificial" media that lack serum or other natural fluids. As would be expected, the composition of these chemically defined media is relatively complex; they consist of a mixture of nutrients and vitamins, together with a variety of purified proteins, including insulin, epidermal growth factor, and transferrin (which provides the cells with iron).

Because they are so rich in nutrients, tissue culture media are a very inviting habitat for the growth of microorganisms. To prevent bacteria from contaminating cell cultures, tissue culturists must go to great lengths to maintain sterile conditions within their working space. This is accomplished by using sterile gloves, sterilizing all supplies and instruments, employing low levels of antibiotics in the media, and conducting activities within a sterile hood.

The first step in cell culture is to obtain the cells. In most cases, one need only remove a vial of frozen, previously cultured cells from a tank of liquid nitrogen, thaw the vial, and transfer the cells to the waiting medium. A culture of this type is referred to as a **secondary culture** because the cells are derived from a previous culture. In a **primary culture**, on the other hand, the cells are obtained directly from the organism. Most primary cultures of animal cells are obtained from embryos, whose tissues are more readily dissociated into single cells than those of adults. Dissociation is accomplished with the aid of a proteolytic enzyme, such as trypsin, which digests the extracellular domains of proteins that mediate cell adhesion (Chapter 7). The tissue is then washed free of the enzyme and usually suspended in a saline solution that lacks Ca²⁺ ions and contains a substance, such as ethylenediamine tetraacetate (EDTA), that binds (chelates) calcium ions. As discussed in Chapter 7, calcium ions play a key role in cell–cell adhesion, and their removal from tissues greatly facilitates the separation of cells.

Normal (nonmalignant) cells can divide a limited number of times (typically 50 to 100) before they undergo senescence and death (page 495). Because of this, many of the cells that are commonly used in tissue culture studies have undergone genetic modifications that allow them to be grown indefinitely. Cells of this type are referred to as a **cell line**, and they typically grow into malignant tumors when injected into susceptible laboratory animals. The frequency with which a nor-

(a)

(b)
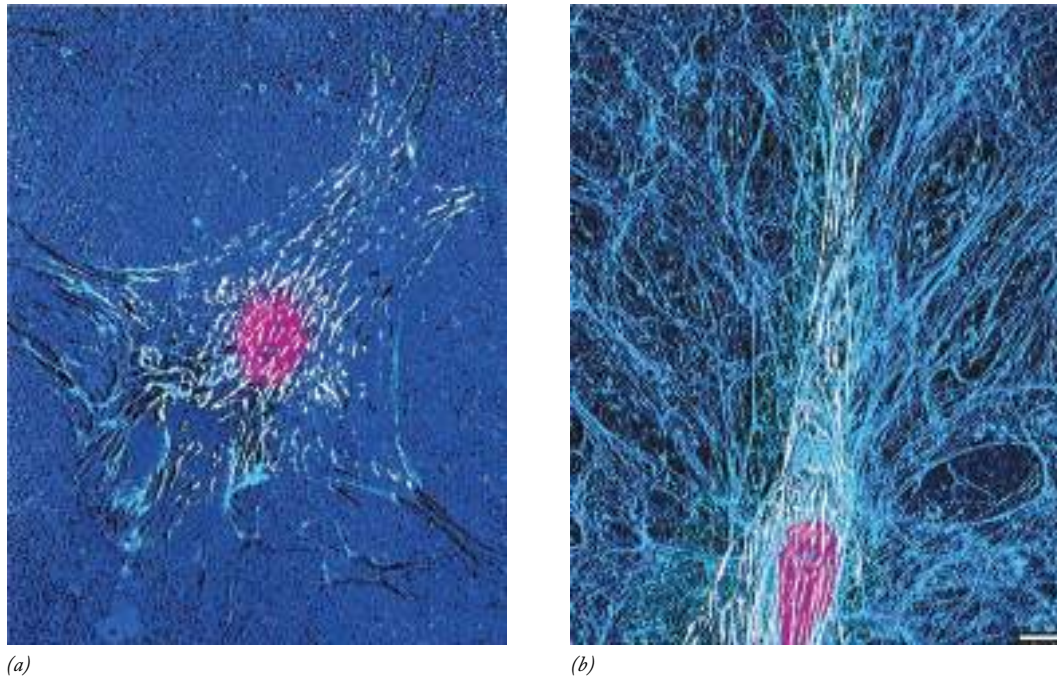
FIGURE 18.22 A comparison of the morphology of cells growing in 2D versus 3D cultures, (*a*) A human fibroblast growing on a flat, fibronectin-coated substratum in 2D culture assumes a highly flattened shape with broad lamellipodia. Integrin-containing adhesions appear white. (*b*) The same type of cell growing in a 3D matrix assumes a non-flattened, spindle-shaped morphology. The extracellular fibronectin matrix appears in blue. Bar equals 10 μm. (FROM KENNETH M. YAMADA AND EDNA CUKIERMAN, CELL 130:603, 2007, BY PERMISSION OF CELL PRESS.)

mal cell growing in culture spontaneously transforms into a cell line depends on the organism from which it was derived. Mouse cells, for example, transform at a relatively high frequency; human cells transform only rarely, if ever. Human cell lines (e.g., HeLa cells) are typically derived from human tumors or from cells treated with cancer-causing viruses or chemicals.

A number of laboratories are moving away from traditional *two-dimensional* culture systems, where cells are grown on the flat surface of a culture dish, to *three-dimensional* cultures, in which cells are grown in a three-dimensional matrix consisting of synthetic and/or natural extracellular materials. These materials can be purchased as products containing proteins and other components derived from natural basement membranes (Figures 7.8 and 7.12). Cells within the body do not live on flat, hard-plastic surfaces, and, consequently, it is thought that three-dimensional matrices provide a much more natural environment for cultured cells. Consequently, the morphology and behavior of cells in these three-dimensional environments are more like those of cells observed within the tissues of the body. This is reflected in differences in the cytoskeletal organization, types of cell adhesions, signaling activities, and states of differentiation of cells in the two types of culture systems. In addition, 3D culture systems are also better suited to study cell–cell interactions because they allow cells to come into contact with one another at any point along their surface. Figure 18.22 shows images of human fibroblasts grown on either a flat surface or within a three-dimensional matrix. The cell in Figure 18.22*a* has pressed itself against the substratum, creating a highly unnatural upper (dorsal) and lower (ventral) surface with unusually broad and flattened lamellipodia. In contrast, the same type of cell grown in 3D culture (Figure 18.22*b*) has a typical spindle-shaped structure without any obvious surface differentiation.

Many different types of plant cells can also grow in culture. In one approach, plant cells are treated with the enzyme cellulase, which digests away the surrounding cell wall, releasing the naked cell, or **protoplast**. Protoplasts can then be grown in a chemically defined medium that promotes their growth and division. Under suitable conditions, the cells can grow into an undifferentiated clump of cells, called a *callus*, which can be induced to develop shoots from which a plant can regenerate. In an alternate approach, the cells in leaf tissue can be induced by hormone treatment to lose their differentiated properties and develop into callus material. The callus can then be transferred to liquid media to start a cell culture.

18.6 THE FRACTIONATION OF A CELL'S CONTENTS BY DIFFERENTIAL CENTRIFUGATION

Most cells contain a variety of different organelles. If one sets out to study a particular function of mitochondria or to isolate a particular enzyme from the Golgi complex, it is useful to first isolate the relevant organelle in a purified state. Isolation of a particular organelle in bulk quantity is generally accom-

plished by the technique of **differential centrifugation**, which depends on the principle that, as long as they are more dense than the surrounding medium, particles of different size and shape travel toward the bottom of a centrifuge tube at different rates when placed in a centrifugal field.

To carry out this technique, cells are first broken open by mechanical disruption using a mechanical *homogenizer*. Cells are homogenized in an isotonic buffered solution (often containing sucrose), which prevents the rupture of membrane vesicles due to osmosis. The homogenate is then subjected to a series of sequential centrifugations at increasing centrifugal forces. The steps in this technique were discussed in Chapter 8 and illustrated in Figure 8.5. Initially, the homogenate is subjected to low centrifugal forces for a short period of time so that only the largest cellular organelles, such as nuclei (and any remaining whole cells), are sedimented into a pellet. At greater centrifugal forces, relatively large cytoplasmic organelles (mitochondria, chloroplasts, lysosomes, and peroxisomes) can be spun out of suspension (Figure 8.5). In subsequent steps, microsomes (the fragments of vacuolar and reticular membranes of the cytosol) and ribosomes are removed from suspension. This last step requires the ultracentrifuge, which can generate speeds of 75,000 revolutions per minute, producing forces equivalent to 500,000 times that of gravity. Once ribosomes have been removed, the supernatant consists of the cell's soluble phase and those particles too small to be removed by sedimentation.

The initial steps of differential centrifugation do not yield pure preparations of a particular organelle, so that further steps are usually required. In many cases, further purification is accomplished by centrifugation of one of the fractions through a density gradient, as shown in Figure 18.23, which distributes the contents of the sample into various layers according to their density. The composition of various fractions can be determined by microscopic examination, or by measuring the amounts of particular proteins known to be specific for particular organelles.

Cellular organelles isolated by differential centrifugation retain a remarkably high level of normal activity, as long as they are not exposed to denaturing conditions during their isolation. Organelles isolated by this procedure can be used in *cell-free systems* to study a wide variety of cellular activities, including the synthesis of membrane-bound proteins (page 270), the formation of transport vesicles (page 287), and the transport of solutes and development of ionic gradients (see Figure 5.25*a*).

18.7 ISOLATION, PURIFICATION, AND FRACTIONATION OF PROTEINS

During the course of this book, we have considered the properties of many different proteins. Before information about the structure or function of a particular protein can be obtained, that protein must be isolated in a relatively pure state. Because most cells contain thousands of different pro-
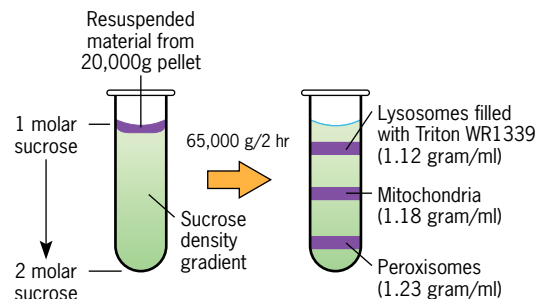
FIGURE 18.23 Purification of subcellular fractions by density-gradient equilibrium centrifugation. In this particular example, the medium is composed of a continuous sucrose-density gradient, and the different organelles sediment until they reach a place in the tube equal to their own density, where they form bands. The 20,000g pellet is obtained as shown in Figure 8.5.

teins, the purification of a single species can be a challenging mission, particularly if the protein is present in the cell at low concentration. In this section, we will briefly survey a few of the techniques used to purify proteins.

Purification of a protein is generally performed by the stepwise removal of contaminants. Two proteins may be very similar in one property, such as overall charge, and very different in another property, such as molecular size or shape. Consequently, the complete purification of a given protein usually requires the use of successive techniques that take advantage of different properties of the proteins being separated. Purification is measured as an increase in **specific activity**, which is the ratio of the amount of that protein to the total amount of protein present in the sample. Some identifiable feature of the specific protein must be utilized as an **assay** to determine the relative amount of that protein in the sample. If the protein is an enzyme, its catalytic activity may be used as an assay to monitor purification. Alternatively, assays can be based on immunologic, electrophoretic, electron microscopic, or other criteria. Measurements of total protein in a sample can be made using various properties, including total nitrogen, which can be very accurately measured and is quite constant at about 16 percent of the dry weight for all proteins.

Selective Precipitation

The first step in purification should be one that can be carried out on a highly impure preparation and yield a large increase in specific activity. The first step usually takes advantage of solubility differences among proteins by selectively precipitating the desired protein. The solubility properties of a protein are determined largely by the distribution of hydrophilic and hydrophobic side chains on its surface. A protein's solubility in a given solution depends on the relative balance between protein–solvent interactions, which keep it in solution, and protein–protein interactions, which cause it to aggregate and precipitate from solution. The salt most commonly employed for selective protein precipitation is ammonium sulfate, which is extremely soluble in water and has high ionic strength. Pu-

rification is achieved by gradually adding a solution of saturated ammonium sulfate to the crude protein extract. As addition of salt continues, precipitation of contaminating proteins increases, and the precipitate can be discarded. Ultimately, a point is reached at which the protein being sought comes out of solution. This point is recognized by the loss of activity in the soluble fraction when tested by the particular assay being used. Once the desired protein is precipitated, contaminating proteins are left behind in solution, while the protein being sought can be redissolved.

Liquid Column Chromatography

Chromatography is a term for a variety of techniques in which a mixture of dissolved components is fractionated as it moves through some type of porous matrix. In liquid chromatographic techniques, components in a mixture can become associated with one of two alternative phases: a mobile phase, consisting of a moving solvent, and an immobile phase, consisting of the matrix through which the solvent is moving.[4] In the chromatographic procedures described below, the immobile phase consists of materials that are packed into a column. The proteins to be fractionated are dissolved in a solvent and then passed through the column. The materials that make up the immobile phase contain sites to which the proteins in solution can bind. As individual protein molecules interact with the materials of the matrix, their progress through the column is retarded. Thus the greater the affinity of a particular protein for the matrix material, the slower its passage through the column. Because different proteins in the mixture have different affinity for the matrix, they are retarded to different degrees. As the solvent passes through the column and drips out the bottom, it is collected as *fractions* in a series of tubes. Those proteins in the mixture with the least affinity for the column appear in the first fractions to emerge from the column. The resolution of many chromatographic procedures has been improved in recent years with the development of *high performance liquid chromatography* (*HPLC*), in which long, narrow columns are used, and the mobile phase is forced through a tightly packed noncompressible matrix under high pressure.

Ion-Exchange Chromatography

Proteins are large, polyvalent electrolytes, and it is unlikely that many proteins in a partially purified preparation have the same overall charge. The overall charge of a protein is a summation of all the individual charges of its component amino acids. Because the charge of each amino acid depends on the pH of the medium (see Figure 2.27), the charge of each protein also depends on the pH. As the pH is lowered, negatively charged groups become neutralized and positively charged groups become more numerous. The opposite occurs as the pH is increased. A pH exists for each protein at which the total number of negative charges equals the total number of positive charges. This pH
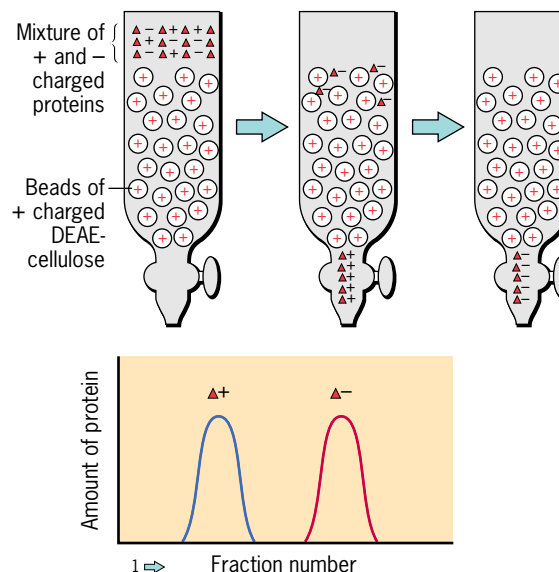
FIGURE 18.24 Ion-exchange chromatography. The separation of two proteins by DEAE-cellulose. In this case, a positively charged ion-exchange resin is used to bind the negatively charged protein.

is the **isoelectric point**, at which the protein is neutral. The isoelectric point of most proteins is below pH 7.

Ionic charge is used as a basis for purification in a variety of techniques, including **ion-exchange chromatography**. Ion-exchange chromatography depends on the ionic bonding of proteins to an inert matrix material, such as cellulose, containing covalently linked charged groups. Two of the most commonly employed ion-exchange resins are diethylaminoethyl (DEAE) cellulose and carboxymethyl (CM) cellulose. DEAE-cellulose is positively charged and therefore binds negatively charged molecules; it is an *anion exchanger*. CM-cellulose is negatively charged and acts as a *cation exchanger*. The resin is packed into a column, and the protein solution is allowed to percolate through the column in a buffer whose composition promotes the binding of some or all of the proteins to the resin. Proteins are bound to the resin reversibly and can be displaced by increasing the ionic strength of the buffer (which adds small ions to compete with the charged groups of the macromolecules for sites on the resin) and/or changing its pH. Proteins are eluted from the column in order from the least strongly bound to the most strongly bound. Figure 18.24 shows a schematic representation of the separation of two protein species by stepwise elution from an ion-exchange column.

Gel Filtration Chromatography

Gel filtration separates proteins (or nucleic acids) primarily on the basis of their effective size (*hydrodynamic radius*). Like ion-exchange chromatography, the separation material consists of tiny beads that are packed into a column through which the protein solution slowly passes. The materials used in gel filtration are composed of cross-linked polysaccharides (dextrans or agarose) of different porosity, which allow proteins to diffuse in and out of

---

[4]Liquid chromatography is distinguished from gas chromatography in which the mobile phase is represented by an inert gas.
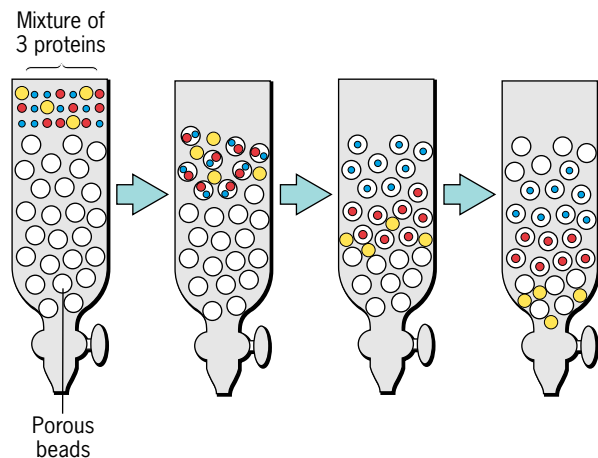
FIGURE 18.25 Gel filtration chromatography. The separation of three globular proteins having different molecular mass, as described in the text. Among proteins of similar basic shape, larger molecules are eluted before smaller molecules.

FIGURE 18.26 Affinity chromatography. (*a*) Schematic representation of the coated agarose beads to which only a specific protein can combine. (*b*) Steps in the chromatographic procedure.

the beads. The technique is best illustrated by example (Figure 18.25).

Suppose one is attempting to purify a globular protein having a molecular mass of 125,000 daltons. This protein is present in solution with two contaminating proteins of similar shape, one much larger at 250,000 daltons and the other much smaller at 75,000 daltons. One way the protein might be purified is to pass the mixture through a column of Sephadex G-150 beads, which allows entry to globular proteins that are less than about 200 kDa. When the protein mixture passes through the column bed, the 250 kDa protein is unable to enter the beads and remains dissolved in the moving solvent phase. As a result, the 250 kDa protein is eluted as soon as the preexisting solvent in the column (the bed volume) has dripped out. In contrast, the other two proteins can diffuse into the interstices within the beads and are retarded in their passage through the column. As more and more solvent moves through the column, these proteins move down its length and out the bottom, but they do so at different rates. Among those proteins that enter the beads, smaller species are retarded to a greater extent than larger ones. Consequently, the 125-kDa protein is eluted in a purified state, while the 75-kDa protein remains in the column.

Affinity Chromatography   Techniques described up to this point utilize the bulk properties of a protein to effect purification or fractionation. Another purification technique called **affinity chromatography** takes advantage of the unique structural properties of a protein, allowing one protein species to be specifically withdrawn from solution while all others remain behind in solution (Figure 18.26). Proteins interact with specific substances: enzymes with substrates, receptors with ligands, antigens with antibodies, and so forth. Each of these types of proteins can be removed from solution by passing a mixture of proteins through a column in which the specific interacting molecule (substrate, ligand, antibody, etc.) is covalently linked to an inert, immobilized material (the matrix). If,
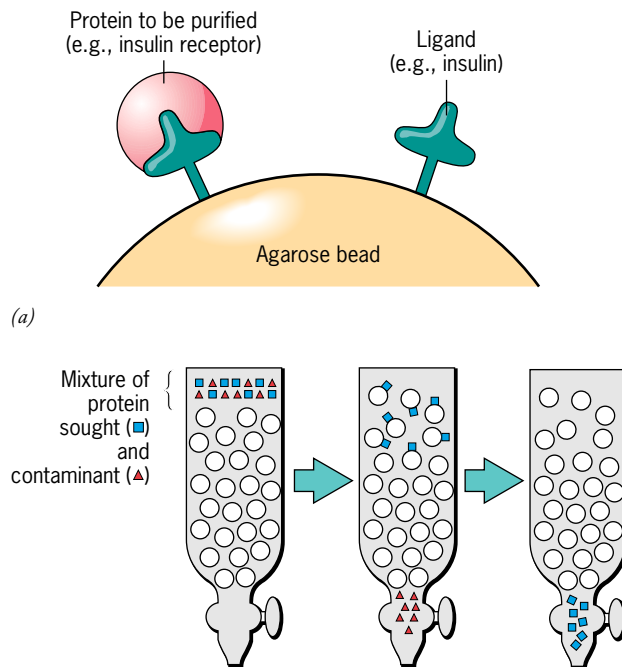
for example, an impure preparation of an acetylcholine receptor is passed through a column containing agarose beads to which an acetylcholine analogue is attached, the receptor binds specifically to the beads as long as the conditions in the column are suitable to promote the interaction (page 167). Once all the contaminating proteins have passed through the column and out the bottom, the acetylcholine receptor molecules can be displaced from their binding sites on the matrix by changing the ionic composition and/or pH of the solvent in the column. Thus, unlike the other chromatographic procedures that separate proteins on the basis of size or charge, affinity chromatography can achieve a near-total purification of the desired molecule in a single step.

Determining Protein–Protein Interactions   One of the ways to learn about the function of a protein is to identify the proteins with which it interacts. Several techniques are available for determining which proteins in a cell might be capable of interacting with a given protein that has already been identified. One of these techniques was just discussed: affinity chromatography. Another technique uses antibodies. Consider, for example, that protein A, which has already been identified and purified, is part of a complex with two other proteins in the cytoplasm, proteins B and C. Once protein A has been purified, an antibody against this protein can be obtained and used as a probe to bind and remove protein A from solution. If a cell extract is prepared that contains the A–B–C protein complex, and the extract is incubated with the anti-A antibody, then binding of the antibody to the A protein will result in the *coprecipitation* of other proteins that are bound to A, in this case

proteins B and C, which can then be identified. Coprecipitation of DNA fragments was discussed on page 513.

The technique most widely used to search for protein–protein interactions is the **yeast two-hybrid system**, which was invented in 1989 by Stanley Fields and Ok-kyu Song at the State University of New York in Stony Brook. The technique is illustrated in Figure 18.27 and depends on the expression of a reporter gene, such as β-galactosidase (*lacZ*), whose activity is readily monitored by a test that detects a color change when the enzyme is present in a population of yeast cells. Expression of the *lacZ* gene in this system is activated by a particular protein—a transcription factor—that contains two domains, a DNA-binding domain and an activation domain (Figure 18.27a). The DNA-binding domain mediates binding to the promoter, and the activation domain mediates interaction with other proteins involved in the activation of gene expression. Both domains must be present for transcription to occur. To employ the technique, two different types of recombinant DNA molecules are prepared. One DNA molecule contains a segment of DNA encoding the DNA-binding domain of the transcription factor linked to a segment of DNA encoding the "bait" protein (X). The bait protein is the protein that has been characterized and the one for which potential binding partners are being sought. When this recombinant DNA is expressed in a yeast cell, a hybrid protein such as that depicted in Figure 18.27b is produced in the cell. The other DNA molecule contains a portion of the transcription factor encoding the activation domain linked to DNA encoding an unknown protein (Y). Such DNAs (or cDNAs as they are called) are prepared from mRNAs by reverse transcriptase as described on page 756. Let's assume that Y is a protein capable of binding to the bait protein. When a recombinant DNA encoding Y is expressed in a yeast cell, a hybrid protein such as that depicted in Figure 18.27c is produced in the cell. When produced in a cell alone, neither the X- nor Y-containing hybrid protein is capable of activating transcription of the *lacZ* gene (Figure 18.27b,c). However, if both of these particular recombinant DNA molecules are introduced into the same yeast cell (as in Figure 18.27d), the X and Y proteins can interact with one another to reconstitute a functional transcription factor, an event that can be detected by the cell's ability to produce β-galactosidase. Using this technique, researchers are able to "fish" for proteins encoded by unknown genes that are capable of interacting with the "bait" protein. The use of this technique in proteomic studies is discussed on page 61.

Polyacrylamide Gel Electrophoresis

Another powerful technique that is widely used to fractionate proteins is **electrophoresis**. Electrophoresis depends on the ability of charged molecules to migrate when placed in an electric field. The electrophoretic separation of proteins is usually accomplished using **polyacrylamide gel electrophoresis (PAGE)**, in which the proteins are driven by an applied current through a gelated matrix. The matrix is composed of polymers of a small organic molecule (acrylamide) that is cross-linked to form a molecular sieve. A polyacrylamide gel may be formed as a thin slab between two glass plates or as a
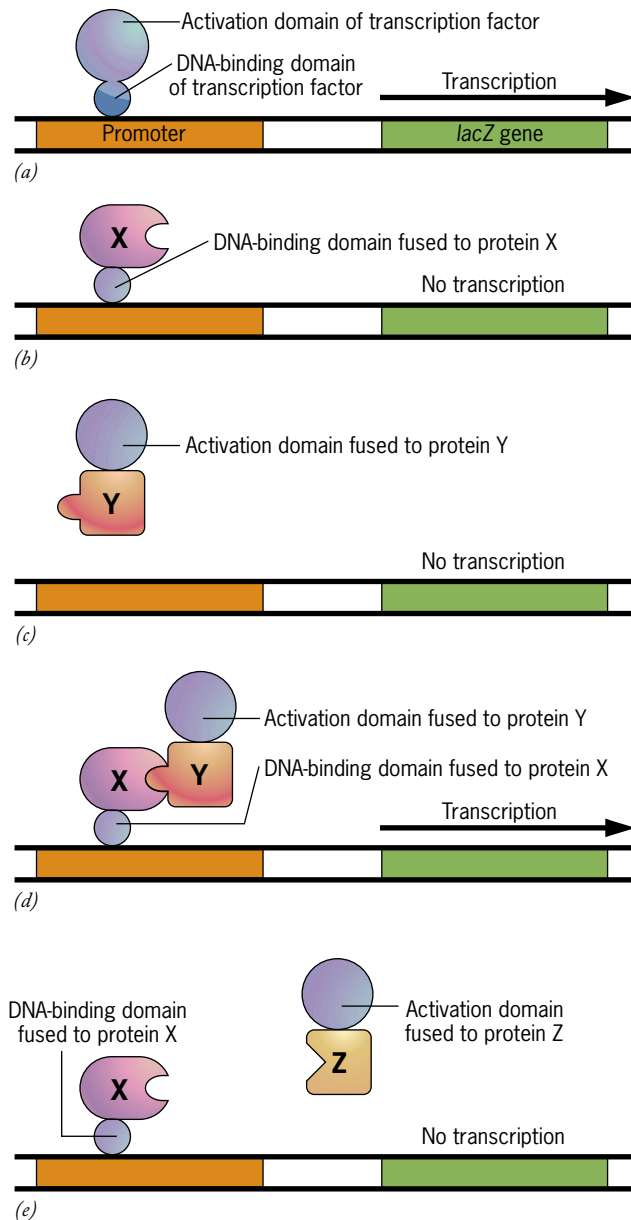
FIGURE 18.27 Use of the yeast two-hybrid system. This test for protein–protein interaction depends on a cell being able to put together two parts of a transcription factor. (*a*) The two parts of the transcription factor—the DNA-binding domain and the activation domain—are seen here as the transcription factor binds to the promoter of a gene (*lacZ*) encoding β-galactosidase. (*b*) In this case, a yeast cell has synthesized the DNA-binding domain of the transcription factor linked to a known "bait" protein X. This complex cannot activate transcription. (*c*) In this case, a yeast cell has synthesized the activation domain of the transcription factor linked to an unknown "fish" protein Y. This complex cannot activate transcription. (*d*) In this case, a yeast cell has synthesized both X and Y protein constructs, which reconstitutes a complete transcription factor, allowing *lacZ* expression, which is readily detected. (*e*) If the second DNA had encoded a protein, for example, Z, that could not bind to X, no expression of the reporter gene would have been detected.

cylinder within a glass tube. Once the gel has polymerized, the slab (or tube) is suspended between two compartments containing buffer in which opposing electrodes are immersed. In

a slab gel, the concentrated, protein-containing sample is layered in slots along the top of the gel, as shown in step 1 of Figure 18.28. The protein sample is prepared in a solution containing sucrose or glycerol, whose density prevents the sample from mixing with the buffer in the upper compartment. A voltage is then applied between the buffer compartments, and current flows across the slab, causing the proteins to move toward the oppositely charged electrode (step 2). Separations are typically carried out using alkaline buffers, which make the proteins negatively charged and cause them to migrate toward the positively charged anode at the opposite end of the gel. Following electrophoresis, the slab is removed from the glass plates and stained (step 3).

The relative movement of proteins through a polyacrylamide gel depends on the *charge density* (charge per unit of mass) of the molecules. The greater the charge density, the more forcefully the protein is driven through the gel, and thus the more rapid its rate of migration. But charge density is only one important factor in PAGE fractionation; size and shape also play a role. Polyacrylamide forms a cross-linked molecular sieve that entangles proteins passing through the gel. The larger the protein, the more it becomes entangled, and the slower it migrates. Shape is also a factor because compact globular proteins move more rapidly than elongated fibrous proteins of comparable molecular mass. The concentration of acrylamide (and cross-linking agent) used in making the gel is also an important factor. The lower the concentration of acrylamide, the less the gel becomes cross-linked, and the more rapidly a given protein molecule migrates. A gel containing 5 percent acrylamide might be useful for separating proteins of 60 to 250 kDa, whereas a gel of 15 percent acrylamide might be useful for separating proteins of 10 to 50 kDa.

The progress of electrophoresis is followed by watching the migration of a charged *tracking dye* that moves just ahead of the fastest proteins (step 2, Figure 18.28). After the tracking dye has moved to the desired location, the current is turned off, and the gel is removed from its container. Gels are typically stained with Coomassie Blue or silver stain to reveal the location of the proteins. If the proteins are radioactively labeled, their location can be determined by pressing the gel against a piece of X-ray film to produce an autoradiograph, or the gel can be sliced into fractions and individual proteins isolated. Alternatively, the proteins in the gel can be transferred by a second electrophoretic procedure to a nitrocellulose membrane to form a blot (page 745). Proteins become absorbed onto the surface of the membrane in the same relative positions they occupied in the gel. In a *Western blot*, the proteins on the membrane are identified by their interaction with specific antibodies.

SDS–PAGE　Polyacrylamide gel electrophoresis (PAGE) is usually carried out in the presence of the negatively charged detergent sodium dodecyl sulfate (SDS), which binds in large numbers to all types of protein molecules (page 128). The electrostatic repulsion between the bound SDS molecules causes the proteins to unfold into a similar rod-like shape, thus eliminating differences in shape as a factor in separation. The number of SDS molecules that bind to a protein is
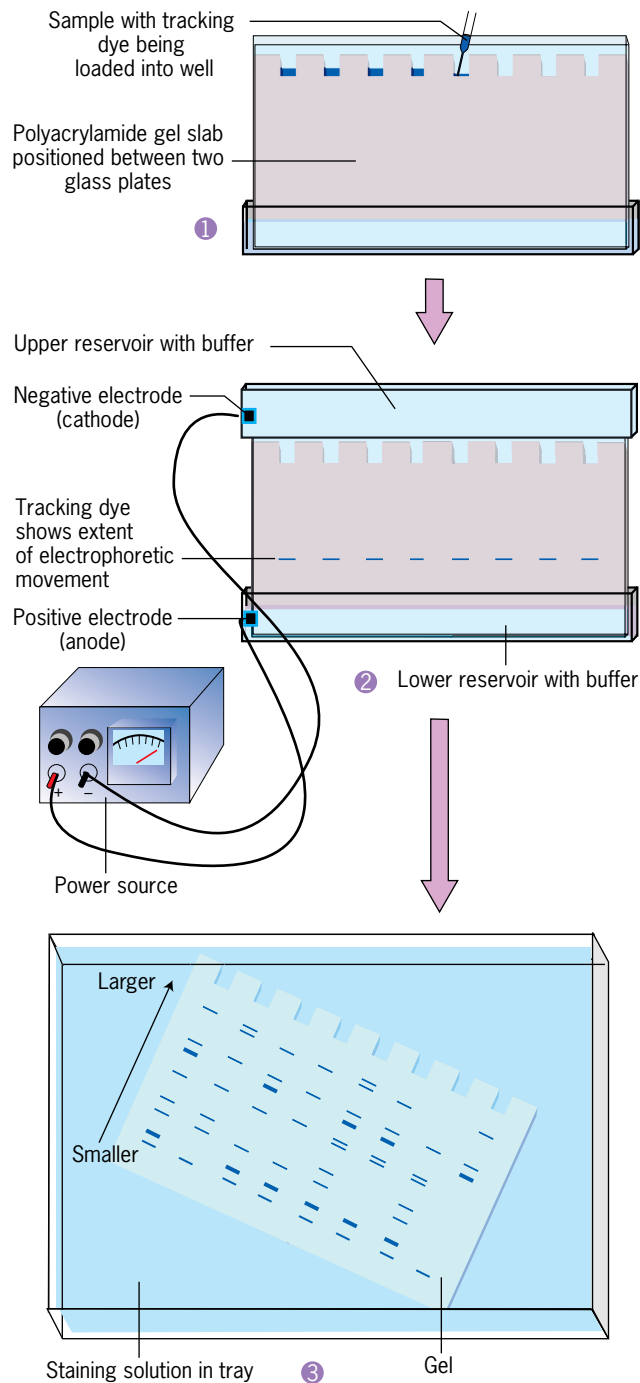
FIGURE 18.28　Polyacrylamide gel electrophoresis. The protein samples are typically dissolved in a sucrose solution whose density prevents the sample from mixing with the buffer and then loaded into the wells with a fine pipette as shown in step 1. In step 2, a direct current is applied across the gel, which causes the proteins to move into the polyacrylamide along parallel lanes. When carried out in the detergent SDS, which is usually the case, the proteins move as bands at rates that are inversely proportional to their molecular mass. Once electrophoresis is completed, the gel is removed from the glass frame and stained in a tray (step 3).

roughly proportional to the protein's molecular mass (about 1.4 g SDS/g protein). Consequently, each protein species, regardless of its size, has an equivalent charge density and is

driven through the gel with the same force. However, because the polyacrylamide is highly cross-linked, larger proteins are held up to a greater degree than smaller proteins. As a result, proteins become separated by SDS–PAGE on the basis of a single property—their molecular mass. In addition to separating the proteins in a mixture, SDS–PAGE can be used to determine the molecular mass of the various proteins by comparing the positions of the bands to those produced by proteins of known size. Examples of SDS–PAGE are shown and on pages 141 and 168.

Two-Dimensional Gel Electrophoresis   In 1975, a technique called *two-dimensional gel electrophoresis* was developed by Patrick O'Farrell at the University of California, San Francisco, to fractionate complex mixtures of proteins using two different properties of the molecules. Proteins are first separated in a tubular gel according to their isoelectric point (page 735) by a technique called *isoelectric focusing*. After separation, the gel is removed and placed on top of a slab of SDS-saturated polyacrylamide and subjected to SDS–PAGE. The proteins move into the slab gel and become separated according to their molecular mass (Figure 18.29). Once separated, individual proteins can be removed from the gel and digested into peptide fragments that can be analyzed by mass spectrometry. The resolution of the technique is sufficiently high to distinguish most of the proteins in a cell. Because of its great resolving power, two-dimensional gel electrophoresis is ideally suited to detect changes in the proteins present in a cell under different conditions, at different stages in development or the cell cycle, or in different organisms (see Figure 2.48). The technique, however, is
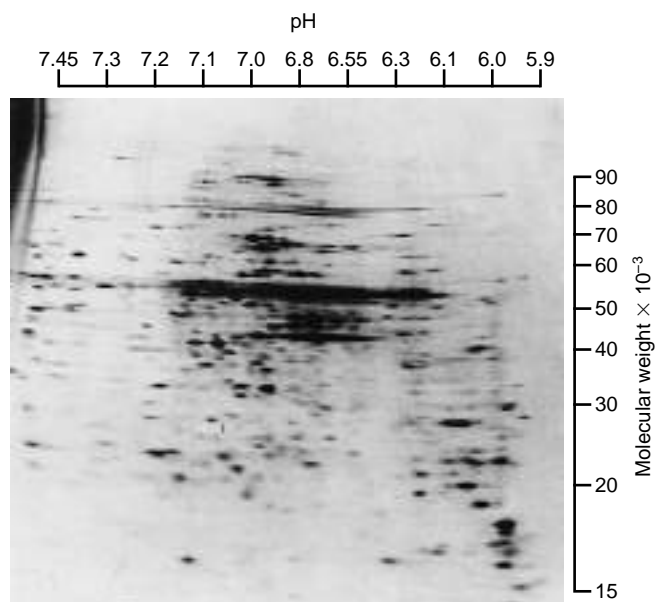
not suitable for distinguishing among proteins that have high molecular mass, that are highly hydrophobic, or that are present at very low copy numbers per cell.

Protein Measurement and Analysis

One of the simplest and most widely used methods to determine the amount of protein (or nucleic acid) present in a given solution is to measure the amount of light of a specific wavelength that is absorbed by that solution. The instrument used for this measurement is a **spectrophotometer**. To make this type of measurement, the solution is placed in a special, flat-sided, quartz container (quartz is used because, unlike glass, it does not absorb ultraviolet light), termed a *cuvette*, which is then placed in the light beam of the spectrophotometer. The amount of light that passes through the solution unabsorbed (i.e., the transmitted light) is measured by photocells on the other side of the cuvette.

Of the 20 amino acids incorporated into proteins, two of them, tyrosine and phenylalanine, absorb light in the ultraviolet range with an absorbance maximum at about 280 nm. If the proteins being studied have a typical percentage of these amino acids, then the absorbance of the solution at this wavelength provides a measure of protein concentration. Alternatively, one can use a variety of chemical assays, such as the Lowry or Biuret technique, in which the protein in solution is engaged in a reaction that produces a colored product whose concentration is proportional to the concentration of protein.

Mass Spectrometry   As discussed on page 69, the emerging field of proteomics depends heavily on the analysis of proteins by *mass spectrometry*. Mass spectrometers are analytical instruments used primarily to measure the masses of molecules, determine chemical formulas and molecular structure, and identify unknown substances. Mass spectrometers accomplish these feats by converting the substances in a sample into positively charged, gaseous ions, which are accelerated through a curved tube toward a negatively charged plate (Figure 18.30). As the ions pass through the tube, they are subjected to a magnetic field that causes them to separate from one another ac-

pH

7.45 7.3 7.2 7.1 7.0 6.8 6.55 6.3 6.1 6.0 5.9

Molecular weight $\times 10^{-3}$

90
80
70
60
50
40
30
20
15

FIGURE 18.29 Two-dimensional gel electrophoresis. A two-dimensional polyacrylamide gel of HeLa cell nonhistone chromosomal proteins labeled with [^{35}S]methionine. Over a thousand different proteins can be resolved by this technique. (FROM J. L. PETERSON AND E. H. McCONKEY, J. BIOL. CHEM. 251:550, 1976.)

Positive ions formed in electrical discharge

Detector

Beam of positive ions

+
–
N
S

Beam is divided into several beams, each containing ions of the same mass

Magnet whose strength can be varied

FIGURE 18.30 Principles of operation of a mass spectrometer. (FROM J. E. BRADY, J. RUSSELL, AND J. R. HOLUM, CHEMISTRY, 3D ED.; COPYRIGHT © 2000, JOHN WILEY AND SONS, INC. REPRINTED WITH PERMISSION OF JOHN WILEY AND SONS, INC.)

cording to their molecular mass [or more precisely according to their mass-to-charge (m/z) ratio]. The ions strike an electronic detector located at the end of the tube. Smaller ions travel faster and strike the detector more rapidly than larger ions. The input to the detector is converted into a series of peaks of ascending *m/z* ratio (as in Figure 2.49).

Mass spectrometers have been a favorite instrument of chemists for many years, but it has only been in the past decade or so that biologists have discovered their remarkable analytic powers. Now, using mass spectrometry (MS), protein biochemists are able to identify unknown proteins in a matter of hours. To carry out this analysis, proteins are generally digested with trypsin, fractionated by liquid chromatography, and the peptides introduced into a mass spectrometer where they are gently ionized and made gaseous by one of two procedures. Development of these peptide ionization techniques has been the key in adapting MS to the study of proteins. In one procedure, called *matrix-assisted laser desorption ionization* (*MALDI*), the protein sample is applied as part of a crystalline matrix, which is irradiated by a laser pulse. The energy of the laser excites the matrix and the absorbed energy converts the peptides into gaseous ions. In an alternate procedure, called *electrospray ionization* (*ESI*), an electric potential is applied to a peptide solution, causing the peptides to ionize and the liquid to spray as a fine mist of charged particles that enter the spectrometer. Because it acts on molecules in solution, ESI is well suited to ionize peptides prepared from proteins fractionated by a widely employed liquid chromatography technique.

Once the molecular masses of the peptides in the sample have been determined, the complete protein can be identified by a database search as discussed on page 70. If the protein is not identified unambiguously, one or more of the peptides generated by tryptic digestion can be fragmented[5] in a second step and subjected to another round of mass spectrometry. This two-step procedure (called tandem MS, or MS/MS) yields the amino acid sequence of the peptide(s) and hence the unmistakable identification of the protein. MS/MS is so powerful that complex mixtures of hundreds of unknown proteins can be digested and subjected to mass spectrometry and the identity of each of the proteins in the mixture determined.

18.8 DETERMINING THE STRUCTURE OF PROTEINS AND MULTISUBUNIT COMPLEXES

X-ray crystallography (or **X-ray diffraction**) utilizes protein crystals, which are bombarded with a fine beam of X-rays (Figure 18.31). The radiation that is scattered (diffracted) by the electrons of the protein's atoms strikes an electron-sensitive

---

[5]Fragmentation is accomplished within the mass spectrometer by collision of the peptides with an inert gas. The energy of collision breaks peptide bonds to produce a random collection of fragments of the original peptide. The amino acid sequence of each fragment, and hence that of the original peptide, can be determined by searching a database containing the masses of theoretical fragments having every possible sequence of amino acids that can be formed from the proteins encoded by that genome.

FIGURE 18.31   X-ray diffraction analysis. Schematic diagram of the diffraction of X-rays by atoms of one plane of a crystal onto a photographic plate. The ordered array of the atoms within the crystal produces a repeating series of overlapping circular waves that spread out and intersect the film. As with diffraction of visible light, the waves form an interference pattern, reinforcing each other at some points on the film and canceling one another at other points.

detector placed behind the crystal. The diffraction pattern produced by a crystal is determined by the structure within the protein; the large number of molecules in the crystal reinforces the reflections, causing it to behave as if it were one giant molecule. The positions and intensities of the reflections, such as those on the photographic plate of Figure 2.33, can be related mathematically to the electron densities within the protein, because it is the electrons of the atoms that produced them. The resolution obtained by X-ray diffraction depends on the number of spots that are analyzed.

Myoglobin was the first protein whose structure was determined by X-ray diffraction. The protein was analyzed successively at 6, 2, and 1.4 Å, with years elapsing between each completed determination. Considering that covalent bonds are between 1 and 1.5 Å in length and noncovalent bonds between 2.8 and 4 Å in length, the information gathered for a protein depends on the resolution achieved. This is illustrated by a comparison of the electron density of a small organic molecule at four levels of resolution (Figure 18.32). In myoglobin, a resolution of 6 Å was sufficient to show the manner in which the polypeptide chain is folded and the location of the heme moiety, but it was not sufficient to show structure within the chain. At a resolution of 2 Å, groups of atoms could be separated from one another, whereas at 1.4 Å, the positions of individual atoms were determined. To date, the structures of several hundred proteins have been determined at atomic resolution (<1.2 Å) and a few as low as 0.66 Å.

Over the years, X-ray diffraction technology has improved greatly. It took Max Perutz 22 years to solve the structure of hemoglobin (Figure 2.38), a task that today might take a few weeks. In most current studies: (1) intense, highly focused X-ray beams are generated by synchrotrons (see p. 98), which are high-energy particle accelerators that produce X-rays as a by-product, and (2) highly sensitive electronic detectors (charge-coupled devices, or CCDs) that provide a digital readout of the diffraction data have replaced photographic plates. Use of these instruments in conjunction with increasingly powerful computers allows researchers to collect and analyze sufficient data to determine the tertiary structure of most pro-

a   6-Å resolution   b   2-Å resolution   c   1.5-Å resolution   d   1.1-Å resolution

FIGURE 18.32 Electron density distribution of a small molecule (diketopiperazine) calculated at several levels of resolution. At the lowest resolution (*a*), only the ring nature of the molecule can be distinguished, whereas at the highest resolution (*d*), the electron density around each atom (indicated by the circular contour lines) is revealed. (AFTER D. HODGKIN. REPRINTED WITH PERMISSION FROM NATURE 188:445, 1960; © COPYRIGHT 1960, MACMILLAN MAGAZINES LTD.)

teins in a matter of hours. As a result of these advances, X-ray crystallography has been applied to the analysis of larger and larger molecular structures. This is probably best illustrated by the success attained in the determination of the structure of the ribosome, which is discussed in Chapter 11. In most cases, as was true in the study of the ribosome, the greatest challenge in this field is obtaining usable crystals.

X-ray crystallography is ideally suited for determining the structure of soluble proteins that lend themselves to crystallization, but can be very challenging for the study of complex multisubunit structures, such as ribosomes or proteasomes, or for membrane proteins, where it is difficult to obtain the three-dimensional crystals required for analysis. Structural analysis of these types of specimens is often conducted using an alternate technique that takes advantage of the tremendous resolving power of the electron microscope and computer-based image processing techniques. There are two general approaches to the study of single particles with the electron microscope. In one approach, the particles are placed on an electron microscope grid, negatively stained, and then air-dried (as discussed on page 726). In the other approach, which is referred to as *electron cryomicroscopy*, or *cryo-EM*, the particles are placed on a grid and rapidly frozen in a hydrated state in liquid nitrogen without being fixed or stained. For higher resolution studies, the use of negatively stained specimens has largely been replaced by frozen-hydrated particles, which are much less likely to generate artifacts and are much better suited for the study of internal features within the particle's structure. In either case, the grids are placed in the column of the microscope and photographs of the particles are taken. Each photograph is a two-dimensional image of an individual particle in the orientation that it happened to assume as it rested on the grid. When two-dimensional images of tens of thousands of different specimens in every conceivable orientation are averaged by high-powered computer analysis, a three-dimensional reconstruction of the particle at a resolution of ~10 Å can be generated. At this resolution, investigators can trace the polypeptide chain of a protein and even identify the location of bulky amino acid side chains. A model of a eukaryotic ribosome based on cryo-EM is illustrated in Figure 2.56 and a model of a U1 snRNP particle in Figure 11.33. This technique is also useful for capturing images of a structure, such as a

ribosome, at different stages during a dynamic process, such as the elongation step of protein synthesis. Using this approach, several of the major conformational changes that occur during each step of translation have been revealed.

In addition, atomic-resolution structures determined by X-ray crystallography can be fitted into the lower resolution electron microscopic reconstructions to show how the individual molecules that make up a multisubunit complex interact and how they might work together to carry out a particular activity. Figure 18.33 shows the tertiary structure of a filament composed of actin and ADF (a member of the cofilin family,
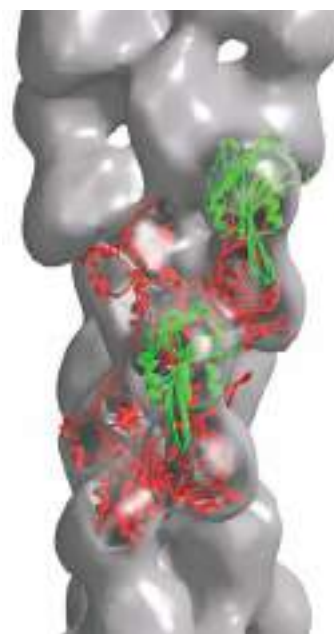
FIGURE 18.33 Combining data from electron microscopy and X-ray crystallography provides information on protein–protein interactions and the structure of multisubunit complexes. The electron microscopic reconstruction of an actin-ADF filament is shown in grey. The high-resolution X-ray crystal structures of individual actin monomers (red) and ADF molecules (green) have been fitted into the lower resolution EM structure. (FROM VITOLD E. GALKIN ET AL, COURTESY OF EDWARD H. EGELMAN, J. CELL BIOL. 163:1059, 2003; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

page 367). The structures of the two proteins were determined by separate X-ray crystallographic studies and then fitted into an electron microscopic model of an actin-ADF filament. The reconstruction shown in this figure served as the basis for a proposed mechanism by which cofilin family proteins can induce severing and depolymerization of an actin filament (page 370).

Electron microscopic analysis of frozen specimens can also be utilized in the study of membrane proteins, such as the nicotinic acetylcholine receptor (see the Experimental Pathway for Chapter 4). This type of analysis requires that the membrane proteins be closely packed at very low temperatures (e.g., −195°C) into two-dimensional crystalline arrays within the plane of the membrane. This technique offers an advantage over X-ray crystallographic studies of membrane proteins in that the protein remains during the entire process within its own natural membrane rather than being extracted in detergent and crystallized in a nonmembranous environment. The structures illustrated on page 169 were determined from combined, high-resolution, electron-microscopic images of many different protein molecules taken at various angles. This technique is known as *electron crystallography*.

18.9 PURIFICATION OF NUCLEIC ACIDS

The steps taken in the purification of nucleic acids are very different from those used in the purification of proteins, which reflects the basic difference in structure of these two types of macromolecules. The first step in DNA purification is generally to homogenize the cells and isolate the nuclei from which the DNA is extracted. The nuclei are then extracted with a buffered salt solution containing a detergent, such as SDS, that serves to lyse the nuclei and release the DNA. As the DNA is released, the viscosity of the solution rises markedly. The detergent also inhibits any nuclease activity present in the preparation.

The major goal of subsequent purification steps is to separate the DNA from contaminating materials, such as RNA and protein. Deproteinization is usually accomplished by shaking the mixture with a volume of phenol. Phenol (or alternatively phenol/chloroform) is an active protein denaturant that causes proteins in the preparation to lose their solubility and precipitate from solution. Because phenol and buffered saline solutions are immiscible, the suspension is simply centrifuged to separate the phases, which leaves the DNA (and RNA) in solution in the upper aqueous phase and the protein present as a precipitate at the boundary between the two phases. The aqueous phase is removed from the tube and subjected to repeated cycles of shaking with phenol and centrifugation until no further protein is removed from solution. The nucleic acids are then precipitated from solution by addition of cold ethanol. The cold ethanol is often layered on top of the aqueous DNA solution, and the DNA is wound on a glass rod as it comes out of solution at the interface between the ethanol and saline. In contrast, RNA comes out of solution as a flocculent precipitate that settles at the bottom of the vessel. Following this initial purification procedure, the DNA is re-

dissolved and treated with ribonuclease to remove contaminating RNA. The ribonuclease is then destroyed by treatment with a protease, the protease is subsequently removed by deproteinization with phenol, and the DNA is reprecipitated with ethanol.

RNA can be purified in a similar manner using DNase in the final purification steps rather than ribonuclease. An alternative procedure for isolating RNA in a single step was published in 1987. In this technique, tissues are homogenized in a solution containing 4 M guanidine thiocyanate, and the RNA extract is mixed with phenol and shaken with chloroform (or bromochloropropane). The suspension is then centrifuged, leaving RNA in the upper aqueous phase and both DNA and protein at the interface between the two phases.

An alternative approach to DNA purification involves the use of membranes or matrices to which DNA will bind under specific conditions. To purify DNA using one of these materials, cells are lysed in a solution that facilitates selective binding of DNA to the matrix. The lysate is applied to the matrix, and contaminants are removed from the DNA by washing. Finally the DNA is rinsed from the matrix with an elution buffer. These matrices are often set up in tiny columns inside centrifuge tubes, so that the binding, washing, and elution steps can be accomplished efficiently through the application of centifugal force.

18.10 FRACTIONATION OF NUCLEIC ACIDS

Any systematic fractionation method must exploit differences between members of a mixture to effect separation. Nucleic acid molecules can differ from one another in overall size, base composition, topology, and nucleotide sequence. Accordingly, fractionation methods for nucleic acids are based on these features.

Separation of DNAs by Gel Electrophoresis

Of the various techniques used in the fractionation of proteins discussed earlier, one of them, gel electrophoresis, is also widely used to separate nucleic acids of different molecular mass (i.e., nucleotide length). Small RNA or DNA molecules of a few hundred nucleotides or less are generally separated by polyacrylamide gel electrophoresis. Larger molecules have trouble making their way through the cross-linked polyacrylamide and are generally fractionated on agarose gels, which have greater porosity. Agarose is a polysaccharide extracted from seaweed; it is dissolved in hot buffer, poured into a mold, and caused to gelate simply by lowering the temperature. Separation of DNA molecules greater than about 25 kb is generally accomplished using the technique of pulsed-field electrophoresis in which the direction of the electric field in the gel is periodically changed, which causes the DNA molecules to reorient themselves during their migration.

Following electrophoresis, the DNA fragments in the gel are visualized by soaking the gel in a solution of stain such as ethidium bromide. Ethidium bromide intercalates into the double helix and causes the DNA bands to appear fluorescent

when viewed under ultraviolet light (Figure 18.34). The sensitivity of gel electrophoresis is so great that DNA or RNA molecules that differ by only a single nucleotide can be sepa-

(a)

(b)

FIGURE 18.34 Separation of DNA restriction fragments by gel electrophoresis. (*a*) DNA is incubated with a restriction enzyme, which cuts it into fragments (page 746). The mixture of fragments is introduced into a slot, or well, in a slab of agarose and an electric current is applied. The negatively charged DNA molecules migrate toward the positive electrode and separate by size. (*b*) All of the DNA fragments that are present in a gel can be revealed by immersing the gel in a solution of ethidium bromide and then viewing the gel under an ultraviolet light. (B: PHOTOGRAPH BY PHILLIPE PLAILLY/SCIENCE PHOTO LIBRARY/PHOTO RESEARCHERS.)

rated from one another, a feature that gave rise to an invaluable method for DNA sequencing (page 754). Since the rate of migration through a gel can also be affected by the shape of the molecule, electrophoresis can be used to separate molecules with different conformations, such as circular and linear or relaxed and supercoiled forms (see Figure 10.12).

Separation of Nucleic Acids by Ultracentrifugation

Common experience tells us that the stability of a solution (or suspension) depends on the components. Cream floats to the top of raw milk, a fine precipitate gradually settles to the bottom of a container, and a solution of sodium chloride remains stable indefinitely. Numerous factors determine whether or not a given component will settle through a liquid medium, including the size, shape, and density of the substance and the density and viscosity of the medium. If a component in a solution or suspension is denser than its medium, then centrifugal force causes it to become concentrated toward the bottom of a centrifuge tube. Larger particles sediment more rapidly than smaller particles of similar shape and density. The tendency for molecules to become concentrated during centrifugation is counteracted by the effects of diffusion, which causes the molecules to become redistributed in a more uniform (random) arrangement. With the development of ultracentrifuges, it has become possible to generate centrifugal forces as high as 500,000 times the force of gravity, which is great enough to overcome the effects of diffusion and cause macromolecules to sediment toward the bottom of a centrifuge tube. Centrifugation proceeds in a near vacuum to minimize frictional resistance. DNA (and RNA) molecules are extensively analyzed by techniques utilizing the ultracentifuge. For the present purpose, we will consider two centrifugation techniques used in the study of nucleic acids, which are illustrated in Figure 18.35.

Velocity Sedimentation   The rate at which a given molecule moves in response to centrifugal force in a centrifuge is known as its *sedimentation velocity*. Since the sedimentation velocity changes as the centrifugal force changes, a given molecule is characterized by a sedimentation coefficient, which is its sedimentation velocity divided by the force. Throughout this book we have referred to various macromolecules and their complexes as having a particular S value. The unit S (or Svedberg, after the inventor of the ultracentrifuge) is equivalent to a sedimentation coefficient of 10^{-13} sec. Because the velocity at which a particle moves through a liquid column depends on a number of factors, including shape, the determination of sedimentation coefficients does not by itself provide the molecular mass. However, as long as one is dealing with the same type of molecule, the S value provides a good measure of relative size. For example, the three ribosomal RNAs of *E. coli*, the 5S, 16S, and 23S molecules, have lengths of 120, 1600, and 3200 nucleotides, respectively.

   In *velocity* (or *rate-zonal*) *sedimentation*, nucleic acid molecules are separated according to nucleotide length. The sample containing the mixture of nucleic acid molecules is carefully layered over a solution containing an increasing concentration of

(a)

(b)

(c)

FIGURE 18.35 Techniques of nucleic acid sedimentation. (*a*) Separation of different-sized DNA molecules by velocity sedimentation. The sucrose density gradient is formed within the tube (step 1) by allowing a sucrose solution of increasing concentration to drain along the wall of the tube. Once the gradient is formed, the sample is carefully layered over the top of the gradient (steps 2 and 3), and the tube is subjected to centrifugation (e.g., 50,000 rpm for 5 hours) as illustrated in step 4. The DNA molecules are separated on the basis of their size (step 5). (*b*) Separation of DNA molecules by equilibrium sedimentation on the basis of differences in base composition. The DNA sample is mixed with the CsCl solution (step 1) and subjected to extended centrifugation (e.g., 50,000 rpm for 72 hours). The CsCl gradient forms during the centrifugation (step 2), and the DNA molecules band in regions of equivalent density (step 3). (*c*) The tube from the experiment of *b* is punctured and the contents are allowed to drip into successive tubes, thereby fractionating the tube's contents. The absorbance of the solution in each fraction is measured and plotted as shown.

cleic acid molecules, even at the bottom of the tube (approximately 1.2 g/ml for the sucrose solution and 1.7 g/ml for the nucleic acid), these molecules continue to sediment as long as the tube is centrifuged. In other words, centrifugation never reaches equilibrium. After a prescribed period, the tube is removed from the centrifuge, its contents are fractionated (as shown in Figure 18.35*c*), and the relative positions of the various molecules are determined. The presence of the viscous sucrose prevents the contents of the tube from becoming mixed due to either convection or handling, allowing molecules of identical S value to remain in place in the form of a band. If marker molecules of known sedimentation coefficient are present, the S values of unknown components can be determined. Experimental results obtained by sucrose-density gradient centrifugation are shown in Figures 11.13 and 11.17.

Equilibrium Centrifugation   In the other type of centrifugation technique, *equilibrium* (or *isopycnic*) *centrifugation* (Figure 18.35*b*), nucleic acid molecules are separated on the basis of their buoyant density. In this procedure, one generally employs a highly concentrated solution of the salt of the heavy-metal cesium. The analysis is begun by mixing the DNA with the solution of cesium chloride or cesium sulfate in the centrifuge tube and then subjecting the tube to extended centrifugation (e.g., 2 to 3 days at high forces). During the centrifugation, the heavy cesium ions are slowly driven toward the bottom of the tube, forming a continuous density gradient through the liquid column. After a time, the tendency for cesium ions to be concentrated toward the bottom of the tube is counterbalanced by the opposing tendency for them to become redistributed by diffusion, and the gradient becomes stabilized. As the cesium gradient is forming, individual DNA molecules are driven downward or move buoyantly upward in the tube until they reach a position that has a buoyant density equivalent to their own, at which point they are no longer subject to further movement. Molecules of equivalent density form narrow bands within the tube. This technique is sensitive enough to separate DNA molecules having different base composition (as illustrated in Figure 18.35*b*) or ones containing different isotopes of nitrogen (^{15}N vs. ^{14}N, as shown in Figure 13.3*b*).

sucrose (or other suitable substance). This preformed gradient increases in density (and viscosity) from the top to the bottom. When subjected to high centrifugal forces, the molecules move through the gradient at a rate determined by their sedimentation coefficient. The greater the sedimentation coefficient, the farther a molecule moves in a given period of centrifugation. Because the density of the medium is less than that of the nu-

18.11 NUCLEIC ACID HYBRIDIZATION

Nucleic acid hybridization includes a variety of related techniques that are based on the observation that two single-stranded nucleic acid molecules of complementary base sequence can form a double-stranded hybrid. Consider a situation in which one has a mixture of hundreds of fragments of DNA of identical length and overall base composition that differ from one another solely in their base sequence. Assume, for example, that one of the DNA fragments constitutes a portion of the β-globin gene and all the other fragments contain unrelated genes. The only way to distinguish between the fragment encoding the β-globin polypeptide and all the others is to carry out a molecular hybridization experiment using complementary molecules as probes.

In the present example, incubation of the mixture of denatured DNA fragments with an excess number of β-globin mRNAs would drive the globin fragments to form double-stranded DNA–RNA hybrids, leaving the other DNA fragments single-stranded. There are a number of ways one could separate the DNA–RNA hybrids from the single-stranded fragments. For example, the mixture could be passed through a column of hydroxylapatite under ionic conditions in which the hybrids would bind to the calcium phosphate salts in the column, while the nonhybridized DNA molecules would pass through unbound. The hybrids could then be released from the column by increasing the concentration of the elution buffer.

Experiments using nucleic acid hybridization require the incubation of two populations of complementary single-stranded nucleic acids under conditions (ionic strength, temperature, etc.) that promote formation of double-stranded molecules. Depending on the type of experiment being conducted, the two populations of reacting molecules may both be present in solution, or one of the populations may be immobilized, for example, by localization within a chromosome (as in Figure 10.19).

In many cases, one of the populations of single-stranded nucleic acids to be employed in a hybridization experiment is present within a gel. Consider a population of DNA fragments that have been prepared from genomic DNA and fractionated by gel electrophoresis (Figure 18.36). To carry out the hybridization, the gel is treated to render the DNA single-stranded. The single-stranded DNA is transferred from the gel to a nitrocellulose membrane and fixed onto the membrane by heating it to 80°C in a vacuum. The procedure by which DNA is transferred to a membrane is termed *blotting*. Once the DNA is attached, the membrane is incubated with a labeled, single-stranded DNA (or RNA) probe capable of hybridizing to a complementary group of fragments. Unbound probe is then washed away, and the location of the bound probe is determined autoradiographically as shown in Figure 18.36. The experiment just described and depicted in Figure 18.36 using a radioactive probe is called a *Southern blot* (named after Edwin Southern, its developer). One or a few DNA restriction fragments that contain a particular nucleotide sequence can be identified in a Southern blot, even if there are thousands of unrelated fragments present
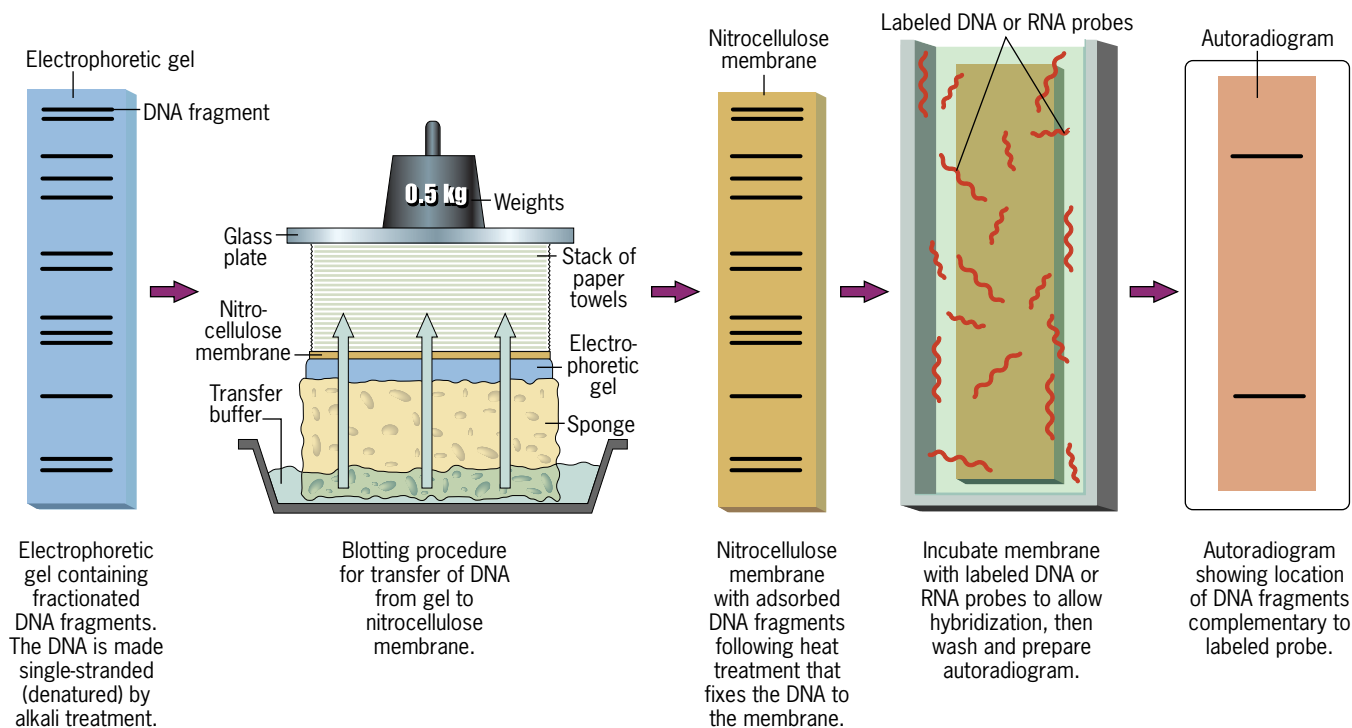
FIGURE 18.36 Determining the location of specific DNA fragments in a gel by a Southern blot. As described in the figure, the fractionated DNA fragments are washed out of the gel and trapped onto a nitrocellulose membrane, which is incubated with radioactively labeled DNA (or RNA) probes. The location of the hybridized fragments is deter-

mined autoradiographically. During the blotting procedure, capillary action draws the buffer upward into the paper towels. As the buffer moves through the electrophoretic gel, it dissolves the DNA fragments and transfers them to the surface of the adjacent membrane.

in the gel. An example of a Southern blot is shown in Figure 10.18. RNA molecules can also be separated by electrophoresis and identified with a labeled DNA probe after being blotted onto a membrane. An example of this procedure, which is called a *Northern blot*, is shown in Figure 11.36.

DNA probes can be labeled in a variety of ways. A radioactive probe incorporates a radioactive isotope (such as ^{32}P) at one or more locations in the molecule. The presence of the probe is detected by autoradiography as shown in Figure 18.36. Probes can also be labeled with fluorophores and detected by fluorescence. Another commonly used label is biotin, a small organic molecule that can be covalently linked to the DNA backbone. Biotin is detected by the protein avidin (or streptavidin), which binds tightly to it. The avidin itself must be labeled for detection, such as with a fluorophore as shown in Figures 10.19 and 10.20.

Nucleic acid hybridization can also provide a measure of the similarity in nucleotide sequence between two samples of DNA, as might be obtained, for example, from two different organisms. The more distant the evolutionary relationship between the two species, the greater the divergence of their DNA sequences. If purified DNA from species A and B is mixed together, denatured, and allowed to reanneal, a percentage of the DNA duplexes are formed by DNA strands from the two species. Because they contain mismatched bases, such duplexes are less stable than those formed by DNA strands of the same species, and this instability is reflected in the lower temperature at which they melt. When DNAs from different species are allowed to reanneal in different combinations, the melting temperature (T_m, page 393) of the hybrid duplexes provides a measure of the evolutionary distance between the organisms. Two other important types of nucleic acid hybridization protocols are discussed in detail in the text: *in situ hybridization* on page 397 and hybridization to *cDNA microarrays* on page 505.

18.12 CHEMICAL SYNTHESIS OF DNA

Hybridization analysis requires single-stranded nucleic acid molecules for use as probes. Other fundamental techniques for the manipulation and analysis of DNA in the laboratory also require short single-stranded nucleic acid molecues, or oligonucleotides. Chemical synthesis of DNA and RNA is therefore a key supporting technology for many procedures.

The development of chemical techniques to synthesize polynucleotides having a specific base sequence was begun by H. Gobind Khorana in the early 1960s as part of an attempt to decipher the genetic code. Khorana and co-workers continued to refine their techniques, and a decade after their initial work on the code, they succeeded in synthesizing a complete bacterial tyrosine tRNA gene, including the nontranscribed promoter region. The gene totaling 126 base pairs was put together from over 20 segments, each of which was individually synthesized and later joined enzymatically. This artificial gene was then introduced into bacterial cells carrying mutations for this tRNA, and the synthetic DNA was able to replace the previously deficient function. The first chemically synthesized gene encoding an average-sized protein, human interferon,

was prepared in 1981, an effort that required the synthesis and assembly of 67 different fragments to produce a single duplex of 514 base pairs containing initiation and termination signals recognized by the bacterial RNA polymerase.

The chemical reactions that link nucleotides have been automated, and oligonucleotide synthesis is now carried out by computer-controlled machines hooked to reservoirs of reagents. The operator enters the desired nucleotide sequence into the computer and keeps the instrument supplied with materials. The oligonucletide is assembled one nucleotide at a time from the 3′ to the 5′ end of the molecule, up to a total of about 100 nucleotides. Modifications such as biotin and fluorophores can be incorporated into the molecules. If a double-stranded molecule is needed, it is synthesized as two complementary single strands that can be hybridized together. Longer synthetic molecules are made in segments which are joined together.

18.13 RECOMBINANT DNA TECHNOLOGY

Over the past 30 years, tremendous advances have been made in the analysis of eukaryotic genomes. This progress began as molecular biologists learned to construct **recombinant DNA** molecules, which are molecules containing DNA sequences derived from more than one source. Recombinant DNAs can be used in myriad ways. We will begin by considering one of the most important applications: the isolation from the genome of a particular segment of DNA that encodes a particular polypeptide. First, however, it is necessary to consider a class of enzymes whose discovery and use has made the formation of recombinant DNA molecules possible.

Restriction Endonucleases

During the 1970s, it was found that bacteria contained nucleases that would recognize short nucleotide sequences within duplex DNA and cleave the DNA backbone at specific sites on both strands of the duplex. These enzymes are called type II **restriction endonucleases**, or simply *restriction enzymes*. They were given this name because they function in bacteria to destroy viral DNAs that might enter the cell, thereby *restricting* the growth of the viruses. The bacterium protects its own DNA from nucleolytic attack by methylating the bases at susceptible sites, a chemical modification that blocks the action of the enzyme.

Enzymes from several hundred different prokaryotic organisms have been isolated that, together, recognize over 100 different nucleotide sequences. The sequences recognized by most of these enzymes are four to six nucleotides long and are characterized by a particular type of internal symmetry. Consider the particular sequence recognized by the enzyme *Eco*R1:

$$\downarrow$$
$$3′—CTTAAG—5′$$
$$5′—GAATTC—3′$$
$$\uparrow$$
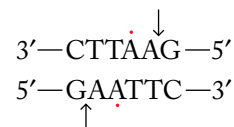
This segment of DNA is said to have *twofold rotational symmetry* because it can be rotated 180° without change in base

sequence. Thus, if one reads the sequence in the same direction (3′ to 5′ or 5′ to 3′) on either strand, the same order of bases is observed. A sequence with this type of symmetry is called a *palindrome*. When the enzyme *Eco*R1 attacks this palindrome, it breaks each strand at the same site *in the sequence*, which is indicated by the arrows between the A and G residues. The red dots indicate the methylated bases in this sequence that protect the host DNA from enzymatic attack. Some restriction enzymes cleave bonds directly opposite one another on the two strands producing blunt ends, whereas others, such as *Eco*R1, make staggered cuts.

The discovery and purification of restriction enzymes have been invaluable in the advances made by molecular biologists in recent years. Since a particular sequence of four to six nucleotides occurs quite frequently simply by chance, any type of DNA is susceptible to fragmentation by these enzymes. The use of restriction enzymes allows the DNA of the human genome, or that of any other organism, to be dissected into a precisely defined set of specific fragments. Once the DNA from a particular individual is digested with one of these enzymes, the fragments generated can be fractionated on the basis of length by gel electrophoresis (as in Figure 18.37*a*). Different enzymes cleave the same preparation of DNA into different sets of fragments, and the sites within the genome that are cleaved by various enzymes can be identified and ordered into a restriction map such as that depicted in Figure 18.37*b*.
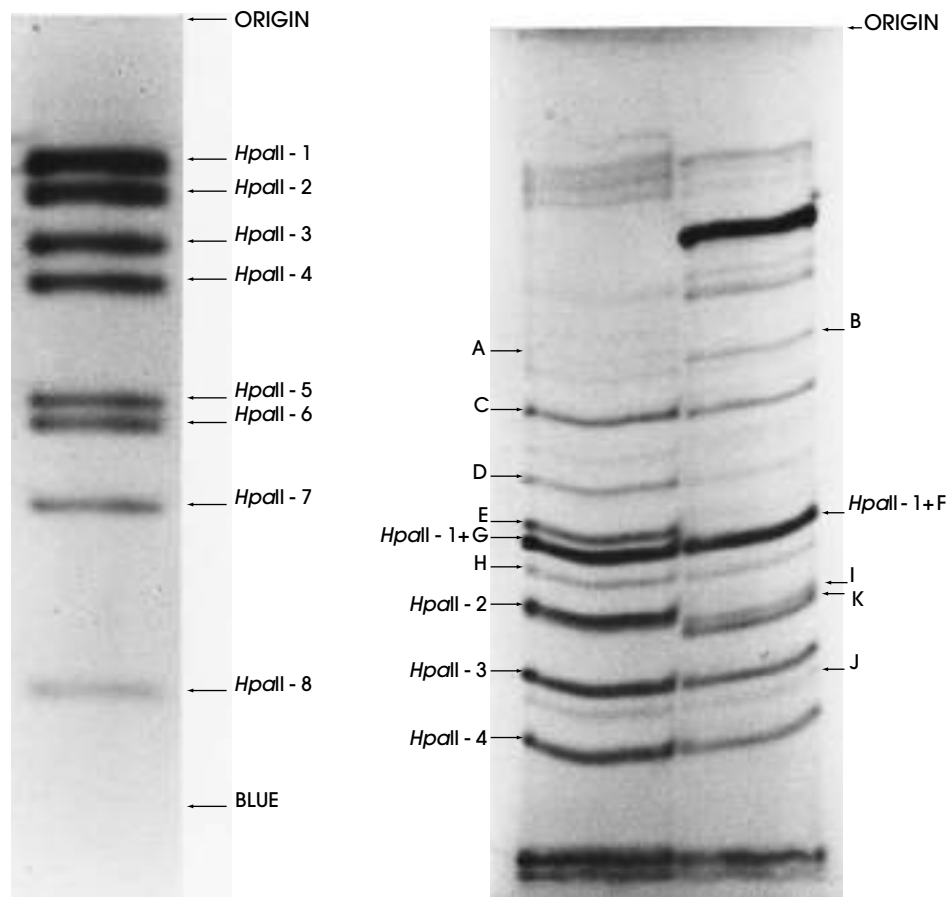
(a)
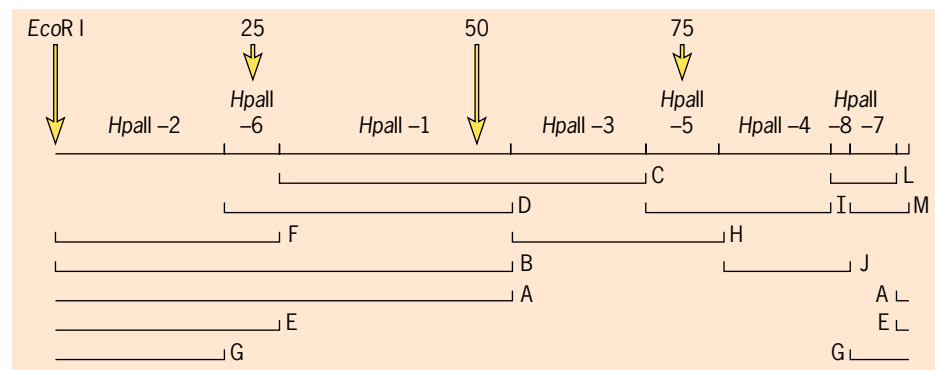
(b)

FIGURE 18.37 The construction of a restriction map of the small circular genome of the DNA tumor virus polyoma. (*a*) Autoradiographs of ^{32}P-labeled DNA fragments that have been subjected to gel electrophoresis. The gel on the left shows the pattern of DNA fragments obtained after a complete digestion of the polyoma genome with the enzyme *Hpa*II. To determine how these eight fragments are pieced together to make up the intact genome, it is necessary to treat the DNA in such a way that overlapping fragments are generated. Overlapping fragments can be produced by treating the intact genome with a second enzyme that cleaves the molecule at different sites, or by treating the genome with the same enzyme under conditions where the DNA is not fully digested as it was in the left gel. The two gels on the right represent examples of partial digests of the polyoma genome with *Hpa*II. The middle gel shows the fragments generated by partial digestion of the superhelical circular DNA, and the gel on the right shows the *Hpa*II fragments formed after the circular genome is converted into a linear molecule by *Eco*R1 (an enzyme that makes only one cut in the circle). (*b*) The restriction map of the linearized polyoma genome based on cleavage by *Hpa*II. The eight fragments from the complete digest are shown along the DNA at the top. The overlapping fragments from the partial digest are shown in their ordered arrangement below the map. (Fragments L and M migrate to the bottom of the gel in part *a*, right side.) (FROM BEVERLY E. GRIFFIN, MIKE FRIED, AND ALLISON COWIE, PROC. NAT'L. ACAD. SCI. U.S.A. 71:2078, 1974.)

Formation of Recombinant DNAs

Recombinant DNAs can be formed in a variety of ways. In the method shown in Figure 18.38, DNA molecules from two different sources are treated with a restriction enzyme that makes staggered cuts in the DNA duplex. Staggered cuts leave short, single-stranded tails that act as "sticky ends" that can bind to a complementary single-stranded tail on another DNA molecule to restore a double-stranded molecule. In the case depicted in Figure 18.38, one of the DNA fragments that will make up the recombinant molecule is a bacterial plasmid. Plasmids are small, circular, double-stranded DNA molecules that are separate from the main bacterial chromosome. The other
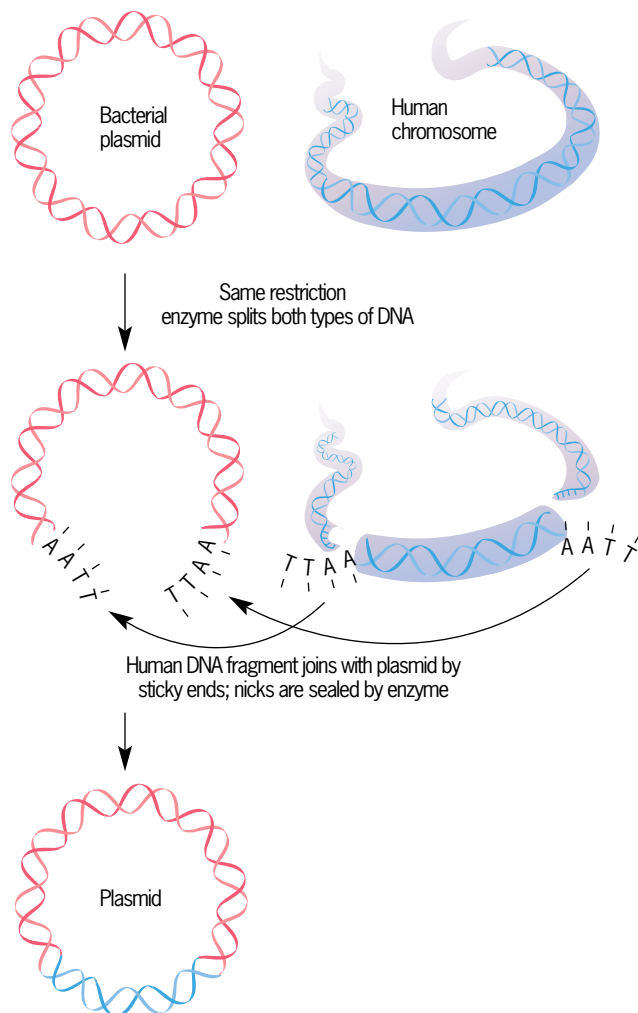
FIGURE 18.38   Formation of a recombinant DNA molecule. In this example, a preparation of bacterial plasmids is treated with a restriction enzyme that makes a single cut within each bacterial plasmid. This same restriction enzyme is used to fragment a preparation of human genomic DNA into small fragments. Because they have been treated with the same restriction enzyme, the cleaved plasmid DNA and the human DNA fragments will have sticky ends. When these two populations are incubated together, the two DNA molecules become noncovalently joined to each other and are then covalently sealed by DNA ligase, forming a recombinant DNA molecule.

DNA fragment in Figure 18.38 is obtained from human cells following treatment with the same restriction enzyme used to open the plasmid. When the human DNA fragments and the treated plasmid are incubated together in the presence of DNA ligase, the two types of DNAs become hydrogen bonded to one another by their sticky ends and are then ligated to form circular DNA recombinants, as in Figure 18.38. The first recombinant DNA molecules were formed by this basic method in 1972 by Paul Berg, Herbert Boyer, Annie Chang, and Stanley Cohen of Stanford University and the University of California, San Francisco, marking the birth of modern genetic engineering.

By following the procedure just described, a large number of different recombinant molecules are produced, each of which contains a bacterial plasmid with a segment of human DNA incorporated into its circular structure (see Figure 18.39). Suppose you were interested in isolating a single gene from the human genome, for example, the gene that codes for insulin. Because your goal is to obtain a purified preparation of the one type of recombinant DNA that contains the insulin-coding fragment, you must separate this one fragment from all of the others. This is done by a process called DNA cloning. We will return to the search for the insulin gene after describing the basic methodology behind DNA cloning.

DNA Cloning

DNA cloning is a technique to produce large quantities of a specific DNA segment. The DNA segment to be cloned is first linked to a *vector DNA*, which is a vehicle for carrying foreign DNA into a suitable host cell, such as the bacterium *E. coli*. The vector contains sequences that allow it to be replicated within the host cell. Two types of vectors are commonly used to clone DNAs within bacterial hosts. In one approach, the DNA segment to be cloned is introduced into the bacterial cell by joining it to a plasmid, as described previously, and then causing the bacterial cells to take up the plasmid from the medium. In an alternate approach, the DNA segment is joined to a portion of the genome of the bacterial virus lambda (λ), which is then allowed to infect a culture of bacterial cells, producing large numbers of viral progeny, each of which contains the foreign DNA segment. Either way, once the DNA segment is inside a bacterium, it is replicated along with the bacterial (or viral) DNA and partitioned to the daughter cells (or progeny viral particles). Thus, the number of recombinant DNA molecules increases in proportion to the number of bacterial cells (or viral progeny) that are formed. From a single recombinant plasmid or viral genome inside a single bacterial cell, millions of copies of the DNA can be formed within a short period of time. Once the amount of DNA has been sufficiently amplified, the recombinant DNA can be purified and used in other procedures. In addition to providing a means to amplify the amount of a particular DNA sequence, cloning can also be used as a technique to isolate a pure form of any specific DNA fragment among a large, heterogeneous population of DNA molecules. We will begin by discussing DNA cloning using bacterial plasmids.
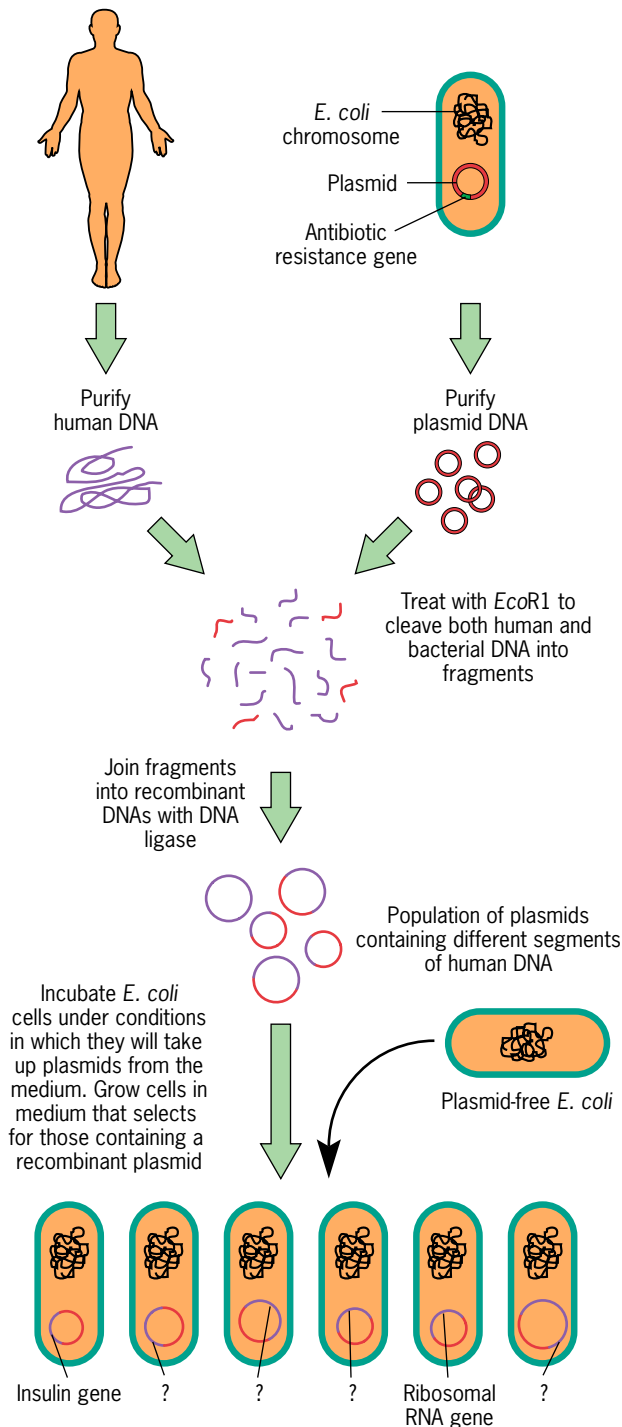
FIGURE 18.39 **An example of DNA cloning using bacterial plasmids.** DNA is extracted from human cells, fragmented with *Eco*R1, and the fragments are inserted into a population of bacterial plasmids. Techniques are available to prevent the formation of plasmids that lack a foreign DNA insert. Once a bacterial cell has picked up a recombinant plasmid from the medium, the cell gives rise to a colony of cells containing the recombinant DNA molecule. In this example, most of the bacteria contain eukaryotic DNAs of unknown function (labeled as ?), while one contains a portion of the DNA encoding ribosomal RNA, and another contains DNA encoding insulin.

Cloning Eukaryotic DNAs in Bacterial Plasmids   The foreign DNA to be cloned is first inserted into the plasmid to form a recombinant DNA molecule. Plasmids used for DNA cloning are modified versions of those found in bacterial cells. Like the natural counterparts from which they are derived, these plasmids contain an origin of replication and one or more genes that make the recipient cell resistant to one or more antibiotics. Antibiotic resistance allows investigators to select for those cells that contain the recombinant plasmid.

As first demonstrated by Avery, Macleod, and McCarty (page 414), bacterial cells can take up DNA from their medium. This phenomenon forms the basis for cloning plasmids in bacterial cells (Figure 18.39). In the most commonly employed technique, recombinant plasmids are added to a bacterial culture that has been pretreated with calcium ions. When subjected to a brief heat shock, such bacteria are stimulated to take up DNA from their surrounding medium. Usually only a small percentage of cells are competent to pick up and retain one of the recombinant plasmid molecules. Once it is taken up, the plasmid replicates autonomously within the recipient and is passed on to the progeny during cell division. Those bacteria containing a recombinant plasmid can be selected from the others by growing the cells in the presence of the antibiotic to which resistance is conferred by a gene on the plasmid.

We began this discussion with the goal of isolating a small DNA fragment that contains the sequence for the insulin gene. To this point, we have formed a population of bacteria containing many different recombinant plasmids, very few of which contain the gene being sought (as in Figure 18.39). One of the great benefits of DNA cloning is that, in addition to producing large quantities of particular DNAs, it allows one to separate different DNAs from a mixture. It was noted above that plasmid-containing bacteria can be selected from those without plasmids by treatment with antibiotics. Once this has been done, the plasmid-bearing cells can be grown at low density on petri dishes so that all of the progeny of each cell (a clone of cells) remain physically separate from the progeny of other cells. Because a large number of different recombinant plasmids were present initially in the medium, the different cells plated on the dish contain different foreign DNA fragments. Once the cells containing the various plasmids have grown into separate colonies, the investigator can search through the colonies for the few that contain the gene being sought—in this case, the insulin gene.

Culture dishes that contain bacterial colonies (or phage plaques) are screened for the presence of a particular DNA sequence using the combined techniques of *replica plating* and *in situ hybridization*. As illustrated in Figure 18.40*a*, replica plating allows one to prepare numerous dishes containing representatives of the same bacterial colonies in precisely the same position on each dish. One of the replica plates is then used to localize the DNA sequence being sought (Figure 18.40*b*), a procedure that requires that the cells be lysed and the DNA fixed onto the surface of a nylon or nitrocellulose membrane. Once the DNA is fixed in place, the DNA is denatured to prepare it for in situ hybridization, during which

(a)

Culture dish with bacterial colonies (or phage plaques) containing recombinant DNA

Bacterial colony → Transfer representative cells to nitrocellulose membrane by replica plating technique → Bacterial colony

Lyse cells and treat to denature DNA and cause single-stranded DNA to adhere to the filter in place

Denatured DNA adhering to membrane

Incubate with radioactively labeled probe and make autoradiograph

Silver grains on X-ray film showing position of labeled hybrid

Autoradiograph

(b)

FIGURE 18.40 Locating a bacterial colony containing a desired DNA sequence by replica plating and in situ hybridization. (*a*) Once the bacterial cells that were plated on the culture dish have grown into colonies, the dish is inverted over a piece of filter paper, allowing some of the cells from each colony to become adsorbed to the paper. Empty culture dishes are then inoculated by pressing them against the filter paper to produce replica plates. (*b*) Procedure for screening the cells of a culture dish for those colonies that contain the recombinant DNA of interest. Once the relevant colonies are identified, cells can be removed from the original dish and grown separately to yield large quantities of the DNA fragments being sought.

FIGURE 18.41 Separation of plasmid DNA from that of the main bacterial chromosome by CsCl equilibrium centrifugation. This centrifugation tube can be seen to contain two bands, one of plasmid DNA carrying a foreign DNA segment that has been cloned within the bacteria, and the other containing chromosomal DNA from these same bacteria. The two types of DNA have been separated during centrifugation (Figure 18.35*b*). The researcher is removing the DNA from the tube with a needle and syringe. The DNA in the tube is made visible using the DNA-binding compound ethidium bromide, which fluoresces under ultraviolet light. (PHOTOGRAPH BY TED SPEIGEL/CORBIS IMAGES.)

the membrane is incubated with a labeled, single-stranded DNA probe containing the complementary sequence to that being sought. Following incubation, the unhybridized probe is washed from the membrane, and the location of the labeled

hybrids is determined by autoradiography. Live representatives of the identified clones can be found at corresponding sites on the original plates, and these cells can be grown into large colonies, which serves to amplify the recombinant DNA plasmid.

After the desired amount of amplification, the cells are harvested, the DNA is extracted, and the recombinant plasmid DNA is readily separated from the much larger chromosome by various techniques, including equilibrium centrifugation (Figure 18.41). The isolated recombinant plasmids can then be treated with the same restriction enzyme used in their formation, which releases the cloned DNA segments from the remainder of the DNA that served as the vector. The cloned DNA can then be separated from the plasmid by centrifugation.

Cloning Eukaryotic DNAs in Phage Genomes Another major cloning vector is the bacteriophage lambda, which is depicted in Figure 18.42. The lambda genome is a linear, double-stranded DNA molecule approximately 50 kb in length. The modified strain used in most cloning experiments contains two cleavage sites for the enzyme *Eco*R1, which fragments the genome into three large segments. Conveniently, all

FIGURE 18.42 Protocol for cloning eukaryotic DNA fragments in lambda phage. The steps are described in the text.

"lawn" is visible at the site of infection. Each plaque contains millions of phage particles, each carrying a single copy of the same eukaryotic DNA fragment.

Lambda phage is appealing as a cloning vector for a number of reasons: (1) the DNA is nicely packaged in a form in which it can be readily stored and easily extracted; (2) virtually every phage containing a recombinant DNA is capable of infecting a bacterial cell; and (3) a single petri dish can accommodate more than 100,000 different plaques. Once the phage plaques have been formed, the particular DNA fragment being sought is identified by a similar process of replica plating and in situ hybridization described in Figure 18.40 for cloning of recombinant plasmids.

18.14 ENZYMATIC AMPLIFICATION OF DNA BY PCR

In 1983, a new technique was conceived by Kary Mullis of Cetus Corporation that has become widely used to amplify specific DNA fragments without the need for bacterial cells. This technique is known as **polymerase chain reaction (PCR)**. There are many different PCR protocols used for a multitude of different applications in which anywhere from one to a large population of related DNAs can be amplified. PCR amplification is readily adapted to RNA templates by first converting them to complementary DNAs using reverse transcriptase.

The basic procedure used in PCR is depicted in Figure 18.43. The technique employs a heat-stable DNA polymerase, called *Taq polymerase*, that was originally isolated from *Thermus aquaticus*, a bacterium that lives in hot springs at temperatures above 90°C. In the simplest protocol, a sample of DNA is mixed with an aliquot of *Taq* polymerase and all four deoxyribonucleotides, along with a large excess of two short, synthetic DNA fragments (oligonucleotides) that are complementary to DNA sequences at the 3′ ends of the region of the DNA to be amplified. The oligonucleotides serve as primers (page 538) to which nucleotides are added during the following replication steps. The mixture is then heated to about 95°C, which is hot enough to cause the DNA molecules in the sample to separate into their two component strands. The mixture is then cooled to about 60°C to allow the primers to hybridize to the strands of the target DNA, and the temperature is raised to about 72°C, to allow the thermophilic polymerase to add nucleotides to the 3′ end of the primers. As the polymerase extends the primer, it selectively copies the target DNA, forming new complementary DNA strands. The temperature is raised once again, causing the newly formed and the original DNA strands to separate from each other. The sample is then cooled to allow the synthetic primers in the mixture to bind once again to the target DNA, which is now present at twice the original amount. This cycle is repeated over and over again, each time doubling the amount of the specific region of DNA that is flanked by the bound primers. Billions of copies of this one specific region can be generated in just a few hours using a *thermal cycler* that automatically

the information essential for infection and cell lysis is contained in the two outer segments, so that the dispensable middle fragment can be replaced by a piece of eukaryotic DNA up to approximately 25 kb. Recombinant DNA molecules can be packaged into phage heads in vitro, and these genetically engineered phage particles can be used to infect host bacteria. (Phage DNA molecules lacking the insert are too short to be packaged.) Once in a bacterium, the eukaryotic DNA segment is amplified along with the viral DNA and then packaged into a new generation of virus particles, which are released when the cell is lysed. The released particles infect new cells, and soon a clear spot (or *plaque*) in the bacterial

FIGURE 18.43 Polymerase chain reaction (PCR). As discussed in the text, the procedure takes advantage of a heat-resistant DNA polymerase whose activity is not destroyed when the temperature is raised to separate the two strands of the double helix. With each cycle of duplication, the strands are separated, flanking segments (primers) bind to the ends of the selected region, and the polymerase copies the intervening segment.

changes the temperature of the reaction mixture to allow each step in the cycle to take place.

Applications of PCR

Amplifying DNA for Cloning or Analysis. Since its invention, PCR has found many uses. It can generate many copies of a specific DNA fragment prior to cloning the fragment, an effi-

cient approach if the target sequence is known in sufficient detail that the nucleotide sequence of two complementary primers can be specified. This is particularly helpful in cases where the source DNA is very scarce, since PCR can generate large amounts of DNA from minuscule samples, such as that in a single cell. PCR has been used in criminal investigations to generate quantities of DNA from a spot of dried blood left on a crime suspect's clothing or even from the DNA present

in part of a single hair follicle left at the scene of a crime. For this purpose, one selects regions of the genome for amplification that are highly polymorphic (i.e., vary at high frequency within the population), so that no two individuals will have the same-sized DNA fragments (as in Figure 10.18). This same procedure can be used to study DNA fragments from well-preserved fossil remains that may be millions of years old. The activity of DNA polymerase in PCR is also employed in DNA sequencing (see Section 18.15).

Testing for the Presence of Specific DNA Sequences. Suppose you wanted to determine whether or not a tissue sample contains a particular virus. You could answer this question by a Southern hybridization (page 745) or you could use PCR. In the PCR approach, nucleic acid is isolated from the sample and PCR primers complementary to the viral DNA are added, along with the other PCR reagents. The reaction is then allowed to proceed. If the virus genome is present in the sample, the PCR primers will hybridize to it and the PCR reaction will generate a product. If the virus is not present, the PCR primers will not hybridize and no product will be generated. Thus, in this application, the PCR reaction itself serves as the detection system.

Comparing DNA Molecules. If two DNA molecules have the same base sequence, they will yield the same PCR products in reactions with identical primers. This is the premise for quick assays that compare the similarity of two DNA samples such as genomic DNA from bacterial isolates. PCR is performed on the samples using several primers, which can be specifically designed or randomly generated. The products are separated by gel electrophoresis and compared. The more similar the sequences of the bacterial genomes, the more similar their PCR products will be.

Quantifying DNA or RNA Templates. PCR can also be used to determine how much of a specific nucleotide sequence (DNA or RNA) is present in a mixed sample. One approach to this quantitative PCR uses the binding of a dye specific for double-stranded DNA to quantify the amount of double-stranded product being generated. The rate of accumulation of product is proportional to the amount of template present in the sample.

Another approach uses what have been called "molecular beacons." These are short reporter oligonucleotides with a fluorochrome bound to one end and a quencher molecule on the other end that hybridize in the middle of the target sequence to be amplified. As long as the short oligonucleotide is intact, the fluorochrome and quencher are close enough in proximity that fluorescence is quenched. When the DNA polymerase synthesizes a new strand of DNA complementary to the template, its exonuclease activity (page 544) degrades the reporter oligonucleotide. The fluorochrome is thus separated from the quencher and fluoresces. The amount of fluorochrome liberated in a given PCR cycle is directly proportional to the number of template molecules being copied by the polymerase.

18.15 DNA SEQUENCING

By 1970, the amino acid sequence of a long list of proteins had been determined, yet virtually no progress had been made toward the sequencing of nucleotides in DNA. There were several reasons for this state of affairs. Unlike DNA molecules, polypeptides come in defined and manageable lengths; a given polypeptide species could be readily purified; a variety of techniques were available to cleave the polypeptide at various sites to produce overlapping fragments; and the presence of 20 different amino acids having widely varying properties made separation and sequencing of small peptides a straightforward task. Then in the mid-1970s, a revolution in DNA sequence technology took place. By 1977, the complete nucleotide sequence of an entire viral genome was reported, that of φX174, some 5375 nucleotides in length. This milestone in molecular biology was accomplished in the laboratory of Frederick Sanger, who had determined the first amino acid sequence of a polypeptide (insulin) some 25 years earlier.

In 1970 it might have taken over a year to sequence a DNA fragment a few dozen nucleotides in length; 15 years later the same task could be accomplished in a single day. In the early days, sequencing was typically performed manually by an individual working at a lab bench who also personally read and interpreted the results of the reactions. Today, sequencing is usually performed in a series of automated procedures at centralized facilities that sequence hundreds of samples every day. The results are interpreted by computer and stored in databases that can be readily analyzed with commonly available software. Not surprisingly, with these technological advances has come an avalanche of sequence data, including complete genomic sequences for the human, dog, domestic cat, chicken, mouse, rat, several insects, fungi and many plants and bacteria. The number and list are growing continuously.

These advances in DNA sequencing were made possible by developments in several areas: molecular approaches to DNA sequencing, instrumentation that could be automated, more powerful and widely available computers, and software for data analysis. The initial key was the development of approaches for determining the sequence of DNA fragments. This advance itself became possible because of the discovery of restriction enzymes and the development of cloning technologies, which provided the means necessary to prepare a defined DNA fragment in sufficient quantity to carry out the necessary biochemical procedures.

In the mid-1970s, two DNA sequencing methodologies were developed almost simultaneously: a chemical approach taken by Alan Maxam and Walter Gilbert of Harvard University, and an enzymatic approach taken by Sanger and A. R. Coulson of the Medical Research Council in Cambridge, England. The Sanger-Coulson method became the most widely used. Following the development of PCR, the Sanger-Coulson sequencing approach and PCR were merged into a sequencing procedure that combines the biochemistry of Sanger-Coulson with the repetitive cycles and automation of PCR. This so-called cycle sequencing became widely used

(a)

(b)

(c)

FIGURE 18.44 **DNA sequencing.** (*a*) The basic steps in sequencing a small hypothetical fragment by the Sanger-Coulson (dideoxy) technique, as described in the text. (*b*) Gel lanes in which fluorescently labeled daughter molecules have been separated. The color of the band is determined by the identity of the dideoxynucleotide at the 3′ end of the DNA strand. (*c*) The sequence of nucleotides in the template strand is interpreted by a computer that "reads" from the bottom to the top of the gel, using the intensity and wavelength of the fluorescent light as input. The computer generates an "electropherogram" showing the intensity and color of the detected fluorescence, along with the DNA sequence interpretation. (B,C: Reprinted with permission from Leroy Hood and David Galas, Nature 421:445, 2003; © copyright 2003, Macmillan Magazines Ltd.)

in genome sequencing, and its basic steps are outlined in Figure 18.44.

In this procedure, one begins with a population of identical template molecules, either a PCR product or a cloned DNA fragment. The template DNA is mixed with a primer that is complementary to the 3′ end of one strand of the region to be sequenced. If the template is a PCR product, the sequencing primer can be one (but only one) of the PCR primers. The reaction mixture also contains the heat stable *Taq* DNA polymerase, all four deoxyribonucleoside triphosphate precursors (dNTPs), and a low concentration of modified precursors called *dideoxyribonucleoside triphosphates*, or *ddNTPs*. Each ddNTP (ddATP, ddGTP, ddCTP, and ddTTP) has been modified by the addition of a different-colored fluorescent dye to its 3′ end.

The sequencing reaction begins, like PCR, by heating the mixture to a temperature that causes the two template strands to denature (Figure 18.44*a*, step 1). Next, the reaction is cooled so that the primer can hybridize to the template DNA (step 2). Note that, in contrast to PCR, only one primer is present, so that only one of the two strands of template DNA can hybridize to a primer. In step 3, the *Taq* polymerase adds dNTPs to the end of the primer that are complementary to the template molecule, synthesizing a new complementary strand of DNA. Every now and then, the polymerase inserts a ddNTP instead of a dNTP. Dideoxynucleotides lack a hy-

droxyl group at both their 2′ and 3′ positions. When one of these nucleotides has been incorporated onto the end of a growing chain, the lack of the 3′ OH makes it impossible for the polymerase to add another nucleotide, thus causing chain termination (Figure 18.44, step 4). Because the ddNTP is present at much lower concentration in the reaction mixture than the corresponding dNTP, the incorporation of the ddNTP is infrequent and random; it may be incorporated near the beginning of one chain, near the middle of another, or not until the end of a third chain. Regardless, when the ddNTP is incorporated, growth of the chain ceases at that point.

After the chain extension phase of the reaction is complete, the temperature is raised again to denature the new double-stranded DNA molecules. The cycle of hybridization, synthesis, and denaturation is repeated many times, generating a large population of daughter DNA strands that by now includes molecules in which a ddNTP has been incorporated at every position. For every A on the template strand, for example, there will be daughter molecules that terminate in a ddTTP at that position. When all cycles are complete, the reaction products are separated by electrophoresis on very thin capillary gels (Figure 18.44, step 5).

High-resolution gel electrophoresis can separate fragments that differ by only one nucleotide in length. If, for example, the initial region to be sequenced contained 100

nucleotides, then each of the 100 labeled daughter molecules would migrate to a different point in the gel, with the shortest daughter migrating the furthest. Each successive band in the gel would contain the daughter molecules that are one nucleotide longer than those in the previous band. Since each ddNTP was labeled with a unique fluorescent dye, the color of the band (viewed under UV light) reveals the identity of the terminal nucleotide on each daughter molecule (Figure 18.44*b*). The order of the colors in the gel therefore corresponds to the base sequence of the template molecule (Figure 18.44*c*).

Over the past few years, a revolution in DNA sequencing technology has taken place, driven by the goal of sequencing genomes—both human and others—rapidly and inexpensively. At the present time, the goal is the "$1000 human genome." Costs for sequencing massive amounts of DNA have decreased dramatically over the past few years, and by 2009 it became possible to sequence the genome of a given person for a cost somewhere in the range of about $10,000, which is orders of magnitude cheaper than just a few years earlier. This progress was made possible by the development of an entirely new strategy of DNA sequencing that is commonly referred to as *next-generation sequencing* (or *NGS*). Unlike the sequencing steps used for the original Human Genome Project, the NGS technologies do not require that pieces of the genome be cloned in yeast or bacteria. Instead, fragments to be sequenced are prepared directly from the whole genome. Like the Sanger approach, NGS is also based on polymerase-dependent DNA synthesis, but does not depend on premature strand termination, nor does it require separation of strands by electrophoresis. Instead, it accomplishes the direct identification of individual nucleotides as they are being incorporated by the polymerase in real time. At the time of this writing three different instruments, each the product of a different manufacturer, are capable of carrying out a variation of this type of rapid automated DNA sequencing. In all of these cases, huge numbers (hundreds of thousands to tens of millions) of DNA molecules are immobilized on a surface and then incubated with a single sample of DNA polymerase in the presence of the four different NTPs, which are added one at a time in repetitive cycles. Various strategies are used to identify the nucleotides that are sequentially incorporated into each of the complementary strands as they are synthesized on the massive numbers of "parallel" DNA templates. The strands that are synthesized in these platforms are relatively short, but there are so many DNA strands being read simultaneously that the total number of nucleotides that is being identified in a single run is enormous.

At the time of this writing, two new instruments have been introduced into the market that utilize a revolutionary new technology, referred to as *third-generation* (or *single-molecule*) *sequencing*. Unlike all previous sequencing strategies, third-generation sequencers do not require amplified DNA, nor do they involve enzymatic activities. Instead, they are able to determine the nucleotide sequence of individual DNA molecules that are isolated directly from a genome. They accomplish sequencing by pulling each DNA molecule through a tiny hole, or "nanopore" and identifying each nucleotide, one at a time, as it passes through the opening. Nucleotide identification is based on differences in properties among the four nucleotides that can be detected as each nucleotide passes one after another through the nanopore. With large numbers of nanopores operating simultaneously, the potential for very rapid sequencing is possible. Whether or not nanopore DNA sequencing will prove to be as accurate as NGS sequencing remains to be established, but, because it doesn't rely on "sequencing by synthesis," it has the potential to be considerably less expensive.

Once the nucleotide sequence of a segment of DNA has been determined, various software tools can be employed to analyze it. For example, the amino acid sequence encoded by the DNA can be determined and compared to other known amino acid sequences to provide information about the polypeptide's possible function. The amino acid sequence also provides clues as to the tertiary structure of the protein, particularly those parts of the polypeptide that may act as membrane-spanning segments of integral membrane proteins. The nucleotide sequence itself can also be compared to other known nucleotide sequences. Such comparisons can be used to assess the evolutionary relatedness or history of the DNA sequences, to identify the DNA fragment just sequenced, or to compare the genomic features of various organisms or individuals.

18.16 DNA LIBRARIES

DNA cloning is often used to produce **DNA libraries**, which are collections of cloned DNA fragments. Two basic types of DNA libraries can be created: genomic libraries and cDNA libraries. *Genomic libraries* are produced from total DNA extracted from nuclei and contain all of the DNA sequences of the species. Once a genomic library of a species is available, researchers can use the collection to isolate specific DNA sequences, such as those containing the human insulin gene. *cDNA libraries*, on the other hand, are derived from DNA copies of an RNA population. cDNA libraries are typically produced from messenger RNAs present in a particular cell type and thus correspond to the genes that are active in that type of cell.

Genomic Libraries

We will first examine the production of a genomic library. In one approach, genomic DNA is treated with one or two restriction enzymes that recognize very short nucleotide sequences under conditions of low enzyme concentration, such that only a small percentage of susceptible sites are actually cleaved. Two commonly used enzymes that recognize tetranucleotide sequences are *Hae*III (recognizes GGCC) and *Sau*3A (recognizes GATC). A given tetranucleotide would be expected to occur by chance with such a high frequency that any sizable segment of DNA is sensitive to fragmentation. Once the DNA is partially digested, the digest is fractionated by gel electrophoresis or density gradient centrifugation, and those fragments of suitable size (e.g., 20 kb in length) are incorpo-

rated into lambda phage particles. These phage are used to generate the million or so plaques needed to ensure that every single segment of a mammalian genome is represented.

Because the DNA is treated with enzymes under conditions in which most susceptible sites are not being cleaved, for all practical purposes, the DNA is randomly fragmented. Once the phage recombinants are produced, they can be stored for later use and, as such, constitute a permanent collection of all the DNA sequences present in the genome of the species. Whenever an investigator wants to isolate a particular sequence from the library, the phage particles can be grown in bacteria and the various plaques (each originating from the infection of a single recombinant phage) screened for the presence of that sequence by in situ hybridization.

The use of randomly cleaved DNA has an advantage in the construction of a library because it generates overlapping fragments that can be used in the analysis of regions of the chromosome extending out in both directions from a particular sequence, a technique known as *chromosome walking*. For example, if one isolates a fragment containing the coding region of a globin gene, that particular fragment can be labeled and used as a probe to screen the genomic library for fragments with which it overlaps. The process is then repeated using the new fragments as labeled probes in successive screening steps, as one gradually isolates a longer and longer part of the original DNA molecule. Using this approach, one can study the organization of linked sequences in an extended region of a chromosome.

Cloning Larger DNA Fragments in Specialized Cloning Vectors

Neither plasmid nor lambda phage vectors are suitable for cloning DNAs larger than about 20 to 25 kb in length. Several vectors have been developed that allow investigators to clone much larger pieces of DNA. One of the most important of these vectors is the **yeast artificial chromosome (YAC)**, which can accept segments of foreign DNA as large as 1000 kb (one million base pairs). As the name implies, YACs are artificial versions of a normal yeast chromosome. They contain all of the elements of a yeast chromosome that are necessary for the structure to be replicated during S phase and segregated to daughter cells during mitosis, including one or more origins of replication, telomeres at the ends of the chromosomes, and a centromere to which the spindle fibers can attach during chromosome separation. In addition to these elements, YACs are constructed to contain (1) a gene whose encoded product allows those cells containing the YAC to be selected from those that lack the element and (2) the DNA fragment to be cloned. Like other cells, yeast cells can take up DNA from their medium, which provides the means by which YACs are introduced into the cells. Over the past few years, laboratories involved in sequencing genomes have relied heavily on an alternate cloning vector, called a **bacterial artificial chromosome (BAC)**, which is also able to accept large foreign DNA fragments (up to about 500 kb). BACs are specialized bacterial plasmids (F factors) that contain a bacterial origin of replication and the genes required to regulate their own replication. BACs have the advantage over YACs in high-speed sequencing projects because they can be cloned in

E. coli, which readily picks up exogenous DNA, has an extremely short generation time, can be grown at high density in simple media, and does not "corrupt" the cloned DNA through recombination.

DNA fragments cloned in YACs and BACs are typically greater than 100 kb in length. Fragments of such large size are usually produced by treatment of DNAs with restriction enzymes that recognize particularly long nucleotide sequences (7–8 nucleotides) containing CG dinucleotides. As noted on page 520, CG dinucleotides have special functions in the mammalian genome, and presumably because of this they do not appear nearly as often as would be predicted by chance. The restriction enzyme *Not*1, for example, which recognizes the 8-nucleotide sequence GCGGCCGC, typically cleaves mammalian DNA into fragments several hundred thousand base pairs long. These fragments can then be incorporated into YACs or BACs and cloned within host yeast or bacterial cells.

cDNA Libraries

Up to this point, the discussion has been restricted to cloning DNA fragments isolated from extracted DNA, that is, genomic fragments. When working with genomic DNA, one is generally seeking to isolate a particular gene or family of genes from among hundreds of thousands of unrelated sequences. In addition, the isolation of genomic fragments allows one to study a variety of topics, including (1) regulatory sequences flanking the coding portion of a gene; (2) noncoding intervening sequences; (3) various members of a multigene family, which often lie close together in the genome; (4) evolution of DNA sequences, including their duplication and rearrangement as seen in comparisons of the DNA of different species; and (5) interspersion of transposable genetic elements.

The cloning of cDNAs, as opposed to genomic DNAs, has been especially important in the analysis of gene expression and alternative splicing. To produce a cDNA library, a population of mRNAs is isolated and used as a template by reverse transcriptase to form a population of DNA–RNA hybrids, as shown in Figure 18.45a. The DNA–RNA hybrids are converted to a double-stranded cDNA population by nicking the RNA with RNase H and replacing it with DNA by DNA polymerase I. The double-stranded cDNA is then combined with the desired vector (in this case a plasmid) and cloned as shown in Figure 18.45b. Messenger RNA populations typically contain thousands of different messages, but individual species may be present in markedly different numbers (abundance). As a result, a cDNA library has to contain a million or so different cDNA clones to be certain that the rarer mRNAs will be represented. In addition, reverse transcriptase is not a very efficient enzyme and tends to fall off its template mRNA before the copying job is completed. As a result, it can be difficult to obtain a population of full-length cDNAs. As with experiments using genomic DNA fragments, clones must be screened to isolate one particular sequence from a heterogeneous population of recombinant molecules.

The analysis of cloned cDNAs serves several functions. It is generally easier to study a diverse population of cDNAs than the corresponding population of mRNAs, so one can use the

(a)

(b)
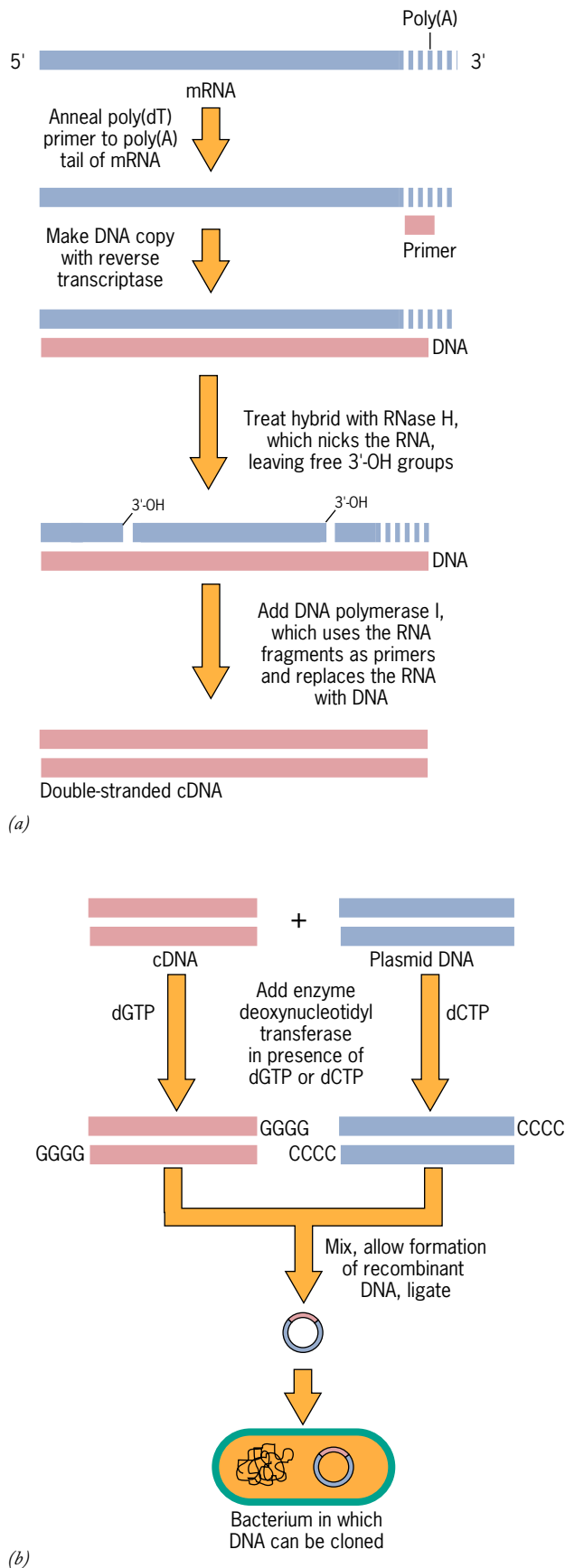
FIGURE 18.45 Synthesizing cDNAs for cloning in a plasmid. (*a*) In this method of cDNA formation, a short poly(dT) primer is bound to the poly(A) of each mRNA, and the mRNA is transcribed by reverse transcriptase (which requires the primer to initiate DNA synthesis). Once the DNA–RNA hybrid is formed, the RNA is nicked by treatment with RNase H, and DNA polymerase is added to digest the RNA and replace it with DNA, just as it does during DNA replication in a bacterial cell (page 545). (*b*) To prepare the blunt-ended cDNA for cloning, a short stretch of poly(G) is added to the 3′ ends of the cDNA and a complementary stretch of poly(C) is added to the 3′ ends of the plasmid DNA. The two DNAs are mixed and allowed to form recombinants, which are sealed and used to transform bacterial cells in which they are cloned.

cDNAs to learn about the diversity and abundance of mRNAs present in a cell, the percentage of mRNAs shared by two different types of cells, or the variety of alternatively spliced mRNAs generated from a specific primary transcript. A single cloned and amplified cDNA molecule is also very useful. The cDNA contains only that information present in the mRNA, so comparisons between a cDNA and its corresponding genomic locus can provide information on the precise locations of the noncoding intervening sequences within the DNA. In more practical endeavors, the purified cDNA can be readily sequenced to provide a shortcut in determining the amino acid sequence of a polypeptide; labeled cDNAs are used as probes to screen for complementary sequences among recombinant clones; and because they lack introns, cDNAs have an advantage over genomic fragments when one is attempting to synthesize eukaryotic proteins in bacterial cell cultures.

18.17 DNA TRANSFER INTO EUKARYOTIC CELLS AND MAMMALIAN EMBRYOS

We have discussed in previous sections how eukaryotic genes can be isolated, modified, and amplified. In this section we will consider some of the ways in which genes can be introduced into eukaryotic cells, where they are generally transcribed and translated. One of the most widely used strategies to accomplish this goal is to incorporate the DNA into the genome of a nonreplicating virus and allow the virus to infect the cell. Viral-mediated gene transfer is called **transduction**. Depending on the type of virus used, the gene of interest can be expressed transiently for a period of hours to days, or it can be stably integrated into the genome of the host cell. Stable integration is usually accomplished using modified retroviruses, which contain an RNA genome that is reverse transcribed into DNA inside the cell. The DNA copy is then inserted into the DNA of the host chromosomes. Retroviruses have been used in many of the recent attempts at gene therapy to transfer a gene into the cells of a patient who is lacking that gene. Generally, these clinical trials have not had much success due to the low efficiency of infection of current viral vectors.

A number of procedures are available to introduce naked DNA into cultured cells, a process called **transfection**. Most often the cells are treated with either calcium phosphate or

DEAE-dextran, both of which form a complex with the added DNA that promotes its adherence to the cell surface. It is estimated that only about 1 in 10^5 cells takes up the DNA and incorporates it stably into the chromosomes. It is not known why this small percentage of cells in the population are competent to be transfected, but those that are transfected generally pick up several fragments. One way to select for those cells that have taken up the foreign DNA is to include a gene that allows the transfected cells to grow in a particular medium in which nontransfected cells cannot survive. Because transfected cells typically take up more than one fragment, the gene used for selection need not be located on the same DNA fragment as the gene whose role is being investigated (called the **transgene**).

Two other procedures to transfect cells are electroporation and lipofection. In *electroporation*, the cells are incubated with DNA in special vials that contain electrodes that deliver a brief electric pulse. Application of the current causes the plasma membrane to become transiently permeable to DNA molecules, some of which find their way into the nucleus and become integrated into the chromosomes. In *lipofection*, cells are treated with DNA that is bound to positively charged lipids (cationic liposomes) that are capable of fusing with the lipid bilayer of the cell membrane and delivering the DNA to the cytoplasm.

One of the most direct ways to introduce foreign genes into a cell is to microinject DNA directly into the cell's nucleus. The nuclei of oocytes and eggs are particularly well suited for this approach. *Xenopus* oocytes, for example, have long been used to study the expression of foreign genes. The nucleus of the oocyte contains all the machinery necessary for RNA synthesis; when foreign DNA is injected into the nucleus, it is readily transcribed. In addition, the RNAs synthesized from the injected templates are processed normally and transported to the cytoplasm, where they are translated into proteins that can be detected immunologically or by virtue of their specific activity.

Another favorite target for injected DNA is the nucleus of a mouse embryo (Figure 18.46). The goal in such experiments is not to monitor the expression of the gene in the injected cell, but to have the foreign DNA become integrated into the egg's chromosomes to be passed on to all the cells of the embryo and subsequent adult. Animals that have been genetically engineered so that their chromosomes contain foreign genes are called **transgenic animals**, and they provide a means to monitor when and where in the embryo particular genes are expressed (as in Figure 12.33), and to determine the impact of extra copies of particular genes on the development and life of the animal.

Transgenic Animals and Plants   In 1981, Ralph Brinster of the University of Pennsylvania and Richard Palmiter of the University of Washington succeeded in introducing a gene for rat growth hormone (GH) into the fertilized eggs of mice. The injected DNA was constructed so as to contain the coding portion of the rat GH gene in a position just downstream from the promoter region of the mouse metallothionein gene. Under normal conditions, the synthesis of metallothionein

FIGURE 18.46 Microinjection of DNA into the nucleus of a recently fertilized mouse egg. The egg is held in place by a suction pipette shown at the right, while the injection pipette is shown penetrating the egg at the left. (FROM THOMAS E. WAGNER ET AL., PROC. NAT'L. ACAD. SCI. U.S.A. 78:6377, 1981.)

is greatly enhanced following the administration of metals, such as cadmium or zinc, or glucocorticoid hormones. The metallothionein gene carries a strong promoter, and it was hoped that placing the GH gene downstream from it might allow the GH gene to be expressed following treatment of the transgenic mice with metals or glucocorticoids. As illustrated in Figure 18.47, this expectation was fully warranted.

FIGURE 18.47 Transgenic mice. This photograph shows a pair of littermates at age 10 weeks. The larger mouse developed from an egg that had been injected with DNA containing the rat growth hormone gene placed downstream from a metallothionein promoter. The larger mouse weighs 44 g; the smaller, uninjected control weighs 29 g. The rat GH gene was transmitted to offspring that also grew larger than the controls. (COURTESY OF RALPH L. BRINSTER, FROM WORK REPORTED IN R. D. PALMITER ET AL., NATURE 300:611, 1982.)

On the practical side, transgenic animals provide a mechanism for creating **animal models**, which are laboratory animals that exhibit a particular human disease they would normally not be subject to. This approach is illustrated by the transgenic mice that carry a gene encoding a mutant version of the human amyloid precursor protein (APP). As discussed on page 66, these mice develop neurological and behavioral symptoms reminiscent of Alzheimer's disease and are an important resource in the development of therapies for this terrible disease. Transgenic animals are also being developed as part of agricultural biotechnology. For example, pigs born with foreign growth hormone genes incorporated into their chromosomes grow much leaner than control animals lacking the genes. The meat of the transgenic animals is leaner because the excess growth hormone stimulates the conversion of nutrients into protein rather than fat.

Plants are also prime candidates for genetic engineering. It was indicated on page 503 that whole plants can be grown from individual cultured plant cells. This provides the opportunity to alter the genetic composition of plants by introducing DNA into the chromosomes of cultured cells that can subsequently be grown into mature plants. One way to introduce foreign genes is to incorporate them into the *Ti plasmid* of the bacterium *Agrobacterium tumefaciens*. Outside of the laboratory, this bacterium lives in association with dicotyledonous plants, where it causes the formation of tumorous lumps called crown galls. During an infection, a section of the Ti plasmid termed the *T-DNA region* is passed from the bacterium into the plant cell. This part of the plasmid becomes incorporated into the plant cell's chromosomes and induces the cell to proliferate and provide nutrients for the bacterium.

The Ti plasmid can be isolated from bacteria and linked with foreign genes to produce a recombinant plasmid that is taken up in culture by undifferentiated dicotyledonous plant cells, including those of carrots and tobacco. Figure 18.48 shows the use of recombinant Ti plasmids to introduce foreign DNA into meristem cells at the tip of newly formed shoots. The shoots can then be rooted and grown into mature plants. This technique, which is called *T-DNA transformation*, has been used to transform plant cells with genes derived from bacteria that code for insect-killing toxins, thus protecting the plants from insect predators.

Other procedures have been developed to introduce foreign genes into cells of monocotyledonous plants. In one of the more exotic approaches, plant cells can serve as targets for microscopic, DNA-coated tungsten pellets fired by a "gene gun." This technique has been used to genetically alter a number of different types of plant cells.

The two most important activities that plant geneticists might improve by genetic engineering are photosynthesis and nitrogen fixation, both crucial bioenergetic functions. Any significant improvement in photosynthetic efficiency would mean great increases in crop production. It is hoped that a modified version of the CO_2-fixing enzyme might be engineered that is less susceptible to photorespiration (page 224). Nitrogen fixation is an activity carried out by certain genera of
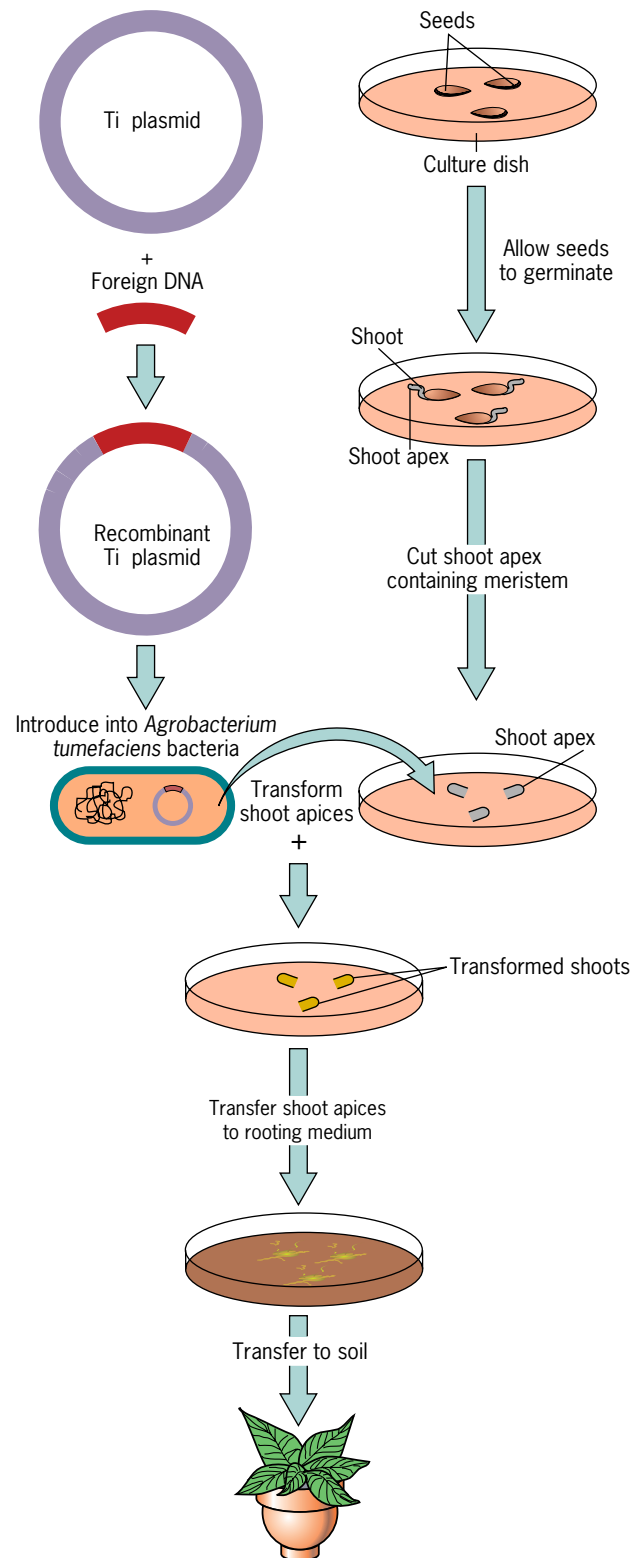
FIGURE 18.48 Formation of transgenic plants using the Ti plasmid. The transgene is spliced into the DNA of the Ti plasmid, which is reintroduced into host bacteria. Bacteria containing the recombinant plasmid are then used to transform plant cells, in this case cells of the meristem at the tip of a dissected shoot apex. The transformed shoots are transferred to a selection medium where they develop roots. The rooted plants can then be transferred to potting soil.

bacteria (e.g., *Rhizobium*) that live in a symbiotic relationship with certain plants (e.g., soybean, peanut, alfalfa, and pea). The bacteria reside in swellings, or *leguminous nodules*, located on the roots, where they remove N_2 from the atmosphere, reduce it to ammonia, and deliver the product to the cells of the plant. Geneticists are looking for a way to isolate genes involved in this activity from the bacterial genome and introduce them into the chromosomes of nonleguminous plants that presently depend heavily on added fertilizer for their reduced nitrogenous compounds. Alternatively, it might be possible to alter the genome of either plant or bacteria so that new types of symbiotic relationships can be developed.

18.18 DETERMINING EUKARYOTIC GENE FUNCTION BY GENE ELIMINATION OR SILENCING

Until fairly recently, investigators discovered new genes and learned of their function by screening for mutants that exhibited abnormal phenotypes (see page 271 on the study of protein secretion). It was only through the process of random mutation that the existence of genes became apparent. This process of learning about genotypes by studying mutant phenotypes is known as *forward genetics*. Since the development of techniques for gene cloning and DNA sequencing, researchers have been able to identify and study genes directly without knowing anything about the function of their encoded protein. This condition has become especially familiar in recent years with the sequencing of entire genomes and the identification of thousands of genes whose function remains unknown. Over the past two decades, researchers have developed means to carry out *reverse genetics*, which is a process of determining phenotype (i.e., function) based on the knowledge of genotype. The basic approach of reverse genetics is to eliminate the function of a specific gene, and then determine what effect the elimination of that function has on phenotype. We will first consider how researchers introduce specific mutations into genes in vitro, then discuss two widely used techniques for eliminating gene function in vivo.

In Vitro Mutagenesis

As is evident throughout this book, the isolation of naturally occurring mutants has played an enormously important role in determining the function of genes and their products. But natural mutations are rare events, and it isn't feasible to use such mutations to study the role of particular amino acid residues in the function of a particular protein. Rather than waiting for an organism to appear with an unusual phenotype and identifying the responsible mutation, researchers can mutate the gene (or its associated regulatory regions) in a desired way and observe the resulting phenotypic change. These techniques, collectively termed *in vitro mutagenesis*, require that the gene, or at least the gene segment, to be mutated has been cloned.

A procedure developed by Michael Smith of the University of British Columbia called **site-directed mutagenesis (SDM)** allows researchers to make very small, specific changes in a DNA sequence, such as the substitution of one base for another, or the deletion or insertion of a very small number of bases. SDM is usually accomplished by first synthesizing a DNA oligonucleotide containing the desired change, allowing this oligonucleotide to hybridize to a single-stranded preparation of the normal DNA, and then using the oligonucleotide as a primer for DNA polymerase. The polymerase elongates the primer by adding nucleotides that are complementary to the normal DNA. The modified DNA can then be cloned and the effect of the genetic alteration determined by introducing the DNA into an appropriate host cell. Scientists typically use SDM to ask very targeted questions about the function of a gene or protein. For example, they might change one amino acid into another to gain insight into the role of that specific site in the overall function of a protein. Alternatively, they might introduce small changes in the regulatory region of a gene and determine the effect on gene expression. If the intent of site-directed mutagenesis is simply to eliminate the function of a gene, less specific methods can be used. For example, cutting a gene sequence at a restriction site, using DNA polymerase to make the single-stranded regions of the sticky ends into double-stranded DNA, and then ligating those ends back together can destroy the reading frame of a protein. In other instances, an entire restriction fragment might be removed from the gene.

Constructing a mutation in vitro is only one aspect of reverse genetics. To study the effect of an engineered mutation on phenotype, it is necessary to substitute the mutant allele for the normal gene in the organism in question. The development of a technique for introducing mutations into the genome of the mouse opened the door for reverse genetic studies in mammals, and literally revolutionized the study of mammalian gene function.

Knockout Mice

We have described in many places in this text the phenotypes of mice that lack a functional copy of a particular gene. For example, it was noted that mice lacking a functional copy of the *p53* gene invariably develop malignant tumors (page 662). These animals, which are called **knockout mice**, can provide a unique insight into the genetic basis of a human disease, as well as a basis for studying the various cellular activities in which the product of a particular gene might be engaged. Knockout mice are born as the result of a series of experimental procedures depicted in Figure 18.49.

The various procedures used to generate knockout mice were developed in the 1980s by Mario Capecchi at the University of Utah, Oliver Smithies at the University of Wisconsin, and Martin Evans at Cambridge University. The first step is to isolate an unusual type of cell that has virtually unlimited powers of differentiation. These cells, called **embryonic stem cells** (page 19), are found in the mammalian blastocyst, which is an early stage of embryonic development comparable to the blastula stage of other animals. A mammalian blastocyst (Figure 18.49) is composed of two distinct parts. The outer layer makes up the *trophectoderm*, which
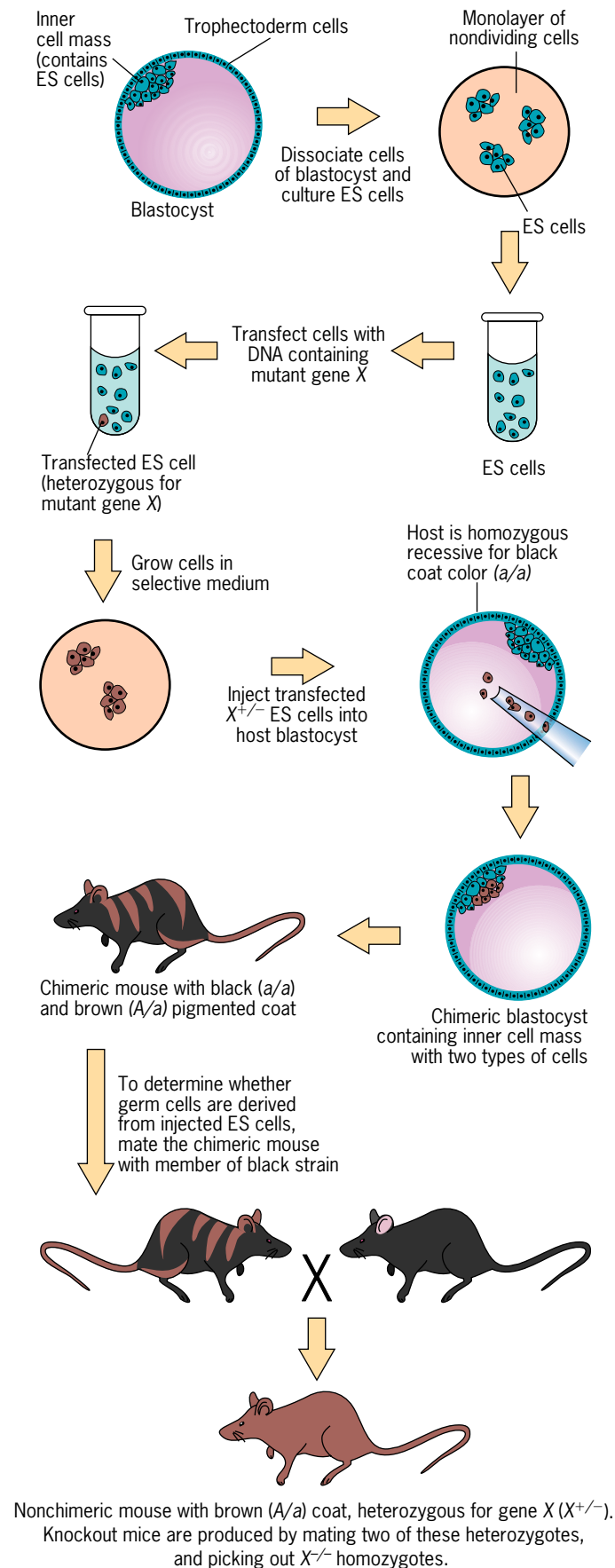
FIGURE 18.49   Formation of knockout mice. The steps are described in the text.

gives rise to most of the extraembryonic membranes characteristic of a mammalian embryo. The inner surface of the trophectoderm contacts a cluster of cells called the *inner cell mass* that projects into the spacious cavity (the blastocoel). The inner cell mass gives rise to the cells that make up the embryo. The inner cell mass contains the embryonic stem (ES) cells, which differentiate into all of the various tissues of which a mammal is composed.

ES cells can be isolated from blastocysts and cultured in vitro under conditions where the cells grow and proliferate. The ES cells are then transfected with a DNA fragment containing a nonfunctional, mutant allele of the gene to be knocked out, as well as antibiotic-resistance genes that can be used to select for cells that have incorporated the altered DNA into their genome. Of those cells that take up the DNA, approximately one in 10^4 undergo a process of *homologous recombination* in which the transfected DNA replaces the homologous DNA sequence. Through use of this procedure, ES cells that are heterozygous for the gene in question are produced and then selected for on the basis of their antibiotic resistance. In the next step, a number of these donor ES cells are injected into the blastocoel of a recipient mouse embryo. In the protocol depicted in Figure 18.49, the recipient embryo is obtained from a black strain. The injected embryo is then implanted into the oviduct of a female mouse that has been hormonally prepared to carry the embryo to term. As the embryo develops in its surrogate mother, the injected ES cells join the embryo's own inner cell mass and contribute to formation of embryonic tissues, including the germ cells of the gonads. These chimeric mice can be recognized because their coat contains characteristics of both the donor and recipient strains. To determine whether the germ cells also contain the knockout mutation, the chimeric mice are mated to a member of an inbred black strain. If the germ cells do contain the knockout mutation, the offspring will be heterozygous for the gene in all of their cells. Heterozygotes can be distinguished by their brown coat coloration. These heterozygotes are then mated to one another to produce offspring that are homozygous for the mutant allele. These are the knockout mice that lack a functional copy of the gene. Any gene within the genome, or any DNA sequence for that matter, can be altered in any desired manner using this experimental approach.

In some cases, the deletion of a particular gene can lead to the absence of a particular process, which provides convincing evidence that the gene is essential for that process. Often, however, the deletion of a gene that is thought to participate in an essential process causes little or no alteration in the animal's phenotype. Such results can be difficult to interpret. It is possible, for example, that the gene is not involved in the process being studied or, as is usually the case, the absence of the gene product is compensated by the product of an entirely different gene. Compensation by one gene for another can be verified by producing mice that lack both of the genes in question (i.e., a double knockout).

In other cases, the absence of a gene leads to the death of the mouse during early development, which also makes it difficult to determine the role of the gene in cellular function.

Researchers are often able to get around this problem by using a technique that allows a particular gene to be switched off in only one or several desired tissues, while the gene is expressed in the remainder of the animal. These *conditional knockouts*, as they are called, allow researchers to study the role of the gene on the development or function of the affected tissue.

RNA Interference

Knockout mice have been an invaluable strategy to learn about gene function, but the generation of these animals is laborious and costly. In the past few years a new technique in the realm of reverse genetics has come into widespread use. As discussed on page 449, RNA interference (RNAi) is a process in which a specific mRNA is degraded due to the presence of a small double-stranded siRNA whose sequence is contained within the mRNA sequence. The functions of plant, nematode, or fruit fly genes can be studied by simply introducing an siRNA into one of these organisms by various means and examining the phenotype of the organism that results from depletion of the corresponding mRNA. Determining the roles of genes at the cellular level, rather than the organismal level, can be accomplished by studying the effects of these same siRNAs on the activities of cultured cells. Using these approaches, researchers can gather information about the functions of large numbers of genes in a relatively short period of time. As discussed on page 451, RNAi can be used to study gene function in mammalian cells by incubating the cells with small dsRNAs encapsulated in lipids or by genetically engineering the cells to produce the siRNAs themselves. Once inside the cell, the siRNA guides the degradation of the target mRNA, leaving the cell unable to produce additional protein encoded by that gene. Any deficiencies in the phenotype of the cell can be attributed to a marked reduction in the level of the protein being investigated. As in the case of gene knockouts, the absence of a phenotype following treatment with an siRNA does not necessarily mean that the gene in question is not involved in a particular process, as another gene may be capable of compensating for the absence of expression of the gene being targeted.

Libraries containing thousands of siRNAs, or vectors containing DNA encoding these RNAs, are also available for the study of gene function in a number of model organisms and in humans. Researchers using these libraries can study the effects of the elimination of a gene's expression on virtually any cellular process. It should be possible, for example, to determine which genes are involved in mitotic spindle assembly by treating cells with a library of siRNAs from across the entire genome and then screening all of the treated cells at the time of cell division for the presence of an abnormal mitotic spindle. Figure 18.50 shows a gallery of images of cultured *Drosophila* cells in the metaphase stage of mitosis. Each of the cells depicted in this gallery had been treated with an siRNA directed against a different gene in the *Drosophila* genome. Over 14,000 different siRNAs were tested in this study, of which approximately 200 caused the treated cells to have an abnormal metaphase spindle. More than half of the siRNAs that affected mitotic spindle assembly targeted genes that were not previ-
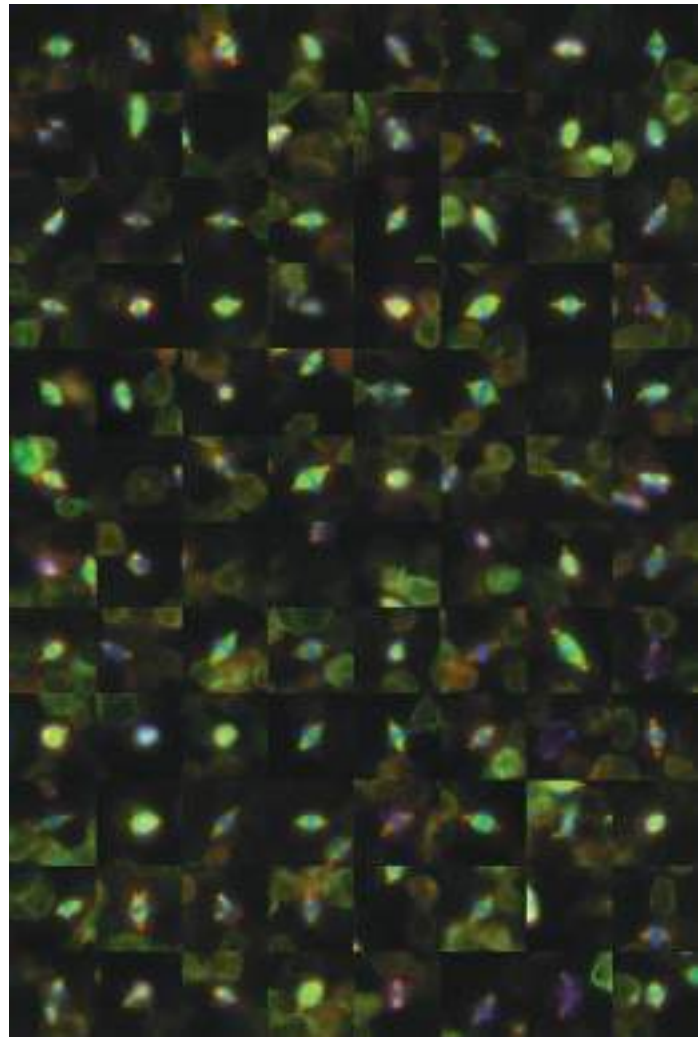
FIGURE 18.50 Determining gene function by RNA interference. The figure shows a gallery of images of cultured *Drosophila* cells that had been incubated with various double-stranded siRNAs. The cells shown here had accumulated in metaphase of mitosis as the result of a separate treatment that blocked progression into anaphase. During the course of this study, over 4 million cells were examined. It was possible to screen such a large number of cells by having a computer select the images of cells that were in metaphase and then crop and assemble the images into panels as shown in this photograph. The images could then be screened for the presence of abnormal mitotic spindles either by human observers or by computer analysis using programs designed to detect specific features of an abnormal spindle. (COURTESY OF GOHTA GOSHIMA AND RONALD D. VALE FROM A STUDY IN SCIENCE 316:417, 2007.)

ously known to be involved in the formation of this structure, thus providing new insights into the functions of these genes. The images in Figure 18.51 show the effects of a single siRNA that targets a gene known to be involved in mitotic spindle assembly and other activities during mitosis. In this case, the cultured mammalian cell seen on the right had been transfected with an siRNA that targets mRNAs encoding the Aurora B kinase (page 584). The resulting knockdown of this enzyme has seriously disturbed chromosome segregation.
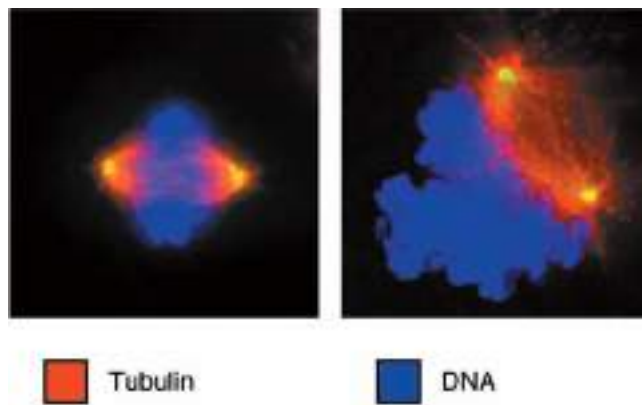
FIGURE 18.51 Determining gene function by RNA interference. (Left) An untreated, cultured mammalian cell at metaphase of mitosis. (Right) The same type of cell that has been treated with a 21-nucleotide dsRNA designed to induce the destruction of mRNAs encoding Aurora B kinase, a protein involved in the metaphase spindle checkpoint. The chromosomes in this cell lie adjacent to the mitotic spindle, suggesting the absence of kinetochore–microtubule interactions. (FROM CLAIRE DITCHFIELD ET AL., COURTESY OF STEPHEN S. TAYLOR, J. CELL BIOL. 161:276, 2003; BY COPYRIGHT PERMISSION OF THE ROCKEFELLER UNIVERSITY PRESS.)

18.19 THE USE OF ANTIBODIES

As discussed in Chapter 17, antibodies (or immunoglobulins) are proteins produced by lymphoid tissues in response to the presence of foreign materials, or antigens. One of the most striking properties of antibodies, and one that makes them so useful to biological researchers, is their remarkable specificity. A cell may contain thousands of different proteins, yet a preparation of antibodies binds only to those select molecules within the cell having a small part that fits into the antigen-binding sites of the antibody molecules. Antibodies can often be obtained that distinguish between two polypeptides that differ by only a single amino acid or by a single posttranslational modification (as on page 516).

Biologists have long taken advantage of antibodies and have developed a wide variety of techniques involving their use. There are basically two different approaches to the preparation of antibody molecules that interact with a given antigen. In the traditional approach, an animal (typically a rabbit or goat) is repeatedly injected with the antigen, and after a period of several weeks, blood is drawn that contains the desired antibodies. Whole blood is treated to remove cells and clotting factors so as to produce an **antiserum**, which can be tested for its antibody titer and from which the immunoglobulins can be purified. Although this method of antibody production remains in use, it suffers from certain inherent disadvantages. Because of the mechanism of antibody synthesis, an animal invariably produces a variety of different species of immunoglobulin, that is, immunoglobulins with different V regions in their polypeptide chains, even when challenged with a highly purified preparation of antigen. An antiserum that contains a variety of immunoglobulins that bind to the same antigen is said to be *polyclonal*. Because immunoglobu-

lins are too similar in structure to be fractionated, it becomes impossible to obtain a preparation of a single purified species of antibody by this technique.

In 1975, Cesar Milstein and Georges Köhler of the Medical Research Council in Cambridge, England, carried out a far-reaching set of experiments that led to the preparation of antibodies directed against specific antigens. To understand their work, it is necessary to digress briefly. A given clone of antibody-producing cells (which is derived from a single B lymphocyte) synthesizes antibodies having identical antigen-combining sites. The heterogeneity of the antibodies produced when a single purified antigen is injected into an animal results from the fact that many different B lymphocytes are activated, each having membrane-bound antibodies with an affinity for a different part of the antigen. An important question arose: was it possible to circumvent this problem and obtain a single species of antibody molecule? Consider for a moment the results of a procedure in which one injects an animal with a purified antigen, waits a period of weeks for antibodies to be produced, removes the spleen or other lymphoid organs, prepares a suspension of single cells, isolates those cells producing the desired antibody, and grows these particular cells as separate colonies so as to obtain large quantities of this particular immunoglobulin. If this procedure were followed, one should obtain a preparation of antibody molecules produced by a single colony (or clone) of cells, which is referred to as a **monoclonal antibody**. But antibody-producing cells do not grow and divide in culture, so an additional manipulation had to be introduced to obtain monoclonal antibodies.

Malignant myeloma cells are a type of cancer cell that grows rapidly in culture and produces large quantities of antibody. But myeloma cells are of little use as analytic tools because they are not formed in response to a particular antigen. Instead, myeloma cells develop from the random conversion of a normal lymphocyte to the malignant state, and therefore, they produce the antibody that was being synthesized by the particular lymphocyte before it became malignant.

Milstein and Köhler combined the properties of these two types of cells, the normal antibody-producing lymphocyte and the immortal myeloma cell. They accomplished this feat by fusing the two types of cells to produce hybrid cells, called **hybridomas**, that grow and proliferate indefinitely, producing large amounts of a single (monoclonal) antibody. The antibody produced is one that was being synthesized by the normal lymphocyte prior to fusion with the myeloma cell.

The procedure for the production of monoclonal antibodies is illustrated in Figure 18.52. The antigen (whether present in a soluble form or as part of a cell) is injected into the mouse to cause the proliferation of specific antibody-forming cells (step 1, Figure 18.52). After a period of several weeks, the spleen is removed and dissociated into single cells (step 2), and the antibody-producing lymphocytes are fused with a population of malignant myeloma cells (step 3), which makes the hybrid cells immortal, that is, capable of unlimited cell division. Hybrid cells (hybridomas) are selected from nonfused cells by placing the cell mixture in a medium in which only the hybrids can survive (step 4). Hybridomas are then grown clonally in

① Mouse being immunized
with antigen

Myeloma cells in culture.
Cells are HGPRT-mutants
and unable to grow in
HAT medium.

Remove spleen

② Spleen-cell
suspension

③ Fusion of cells in
polyethylene glycol

Add cells to HAT
selection medium

④ Hybrid cells (hybridomas)
grow while unfused cells die

⑤ Test supernatant in which cells are
growing for presence of desired
antibody. Then culture cells that
produce this antibody.

⑥ Culture clone of cells that produces
antibody against injected antigen.

FIGURE 18.52 **Formation of monoclonal antibodies.** The steps are described in the text. The HAT medium is so named because it contains hypoxanthine, aminopterin, and thymidine. This medium allows cells with a functional hypoxanthine–guanine phosphoribosyl transferase (HGPRT) to grow, but it does not support the growth of cells lacking this enzyme, such as the unfused myeloma cells used in this procedure.

separate wells (step 5) and individually screened for production of antibody against the antigen being studied. Hybrid cells containing the appropriate antibody (step 6) can then be cloned in vitro or in vivo (by growing as tumor cells in a recip-

ient animal), and unlimited amounts of the monoclonal antibody can be prepared. Once the hybridomas have been produced, they can be stored indefinitely in a frozen state, and aliquots can be made available to researchers around the world.

One of the most important features of this methodology is that one does not have to begin with a purified antigen to obtain an antibody. In fact, the antigen to which the monoclonal antibody is sought may even be a minor component of the entire mixture. In addition to their use in research, monoclonal antibodies are playing a valuable role in diagnostic medicine in tests to determine the concentration of specific proteins in the blood or urine. In one example, monoclonal antibodies form the basis of certain home-pregnancy tests that monitor the presence of a protein (chorionic gonadotrophin) that appears in the urine a few days after conception.

Once a preparation of antibody molecules is in hand, whether obtained by conventional immunologic techniques or by formation of hybridomas, these molecules can be used as highly specific probes in a variety of analytic techniques. For example, antibodies can be used in protein purification. When a purified antibody is added to a crude mixture of proteins, the specific protein being sought selectively combines with the antibody and precipitates from solution. Antibodies can also be used in conjunction with various types of fractionation procedures to identify a particular protein (antigen) among a mixture of proteins. In a *Western blot*, for example, a mixture of proteins is first fractionated by two-dimensional gel electrophoresis (as in Figure 18.29). The fractionated proteins are then transferred to a sheet of nitrocellulose filter, which is then incubated with a preparation of antibodies that have been labeled either radioactively or fluorescently. The location on the filter of the specific protein bound by the antibody can be determined from the location of the bound radioactivity or fluorescence.

Monoclonal antibodies prepared by the method just described have also become very useful as therapeutic agents in treating human disease (page 672). To date, efforts to develop human hybridomas that produce human monoclonal antibodies have not been successful. To get around this stumbling block, mice have been genetically engineered so that the antibodies they produce are increasingly human in amino acid sequence. A number of these humanized monoclonal antibodies have been approved for the treatment of several diseases. More recently, mice have been engineered so that their immune system is essentially "human" in nature. These animals produce antibodies that are fully human in structure.

The first fully human antibody (Humira), which is approved for the treatment of rheumatoid arthritis (page 708), was produced by a very different technique that uses bacteriophage rather than hybridomas as the basis for monoclonal antibody production. The technique is known as *phage display*. In this technique, billions of different phage particles are generated in which each phage carries a gene encoding a human antibody molecule that possesses a unique variable region (page 696). Different phage within this vast library encode different antibodies, that is, antibodies with different variable regions. In each case, the antibody gene is fused to a gene encoding one of the viral coat proteins so that when the phage is assem-

bled within a host cell, the antibody molecule is displayed on the surface of the viral particle. Suppose that you have a protein (antigen) that you believe would be a good target for a particular therapeutic antibody. The antigen is purified and allowed to interact with a sample of each of the billions of phage particles that make up the phage library. Those phage that bind to the antigen with high affinity can be identified and allowed to multiply with an appropriate host cell. Once it has been amplified this way, the DNA encoding the antibody gene can be isolated and used to transfect an appropriate mammalian cell. The genetically engineered cells can then be grown in large cultures to produce therapeutic quantities of the antibody. Production of antibodies in cultured mammalian cells is an expensive venture and alternative "living factories" are being pursued. Among the alternatives being considered for this purpose are goats, rabbits, and tobacco cells.

A quick scan through this text will reveal numerous micrographs depicting the immunolocalization of a particular protein within a cell as visualized in the light or electron microscope. Immunolocalization of proteins within a cell depends on the use of antibodies that have been specifically prepared against that particular protein. Once prepared, the antibody molecules are linked (conjugated) to a substance that makes them visible under the microscope, but does not interfere with the specificity of their interactions. For use with the light microscope, antibodies are generally complexed with small fluorescent molecules, such as fluorescein or rhodamine, to generate derivatives that are then incubated with the cells or sections of cells. The binding sites are then visualized with the fluorescence microscope. This technique is called **direct immunofluorescence**. It is often preferable to carry out antigen localization by a variation of this technique called **indirect immunofluorescence**. In indirect immunofluorescence, the cells are incubated with *unlabeled* antibody, which is allowed to form a complex with the corresponding antigen. The location of the antigen–antibody couple is then revealed in a second step using a preparation of fluorescently labeled antibodies whose combining sites are directed against the *antibody* molecules used in the first step. Indirect immunofluorescence results in a brighter image because numerous secondary antibody molecules can bind to a single primary antibody. Indirect immunofluorescence also has a practical advantage; the conjugated (fluorescent) antibody is readily purchased from vendors. Immunofluorescence provides remarkable clarity because only the proteins bound by the antibody are revealed to the eye; all of the unlabeled materials remain invisible. Localization of antigens in the electron microscope is accomplished using antibodies that have been tagged with electron-dense materials, such as the iron-containing protein ferritin or gold particles. An example of this technique is shown in Figure 8.23*c,d*.

Glossary

Many key terms and concepts are defined here. The parenthetical numbers following most definitions identify the chapter and section in which the term is first defined. For example, a term followed by (3.2) is first defined in Chapter 3, Section 2: "Enzymes." Terms defined in *The Human Perspective* or *Experimental Pathways* essays are identified parenthetically with HP or EP. For example, a term followed by (EP1) is first defined in the *Experimental Pathways* essay of Chapter 1.

A (aminoacyl) site Site into which aminoacyl-tRNAs enter the ribosome-mRNA complex. (11.8)

Absorption spectrum A plot of the intensity of light absorbed relative to its wavelength. (6.3)

Acetyl CoA A metabolic intermediate produced through catabolism of many compounds, including fatty acids, and used as the initial substrate for the central respiratory pathway, the TCA cycle. (5.2)

Acid A molecule that is capable of releasing a hydrogen ion. (2.3)

Acid hydrolases Hydrolytic enzymes with optimal activity at an acid pH. (8.6)

Actin A globular, cytoskeletal protein that polymerizes to form a flexible, helical filament capable of interacting with myosin. Actin filaments provide mechanical support for eukaryotic cells, determine the cell's shape, and enable cell movements. (9.5)

Actin-binding proteins Any of nearly 100 different proteins belonging to numerous families that affect the assembly of actin filaments, their physical properties, and their interactions with one another and with other cellular organelles. (9.7)

Action potential The collective changes in membrane potential, beginning with depolarization to threshold and ending with return to resting potential, that occur with stimulation of an excitable cell and act as the basis for neural communication. (4.8)

Action spectrum A plot of the relative rate (or efficiency) of a process produced by light of various wavelengths. (6.3)

Activation energy The minimal kinetic energy needed for a reactant to undergo a chemical reaction. (3.2)

Active site The part of an enzyme molecule that is directly involved in binding the substrate. (3.2)

Active transport The energy-requiring process in which a substance binds to a specific transmembrane protein, changing its conformation to allow passage of the substance through the membrane against the electrochemical gradient for that substance. (4.7)

Adaptive immune response A specific response to a pathogen that requires previous exposure to that agent. Includes responses mediated by antibodies and T lymphocytes. (17)

Adenosine triphosphate (ATP) Nucleotide consisting of adenosine bonded to three phosphate groups; it is the principal immediate-energy source for prokaryotic and eukaryotic cells. (2.5, 3)

Adherens junctions (zonulae adherens) Adherens junctions are a type of specialized adhesive junction particularly common in epithelia. The plasma membranes in this region are separated by 20 to 35 nm and are sites where cadherin molecules are concentrated. The cells are held together by linkages between the extracellular domains of cadherin molecules that bridge the gap between neighboring cells. (7.3)

Aerobes Organisms dependent on the presence of oxygen to metabolize energy-rich compounds. (5.1)

Affinity chromatography Protein purification technique that utilizes the unique structural properties of a protein that allow the molecule to be specifically withdrawn from solution while all other molecules remain in solution. The solution is passed through a column in which a specific interacting molecule (such as a substrate, ligand, or antigen) is immobilized by linkage to an inert material (the matrix). (18.7)

Alleles Alternate forms of the same gene. (10.1)

Allosteric modulation Modification of the activity of an enzyme through interaction with a compound that binds to a site (i.e., allosteric site) other than the active site. (3.3)

Alpha (α) helix One possible secondary structure of polypeptides, in which the backbone of the chain forms a spiral (i.e., helical) conformation. (2.5)

Alternative splicing Widespread mechanism by which a single gene can encode two or more related proteins. (12.5)

Amide bond The chemical bond that forms between carboxylic acids and amines (or acidic and amino functional groups) while producing a molecule of water. (2.4)

Amino acids The monomeric units of proteins; each is composed of three functional groups attached to a central, α carbon: an amino group, a defining side chain, and a carboxyl group. (2.5)

Aminoacyl-tRNA synthetase (AARS) An enzyme that covalently links amino acids to the 3′ ends of their cognate tRNA(s). Each amino acid is recognized by a specific aminoacyl-tRNA synthetase. (11.7)

Amphipathic The biologically important property of a molecule having both hydrophobic and hydrophilic regions. (2.5, 4.3)

Amphoteric Structural property allowing the same molecule to act as an acid or as a base. (2.3)

Anabolic pathway A metabolic pathway resulting in the synthesis of relatively complex products. (3.3)

Anaerobes Organisms that utilize energy-rich compounds through oxygen-independent metabolic pathways such as glycolysis and fermentation. (5.1)

Anaphase The stage of mitosis during which sister chromatids separate from one another. (14.2)

Anaphase A The movement of the chromosomes toward the poles during mitosis. (14.2)

Anaphase B Elongation of the mitotic spindle, which causes the two spindle poles to move farther apart. (14.2)

Aneuploidy A condition in which a cell has an abnormal number of chromosomes that is not a multiple of the haploid number. (16.1)

Angiogenesis The formation of new blood vessels. (16.4)

Angstrom The unit, equivalent to 0.1 nm, used to describe atomic and molecular dimensions. (1.3)

Animal models Laboratory animals that exhibit characteristics of a particular human disease. (HP2)

Anion An ionized atom or molecule with a net negative charge. (2.1)

Antenna The light harvesting molecules of a photosynthetic unit that absorbs photons of varying wavelengths and transfers the energy to the pigment molecules at the reaction center. (6.4)

Antibody An immunoglobulin protein produced by plasma cells derived from B lymphocytes that interacts with the surface of a pathogen or foreign substance to facilitate its destruction. (17.4)

Anticodon A three-nucleotide sequence in each tRNA that functions in the recognition of the complementary mRNA codon. (11.7)

Antigen Any substance recognized by an immune system as foreign to the organism. (17.2)

Antigen-presenting cells (APCs) Cells that display portions of protein antigens at their surface. The portions are derived by proteolysis of larger antigens. The antigenic peptides are displayed in conjunction with MHC molecules. Virtually any cell in the body can function as an APC by displaying peptides in combination with MHC class I molecules, which provides a mechanism for the destruction of infected cells. In contrast, macrophages, dendritic cells, and B cells are referred to as "professional" APCs because they function to ingest materials, process them, and display them to T_H cells in combination with MHC class II molecules. (17.4)

Antiserum Fluid containing desired antibodies that remains after the removal of the cells and clotting factors from whole blood that has been exposed to given antigen. (18.19)

Apoptosis A type of orderly, or programmed, cell death in which the cell responds to certain signals by initiating a normal response that leads to the death of the cell. Death by apoptosis is characterized by the overall compaction of the cell and its nucleus, the orderly dissection of the chromatin into pieces at the hands of a special DNA-splitting endonuclease, and the rapid engulfment of the dying cell by phagocytosis. (15.8)

Artifact A structure seen in a microscopic image that results from the coagulation or precipitation of materials that had no existence in the living cell. (18.2)

Assay Some identifiable feature of a specific protein, such as the catalytic activity of an enzyme, used to determine the relative amount of that protein in a sample. (18.7)

Aster "Sunburst" arrangement of microtubules around each centrosome during mitosis. (14.2)

Atomic force microscope (AFM) A high-resolution scanning instrument that is becoming increasingly important in nanotechnology and molecular biology. The AFM operates by scanning a delicate probe over the surface of the specimen. (18.3)

ATP synthase The ATP-synthesizing enzyme of the inner mitochondrial membrane, which is composed of two chief components: the F1 headpiece and the F_0 basal piece, the latter of which is embedded in the membrane itself. (5.5)

Autoantibodies Antibodies capable of reacting with the body's own tissues. (HP17)

Autoimmune diseases Diseases characterized by an attack of the immune system against the body's own tissues. Includes multiple sclerosis, insulin-dependent diabetes mellitus, and rheumatoid arthritis. (HP17)

Autophagy The destruction of organelles and their replacement during which an organelle is surrounded by a double membrane. The membrane surrounding the organelle then fuses with a lysosome. (8.6)

Autoradiography A technique for visualizing biochemical processes by allowing an investigator to determine the location of radioactively labeled materials within a cell. Tissue sections containing radioactive isotopes are covered with a thin layer of photographic emulsion, which is exposed by radiation emanating from the tissue. Sites in the cells containing radioactivity are revealed under the microscope by silver grains after development of the overlying emulsion. (8.2, 18.4)

Autotroph Organism capable of surviving on CO_2 as its principal carbon source. (6.9)

Axon A single, prominent extension that emerges from the cell body and conducts outgoing impulses away from the cell body and toward the target cell(s). (4.8)

Axonal transport Process by which vesicles, cytoskeletal polymers, and macromolecules are moved along microtubules within the axon of a neuron. Anterograde transport moves materials from the cell body toward the synaptic terminals, whereas retrograde transport moves materials in the opposite direction. (9.3)

Axoneme The central, microtubule-containing core of a cilium or flagellum. Most axonemes consist of nine peripheral doublets, two central microtubules, and numerous accessory structures. (9.3)

B lymphocytes (B cells) Lymphocytes that respond to antigen by proliferating and differentiating into plasma cells that secrete blood-borne antibodies. These cells attain their differentiated state in the bone marrow. (17.2)

Bacterial artificial chromosome (BAC) Cloning vector capable of accepting large foreign DNA fragments that can be cloned in bacteria. Consists of an F plasmid with an origin of replication and the genes required to regulate replication. BACs have played a key role in genome sequencing. (18.16)

Bacteriophages A group of viruses that require bacteria as host cells. (1.4)

Basal body A structure that resides at the base of the cilium or flagellum and which generates their outer microtubules. Basal bodies are identical in structure to centrioles. Both can give rise to one another. (9.3)

Base Any molecule that is capable of accepting a hydrogen ion. (2.3)

Base composition analysis The relative amounts of each base in various samples of DNA. (10.3)

Base excision repair (BER) A cut-and-patch mechanism for the removal from the DNA of altered nucleotides, e.g., uracil (formed from cytosine) and 8-oxoguanine. (13.2)

Basement membrane (basal lamina) Thickened layer of approximately 50 to 200 nm of extracellular matrix that surrounds muscle and fat cells and underlies the basal surface of epithelial tissues such as the skin, the inner lining of the digestive and respiratory tracts, and the inner lining of blood vessels. (7.1)

Benign tumor A tumor composed of cells that are no longer responsive to normal growth controls but lack the capability to invade normal tissues or metastasize to distant sites. (16.3)

Beta (β) clamp One of the noncatalytic components of the replisome that encircles the DNA and keeps the polymerase associated with the DNA template. (13.1)

Beta (β) pleated sheet One possible secondary structure of a polypeptide, in which several β-strands lie parallel to each other, creating the conformation of a sheet. (2.5)

Beta (β) strand One possible secondary structure of a polypeptide, in which the backbone of the chain assumes a folded (or pleated) conformation. (2.5)

Biochemicals Compounds synthesized by living organisms. (2.4)

Bioenergetics The study of the various types of energy transformations that occur in living organisms. (3.1)

Biosynthetic pathway (secretory pathway) Route through the cytoplasm by which materials are synthesized in the endoplasmic reticulum or Golgi complex, modified during passage through the Golgi complex, and transported within the cytoplasm to various destinations such as the plasma membrane, a lysosome, or a large vacuole of a plant cell. The alternative term secretory pathway has been used because many of the materials synthesized in the pathway are destined to be discharged (secreted) outside the cell. (8.1)

Bivalent (tetrad) The complex formed during meiosis by a pair of synapsed homologous chromosomes. (14.3)

Bright-field microscope A microscope in which light from the illuminating source is caused to converge on the specimen by the substage condenser, thereby forming a cone of bright light that can enter the objective lens. (18.1)

Buffers Compounds that can interact with either free hydrogen or hydroxyl ions, minimizing a change in pH. (2.3)

C₃ pathway The metabolic pathway by which carbon dioxide is assimilated into the organic molecules of the cell during photosynthesis. Ribulose 1,5-bisphosphate (RuBP) is used by Rubisco as the initial CO_2 acceptor. The product then fragments into two three-carbon PGA molecules. (6.6)

C₃ plants Plants that depend solely on the C_3 pathway to fix atmospheric CO_2. (6.6)

C₄ pathway Alternate pathway for carbon fixation utilizing phosphoenolpyruvate as the CO_2 acceptor to produce four-carbon compounds (predominantly malate and oxaloacetate). (6.6)

C₄ plants Plants, primarily tropical grasses, that use the C_4 carbon fixation pathway. (6.6)

Cadherins A family of related glycoproteins that mediate Ca^{2+}-dependent cell–cell adhesion. (7.3)

Calcium-binding proteins Proteins, such as calmodulin, that bind calcium and allow the calcium to elicit a variety of cellular responses. (15.5)

Calmodulin A small, calcium-binding protein that is widely distributed. Each molecule of calmodulin contains four binding sites for calcium. (15.5)

Calvin cycle (or Calvin-Benson cycle) The pathway for conversion of CO_2 into carbohydrate; the cycle occurs in cyanobacteria and all eukaryotic photosynthetic cells. (6.6)

CAM plants Plants that utilize PEP carboxylase in fixing CO_2 just as C_4 plants do, but conduct the light-dependent reactions and carbon fixation at different times of the day, so that stomata can be closed during the peak water loss hours of the day. (6.6)

Carbohydrates (Glycans) Organic molecules including simple sugars (monosaccharides) and multisaccharide polymers, which largely serve as energy-storage and structural compounds in cells. (2.5)

Caspases A family of cysteine proteases that are activated at an early stage of apoptosis and are responsible for the degradative events observed during cell death. (15.8)

Catabolic pathway A metabolic pathway in which relatively complex molecules are broken down into simpler products. (3.3)

Cation An ionized atom or molecule with an extra positive charge. (2.1)

Cell culture The technique used to grow cells outside the organism. (18.5)

Cell cycle The stages through which a cell passes from one cell division to the next. (14.1)

Cell division The process by which new cells originate from other living cells. (14)

Cell fractionation Bulk separation of the various cell organelles by differential centrifugation. (8.2)

Cell-free system An experimental system to study cellular activities that does not require whole cells. Such systems typically contain a preparation of purified proteins and/or subcellular fractions and are amenable to experimental manipulation. (8.2)

Cell fusion Technique whereby two different types of cells (from one organism or from different species) are joined to produce one cell with one, continuous plasma membrane. (4.6)

Cell line Cells that are commonly used in tissue culture studies that have undergone genetic modifications that allow them to be grown indefinitely. (18.5)

Cell-mediated immunity Carried out by T lymphocytes, that when activated, can specifically recognize and kill an infected (or foreign) cell. (17.1)

Cell plate Structure between the cytoplasm of two newly formed daughter cells that gives rise to a new cell wall in plant cells. (7.6, 14.3)

Cell signaling Communication in which information is relayed across the plasma membrane to the cell interior and often to the cell nucleus by means of a series of molecular interactions. (15)

Cell theory Theory of biological organization, which has three tenets: all organisms are made up of one or more cells; the cell is the structural unit of life; cells only arise from the division of preexisting cells. (1.1)

Cell wall A rigid, nonliving structure that provides support and protection for the cell it surrounds. (7.6)

Cellulose Unbranched glucose polymer with $\beta(1{\rightarrow}4)$ linkages that assembles into cables and serves as a principal structural element of plant cell walls. (2.5)

Centrioles Cylindrical structures, about 0.2 μm across and typically about twice as long, that contain nine evenly spaced fibrils, each of which appears in cross section as a band of three microtubules. Centrioles are nearly always found in pairs, with each of the members situated at right angles to one another. (9.3)

Centromere Marked indentation on a mitotic chromosome that serves as the site of kinetochore formation. (12.1)

Centrosome A complex structure that contains two barrel-shaped centrioles surrounded by amorphous, electron-dense **pericentriolar material** (or **PCM**) where microtubules are nucleated. (9.3)

Chaperones Proteins that bind to other polypeptides, preventing their aggregation and promoting their folding and/or assembly into multimeric proteins. (EP2)

Chaperonins Members of the Hsp60 class of chaperones, e.g., GroEL, that form a cylindrical complex of 14 subunits within which the polypeptide folding reaction takes place. (EP2)

Checkpoint Mechanisms that halt the progress of the cell cycle if (1) any of the chromosomal DNA is damaged, or (2) certain critical processes, such as DNA replication or chromosome alignment during mitosis, have not been properly completed. (14.1)

Chemiosmotic mechanism The mechanism for ATP synthesis whereby the movement of electrons through the electron-transport chain results in establishment of a proton gradient across the bacterial, thylakoid, or inner mitochondrial membrane, with the gradient acting as a high-energy intermediate, linking oxidation of substrates to the phosphorylation of ADP. (EP5)

Chemoautotroph An autotroph that utilizes the energy stored in inorganic molecules (such as ammonia, hydrogen sulfide, or nitrites) to convert CO_2 into organic compounds. (6)

Chiasmata Specific points of attachment between homologous chromosomes of bivalents, observed as the homologous chromosomes move apart at the beginning of the diplotene stage of meiotic prophase 1. The chiasmata are located at the sites on the chromosomes at which genetic exchange during crossing over had previously occurred. (14.3)

Chlorophyll The most important light-absorbing photosynthetic pigment. (6.3)

Chloroplast A specialized membrane-bound cytoplasmic organelle that is the principal site of photosynthesis in eukaryotic cells. (6)

Cholesterol Sterol found in animal cells that can constitute up to half of the lipid in a plasma membrane, with the relative proportion in any membrane affecting its fluid behavior. (4.3)

Chromatid Paired, rod-shaped members of mitotic chromosomes that together represent the duplicated chromosomes formed during replication in the previous interphase. (14.2)

Chromatin A complex nucleoprotein material that makes up the chromosomes of eukaryotes. (1.3)

Chromatin remodeling complexes Multisubunit protein complexes that use the energy released by ATP hydrolysis to alter chromatin structure and allow binding of transcription factors to regulatory sites in the DNA. (12.4)

Chromatography A term used for a wide variety of techniques in which a mixture of dissolved components is fractionated as it moves through some type of immobile matrix. (18.7)

Chromosome compaction (chromosome condensation) A process in which a cell converts its chromosomes into shorter, thicker structure capable of being segregated during mitosis or meiosis. (14.2)

Chromosomes Threadlike strands that are composed of the nuclear DNA of eukaryotic cells and are the carriers of genetic information. (10.1)

Cilia Hairlike motile organelles that project from the surface of a variety of eukaryotic cells. Cilia tend to occur in large numbers on a cell's surface. (9.3)

Ciliary or axonemal dynein A huge protein (up to 2 million daltons) responsible for conversion of the chemical energy of ATP into the mechanical energy of ciliary locomotion. (9.3)

***Cis* cisternae** The cisternae of the Golgi complex closest to the endoplasmic reticulum. (8.4)

***Cis* Golgi network (CGN)** The *cis*-most face of the organelle that is composed of an interconnected network of tubules located at the entry face closest to the ER. The CGN is thought to function primarily as a sorting station that distinguishes between proteins to be shipped back to the ER and those that are allowed to proceed to the next Golgi station. (8.4)

Cisternal (luminal) space The region of the cytoplasm enclosed by the membranes of the endoplasmic reticulum or Golgi complex. (8.3)

Clonal selection theory Well supported theory that B and T lymphocytes develop their ability to produce specific antibodies or T cell receptors prior to exposure to antigen. Should an antigen then enter the body, it can interact specifically with those B and T lymphocytes bearing complementary receptors. Interaction between the antigen and the B or T lymphocytes leads to proliferation of the lymphocyte to form a clone of cells capable of responding to that specific antigen. (17.2)

Coactivators The intermediaries that help bound transcription factors stimulate the initiation of transcription at the core promoter. (12.4)

Coated pits Specialized domains of the plasma membrane; coated pits serve as collection points for receptors that bind substances that enter a cell by means of endocytosis. (8.8)

Coated vesicles Vesicles that bud from a membrane compartment typically possess a multi-subunit protein coat that promotes the budding process and binds specific membrane proteins. **COPI-, COPII-, and clathrin-coated vesicles** are the best characterized coated vesicles. (8.5)

Codons Sequences of three nucleotides (nucleotide triplets) in mRNAs that specify amino acids. (11.6)

Coenzyme An organic, nonprotein component of an enzyme. (3.2)

Cofactor The nonprotein component of an enzyme, it can be either inorganic or organic. (3.2)

Cohesin A multiprotein complex that keeps replicated chromatids associated with one another until they are separated during cell division. (14.2)

Collagens A family of fibrous glycoproteins known for their high tensile strength that function exclusively as part of the extracellular matrix. (7.1)

Competitive inhibitor An enzyme inhibitor that competes with substrate molecules for access to the active site. (3.2)

Complement A system of blood-plasma proteins that act as part of the innate immune system to destroy invading microorganisms either directly (by making their plasma membrane porous) or indirectly (by making them susceptible to phagocytosis). (17.1)

Complementary The relationship between the sequence of bases in the two strands of the double helix of DNA. Structural restrictions on the configurations of the bases limits bonding to the two pairs: adenine-thymine and guanine-cytosine. (10.2)

Conductance The movement of small ions across membranes. (4.7)

Confocal scanning microscope A microscope in which the specimen is illuminated by a finely focused laser beam that rapidly scans across the specimen at a single depth, thus illuminating only a thin plane (or "optical section") within the object. The microscope is typically used with fluorescently stained specimens, and the light emitted from the illuminated optical section is used to form an image of the section on a video screen. (17.1)

Conformation The three-dimensional arrangement of the atoms within a molecule, often important in understanding the biological activity of proteins and other molecules in a living cell. (2.5)

Conformational change A predictable movement within a molecule that is associated with biological activity. (2.5)

Congression The movement of duplicated chromosomes to the metaphase plate during prometaphase of mitosis. (14.2)

Conjugate acid Paired form created when a base accepts a proton in an acid-base reaction. (2.3)

Conjugate base Paired form created when an acid loses a proton in an acid-base reaction. (2.3)

Conjugated protein Protein linked covalently or noncovalently to substances other than amino acids, such as metals, nucleic acids, lipids, and carbohydrates. (2.5)

Connective tissue Tissue that consists largely of a variety of distinct fibers that interact with each other in specific ways. The deeper layer of the skin (the dermis) is a type of connective tissue. (7.0)

Connexon Multisubunit complex of a gap junction formed from the clustering within the plasma membrane of an integral membrane protein called connexin. Each connexon is composed of six connexin subunits arranged around a central opening (or annulus) about 1.5 nm in diameter. (7.5)

Consensus sequence The most common version of a conserved sequence. The TTGACA sequence of a bacterial promoter (known as the -35 element) is an example of a consensus sequence. (11.2)

Conserved sequences Refers to the amino acid sequences of particular polypeptides or the nucleotide sequences of particular nucleic acids. If two sequences are similar to one another, i.e., homologous, they are said to be conserved, which indicates that they have not diverged very much from a common ancestral sequence over long periods of evolutionary time. (2.4)

Constant regions The portions of light and heavy antibody polypeptide chains that have the same amino acid sequence. (17.4)

Constitutive Occurs in a continual, non-regulated manner. Can relate to a normal process, such as constitutive secretion, or the result of a mutation that leads to a breakdown in regulation, which causes continual activity, such as the constitutive activation of a signaling pathway.

Constitutive heterochromatin Chromatin that remains in the compacted state in all cells at all times and, thus, represents DNA that is permanently silenced. It consists primarily of highly repeated sequences. (12.1)

Constitutive secretion Discharge of materials synthesized in the cell into the extracellular space in a continual manner. (8.1)

Contrast The difference in appearance between adjacent parts of an object or an object and its background. (18.1)

Conventional (or type II) myosins A family of myosins, first identified in muscle tissue, that are the primary motors for muscle

contraction but are also found in a variety of nonmuscle cells. Type II myosins are needed for splitting a cell in two during cell division, generating tension at focal adhesions, and the turning behavior of growth cones. (9.5)

Copper atoms of the electron transport chain A type of electron carrier; these atoms are located within a single protein complex of the inner mitochondrial membrane that accept and donate a single electron as they alternate between Cu^{2+} and Cu^{1+} states. (5.3)

Cotransport A process that couples the movement of two solutes across a membrane, termed symport if the two solutes are moved in the same direction and antiport if the two move in opposite directions. (4.7)

Covalent bond The type of chemical bond in which electron pairs are shared between two atoms. (2.1)

Cristae The many deep folds that are characteristic of the inner mitochondrial membrane, which contain the molecular machinery of oxidative phosphorylation. (5.1)

Crossing over (genetic recombination) Reshuffling of the genes on chromosomes (thereby disrupting linkage groups) that occurs during meiosis as a result of breakage and reunion of segments of homologous chromosomes. (10.1)

Cyanobacteria Evolutionarily important and structurally complex prokaryotes that possess oxygen-producing photosynthetic membranes. (1.3)

Cyclic photophosphorylation The formation of ATP in chloroplasts that is carried out by PSI independent of PSII. (6.5)

Cyclin-dependent kinases (Cdks) Enzymes that control progression of cells through the cell cycle. (14.1)

Cytochromes A type of electron carrier consisting of a protein bound to a heme group. (5.3)

Cytokines Proteins secreted by cells of the immune system that alter the behavior of other immune cells. (17.3)

Cytokinesis The part of the cell cycle during which physical division of the cell into two daughter cells occurs. (14.2)

Cytoplasmic dynein A huge protein (molecular mass over 1 million daltons) composed of numerous polypeptide chains. The molecule contains two large globular heads that act as force-generating engines. Evidence suggests that cytoplasmic dynein functions in the movement of chromosomes during mitosis and also as a minus end-directed microtubular motor for the movement of vesicles and membranous organelles through the cytoplasm. (9.3)

Cytoskeleton An elaborate interactive network composed of three well-defined filamentous structures: microtubules, microfilaments, and intermediate filaments. These elements function to provide structural support; an internal framework responsible for positioning the various organelles within the interior of the cell; as part of the machinery required for movement of materials and organelles within cells; as force-generating elements responsible for the movement of cells from one place to another; as sites for anchoring messenger RNAs and facilitating their translation into polypeptides; and as a signal transducer, transmitting information from the cell membrane to the cell interior. (9)

Cytosol The region of fluid content of the cytoplasm outside of the membranous organelles of a eukaryotic cell. (1.3)

Cytosolic surface The surface of a membrane that faces the cytosol. (8.3)

Dalton A measure of molecular mass, with one dalton equivalent to one unit of atomic mass (e.g., the mass of a ^1H atom). (1.4)

Dehydrogenase An enzyme that catalyzes a redox reaction by removing a hydrogen atom from one reactant. (3.3)

Deletion A chromosomal aberration that occurs when a portion of a chromosome is missing. (HP12)

Deletion Loss of a segment of DNA caused by the misalignment of homologous chromosomes during meiosis. (10.4)

Denaturation Separation of the DNA double helix into its two component strands. (10.4)

Denaturation The unfolding or disorganization of a protein from its **native** or fully folded state. (2.5)

Dendrites Fine extensions from the cell bodies of most neurons; dendrites receive incoming information from external sources, typically other neurons. (4.8)

Deoxyribonucleic acid (DNA) A double-stranded nucleic acid composed of two polymeric chains of deoxyribose-containing nucleotides. The genetic material of all cellular organisms (2.5) DNA may be duplicated as in **DNA replication** (13) and produced in large quantities of a specific segment as in **DNA cloning** (18.12).

Depolarization A decrease in the electrical potential difference across a membrane. (4.8)

Desmosomes (maculae adherens) Disc-shaped adhesive junction containing cadherins found in a variety of tissues, but most notably epithelia, where they are located basal to the adherens junction. Dense cytoplasmic plaques on the inner surface of the plasma membranes in this region serve as sites of anchorage for looping intermediate filaments that extend into the cytoplasm. (7.3)

Differential centrifugation A technique used to isolate a particular organelle in bulk quantity, which depends on the principle that, as long as they are more dense than the surrounding medium, particles of different size and shape travel toward the bottom of a centrifuge tube at different rates when placed in a centrifugal field. (18.6)

Differentiation The process through which unspecialized cells become more complex and specialized in structure and function. (1.3)

Diffusion The spontaneous process in which a substance moves from an area of higher concentration to one of lower concentration, eventually reaching the same concentration in all areas. (4.7)

Diploid Containing both members of each pair of homologous chromosomes, as exemplified by most somatic cells. Diploid cells are produced from diploid parental cells during mitosis. Contrast with haploid. (10.5, 14.3)

Direct immunofluorescence Technique for the intracellular localization of an antigen, in which antibodies are conjugated with small fluorescent molecules to form derivatives that are then incubated with the cells or sections of cells. The binding sites are then visualized with the fluorescence microscope. (18.19)

Disulfide bridge Forms between two cysteines that are distant from one another in the polypeptide backbone or in two separate polypeptides. They help stabilize the intricate shapes of proteins. (2.5)

DNA gyrase A type II topoisomerase that is able to change the state of supercoiling in a DNA molecule by relieving the tension that builds up during replication. It does this by traveling along the DNA and acting like a "swivel," changing the positively supercoiled DNA into negatively supercoiled DNA. (13.1)

DNA ligase The enzyme responsible for joining DNA fragments into a continuous strand. (13.1)

DNA methylation An epigenetic process in which methyl groups are added to cytosine residues in DNA by DNA methyltransferases. In vertebrates, DNA methylation occurs on certain CpG residues in the promoter regions of genes and is associated with inactivation of gene expression. Also associated more widely with preventing transposition of mobile genetic elements. (12.4)

DNA microarrays "DNA chips" prepared by spotting DNAs from different genes in a known, ordered location on a glass slide. The slide is then incubated with

fluorescently labeled cDNAs whose hybridization level provides a measure of the level of expression of each gene in the array. (12.4, 16.3)

DNA polymerases The enzymes responsible for constructing new DNA strands during replication or DNA repair. (13.1)

DNA-dependent RNA polymerases (RNA polymerases) The enzymes responsible for transcription in both prokaryotic and eukaryotic cells. (11.2)

DNA tumor viruses Viruses capable of infecting vertebrate cells, transforming them into cancer cells. DNA viruses have DNA in the mature virus particle. (16.2)

Dolichol phosphate Hydrophobic molecule built from more than 20 isoprene units that assembles the basal, or core, segment of carbohydrate chains within glycoproteins. (8.3)

Domain A region within a protein (or RNA) that folds and functions in a semi-independent manner. (2.5)

Double-strand breaks (DSBs) DNA damage often resulting from ionizing radiation involving a fracture of both strands of the double helix. DSBs can be debilitating to a cell and at least two distinct repair systems are devoted to their repair. (13.2)

Dynamic instability A term that relates to the assembly/disassembly properties of the plus end of a microtubule. The term describes the fact that growing and shrinking microtubules can coexist in the same region of a cell, and that a given microtubule can switch back and forth unpredictably between growing and shortening phases. (9.3)

Dynein An exceptionally large, cargo-carrying, multisubunit motor protein that moves along microtubules toward their minus end. This family of proteins occurs as **cytoplasmic dyneins** and **ciliary** or **axonemal dyneins**. (9.3)

Effector A substance that brings about a cellular response to a signal. (15.1)

Electrochemical gradient The overall difference in electrical charge and in solute concentration that determines the ability of an electrolyte to diffuse between two compartments. (4.7)

Electrogenic Any process that contributes directly to a separation of charge across a membrane. (4.7)

Electron-transfer potential The relative affinity for electrons, such that a compound with a low affinity has a high potential to transfer one or more electrons in a redox reaction (and thus act as a reducing agent). (5.3)

Electron-transport or respiratory chain Membrane-embedded electron carriers that accept high-energy electrons and sequentially lower the energy state of the

electrons as they pass through the chain, with the net results of capturing energy for use in synthesizing ATP or other energy-storage molecules. (5.3)

Electronegative atom The atom with the greater attractive force; the atom that can capture the major share of electrons of a covalent bond. (2.1)

Electrophoresis Fractionation techniques that depend on the ability of charged molecules to migrate when placed in an electric field. (18.7)

Embryonic stem (ES) cells A type of cell that has virtually unlimited powers of differentiation, found in the mammalian blastocyst, which is an early stage of embryonic development comparable to the blastula of other animals. (HP1, 18.18)

Endergonic reactions Reactions that are thermodynamically unfavorable and cannot occur spontaneously, possessing a $+\Delta G$ value. (3.1)

Endocytic pathway Route for moving materials from outside the cell (and from the membrane surface of the cell) to compartments, such as endosomes and lysosomes, located within the cell interior. (8.1, 8.8)

Endocytosis Mechanism for the uptake of fluid and solutes into a cell. Can be divided into two types: bulk-phase endocytosis, which is non-specific, and **receptor-mediated endocytosis**, which requires the binding of solute molecules such as LDL or transferrin to a specific cell-surface receptor. (8.8)

Endomembrane system Functionally and structurally interrelated group of membranous cytoplasmic organelles including the endoplasmic reticulum, Golgi complex, endosomes, lysosomes, and vacuoles. (8)

Endoplasmic reticulum (ER) A system of tubules, cisternae, and vesicles that divides the fluid content of the cytoplasm into a luminal space within the ER membrane and a cytosolic space outside the membranes. (8.3)

Endosomes Organelles of the endocytic pathway. Materials taken up by endocytosis are transported to **early endosomes** where they are sorted, and then on to late endosomes and ultimately to lysosomes. **Late endosomes** also function as destination sites of lysosomal enzymes transported from the Golgi complex. (8.8)

Endosymbiont theory Proposal, which is based on considerable evidence, that mitochondria and chloroplasts arose from symbiotic prokaryotes that took up residence within a primitive host cell. (EP1)

Endothermic reactions Those gaining heat under conditions of constant pressure and volume. (3.1)

Energy The capacity to do work, it exists in two forms: potential and kinetic. (3.1)

Enhancer A regulatory site in the DNA that may be located at considerable distance either upstream or downstream from the promoter that it regulates. Binding of one or more transcriptional factors to the enhancer can dramatically increase the rate of transcription of the gene. (12.4)

Enthalpy change (ΔH) The change during a process in the total energy content of the system. (3.1)

Entropy (S) A measure of the relative disorder of the system or universe associated with random movements of matter; because all movements cease at absolute zero (0 K), entropy is zero only at that temperature. (3.1)

Enzyme inhibitor Any molecule that can bind to an enzyme and decrease its activity, classified as noncompetitive or competitive based on the nature of the interaction with the enzyme. (3.2)

Enzyme–substrate complex The physical association between an enzyme and its substrate(s), during which catalysis of the reaction takes place. (3.2)

Enzymes The vitally important protein catalysts of cellular reactions. (3.2)

Epigenetic inheritance Heritable changes, i.e., changes that can be transmitted from one cell to its progeny, that do not involve changes in DNA sequence. Epigenetic changes can result from DNA methylation, covalent modification of histones, and likely other types of chromatin modifications. (12.1, 12.4)

Epithelial tissue Tissue composed of closely packed cells that line the spaces within the body. The outer layer of skin (the epidermis) is a type of epithelial tissue. (7.0)

Epitope (or antigenic determinant) Portion of an antigen that binds to the antigen-combining site of a specific antibody. (17.4)

Equilibrium constant of a reaction (K_{eq}) The ratio of product concentrations to reactant concentrations when a reaction is at equilibrium. (3.1)

Ester bond The chemical bond that forms between carboxylic acids and alcohols (or acidic and alcoholic functional groups) while producing a molecule of water. (2.4)

Euchromatin Chromatin that returns to its dispersed state during interphase. (12.1)

Eukaryotic cells Cells (e.g., plant, animal, protist, fungal) characterized by an internal structure based on organelles such as the nucleus, derived from *eu-karyon*, or true nucleus. (1.3)

Excitation-contraction coupling The steps that link the arrival of a nerve impulse at the muscle plasma membrane to the shortening of the sarcomeres deep within the muscle fiber. (9.6)

Excited state Electron configuration of a molecule after absorption of a photon has energized an electron to shift from an inner to an outer orbital. (6.3)

Exergonic reactions Reactions that are thermodynamically favorable, possessing a $-\Delta G$ value. (3.1)

Exocytosis The process of membrane fusion and content discharge during which the membrane of a secretory granule or vesicle comes into contact with the overlying plasma membrane with which it fuses, thereby forming an opening through which the contents of the granule or vesicle can be released. (8.5)

Exon shuffling The movement of genetic "modules" among unrelated genes facilitated by the presence of introns; the introns act like inert spacer elements between exons. (11.4)

Exons Those parts of a split gene that contribute to a mature RNA product. (11.4)

Exon-junction complex (EJC) A complex of proteins deposited on the transcript 20–24 nucleotides upstream from the newly formed exon-exon junction. This conglomeration of proteins stays with the mRNA until it is translated. (11.8)

Exonuclease A DNA- or RNA-digesting enzyme that attaches to either the 5′ or 3′ end of the nucleic acid strand and removes one nucleotide at a time from that shrinking end. (e.g., 12.6, 13.1)

Exothermic reactions Those releasing heat under conditions of constant pressure and volume. (3.1)

Extracellular matrix (ECM) An organized network of extracellular materials that is present beyond the immediate vicinity of the plasma membrane. It may play an integral role in determining the shape and activities of the cell. (7.1)

Extracellular messenger molecules The means by which cells usually communicate with each other. Extracellular messengers can travel a short distance and stimulate cells that are in close proximity to the origin of the message, or they can travel throughout the body, potentially stimulating cells that are far away from the source. (15.1)

Facilitated diffusion Process by which the diffusion rate of a substance is increased through interaction with a substance-specific membrane protein. (4.7)

Facilitative transporter A transmembrane protein that binds a specific substance and, in so doing, changes conformation so as to facilitate diffusion of the substance down its concentration gradient. (4.7)

Facultative heterochromatin Chromatin that has been specifically inactivated during certain phases of an organism's life. (12.1)

Families Groupings of proteins that have arisen from a single ancestral gene that underwent a series of duplications and subsequent modifications during the course of evolution. (2.5)

Fats Molecules consisting of a glycerol backbone linked by ester bonds to three fatty acids, also termed triacylglycerols. (2.5)

Fatty acid Long, unbranched hydrocarbon chain with a single carboxylic acid group at one end. (2.5)

Feedback inhibition A mechanism to control metabolic pathways where the end product interacts with an enzyme in the pathway, resulting in inactivation of the enzyme. (3.3)

Fermentation An anaerobic metabolic pathway in which pyruvate is converted to another molecule (often lactate or ethanol, depending on the organism) and $NAD^{\dagger}$ is regenerated for use in glycolysis. (3.3)

Fibrous protein One with a tertiary structure that is greatly elongated, resembling a fiber. (2.5)

First law of thermodynamics The law of conservation of energy, which states that energy can neither be created nor destroyed. (3.1)

Fixative A chemical solution that kills cells by rapidly penetrating the cell membrane and immobilizing all of its macromolecular material in such a way that the structure of the cell is maintained as close as possible to that of the living state. (18.1)

Flagella Hairlike motile organelles that project from the surface of a variety of eukaryotic cells. Essentially the same structure as cilia but present in much fewer numbers. (9.3)

Flavoproteins A type of electron carrier in which a polypeptide is bound to one of two related prosthetic groups, either FAD or FMN. (5.3)

Fluid-mosaic model Model presenting membranes as dynamic structures in which both lipids and associated proteins are mobile and capable of moving within the membrane to engage in interactions with other membrane molecules. (4.2)

Fluorescence in situ hybridization (FISH) Technique in which probe DNAs (or RNAs) are labeled with fluorescent dyes, hybridized to single-stranded DNA that remains within its chromosome, and localized with a fluorescence microscope. (10.4)

Fluorescence recovery after photobleaching (FRAP) Technique to study movement of membrane components that consists of three steps: (1) linking cellular components to a fluorescent dye, (2) irreversibly bleaching (removing the visible fluorescence of) a portion of the cell, (3) monitoring the reappearance of fluorescence (due to random movement of fluorescent-dyed components from outside of the bleached area) in the bleached portion of the cell. (4.6, 9.2)

Fluorescence resonance energy transfer (FRET) A technique to measure changes in distance between two parts of a protein (or between two separate proteins within a larger structure). Based on the transfer of energy from a donor fluorochrome to an acceptor fluorochrome, changing the fluorescence intensity of the two molecules. (Fig. 18.8)

Focal adhesions Adhesive structures characteristic of cultured cells adhering to the surface of a culture dish. The plasma membrane in the region of the focal adhesion contains clusters of integrins that connect the extracellular material that coats the culture dish to the actin-containing microfilament system of the cytoskeleton. (7.2)

Fractionated Disassembly of a preparation into its component ingredients so that the properties of individual species of molecules can be examined. (18.7)

Frameshift mutations Mutations in which a single base pair is either added to or deleted from the DNA, resulting in an incorrect reading frame from the point of mutation through the remainder of the coding sequence. (11.8)

Free energy change (ΔG) The change during a process in the amount of energy available to do work. (3.1)

Free radical Highly reactive atom or molecule that contains a single unpaired electron. (HP2)

Freeze-etching Technique in which tissue is freeze fractured, then exposed briefly to a vacuum so a thin layer of ice can evaporate from above and below the fractured surfaces, exposing features for identification by electron microscopy. (18.2)

Freeze-fracture replication Technique in which a tissue sample is first frozen and then struck with a blade that fractures the tissue block along the lines of least resistance, often resulting in a fracture line between the two leaflets of the lipid bilayer; metals are then deposited on the exposed surfaces to create a shadowed replica that is analyzed by electron microscopy. (14.4, 18.2)

Functional groups Particular groupings of atoms that tend to act as a unit, often affecting the chemical and physical behavior of the larger organic molecules to which they belong. (2.4)

γ-tubulin A type of tubulin that plays a critical component in microtubule nucleation. (9.3)

G protein See GTP-binding protein.

G protein-coupled receptors (GPCRs) A group of related receptors that span the plasma membrane seven times. The binding of the ligand to its specific receptor causes a change in the conformation of the receptor that increases its affinity for a heterotrimeric G protein initiating a response within the cell. (15.3)

G₁ Period within the cell cycle following mitosis and preceding the initiation of DNA synthesis. (14.1)

G₂ Period within the cell cycle between the end of DNA synthesis and the beginning of M phase. (14.1)

Gametophyte The haploid stage of the life cycle of plants that begins with spores generated during the sporophyte stage. During the gametophyte stage, gametes form through the process of mitosis. (14.3)

Gap junctions Sites between animal cells that are specialized for intercellular communication. Plasma membranes of adjacent cells come within about 3 nm of each other, and the gap is spanned by very fine "pipelines," or connexons that allow the passage of small molecules. (7.5)

Gated channel An ion channel that can change conformation between a form open to its solute ion and one closed to the ion; such channels can be voltage gated, chemical gated, or mechanically gated depending on the nature of the process that triggers the change in conformation. (4.7)

Gel filtration Purification technique in which separation of proteins (or nucleic acids) is based primarily on molecular mass. The separation material consists of tiny porous beads that are packed into a column through which the solution of protein slowly passes. (18.7)

Gene duplication The duplication of a small portion of a single chromosome usually by a process of unequal crossing over. (10.5)

Gene regulatory protein A protein that is capable of recognizing a specific sequence of base pairs within the DNA and binding to that sequence with high affinity thereby altering gene expression. (12.3)

Gene therapy The process by which a patient is treated by altering the genotype of diseased cells. (HP4)

General transcription factors (GTFs) Auxiliary proteins necessary for RNA polymerase to initiate transcription. These factors are termed "general" because the same ones are required for transcription of a diverse array of genes by the polymerase. (11.4)

Genes In nonmolecular terms, a unit of inheritance that governs the character of a particular trait. In molecular terms, a segment of DNA containing the information for a single polypeptide or RNA molecule, including transcribed but non-coding regions. (10.1)

Genetic code Manner in which the nucleotide sequences of DNA encode the information for making protein products. (11.6)

Genetic map Assignment of genetic markers to relative positions on a chromosome based on crossover frequency. (10.5)

Genetic polymorphisms Sites in the genome that vary with relatively high frequency among different individuals in a species population. (10.6)

Genetic recombination (crossing over) Reshuffling of the genes on chromosomes (thereby disrupting linkage groups) that occurs as a result of breakage and reunion of segments of homologous chromosomes. (10.1)

Genome The complement of genetic information unique to each species of organism. Equivalent to the DNA of a haploid set of chromosomes from that species. (10.4)

Germ cells Cells (e.g., spermatogonia, oogonia, spermatocytes, oocytes) that have the capability to give rise to gametes.

Globular protein One with a tertiary structure that is compact, resembling a globe. (2.5)

Glycocalyx A layer closely applied to the outer surface of the plasma membrane. It contains membrane carbohydrates along with extracellular materials that have been secreted by the cell into the external space, where they remain closely associated with the cell surface. (7.1)

Glycogen Highly branched glucose polymer that serves as readily available chemical energy in most animal cells. (2.5)

Glycolipids Sphingosine-based lipid molecules linked to carbohydrates, often active components of plasma membranes. (4.3)

Glycolysis The first pathway in the catabolism of glucose, it does not require oxygen and results in the formation of pyruvate. (3.3)

Glycosaminoglycans (GAGs) A group of highly acidic polysaccharides with the structure of —A—B—A—B—, where A and B represent two different sugars. (2.5)

Glycosidic bond The chemical bond that forms between sugar molecules. (2.5)

Glycosylation The reactions by which sugar groups are added to proteins and lipids. (4.3, 8.3, 8.4)

Glycosyltransferases A large family of enzymes that transfer specific sugars from a specific donor (a nucleotide sugar) to a specific receptor (typically the growing end of an oligosaccharide chain). (8.3)

Glyoxysomes Organelles found in plant cells that serve as sites for enzymatic reactions including the conversion of stored fatty acids to carbohydrate. (5.6)

Golgi complex Network of smooth membranes organized into a characteristic morphology, consisting of flattened, disc-like cisternae with dilated rims and associated vesicles and tubules. The Golgi complex functions primarily as a processing plant where proteins newly synthesized in the endoplasmic reticulum are modified in specific ways. (8.4)

GPI-anchored proteins Peripheral membrane proteins that are anchored to the membrane via linkage to a glycosylphosphatidylinositol molecule of the bilayer. (4.4)

Grana Orderly, stacked arrangement of thylakoids. (6.1)

Green fluorescent protein (GFP) A fluorescent protein encoded by the jellyfish *Aequoria victoria* that is widely used to follow events in living cells. In most cases the gene encoding the protein is fused to the gene of interest and the DNA containing the fusion protein is introduced into the cells to be studied. (8.2)

Ground state The unexcited state of an atom or molecule. (6.3)

Growth cone The distal tip of a growing neuron that contains the locomotor activity required for axonal extension. (9.7)

GTP-binding proteins (or G proteins) With key regulatory roles in many different cellular processes, G proteins can be present in at least two alternate conformations, an active form containing a bound GTP molecule, and an inactive form containing a bound GDP molecule. (8.3)

GTPase-activating proteins (GAPs) Proteins that bind to G proteins, activating their GTPase activity. As a result, GAPs shorten the duration of a G protein-mediated response. (15.4)

Guanine nucleotide-dissociation inhibitors (GDIs) Proteins that bind to G proteins, inhibiting the dissociation of the bound GDP, thus maintaining the G protein in the inactive state. (15.4)

Guanine nucleotide-exchange factors (GEFs) Proteins that bind to a G protein, stimulating the exchange of a bound GDP with a GTP, thereby activating the G protein. (15.4)

Guanosine triphosphate (GTP) A nucleotide of great importance in cellular activities. It binds to a variety of proteins (called G proteins) and acts as a switch to turn on their activities. (2.5)

Half-life A measure of the instability of a radioisotope, or, equivalently, the amount of time required for one-half of the radioactive material to disintegrate. (18.4)

Haploid Containing only one member of each pair of homologous chromosomes. Haploid

cells are produced during meiosis, as exemplified by sperm. Contrast with diploid. (10.4, 14.3)

Haplotype A block of the genome that tends to be inherited intact from generation to generation. Haplotypes are generally defined by the presence of a consistent combination of single nucleotide polymorphisms (SNPs). (HP10)

Head group The polar, water-soluble region of a phospholipid that consists of a phosphate group linked to one of several small, hydrophilic molecules. (4.3)

Heat shock response Activation of the expression of a diverse array of genes in response to temperature elevation. The products of these genes, including molecular chaperones, help the organism recover from the damaging effects of elevated temperature. (EP2)

Heavy chain One of the two types of polypeptide chains in an antibody, usually with a molecular mass of 50 to 70 kD. (17.4)

Helicase A protein that unwinds the DNA (or RNA) duplex in a reaction in which energy released by ATP hydrolysis is used to break the hydrogen bonds that hold the two strands together. (13.1)

Hematopoietic stem cells (HSCs) Cells that are situated primarily in the bone marrow that are capable of both self-renewal and of giving rise to all types of blood cells. (HP1, 17.1)

Hemicelluloses Branched polysaccharides of the plant cell wall whose backbone consists of one sugar, such as glucose, and sidechains of other sugars, such as xylose. (7.6)

Hemidesmosome Specialized adhesive structure at the basal surface of epithelial cells that functions to attach the cells to the underlying basement membrane. The hemidesmosome contains a dense plaque on the inner surface of the plasma membrane, with keratin-containing filaments coursing out into the cytoplasm. (7.2)

Hemolysis The permeabilization of red blood cell membranes, experimentally done through placement of cells in hypotonic solution, where they swell before bursting, releasing cellular contents and leaving membrane ghosts. (4.6)

Heterochromatin Chromatin that remains compacted during interphase. (12.1)

Heterogeneous nuclear RNAs (hnRNAs) A large group of RNA molecules that share the following properties: (1) they heave large molecular weights (up to about 80S, or 50,000 nucleotides); (2) they represent many different nucleotide sequences; and (3) they are found only in the nucleus. Includes pre-mRNAs. (11.4)

Heterotrimeric G protein A component of certain signal transduction systems, referred to as G proteins because they bind guanine nucleotides (either GDP or GTP), and described as heterotrimeric because all of them consist of three different polypeptide subunits. (15.3)

Heterotroph Organism that depends on an external source of organic compounds. (6.1)

High-performance liquid chromatography (HPLC) A high-resolution type of chromatography in which long, narrow columns are used, and the mobile phase is forced through a tightly packed matrix under high pressure. (18.7)

Highly repeated fraction Typically short (a few hundred nucleotides at their longest) DNA sequences that are present in at least 10^5 copies per genome. The highly repeated sequences typically account for about 10 percent of the DNA of vertebrates. (10.4)

Histone acetyltransferases (HATs) Enzymes that transfer acetyl groups to lysine and arginine residues of core histones. Histone acetylation is associated with activation of transcription. (12.1, 12.4)

Histone code A concept that the state and activity of a particular region of chromatin depends upon the specific covalent modifications, or combination of modifications, to the histone tails of the nucleosomes in that region. The modifications are created by enzymes that acetylate, methylate, and phosphorylate various amino acid residues within the core histones. (12.1)

Histone deacetylases (HDACs) Enzymes that catalyze the removal of acetyl groups from core histones. Histone deacetylation is associated with the repression of transcription. (12.4)

Histones A collection of small, well-defined, basic proteins of chromatin. (12.1)

hnRNP (heterogeneous nuclear ribonucleoprotein) The result of the transcription of each hnRNA which becomes associated with a variety of proteins; hnRNP represents the substrate for the processing reactions that follow. (11.4)

Homogenize To mechanically rupture cells. (8.2)

Homologous chromosomes Paired chromosomes of diploid cells, each carrying one of the two copies of the genetic material carried by that chromosome. (10.1)

Homologous sequences When the amino acid sequences of two or more polypeptides, or the nucleotide sequences of two or more genes, are similar to one another, they are presumed to have evolved from the same ancestral sequence. Such sequences are said to be *homologous*, a term reflecting an evolutionary relatedness. (2.4)

Humoral immunity Immunity mediated by blood-borne antibodies. (17.1)

Hybridomas Hybrid cells produced by fusion of a normal antibody-producing lymphocyte and a malignant myeloma cell. Hybridomas proliferate, producing large amounts of a single (monoclonal) antibody that was being synthesized by the normal cell prior to fusion with the myeloma cell. (18.19)

Hydrocarbons The simplest group of organic molecules, consisting solely of carbon and hydrogen. (2.4)

Hydrogen bond The weak, attractive interaction between a hydrogen atom covalently bonded to an electronegative atom (thus, with a partial positive charge) and a second electronegative atom. (2.2)

Hydrophilic The tendency of polar molecules to interact with surrounding water molecules, which are also polar; derived from "water loving." (2.2)

Hydrophobic interaction The tendency of nonpolar molecules to aggregate so as to minimize their collective interaction with surrounding polar water molecules; derived from "water fearing." (2.2)

Hypertonic (or Hyperosmotic) Property of one compartment having a higher solute concentration compared with that in a given compartment. (4.7)

Hypervariable regions The subsections of antibody variable regions that vary more significantly in sequence from one molecule to another and are associated with antigen specificity. (17.4)

Hypotonic (or Hypoosmotic) Property of one compartment having a lower solute concentration compared with that in a given compartment. (4.7)

Immune responses Responses elicited by cells of the immune system upon contact with foreign materials, including invading pathogens. Includes innate and adaptive responses. Adaptive immune responses can be divided into primary responses that follow initial exposure to an antigen and secondary responses that follow reexposure to that antigen. (17.1)

Immune system Physiological system consisting of organs, scattered tissues, and independent cells that protect the body from invading pathogens and foreign materials. (17)

Immunity State in which the body is not susceptible to infection by a particular pathogen. (17)

Immunoglobulin superfamily (IgSF) A wide variety of proteins that contain domains composed of 70 to 110 amino acids that are homologous to domains that make up the polypeptide chains of blood-borne antibodies. (7.3, 17.1)

Immunologic tolerance A state in which the body does not react against a particular substance, such as the body's own proteins, because cells that could engage in such a response have either been inactivated or destroyed. (17.1, HP 17)

Immunotherapy Treatment of disease, including cancer and autoimmune disorders, with antibodies or immune cells. (16.4, HP17)

Imprinting Differential expression of genes depending solely on whether they were contributed to the zygote by the sperm or the egg. (12.4)

In situ hybridization Technique to localize a particular DNA or RNA sequence in a cell or on a culture plate or electrophoretic gel. (10.3, 18.11)

Indirect immunofluorescence A variation of direct immunofluorescence in which the cells are treated with unlabeled antibody, which is allowed to form a complex with the corresponding antigen. The location of the antigen-antibody couple is then revealed in a second step using a preparation of fluorescently labeled antibodies whose combining sites are directed against the antibody molecules used in the first step. (18.19)

Induced fit The conformational change in an enzyme after the substrate has been bound that allows the chemical reaction to proceed. (3.2)

Inducer A compound that binds to a protein repressor and activates transcription of a bacterial operon. (12.2)

Inducible operon An operon in which the presence of the key metabolic substance induces transcription of the structural genes. (12.2)

Inflammation The localized accumulation of fluid and leukocytes in response to injury or infection, leading to redness, swelling, and fever. (17.1)

Initiation codon The triplet AUG, the site to which the ribosome attaches to the mRNA to assure that the ribosome is in the proper **reading frame** to correctly read the entire message. (11.8)

Initiation factors Soluble proteins (IFs in bacteria and eIFs in eukaryotes) that make initiation of translation possible. (11.8)

Innate immune response A nonspecific response to a pathogen that does not require previous exposure to that agent, includes responses mediated by NK cells, complement, phagocytes, and interferon. (17.1)

Insulators Specialized boundary sequences that "cordon off" a promoter and its enhancer from other promoter/enhancer elements. According to one model, insulator sequences bind to proteins of the nuclear matrix. (12.4)

Insulin receptor substrates (IRSs) Protein substrates that, when phosphorylated in response to insulin, bind and activate a variety of "downstream" effectors. (15.4)

Integral protein A membrane-associated protein that penetrates or spans the lipid bilayer. (4.4)

Integrins A superfamily of integral membrane proteins that bind specifically to extracellular molecules. (7.2)

Intermediate filaments (IFs) Strong, ropelike cytoskeletal fibers approximately 10 nm in diameter that, depending on the cell type, may be composed of a variety of protein subunits capable of assembling into similar types of filaments. IFs are thought to provide mechanical stability to cells and provide specialized, tissue-specific functions. (9.4)

Intermembrane space The space between the inner and outer mitochondrial membranes. (5.1)

Interphase The portion of the cell cycle between periods of cell division. (14.1)

Intervening sequences Regions of DNA that are between coding sequences of a gene and that are therefore missing from corresponding mRNA. (11.4)

Intraflagellar transport (IFT) A process in which particles are moved in both directions between the base of a flagellum or cilium and its tip. The force that drives IFT is generated by motor proteins that track along the peripheral doublets of the axoneme. (9.3)

Introns Those parts of a split gene that correspond to the intervening sequences. (11.4)

Inversion A chromosomal aberration that results after a chromosome is broken in two places and the resulting center segment is reincorporated into the chromosome in reverse order. (HP12)

In vitro Outside the body. Cells grown in culture are said to be grown in vitro; and studies on cultured cells are an essential tool of cell and molecular biologists. (1.2)

Ion An atom or molecule with a net positive or negative charge because it has lost or gained one or more electrons during a chemical reaction. (2.1)

Ion channel A transmembranous structure (e.g., an integral protein with an aqueous pore) permeable to a specific ion or ions. (4.7)

Ion-exchange chromatography A technique for protein purification in which ionic charge is used to separate different proteins. (18.7)

Ionic bond A noncovalent bond occurring between oppositely charged ions, also called a salt bridge. (2.2)

Iron-sulfur proteins A group of protein electron carriers with an inorganic, iron-sulfur center. (5.3)

Irreversible inhibitor An enzyme inhibitor that binds tightly, often covalently, thus inactivating the enzyme molecule permanently. (3.2)

Isoelectric focusing A type of electrophoresis in which proteins are separated according to isoelectric point. (18.7)

Isoelectric point The pH at which the negative charges of the component amino acids of a protein equal the positive charges of the component amino acids, so the protein is neutral. (18.7)

Isoforms Different versions of a protein. Isoforms may be encoded by separate, closely related genes, or formed as splice variants by alternative splicing from a single gene. (2.5)

Isotonic Property of one compartment having the same solute concentration compared with that in a given compartment. (4.7)

Karyotype Image in which paired homologous chromosomes are ordered in decreasing size. (12.1)

Kinesin A plus end-directed motor protein that moves membranous vesicles and other organelles along microtubules through the cytoplasm. Kinesin is a member of a family of **kinesin-like proteins (KLPs)**. (9.3)

Kinetic energy Energy released from a substance through atomic or molecular movements. (3.1)

Kinetochore A buttonlike structure situated at the outer surface of the centromere to which the microtubules of the spindle attach. (14.2)

Knockout mice Mice, born as the result of a series of experimental procedures, that are lacking a functional gene that they would normally contain. (18.18)

Lagging strand The newly synthesized daughter DNA strand that is synthesized discontinuously, so called because the initiation of each fragment must wait for the parental strands to separate and expose additional template. (13.1)

Lamellipodium The leading edge of a moving fibroblast, which is extended out from the cell as a broad, flattened, veil-like projection that glides over the substratum. (9.7)

Lateral gene transfer Transfer of genes from one species to another. (EP1)

Leading strand The newly synthesized daughter DNA strand that is synthesized continuously, so called because its synthesis continues as the replication fork advances. (13.1)

Ligand Any molecule that can bind to a receptor because it has a complementary structure. (4.1, 15.1)

Light chain Smaller of the two types of polypeptide chains in an antibody, with a molecular mass of 23 kDa. (17.4)

Light-harvesting complex II (LHCII) Pigment-protein complex, located outside the photosystem itself, that contains most of the antenna pigments that collect light for PSII. It can also be associated with PSI. (6.4)

Light-dependent reactions First of two series of reactions that compose photosynthesis. In these reactions, energy from sunlight is absorbed and converted to chemical energy that is stored in ATP and NADPH. (6.2)

Light-independent reactions (dark reactions) Second of two series of reactions that compose photosynthesis. In these reactions, carbohydrates are synthesized from carbon dioxide using the energy stored in the ATP and NADPH molecules formed in the light-dependent reactions. (6.2)

Limit of resolution The resolution attainable by a microscope is limited by the wavelength of the illumination according to the equation $D = 0.61 \lambda/n \sin \alpha$, where D is the minimum distance that two points in the specimen must be separated to be resolved, λ is the wavelength of light, and n is the refractive index of the medium. Alpha is a measure of the light-gathering ability of the lens and is directly related to its aperture. For a light microscope, the limit of resolution is slightly less than 200 nm. (18.1)

Linkage groups Groups of genes that reside on the same chromosome causing nonindependent segregation of traits controlled by these genes. (10.1)

Lipid-anchored protein A membrane-associated protein that is located outside the bilayer but is covalently linked to a lipid molecule within the bilayer. (4.4)

Lipid bilayer Phospholipids self-assembled into a bimolecular structure based on hydrophobic and hydrophilic interactions; biologically important as the core organization of cellular membranes. (4.2)

Lipid rafts Microdomains within a cellular membrane that possess decreased fluidity due to the presence of cholesterol, glycolipids, and phospholipids containing longer, saturated fatty acids. A proposed residence of GPI-anchored proteins and signaling proteins. (4.5)

Lipids Nonpolar organic molecules, including fats, steroids, and phospholipids, whose common property of not dissolving in water contributes to much of their biological activity. (2.5)

Liposome An artificial lipid bilayer that self-assembles into a spherical vesicle or vesicles when in an aqueous environment. (4.3)

Locus (pl. loci) The position of a gene on a chromosome. (10.1)

Luminal (cisternal) space The region of fluid content of the cytoplasm enclosed by the membranes of the endoplasmic reticulum or Golgi complex. (8.3)

Lymphocytes Nucleated leukocytes (white blood cells) that circulate between the blood and lymph organs that mediate acquired immunity. Includes both B cells and T cells. (17)

Lysosomal storage disorders Diseases characterized by the deficiency of a lysosomal enzyme and the corresponding accumulation of undegraded substrate. (HP8)

M phase The part of the cell cycle that includes the processes of mitosis, during which duplicated chromosomes are separated into two nuclei, and cytokinesis, during which the entire cell is physically divided into two daughter cells. (14.1)

Macromolecules Large, highly organized molecules crucial to the structure and function of cells; divided into polysaccharides, certain lipids, proteins, and nucleic acids. (2.4)

Major histocompatibility complex (MHC) A region of the genome that encodes **MHC proteins.** The genes that encode these proteins tend to be highly polymorphic, being represented by a large number of different alleles. These genetic differences between humans account for the tendency of a person to reject a transplant from another person other than an identical twin. (17.4)

Mass spectrometry Methodology to identify molecules (including proteins). A protein or mixture of proteins is fragmented, converted into gaseous ions, and propelled through a tubular component of a mass spectrometer, causing the ions to separate according to their mass/charge (m/z) ratio. Identification of the protein(s) is made by comparison with a computer database of the sequence of proteins encoded by a particular genome. (2.5, 18.7)

Matrix One of two aqueous compartments of a mitochondrion; the matrix is located within the interior of the organelle; the second compartment is called the **intermembrane** space and is located between the outer and inner mitochondrial membrane. (5.1)

Matrix metalloproteinases (MMPs) A family of zinc-containing enzymes that act in the extracellular space to digest various extracellular proteins and proteoglycans. (7.1)

Maximal velocity (V_{max}) The highest rate achieved for a given enzymatically catalyzed reaction, it occurs when the enzyme is saturated with substrate. (3.2)

Medial **cisternae** The cisternae of the Golgi complex between the *cis* and *trans* cisternae. (8.4)

Meiosis The process during which the chromosome number is reduced so that cells are formed that contain only one member of each pair of homologous chromosomes. (14.3)

Membrane fluidity A property of the physical state of the lipid bilayer of a membrane that allows diffusion of membrane lipids and proteins within the plane of the membrane. Inversely related to membrane viscosity. Membrane fluidity increases as the temperature rises and in bilayers with more unsaturated lipids. (4.5)

Membrane potential The electrical potential difference across a membrane. (4.8)

Messenger RNA (mRNA) The intermediate molecule between a gene and the polypeptide for which it codes. Messenger RNA is assembled as a complementary copy of one of the two DNA strands that encodes the gene. (11.1)

Metabolic intermediate A compound produced during one step of a metabolic pathway. (2.4)

Metabolic pathway A series of chemical reactions that results in the synthesis of an end product important to cellular function. (2.4)

Metabolism The total of the chemical reactions occurring within a cell. (1.2)

Metaphase The stage of mitosis during which all of the chromosomes have become aligned at the spindle equator, with one chromatid of each chromosome connected to one pole and its sister chromatid connected to the opposite pole. (14.2)

Metastasis Spread of cancer cells from a primary tumor to distant sites in the body where the formation of secondary tumors may arise. (16.3)

Methylguanosine cap Modification of the 5′ end of an mRNA precursor molecule, so that the terminal "inverted" guanosine is methylated at the 7′ position on its guanine base, while the nucleotide on the internal side of the triphosphate bridge is methylated at the 2′ position of the ribose. This cap prevents the 5′ end of the mRNA from being digested by nucleases, aids in transport of the mRNA out of the nucleus, and plays a role in the initiation of mRNA translation. (11.4)

MHC proteins Proteins encoded by the MHC region of the genome that bind processed antigens (antigenic peptides) and display them on the surface of the cell. Divided into two major classes, **MHC class I** molecules produced by virtually all cells of the body, and **MHC class II** molecules produced by "professional"

APCs, such as macrophages and dendritic cells. (17.4)

Michaelis constant (K_M) In enzyme kinetics, the value equal to the substrate concentration present when reaction rate is one-half of the maximal velocity. (3.2)

Micrometer Measure of length equaling 10^{-6} meters. (1.3)

MicroRNAs (miRNAs) Small RNAs (20-23 nucleotides long) that are synthesized from many sites in the genome and involved in inhibiting translation or increasing degradation of complementary mRNAs. (11.5)

Microfibrils Bundles of cellulose molecules that confer rigidity to the cell wall and provide resistance to pulling forces. (7.6)

Microfilaments Solid, 8-nm thick, cytoskeletal structures composed of a double-helical polymer of the protein actin. They play a key role in virtually all types of contractility and motility within cells. (9, 9.5)

Microscope An instrument that provides a magnified image of a tiny object. (1.1)

Microsomes A heterogeneous collection of vesicles formed from the endomembrane system (primarily the endoplasmic reticulum and Golgi complex) after homogenization. (8.2)

Microtubule-organizing centers (MTOCs) A variety of specialized structures that exert a role in initiating microtubule formation. (9.3)

Microtubule-associated proteins (MAPs) Proteins other than tubulin contained in microtubules obtained from cells. MAPs may interconnect microtubules to form bundles and can be seen as cross-bridges connecting microtubules to each other. Other MAPs may increase the stability of microtubules, alter their rigidity, or influence the rate of their assembly. (9.3)

Microtubules Hollow, cylindrical cytoskeletal structures, 25 nm in diameter, whose wall is composed of the protein tubulin. Microtubules are polymers assembled from αβ-tubulin heterodimers that are arranged in rows, or protofilaments. Because of their rigidity, microtubules often act in a supportive capacity. (9, 9.3)

Mismatch repair DNA repair system that removes mismatched bases that are incorporated by the DNA polymerase and escape the enzyme's proofreading exonuclease. (13.2)

Mitochondrial matrix The aqueous compartment within the interior of a mitochondrion. (5.1)

Mitochondrial membranes The outer membrane serves as a boundary with the cytoplasm and is relatively permeable, and the inner membrane houses respiratory machinery in its many invaginations and is highly impermeable. (5.1)

Mitochondrion The cellular organelle in which aerobic energy transduction takes place, oxidizing metabolic intermediates such as pyruvate to produce ATP. (5.1)

Mitosis Process of nuclear division in which duplicated chromosomes are faithfully separated from one another, producing two nuclei, each with a complete copy of all the chromosomes present in the original cell. (14.2)

Mitotic spindle Microtubule-containing "machine" that functions in the organization and sorting of duplicated chromosomes during mitotic cell division. (14.2)

Model organisms Organisms that have been widely used for research so that a great deal is known about their biology. These organisms have properties that have made them excellent research subjects. Such organisms include the bacterium, *E. coli*; the budding yeast, *S. cerevisiae*; the nematode, *C. elegans*; the fruit fly, *D. melanogaster*; the mustard plant, *A. thaliana*; and the mouse, *M. musculus*. (1.3)

Moderately repeated fraction DNA sequences that are repeated from a few to several hundred thousand times within a eukayotic genome. The moderately repeated fraction of the DNA can vary from about 20 to about 80 percent of the total DNA. These sequences may be identical to each other or nonidentical but related. (10.4)

Molecular chaperones Various families of proteins whose role is to assist the folding and assembly of proteins by preventing undesirable interactions. (EP2)

Monoclonal antibody A preparation of antibody molecules produced from a single colony (or clone) of cells. (18.19)

Monosomy A chromosome complement that lacks one chromosome, i.e., has only one member of one of the pairs of homologous chromosomes. (HP14)

Motif A substructure found among many different proteins, such as the αβ barrel, which consists of β strands connected by an α-helical region. (2.5)

Motor proteins Proteins that utilize the energy of ATP hydrolysis to generate mechanical forces that propel the protein, as well as attached cargo, along one of the components of the cytoskeleton. Three families of motor proteins are known: kinesins and dyneins move along microtubules and myosins move along microfilaments. (9.3)

Multiprotein complex The interaction of more than one complete protein to form a larger, functional complex. (2.5)

Muscle fiber A skeletal muscle cell, referred to as a muscle fiber because of its highly ordered, multinucleated, cablelike structure composed of hundreds of thinner, cylindrical myofibrils. (9.6)

Mutant An individual having an inheritable characteristic that distinguishes it from the wild type. (10.1)

Mutation A spontaneous change in a gene that alters it in a permanent fashion so that it causes heritable change. (10.1)

Myelin sheath The lipid-rich material wrapped around most neurons in the vertebrate body. (4.8)

Myofibrils The thin, cylindrical strands found within muscle fibers. Each myofibril is composed of repeating linear arrays of contractile units, called **sarcomeres**, that give skeletal muscle cells their striated appearance. (9.6)

Myosin A large family of motor proteins that moves along actin-containing microfilaments. Most myosins are plus-end directed motors. **Conventional myosin (myosin II)** is the protein that mediates muscle contractility as well as certain types of nonmuscle motility, such as cytokinesis. **Unconventional myosins** (I and III-XVIII) have many diverse roles including organelle transport. (9.5)

Nanometer Measure of length equaling 10^{-9} meters. (1.3)

nanotechnology A field of engineering involving the development of tiny "nanomachines" capable of performing specific activities in a submicroscopic world. (9.1)

Nascent protein A protein in the process of being synthesized, i.e., not yet complete. (11.8)

Natural killer (NK) cell A type of lymphocyte that engages in a nonspecific attack on an infected host cell, leading to apoptosis. (17.1)

Negative staining Procedures in which heavy-metal deposits are collected everywhere on the specimen grid except in the locations of very small particulate materials, including high-molecular-weight aggregates such as viruses, ribosomes, multisubunit enzymes, cytoskeletal elements, and protein complexes. (18.2)

Nerve impulse The process through which an action potential is propagated along the membrane of a neuron by sequentially triggering action potentials in adjacent stretches of membrane. (4.8)

Neurofilaments Loosely packed bundles of intermediate filaments located within the cytoplasm of neurons. Neurofilaments have long axes that are oriented parallel to that of the nerve cell axon and are composed of three distinct proteins: NF-L, NF-H, and NF-M. (9.4)

Neuromuscular junction The point of contact of a terminus of an axon with a muscle fiber,

the neuromuscular junction is a site of transmission of nerve impulses from the axon across the synaptic cleft to the muscle fiber. (4.8, 9.6)

Neurotransmitter A chemical that is released from a presynaptic terminal and binds to the postsynaptic target cell, altering the membrane potential of the target cell. (4.8)

Nitrogen fixation The process through which nitrogen gas is chemically reduced and converted into a component of organic compounds. (1.3)

Noncompetitive inhibitor An enzyme inhibitor that does not bind at the same site as the substrate, and so the level of inhibition depends only on the concentration of inhibitor. (3.2)

Noncovalent bond A relatively weak chemical bond based on attractive forces between oppositely charged regions within a molecule or between two nearby molecules. (2.2)

Noncyclic photophosphorylation The formation of ATP during the process of oxygen-releasing photosynthesis in which electrons move in a linear path from H_2O to $NADP^+$. (6.5)

Nonpolar molecules Molecules whose covalent bonds have a nearly symmetric distribution of charge because the component atoms have approximately the same electronegativities. (2.1)

Nonrepeated fraction Those DNA sequences in the genome that are present in only one copy per haploid set of chromosomes. These sequences contain the greatest amount of genetic information including the codes for virtually all proteins other than histones. (10.4)

Nonsense mutations Mutations that produce stop codons within genes, thereby causing premature termination of the encoded polypeptide chain. (11.8)

Nonsense-mediated decay (NMD) An mRNA surveillance mechanism that detects mRNAs containing premature termination (nonsense) codons, leading to their destruction. (11.8)

Nontranscribed spacer The region of a gene cluster that is not transcribed. Nontranscribed spacers are present between various types of tandemly repeated genes, including those of tRNAs, rRNAs, and histones. (11.3)

Nuclear envelope The complex, double-membrane structure that divides the eukaryotic nucleus from its cytoplasm. (12.1)

Nuclear envelope breakdown The disassembly of the nuclear envelope at the end of prophase. (14.2)

Nuclear lamina A thin meshwork composed of intermediate filaments that lines the inner surface of the nuclear envelope. (12.1)

Nuclear localization signal (NLS) Sequence of amino acids in a protein that is recognized by a transport receptor leading to translocation of the protein from the cytoplasm to the nucleus. (12.1)

Nuclear matrix Insoluble fibrillar protein network that crisscrosses through nuclear space. (12.1)

Nuclear pore complex (NPC) Complex, basketlike apparatus that fills the nuclear pore like a stopper, projecting outward into both the cytoplasm and the nucleoplasm. (12.1)

Nucleic acid Polymers composed of nucleotides, which in living organisms are based on one of two sugars, ribose or deoxyribose, yielding the terms ribonucleic acid (RNA) and deoxyribonucleic acid (DNA). (2.5)

Nucleic acid hybridization A variety of related techniques that are based on the fact that two single-stranded nucleic acid molecules of complementary base sequence will form a double-stranded hybrid. (18.11)

Nucleoid The poorly defined region of a prokaryotic cell that contains its genetic material. (1.3)

Nucleoli (sing. nucleolus) Irregular-shaped nuclear structures that function as ribosome-producing organelles. (11.3)

Nucleosomes Repeating subunits of chromatin. Each nucleosome contains a nucleosome core particle, consisting of 146 base pairs of supercoiled DNA wrapped almost twice around a disc-shaped complex of eight histone molecules. The nucleosome core particles are connected to one another by a stretch of linker DNA. (12.1)

Nucleotide The monomer of nucleic acids, each consists of three parts: a sugar (either ribose or deoxyribose), a phosphate group, and a nitrogenous base, with the phosphate linked to the sugar at the 5′ carbon and the base at the 1′ carbon. (2.5, 10.3)

Nucleotide excision repair (NER) A cut-and-patch mechanism for the removal from the DNA of a variety of bulky lesions, e.g., pyrimidine dimers, caused by ultraviolet radiation. (13.2)

Nucleus The organelle that contains a eukaryotic cell's genetic material. (1.3)

Objective lens The lens of a light microscope that focuses light rays from the specimen to form a real, enlarged image of the object within the column of the microscope. (18.1)

Oils Fats that are liquid at room temperature. (2.5)

Okazaki fragments Small segments of DNA that are rapidly linked to longer pieces that have been synthesized previously, to form the lagging strand. (13.1)

Oligosaccharides Small chains composed of sugars covalently attached to lipids and proteins; they distinguish one type of cell from another and help mediate interactions of a cell with its surroundings. (2.5)

Oncogenes Genes that encode proteins that promote the loss of growth control and the conversion of the cell to a malignant state. These genes have the ability to transform cells. (16.3)

Operator Binding site for bacterial repressors that is situated between the polymerase binding site and the first structural gene. (12.2)

Operon A functional complex on a bacterial chromosome comprising a cluster of genes including structural genes, a promoter region, an operator region, and a regulatory gene. (12.2)

Organelles The organizationally and functionally diverse, membranous or membrane-bounded, intracellular structures that are the defining feature of eukaryotic cells. (1.3)

Origin of replication The specific site on the bacterial chromosome where replication begins. (13.1)

Osmosis The property of water passing through a semipermeable membrane from a region of lower solute concentration to one of higher solute concentration, with the tendency of eventually equalizing solute concentration in the two compartments. (4.7)

Oxidation The process through which an atom loses one or more electrons to another atom, in which the atom gaining electrons is considered to be reduced. (3.3)

Oxidation-reduction (redox) potential The separation of charge, measured in voltage, for any given pair of oxidizing-reducing agents, such as NAD^+ and NADH, relative to a standard couple (e.g., H^+ and H_2). (5.3)

Oxidation-reduction (redox) reaction One in which a change in the electronic state of the reactants occurs. (3.3)

Oxidative phosphorylation ATP formation driven by energy derived from high-energy electrons removed during substrate oxidation in pathways such as the TCA cycle, with the energy released for ATP formation by passage of the electrons through the electron-transport chain in the mitochondrion. (5.3)

Oxidizing agent The substance in a redox reaction that becomes reduced, causing the other substance to become oxidized. (3.3)

P (peptidyl) site Site in the ribosome from which the tRNA donates amino acids to the growing polypeptide chain. (11.8)

P680 The reaction center of photosystem II. "P" stands for pigment and "680" is the wavelength of light that this molecule absorbs most strongly. (6.4)

P700 The reaction center of photosystem I. "P" stands for pigment, and "700" is the wavelength of light that this molecule absorbs most strongly. (6.4)

Partition coefficient The ratio of a solute's solubility in oil to that in water, it is a measure of the relative polarity of a biological substance. (4.7)

Patch clamping A technique to study ion movement through ion channels that is accomplished by clamping, or maintaining voltage, across a patch of membrane by sealing a micropipette electrode to the surface and then measuring the current across that portion of membrane. (4.5)

Pathogen Any agent capable of causing infection or disease in a cell or organism. (2.5)

Pectins A heterogeneous class of negatively charged polysaccharides that make up the matrix of the plant cell wall. Pectins hold water and form a gel that fills in the spaces between the fibrous elements. (7.6)

Penetrance The probability that a given mutation will result in disease.

Peptide bond The chemical bond linking amino acids in a protein, which forms when the carboxyl group of one amino acid reacts with the amino group of a second amino acid. (2.5)

Peptidyl transferase That portion of the large ribosomal subunit that is responsible for catalyzing peptide bond formation; peptidyl transferase activity resides in the large ribosomal RNA molecule. (11.8)

Pericentriolar material (PCM) Amorphous, electron-dense material that surrounds the centrioles in an animal cell. (9.3)

Peripheral protein A membrane-associated protein that is located entirely outside of the lipid bilayer and interacts with it through noncovalent bonds. (4.4)

Peroxisomes (microbodies) Simple, membrane-bound, multi-functional organelles of the cytoplasm that carry out a diverse array of metabolic reactions, including substrate oxidation leading to formation of hydrogen peroxide. For example, peroxisomes are the site of oxidation of very-long-chain fatty acids, the oxidation of uric acid, and the synthesis of plasmalogens. Plant glyoxysomes, which carry out the glyoxylate cycle, are a type of peroxisome. (5.6)

pH The standard measure of relative acidity, it mathematically equals $-\log[H^+]$. (2.3)

PH domain A protein domain that binds to the phosphorylated inositol rings of membrane-bound phosphoinositides. (15.2)

Phagocytosis Process by which particulate materials are taken into cells. Materials are enclosed within a fold of the plasma membrane, which buds into the cytoplasm to form a vesicle called a phagosome. (8.8)

Phase-contrast microscope A microscope that converts differences in refractive index into differences in intensity (relative brightness and darkness), which are then visible to the eye, making highly transparent objects more visible. (18.1)

Phosphatidylinositol 3-hydroxy kinase [or PI3K] One of the best-studied effectors containing an SH2 domain. The products of the enzyme serve as inositol-containing cellular messengers that have diverse functions in cells. (15.4)

Phosphoglycerides The name given to membrane phospholipids that are built on a glycerol backbone. (4.2)

Phosphoinositides Includes a number of phosphorylated phosphatidylinositol derivatives (e.g., PIPs, PIP_2s, and PIP_3) that serve as second messengers in signaling pathways. (15.3)

Phospholipase C An enzyme that catalyzes a reaction that splits PIP_2 into two molecules: inositol 1,4,5,-trisphosphate (IP_3) and diacylglycerol (DAG), both of which play important roles as second messengers in cell signaling. (15.3)

Phospholipid-transfer proteins Proteins whose function is to transport specific phospholipids through the aqueous cytosol from one type of membrane compartment to another. (8.3)

Phospholipids Phosphate-containing lipids that represent the primary constituents of the lipid bilayer of cellular membranes. Phospholipids include both phosphoglycerides and sphingomyelin. (4.3)

Photoautotroph An autotroph that utilizes the radiant energy of the sun to convert CO_2 into organic compounds. (6)

Photolysis The splitting of water during photosynthesis. (6.4)

Photon Packet of light energy. The shorter the wavelength, the greater the energy of the photons. (6.3)

Photorespiration A series of reactions in which O_2 is attached to RuBP, and eventually resulting in the release of recently fixed CO_2 from the plant. (6.6)

Photosynthesis The pathway converting the energy of sunlight into chemical energy that is usable by living organisms. (3.16)

Photosynthetic unit A group of several hundred chlorophyll molecules acting together to trap photons and transfer energy to the pigment molecule at the reaction center. (6.4)

Photosystem I (PSI) One of two spatially separated pigment complexes, which are necessary to boost the energy of a pair of electrons sufficiently to remove them from a molecule of water and transfer them to $NADP^+$. Photosystem I raises electrons from an energy level at about the midway point to an energy level above $NADP^+$. (6.4)

Photosystem II (PSII) One of two spatially separated pigment complexes, which are necessary to boost the energy of a pair of electrons sufficiently to remove them from a molecule of water and transfer them to $NADP^+$. Photosystem II boosts electrons from an energy level below that of water at the bottom of the energy trough to about the midway point. (6.4)

Phragmoplast Dense material roughly aligned in the equatorial plane of the previous metaphase plate in plant cells, consisting of clusters of interdigitating microtubules oriented perpendicular to the future cell plate, together with vesicles and associated electron-dense material. (14.3)

Pigments Molecules that contain a chromophore, a chemical group capable of absorbing light of particular wavelength(s) within the visible spectrum. (6.3)

PiwiRNAs (piRNAs) Small RNAs (24-32 bases) that are encoded by a small number of large genomic loci and act to suppress the movement of transposable elements in germ cells. piRNAs are derived from single-stranded precursors and do not require Dicer for processing. (11.5)

Plasma cells Terminally differentiated cells that develop from B lymphocytes that synthesize and secrete large amounts of blood-borne antibodies. (17.2)

Plasma membrane The membrane serving as a boundary between the interior of a cell and its extracellular environment. (4.1)

Plasmodesmata Cytoplasmic channels, 30 to 60 nm in diameter, that connect most plant cells and extend between adjacent cells directly through the cell wall. Plasmodesmata are lined with plasma membrane and usually contain a dense central structure, the desmotubule, derived from the endoplasmic reticulum of the two cells. (7.5)

Plasmolysis The shrinkage that occurs when a plant cell is placed into a hypertonic medium; its volume shrinks as the plasma membrane pulls away from the surrounding cell wall. (4.7)

Polar molecules Molecules with an uneven distribution of charge because the component atoms of various bonds have greatly different electronegativities. (2.1)

Poly(A) tail A string of adenosine residues at the 3′ end of an mRNA added posttranscriptionally. (11.4)

Polyacrylamide gel electrophoresis (PAGE) Protein fractionation technique in which the proteins are driven by an applied current through a gel composed of a small organic molecule (acrylamide) that is cross-linked to form a molecular sieve. (18.7)

Polymerase chain reaction (PCR) A technique in which a single region of DNA which may be present in vanishingly small amounts, can be amplified cheaply and rapidly. (18.14)

Polypeptide chain A long, continuous unbranched polymer formed by amino acids joined to one another by covalent peptide bonds. (2.5)

Polyploidization (whole-genome duplication) Phenomenon in which offspring have twice the number of chromosomes in each cell as their diploid parents. Can be an important step in the evolution of a new species. (10.5)

Polyribosome (polysome) The complex formed by an mRNA and a number of ribosomes in the process of translating that mRNA. (11.8)

Polysaccharide A polymer of sugar units joined by glycosidic bonds. (2.5)

Polytene chromosomes Giant chromosomes of insects that contain perfectly aligned, duplicated DNA strands, with as many as 1,024 times the number of DNA strands of normal chromosomes. (10.1)

Porins Integral proteins found in bacterial, mitochondrial and chloroplast outer membranes that act as large, relatively nonselective channels. (5.1)

Posttranslational modifications (PTMs) Alterations to the side chains of the 20 basic amino acids after their incorporation into a polypeptide chain. (2.5)

Potential difference The difference in charge between two compartments, often measured as voltage across the separating membrane. (4.7)

Potential energy Stored energy that can be used to perform work. (3.1)

Preinitiation complex The assembled association of general transcription factors and RNA polymerase, required before transcription of the gene can be initiated. (11.4)

Pre-RNA An RNA molecule that has not yet been processed into its final mature form (e.g., a **pre-mRNA**, **pre-rRNA**, or **pre-tRNA**). (11.4)

Primary cilium A single nonmotile cilium present on many types of cells in vertebrates and thought to have a sensory function. (HP9)

Primary culture Culturing of cells obtained directly from the organism. (18.5)

Primary electron acceptor Molecule that receives the photoexcited electron from reaction-center pigments in both photosystems. (6.4)

Primary structure The linear sequence of amino acids within a polypeptide chain. (2.5)

Primary transcript (or pre-RNA) The initial RNA molecule synthesized from DNA, which is equivalent in length to the DNA from which it was transcribed. Primary transcripts typically have a fleeting existence, being processed into smaller, functional RNAs by a series of "cut-and-paste" reactions. (11.2)

Primary cell walls The walls of a growing plant cell. They allow for extensibility. (7.6)

Primase Type of RNA polymerase that assembles the short RNA primers that begin the synthesis of each Okazaki fragment of the lagging strand. (13.1)

Primer The DNA or RNA strand that provides DNA polymerase with the necessary 3′ OH terminus. (13.1)

Prion An infectious agent associated with certain mammalian neurodegenerative diseases that is composed solely of protein. (HP2)

Processing-level control Regulation of the path by which a primary RNA transcript is **processed** into a messenger RNA that can be translated into a polypeptide. (12.4)

Processive A term applied to proteins (e.g., kinesin or RNA polymerase) that are capable of moving considerable distances along their track or template (e.g., a microtubule or a DNA molecule) without dissociating from it. (9.3, 11.2)

Prokaryotic cells Structurally simple cells, including archaea and bacteria that do not have membrane-bounded organelles; derived from *pro-karyon*, or "before the nucleus." (1.3)

Prometaphase The phase of mitosis during which the definitive mitotic spindle is formed and the chromosomes are moved into position at the center of the cell. (14.2)

Promoter The site on the DNA to which an RNA polymerase molecule binds prior to initiating transcription. The promoter contains information that determines which of the two DNA strands is transcribed and the site at which transcription begins. (11.2, 12.2)

Prophase The first stage of mitosis during which the duplicated chromosomes are prepared for segregation and the mitotic machinery is assembled. (14.2)

Proplastids Nonpigmented precursors of chloroplasts. (6.1)

Prosthetic group A portion of a protein that is not composed of amino acids, such as the heme group within hemoglobin and myoglobin. (2.5)

Proteasome Barrel-shaped, multiprotein complex in which cytoplasmic proteins are degraded. Proteins selected for destruction are linked to ubiquitin molecules and threaded into the central chamber of the proteasome. (12.7)

Protein kinase An enzyme that transfers phosphate groups to other proteins, often having the effect of regulating the activity of the other proteins. (3.3)

Protein tyrosine kinases Enzymes that phosphorylate specific tyrosine residues of other proteins. (15.4)

Proteins Structurally and functionally diverse group of polymers built of amino acid monomers. (2.5)

Proteoglycan A protein-polysaccharide complex consisting of a core protein molecule to which chains of glycosaminoglycans are attached. Due to the acidic nature of the glycosaminoglycans, proteoglycans are capable of binding huge numbers of cations, which in turn draw huge numbers of water molecules. As a result, proteoglycans form a porous, hydrated gel that acts like a "packing" material to resist compression. (7.1)

Proteome The entire inventory of proteins in a particular organism, cell type, or organelle. (2.5)

Proteomics Expanding field of protein biochemistry that performs large-scale studies on diverse mixtures of proteins. (2.5)

Protofilaments Longitudinally arranged rows of globular subunits of a microtubule that are aligned parallel to the long axis of the tubule. (9.3)

Proton-motive force (Δp) An electrochemical gradient that is built up across energy-transducing membranes (inner mitochondrial membrane, thylakoid membrane, bacterial plasma membrane) following the translocation of protons during electron transport. The energy of the gradient, which is comprised of both a pH gradient and a voltage and is measured in volts, is utilized in the formation of ATP. (5.4)

Proto-oncogenes A variety of genes that have the potential to subvert the cell's own activities and push the cell toward the malignant state. Proto-oncogenes encode proteins that have various functions in a cell's normal activities. Proto-oncogenes can be converted to oncogenes. (16.3)

Protoplast A naked plant cell whose cell wall has been digested away by the enzyme cellulase. (18.5)

Provirus The term for viral DNA when it has been integrated into the DNA of its host cell's chromosome(s). (1.4)

Pseudogenes Sequences that are clearly homologous to functional genes, but have accumulated mutations that render them nonfunctional. (10.5)

Pseudopodia Broad, rounded protrusions formed during amoeboid movement as portions of the cell surface are pushed outward by a column of cytoplasm that flows through the interior of the cell toward the periphery. (9.7)

Purine A class of nitrogenous base found in nucleotides that has a double-ring structure, including **adenine** and **guanine**, which are found in both DNA and RNA. (2.5, 10.3)

Pyrimidine A class of nitrogenous base found in nucleotides that has a single-ring structure, including **cytosine** and **thymine**, which are found in DNA, and cytosine and **uracil**, which are found in RNA. (2.5, 10.3)

Quality control Cells contain various mechanisms that ensure that the proteins and nucleic acids they synthesize have the appropriate structure. For example, misfolded proteins are translocated out of the ER and destroyed by proteasomes in the cytosol; mRNAs that contain premature termination codons are recognized and destroyed; and DNA containing abnormalities (lesions) are recognized and repaired. (e.g., 8.3)

Quaternary structure The three-dimensional organization of a protein that consists of more than one polypeptide chain, or subunit. (2.5)

Rabs A family of monomeric G proteins involved in vesicle trafficking. (8.5)

Ran A GTP-binding protein that exists in an active GTP-bound form or an inactive GDP-bound form. Ran regulates nucleocytoplasmic transport. (12.1)

Ras-MAP kinase cascade A cascade that is turned on in response to a wide variety of extracellular signals and plays a key role in regulating vital activities such as cell proliferation and differentiation. (15.4)

rDNA The DNA sequences encoding rRNA that are normally repeated hundreds of times and are typically clustered in one or a few regions of the genome. (11.3)

Reaction-center chlorophyll The single chlorophyll molecule of the several hundred or so in the photosynthetic unit that actually transfers electrons to an electron acceptor. (6.4)

Reannealing (renaturation) Reassociation of complementary single strands of a DNA double helix that had been previously denatured. (10.3)

Receptor Any substance that can bind to a specific molecule (ligand), often leading to uptake or signal transduction. (4.1, 15.1)

Receptor protein-tyrosine kinases (or RTKs) Cell-surface receptors that, following ligand binding, can phosphorylate tyrosine residues on themselves and/or on cytoplasmic substrates. They are involved primarily in the control of cell growth and differentiation. (15.4)

Recombinant DNA Molecules containing DNA sequences derived from more than one source. (18.13)

Reducing agent The substance in a redox reaction that becomes oxidized, causing the other substance to become reduced. (3.3)

Reducing power The potential in a cell to reduce metabolic intermediates into

products, usually measured through the size of the NADPH pool. (3.3)

Reduction The process through which an atom gains one or more electrons from another atom, in which the atom losing electrons is considered to be oxidized. (3.3)

Refractory period The brief period of time following the end of an action potential during which an excitable cell cannot be restimulated to threshold. (4.8)

Regulated secretion Discharge of materials synthesized in the cell that have been stored in membrane-bound secretory granules in the peripheral regions of the cytoplasm, occurring in response to an appropriate stimulus. (8.1)

Regulatory gene Gene that codes for a bacterial repressor protein. (12.2)

Renaturation (reannealing) Reassociation of complementary single-stranded DNA molecules that had been previously denatured. (10.4)

Replica Metal-carbon cast of a tissue surface used in electron microscopy. Variations in the thickness of the metal in different parts of the replica cause variations in the number of penetrating electrons to reach the viewing screen. (18.2)

Replication Duplication of the genetic material. (13)

Replication foci Localization of active replication forks in the cell nucleus. There are about 50 to 250 foci, each of which contains approximately 40 replication forks incorporating nucleotides into DNA strands simultaneously. (13.1)

Replication forks The points at which the pair of replicated segments of DNA come together and join the nonreplicated segments. Each replication fork corresponds to a site where (1) the parental double helix is undergoing strand separation, and (2) nucleotides are being incorporated into the newly synthesized complementary strands. (13.1)

Repressor A **gene regulatory protein** that binds to DNA and inhibits transcription. (12.2)

Resolution The ability to see two neighboring points in the visual field as distinct entities. (18.1)

Response elements The sites at which specific transcription factors bind to the regulatory regions of a gene. (12.4)

Resting potential The electrical potential difference measured for an excitable cell when it is not subject to external stimulation. (4.8)

Restriction endonucleases (restriction enzymes) Nucleases contained in bacteria that recognize short nucleotide sequences within duplex DNA and cleave the backbone at highly specific sites on both strands of the duplex. (18.13)

Restriction map A type of physical map of the chromosome based on the identification

and ordering of sets of fragments generated by restriction enzymes. (10.5)

Retrotransposons Transposable elements that require reverse transcriptase for their movements within the genome. (10.5)

Reverse transcriptase An RNA-dependent DNA polymerase. An enzyme that uses RNA as a template to synthesize a complementary strand of DNA. [An enzyme that is found in RNA-containing viruses and used in the laboratory to synthesize cDNAs.] (10.5)

Ribonucleic acid (RNA) A single-stranded nucleic acid composed of a polymeric chain of ribose-containing nucleotides. (2.5)

Ribosomal RNAs (or rRNAs) The RNAs of a ribosome. rRNAs recognize and bind other molecules, provide structural support, and catalyze the chemical reaction in which amino acids are covalently linked to one another. (11.1)

Riboswitches mRNAs that, once bound to a metabolite, undergo a change in their folded conformation that allows them to alter the expression of a gene involved in production of that metabolite. Most riboswitches suppress gene expression by blocking either termination of transcription or initiation of translation. (12.3)

Ribozyme An RNA molecule that functions as a catalyst in cellular reactions. (2.5)

RNA interference (RNAi) A naturally occurring phenomenon in which double-stranded RNAs (dsRNAs) lead to the degradation of mRNAs having identical sequences. RNAi is believed to function primarily in blocking the replication of viruses and restricting the movement of mobile elements, both of which involve the formation of dsRNA intermediates. Mammalian cells can be made to engage in RNAi by treatment of the cells with small (21 nt) RNAs. These small RNAs (siRNAs) induce the degradation of mRNAs that contain the same sequence. (11.5)

RNA polymerase I The transcribing enzyme found in eukaryotic cells that synthesizes the large (28S, 18S, and 5.8S) ribosomal RNAs. (11.3)

RNA polymerase II The transcribing enzyme found in eukaryotic cells that synthesizes messenger RNAs and most small nuclear RNAs. (11.4)

RNA polymerase III The transcribing enzyme found in eukaryotic cells that synthesizes the various transfer RNAs and the 5S ribosomal RNA. (11.3)

RNA silencing A process in which small, noncoding RNAs, typically derived from longer double-stranded precursors, trigger sequence-specific inhibition of gene expression. (11.5)

RNA splicing The process of removing the intervening DNA sequences (introns) from a primary transcript. (11.4)

RNA tumor viruses Retroviruses capable of infecting vertebrate cells, transforming them into cancer cells. RNA viruses have RNA in the mature virus particle. (16.2)

RNA world A proposed stage in the early evolution of life before the appearance of DNA and proteins in which RNA molecules served both as genetic material and catalysts. (11.4)

Rough endoplasmic reticulum (RER) That part of the endoplasmic reticulum that has ribosomes attached. The RER appears as an extensive membranous organelle composed primarily of flattened sacs (**cisternae**) separated by a cytosolic space. RER functions include synthesis of secretory proteins, lysosomal proteins, integral membrane proteins, and membrane lipids. (8.3)

S phase The phase of the cell cycle in which replication occurs. (14.1)

Saltatory conduction Propagation of a nerve impulse when one action potential triggers another at the adjacent stretch of unwrapped membrane (i.e., propagating by causing the action potentials to jump form one node of Ranvier to the next). (4.8)

Sarcomeres Contractile units of myofibrils that are endowed with a characteristic pattern of bands and stripes that give skeletal muscle cells their striated appearance. (9.6)

Sarcoplasmic reticulum (SR) A system of cytoplasmic, Ca^{2+}-storing SER membranes in muscle cells that forms a membranous sleeve around the myofibril. (9.6)

Saturated fatty acids Those lacking double bonds between carbons. (2.5)

Second law of thermodynamics Events in the universe proceed from a state of higher energy to a state of lower energy. (3.1)

Second messenger A substance that is formed in the cell as the result of the binding of a first messenger—a hormone or other ligand—to a receptor at the outer surface of the cell. (15.3)

Secondary culture Transfer of previously cultured cells to a culture medium. (18.5)

Secondary structure The three-dimensional arrangement of portions of a polypeptide chain. (2.5)

Secondary walls Thicker cell walls found in most mature plant cells. (7.6)

Secreted Discharged outside the cell. (8.1)

Secretory granule Large, densely packed, membrane-bound structure containing highly concentrated secretory materials that are discharged into the extracellular space (secreted) following a stimulatory signal. (8.1, 8.5)

Secretory pathway (biosynthetic pathway) Route through the cytoplasm by which materials are synthesized in the endoplasmic reticulum or Golgi complex, modified during passage through the Golgi complex, and transported within the cytoplasm to various destinations such as the plasma membrane, a lysosome, or a large vacuole of a plant cell. Many of the materials synthesized in the endoplasmic reticulum or Golgi complex are destined to be discharged outside the cell; hence the term secretory pathway has been used. (8.1)

Section A very thin slice of tissue. (18.1)

Selectins A family of integral membrane glycoproteins that recognize and bind to specific arrangements of carbohydrate groups projecting from the surface of other cells. (7.3)

Selectively permeable barrier Any structure, such as a plasma membrane, that allows some substances to freely pass through while denying passage to others. (4.1)

Self-assembly The property of proteins (or other structures) to assume the correct (native) conformation based on the chemical behavior dictated by the amino acid sequence. (2.5)

Semiconservative Replication in which each daughter cell receives one strand of the parent DNA helix. (13.1)

Semipermeable The membrane property of being freely permeable to water while allowing much slower passage to small ions and polar solutes. (4.7)

Serial sections A series of successive sections cut from a block of tissue. (18.1)

SH2 domains Domains with high-affinity binding sites for phosphotyrosine motifs. Found in a variety of proteins involved in cell signaling. (15.4)

side chain or R group The defining functional group of an amino acid, which can range from a single hydrogen to complex polar or non-polar units in the 20 amino acids most commonly found in cells. (2.5)

Signal peptidase Proteolytic enzyme that removes the N-terminal portion including the signal peptide of a nascent polypeptide synthesized in the RER. (8.3)

Signal recognition particle (SRP) A particle consisting of six distinct polypeptides and a small RNA molecule, called the 7S RNA, that recognizes the signal sequence as it emerges from the ribosome. SRP binds to the signal sequence and then to an ER membrane. (8.3)

Signal sequence Special series of amino acids located at the N-terminal portion of newly forming proteins that triggers the attachment of the protein-forming ribosome to an ER membrane and the movement of the nascent polypeptide into the cisternal space of the ER. (8.3)

Signal transduction The overall process in which information carried by extracellular messenger molecules is translated into changes that occur inside a cell. (15.1)

Signaling pathways The information superhighways of the cell. Each consists of a series of distinct proteins that operate in sequence. Each protein in the pathway acts by altering the conformation of the downstream protein in the series. (15.1)

Single nucleotide polymorphisms (SNPs) Sites in the genome where alternate bases are found with high frequency in the population. SNPs are excellent genetic markers for genome mapping studies. (10.6)

Single-particle tracking (SPT) A technique for studying movement of membrane proteins that consists of two steps: (1) linking the protein molecules to visible substances such as colloidal gold particles and (2) monitoring the movements of the individual tagged particles under the microscope. (4.6)

Single-stranded DNA-binding (or SSB) proteins Proteins that facilitate the separation of the DNA strands by their attachment to bare, single DNA strands, keeping them in an extended state and preventing them from becoming rewound. (13.1)

Site-directed mutagenesis A research technique to modify a gene in a predetermined way so as to produce a protein with a specifically altered amino acid sequence. (18.18)

Sliding clamp A ring-shaped protein that plays a key role in DNA replication by encircling the DNA and imparting processivity to the replicative DNA polymerase. (13.1)

Small interfering RNAs (siRNAs) Small (21-23 nucleotide), double-stranded fragments formed when double-stranded RNA initiates the response during RNA silencing. (11.5)

Small nuclear RNAs (snRNAs) RNAs required for mRNA processing that are small (90 to 300 nucleotides long) and that function in the nucleus. (11.4)

Small-nucleolar RNAs (snoRNAs) RNAs required for the methylation and pseudouridylation of pre-rRNAs during ribosome formation in the nucleolus. (11.3)

snoRNPs (small, nucleolar ribonucleoproteins) Particles that are formed when snoRNAs are packaged with particluar proteins; snoRNPs play a role in the maturation and assembly of ribosomal RNAs. (11.3)

Smooth endoplasmic reticulum (SER) That part of the endoplasmic reticulum that is without attached ribosomes. The membranous elements of the SER are

typically tubular and form an interconnecting system of pipelines curving through the cytoplasm in which they occur. The SER functions vary from cell to cell and include the synthesis of steroid hormones, detoxification of a wide variety of organic compounds, mobilization of glucose from glucose 6-phosphate, and sequestration of calcium ions. (8.3)

SNAREs   Key proteins that mediate the process of membrane fusion. **T-SNARE**s are located in the membranes of target compartments. **V-SNARE**s incorporate into the membranes of transport vesicles during budding. (8.5)

snRNPs   Distinct ribonucleoprotein particles contained in spliceosomes, so called because they are composed of snRNAs bound to specific proteins. (11.4)

Sodium-potassium pump (Na^+/K^+-ATPase)   Transport protein that uses ATP as the energy source for transporting sodium and potassium ions, with the result that each conformational change transports three sodium ions out of the cell and two potassium ions into the cell. (4.7)

Somatic cells   Cells of the body, excluding cells of the germ line (i.e., those that can give rise to gametes).

Specific activity   The ratio of the amount of a protein of interest to the total amount of protein present in a sample, which is used as a measure of purification. (18.7)

Specificity   The property of selective interaction between components of a cell that is basic to life. (2.5)

Spectrophotometer   Instrument used to measure the amount of light of a specific wavelength that is absorbed by a solution. If one knows the absorbance characteristics of a particular type of molecule, then the amount of light of the appropriate wavelength absorbed by a solution of that molecule provides a sensitive measure of its concentration. (18.7)

Sphingolipids   A class of membrane lipids—derivations of sphingosine—that consist of sphingosine linked to a fatty acid by its amino group. (4.3)

Spindle checkpoint   A checkpoint that operates at the transition between metaphase and anaphase; the spindle checkpoint is best revealed when a chromosome fails to become aligned properly at the metaphase plate. (14.2)

Splice sites   The 5′ and 3′ ends of each intron. (11.4)

Spliceosome   A macromolecular complex containing a variety of proteins and a number of distinct ribonucleoprotein particles that functions in removal of introns from a primary transcript. (11.4)

Split genes   Genes with intervening sequences. (11.4)

Spontaneous reactions   Reactions that are thermodynamically favorable, capable of occurring without any input of external energy. (3.1)

Sporophyte   A diploid stage of the life cycle of plants that begins with the union of two gametes to form a zygote. During the sporophyte stage, meiosis occurs, producing spores that germinate directly into a haploid gametophyte. (14.3)

SRP receptor   Situated within the ER membrane, the SRP receptor binds specifically with the SRP-ribosome complex. (8.3)

Standard free-energy change ($\Delta G°′$)   The change in free energy when one mole of each reactant is converted to one mole of each product under defined standard conditions: temperature of 298 K and pressure of 1 atm. (3.1)

Starch   Mixture of two glucose polymers, amylose and amylopectin, that serves as readily available chemical energy in most plant cells. (2.5)

Steady state   Metabolic condition in which concentrations of reactants and products are essentially constant, although individual reactions may not be at equilibrium. (3.1)

Stem cells   Cells situated in various tissues of the body that constitute a reserve population capable of giving rise to the various cells of that tissue. Stem cells can be defined as undifferentiated cells that are capable of both (1) self-renewal, that is, production of cells like themselves, and (2) differentiation into two or more mature cell types. (HP1)

Stereoisomers   Two molecules that structurally are mirror images of each other and may have vastly different biological activity. (2.5)

Steroid   Lipid molecule based on a characteristic four-ring hydrocarbon skeleton, including cholesterol and hormones such as testosterone and progesterone. (2.5)

Stomata   Openings at the surface of leaves through which gas and water are exchanged between the plant and the air. (6.6)

Stop codons   Three of the 64 possible trinucleotide codons whose function is to terminate polypeptide assembly. (11.6)

Stroma   Space outside the thylakoid but within the relatively impermeable inner membrane of the chloroplast envelope. (6.1)

Stroma thylakoids   Also called stroma lamellae, these are flattened membranous cisternae that connect some of the thylakoids of one granum with those of another. (6.1)

Structural genes   Genes that code for protein molecules. (12.2)

Structural isomers   Molecules having the same chemical formula but different structures. (2.4)

Subcellular fractionation   An approach that allows different organelles (e.g. nucleus, mitochondrion, plasma membrane, endoplasmic reticulum) having different properties, to be separated from one another. (8.2)

Substrate   The reactant bound by an enzyme. (3.2)

Substrate-level phosphorylation   Direct synthesis of ATP through the transfer of a phosphate group from a substrate to ADP. (3.3)

Subunit   A polypeptide chain that associates with other chains (subunits) to form a complete protein or protein complex. (2.5)

Supercoiled   A molecule of DNA that has greater or fewer than 10 base pairs per turn of the helix. (10.3)

Surface area/volume ratio   The proportion between cellular dimensions that indicates how effectively (or whether) a cell can sustainably exchange substances with its environment. (1.3)

Synapse   The specialized junction of a neuron with its target cell. (4.8)

Synapsis   The process by which homologous chromosomes become joined to one another during meiosis. (14.3)

Synaptic cleft   The narrow gap between two excitable cells. A **presynaptic cell** conducts impulses toward a synapse; a **postsynaptic cell** always lies on the receiving side of a synapse. (4.8)

Synaptic vesicles   The storage sites for neurotransmitter within the terminal knobs of a neuronal axon. (4.8)

Synaptonemal complex (SC)   A ladderlike structure composed of three parallel bars with many cross fibers. The SC holds each pair of homologous chromosomes in the proper position to allow the continuation of genetic recombination between strands of DNA. (14.3)

tDNA   The DNA encoding tRNAs. (11.3)

T-cell receptor (TCR)   Proteins present on the surface of T lymphocytes that mediate interaction with specific cell-bound antigens. Like the immunoglobulin of B cells, these proteins are formed by a process of DNA rearrangement that generates a specific antigen-combining site. TCRs consist of two subunits, each containing both a variable and a constant domain. (17.3)

T lymphocytes (T cells)   Lymphocytes that respond to antigen by proliferating and differentiating into either CTLs (cytotoxic lymphocytes) that attack and kill infected cells or T_H cells that are required for antibody production by B cells. These cells attain their differentiated state in the thymus. (17.2)

Tandem repeats A cluster in which a DNA sequence repeats itself over and over again without interruption. (10.3)

Telomere An unusual stretch of repeated DNA sequences, which forms a "cap" at each end of a chromosome. (12.1)

Telomerase A novel enzyme that can add new repeat units of DNA to the 3′ end of the overhanging strand of a telomere. Telomerase is a reverse transcriptase that synthesizes DNA using an RNA template. (12.1)

Telophase The final stage of mitosis in which daughter cells return to the interphase condition: the mitotic spindle disassembles, the nuclear envelope reforms, and the chromosomes become more and more dispersed until they disappear from view under the microscope. (14.2)

Temperature-sensitive (ts) mutations Mutations that are only expressed phenotypically when the cells (or organism) are grown at a higher (restrictive) temperature. At the lower (permissive) temperature, the encoded protein is able to hold together sufficiently well to carry out its activity, leading to a relatively normal phenotype. ts mutations are particularly useful for studying required activities such as secretion and replication, because "ordinary" mutations affecting these processes are typically lethal. (13.1)

Template A single strand of DNA (or RNA) that contains the information (encoded as a nucleotide sequence) for construction of a complementary strand. (13.1)

Tertiary structure The three dimensional shape of an entire macromolecule. (2.5)

Tetrad (bivalent) The complex formed during meiosis by a pair of synapsed homologous chromosomes that includes four chromatids. (14.3)

Thermodynamics Study of the changes in energy accompanying events in the physical universe. (3.1)

Thermodynamics, first law of Principle of conservation of energy that states energy can neither be created nor destroyed, merely **transduced** (converted) from one form to another. (3.1)

Thermodynamics, second law of Principle that all events move from a higher energy state to a lower energy state, and thus are spontaneous. (3.1)

Thick filaments One of two distinct types of filaments that give sarcomeres their characteristic appearance. Thick filaments consist primarily of myosin and are surrounded by a hexagonal array of thin filaments. (9.6)

Thin filaments One of two distinct types of filaments that give sarcomeres their characteristic appearance. Thin filaments consist primarily of actin and are arranged in a hexagonal array around each thick filament, with each thin filament situated between two thick filaments. (9.6)

Threshold The point during depolarization of an excitable cell where voltage-gated sodium channels open, with the resulting Na^+ influx causing a brief reversal in membrane potential. (4.8)

Thylakoids Flattened, membranous sacs formed by the chloroplast's internal membrane, which contain the energy-transducing machinery for photosynthesis. (6.1)

Tight junctions Specialized contacts that occur at the very apical end of the junctional complex between adjacent epithelial cells. The adjoining membranes make contact at intermittent points, where integral proteins of the two adjacent membranes meet. (7.4)

Toll-like receptors (TLRs) A type of pathogen receptor of the innate immune system. Humans express at least 10 functional TLRs, all of which are transmembrane proteins present on the surfaces of many different types of cells. (17.1)

Tonoplast The membrane that bounds the vacuole of a plant cell. (8.7)

Topoisomerases Enzymes found in both prokaryotic and eukaryotic cells that are able to change the supercoiled state of the DNA duplex. They are essential in processes, such as DNA replication and transcription, that require the DNA duplex to unwind. (10.3)

***Trans* cisternae** The cisternae of the Golgi complex farthest from the endoplasmic reticulum. (8.4)

***Trans* Golgi network (TGN)** A network of interconnected tubular elements at the *trans* end of the Golgi complex that sorts and targets proteins for delivery to their ultimate cellular or extracellular destination. (8.4)

Transcription The formation of a complementary RNA from a DNA template. (11.1)

Transcription factors Auxiliary proteins (beyond the polypeptides that make up the RNA polymerases) that bind to specific sties in the DNA and alter the transcription of nearby genes. (11.2, 12.4)

Transcription unit The corresponding segment of DNA on which a primary transcript is transcribed. (11.4)

Transcriptional-level control Determination whether a particular gene can be transcribed and, if so, how often. (12.4)

Transcriptome The entire inventory of RNAs transcribed by a particular cell, tissue, or organism. (11.5)

Transduction The incorporation of a gene into a cellular genome by means of a virus. (18.7)

Transfection A process by which naked DNA is introduced into cultured cells typically leading to the incorporation of the DNA into the cellular genome and its subsequent expression. (18.7)

Transfer potential A measure of the ability of a molecule to transfer any group to another molecule, with molecules having a higher affinity for the group being the better acceptors and molecules having a lower affinity better donors. (3.3)

Transfer RNAs (tRNAs) A family of small RNAs that translate the information encoded in the nucleotide "alphabet" of an mRNA into the amino acid "alphabet" of a polypeptide. (11.1)

Transformation Uptake of naked DNA into a cell leading to a heritable change in the cell's genome. (HP10)

Transformed Normal cells that have been converted to cancer cells by treatment with a carcinogenic chemical, radiation, or an infective tumor virus. (16.1)

Transgene A gene that has been stably incorporated into a cellular genome by the process of transfection. (18.7)

Transgenic animals Animals that have been genetically engineered so that their chromosomes contain foreign genes. (18.13)

Transition state The point during a chemical reaction at which bonds are being broken and reformed to yield products. (3.2)

Transition temperature The temperature at which a membrane is converted from a fluid state to a crystalline gel in which lipid-molecule movement is greatly reduced. (4.5)

Translation Synthesis of proteins in the cytoplasm using the information encoded by an mRNA. (11.1, 11.8)

Translational-level control Determination whether a particular mRNA is actually translated and, if so, how often and for how long a period. (12.6)

Translesion synthesis Replication that bypasses a lesion in the template strand. Conducted by a special group of DNA polymerases that lack processivity, proofreading capacity, and high fidelity. (13.3)

Translocation A chromosomal aberration that results when all or part of one chromosome becomes attached to another chromosome. (HP12)

Translocation The step in the translation elongation cycle that involves (1) ejection of the uncharged tRNA from the P site and (2) the movement of the ribosome three nucleotides (one codon) along the mRNA in the 3′ direction. (11.8)

Translocon A protein-lined channel embedded in the ER membrane; the nascent polypeptide is able to move through the translocon in its passage from the cytosol to the ER lumen. (8.3)

Transmembrane domain The portion of a membrane protein that passes through the lipid bilayer, often composed of non-polar amino acids in an α-helical conformation. (4.4)

Transmembrane signaling Transfer of information across the plasma membrane. (7.3, 15)

Transmission electron microscopes (TEMs) Microscopes that form images from electrons that are transmitted through a specimen. (18.2)

Transport vesicles The shuttles, formed by budding from a membrane compartment, that carry materials between organelles. (8.1)

Transposable elements DNA segments that move from one place on a chromosome to a completely different site, often affecting gene expression. (10.5)

Transposition Movement of DNA segments from one place on a chromosome to an entirely different site, often affecting gene expression. (10.5)

Transposons DNA segments capable of moving from one place in the genome to another. (10.5)

Transverse (T) tubules Membranous folds along which the impulse generated in a skeletal muscle cell is propagated into the interior of the cell. (9.6)

Triacylglycerols Polymers consisting of a glycerol backbone linked by ester bonds to three fatty acids, commonly called fats. (2.5)

Tricarboxylic acid cycle (TCA cycle) The circular metabolic pathway that oxidizes acetyl CoA, conserving its energy; the cycle is also known as the Krebs cycle or the citric acid cycle. (5.2)

Trisomy A chromosome complement that has one extra chromosome, i.e., a third homologous chromosome. (HP14)

Tubulin The protein that forms the walls of microtubules. Isoforms include α, β, and γ tubulin. (9.3)

Tumor-suppressor genes Genes that encode proteins that restrain cell growth and prevent cells from becoming malignant. (16.3)

Turgor pressure Hydrostatic pressure that builds up in a plant cell due to the hypertonic state of the intracellular compartment. Turgor pressure is exerted against the surrounding cell wall and provides support for plant tissues. (4.7)

Turnover number The maximum number of substrate molecules that can be converted to product by one enzyme molecule per unit of time. (3.2)

Turnover The regulated destruction of cellular materials and their replacement. (8.7)

Ubiquinone A component of the electron transport chain, ubiquinone is a lipid-soluble molecule containing a long hydrophobic chain composed of five-carbon isoprenoid units. (5.3)

Ubiquitin A small, highly conserved protein that is linked to proteins targeted for internalization by endocytosis or degradation in proteasomes. (8.8, 12.7)

Unconventional myosins (See Myosin.)

Unfolded protein response (UPR) A comprehensive response that occurs in cells whose ER cisternae contain an excessively high concentration of unfolded or misfolded proteins. Sensors that detect this situation trigger a pathway that leads to the synthesis of proteins (e.g., molecular chaperones) that can alleviate the stress in the ER. (8.3)

Unsaturated fatty acids Those having one or more double bonds between carbon atoms. (2.5)

Untranslated regions (UTRs) Noncoding segments contained at both 5′ and 3′ ends of mRNAs. (12.6)

Vacuole A single membrane-bound, fluid filled structure that comprises as much as 90% of the volume of many plant cells. (8.7)

Van der Waals force A weak attractive force due to transient asymmetries of charge within adjacent atoms or molecules. (2.2)

Variable regions The portions of light and heavy antibody polypeptide chains that differ in amino acid sequence from one specific antibody to another. (17.4)

V(D)J joining The DNA rearrangements that occur during the development of B cells that limit the cells to the production of a specific antibody species. (17.4)

Vector DNA A vehicle for carrying foreign DNA into a suitable host cell, such as the bacterium *E. coli*. The vector contains sequences that allow it to be replicated inside the host cell. Most often, the vector is a plasmid or the bacterial virus lambda (λ). Once the DNA is inside the bacterium, it is replicated and partitioned to the daughter cells. (18.13)

Virion The form a virus assumes outside of a cell, which consists of a core of genetic material surrounded by a protein or lipoprotein capsule. (1.4)

Viroids Small, obligatory intracellular pathogens, that, unlike viruses, consist only of an uncoated circle of genetic material, RNA. (1.4)

Viruses Small, obligatory intracellular pathogens that are not considered to be alive because they cannot divide directly, which is required by the cell theory of life. (1.4)

Whole mount A specimen to be observed with a microscope that is an intact object, either living or dead, and can be an entire intact organism or a small part of a large organism. (18.1)

Wild type The original strain of a living organism from which other organisms are bred for research. (10.2)

Wobble hypothesis Crick's proposal that the steric requirement between the anticodon of the tRNA and the codon of the mRNA is flexible at the third position, which allows two codons that differ only at the third position to share the same tRNA during protein synthesis. (11.7)

X-ray crystallography (X-ray diffraction) A technique that bombards protein crystals with a thin beam of X-rays of a single (monochromatic) wavelength. The radiation that is diffracted by the electrons of the protein atoms strikes a photographic plate or sensor. The diffraction pattern produced by the crystal is determined by the structure within the protein. (2.5, 18.8)

Yeast artificial chromosomes (YACs) Cloning elements that are artificial versions of a normal yeast chromosome. They contain all of the elements of a yeast chromosome that are necessary for the structure to be replicated during S phase and segregated to daughter cells during mitosis, plus a gene whose encoded product allows those cells containing the YAC to be selected from those that lack the element and the DNA fragment to be cloned. (18.16)

Yeast two-hybrid system A technique used to search for protein-protein interactions. It depends on the expression of a reporter gene such as (β)-galactosidase, whose activity is readily monitored by a test that detects a color change when the enzyme is present in a population of yeast cells. (18.7)

Additional Readings

CHAPTER 1 TEXT READINGS

General References in Microbiology and Virology

Knipe, D. M., et al. 2007. *Fields Virology*, 5th ed. Lippincott.

Madigan, M. T., et al. 2006. *Brock—Biology of Microorganisms*, 11th ed. Prentice-Hall.

Other Readings

Ball, P., 2007. Designs for life. *Nature* 448:32–33. [on generating a synthetic cell]

Blelloch, R. 2008. Short cut to cell replacement. *Nature* 455:604–605.

Cohan, F. M. & Perry, E. B., 2007. A systematics for discovering the fundamental units of bacterial diversity. *Curr. Biol.* 17:R373–R386.

Cyranoski, D. 2008. 5 things to know before jumping on the iPS bandwagon. *Nature* 452:406–408.

Daley, G. Q. & Scadden, D. T. 2008. Prospects for stem cell-based therapy. *Cell* 132:544–548.

Dethlefsen, L., et al. 2007. An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* 449:811–818.

DeWitt, N., et al. 2008. Regenerative medicine. *Nature* 453:301–351.

Esser, C. & Martin, W. 2007. Supertrees and symbiosis in eukaryote genome evolution. *Trends Microbiol.* 15:435–437.

Fischer, W. W. 2008. Life before the rise of oxygen. *Nature* 455:1051–1052.

Gura, T. 2008. Just spit it out. *Nat. Med.* 14:706–709. [oral microbes]

Holt, R. A. 2008. Synthetic genomes brought closer to life. *Nature Biotech.* 26:296–297.

Koonin, E. V. 2007. Metagenomic sorcery and the expanding protein universe. *Nat. Biotech.* 25:540–542.

Mascarelli, A. L. 2009. Subterranean bacterial: low life. *Nature* 459:770–773.

McInerney, J. O. & Pisani, D. 2007. Paradigm for life. *Science* 318:1390–1391. [on LGT]

Nishikawa, S., et al. 2008. The promise of human induced pluripotent stem cells for research and therapy. *Nat. Revs. Mol. Cell Biol.* 9:725–729.

Nobile, C. J. & Mitchell, A. P. 2007. Microbial biofilms: e pluribus unum. *Curr. Biol.* 17:R349–R353.

Poole, A. & Penny, D. 2007. Engulfed by speculation. *Nature* 447:913. [origins of eukaryotes]

Segers, V. F. M. & Lee, R. T. 2008. Stem-cell therapy for cardiac disease. *Nature* 451:937–942.

Sekirov, I. & Finlay, B. B., 2006. Human and microbe: united we stand. *Nat. Med.* 12:736–737.

Sendtner, M. 2009. Tailor-made diseased neurons. *Nature* 457:269–270. [iPS cells]

CHAPTER 2 TEXT READINGS

General Biochemistry

Berg, J. M., Tymoczko, J. L. & Stryer, L. 2007. *Biochemistry*, 6th ed. W. H. Freeman.

Nelson, D. L., et al. 2009. *Lehninger Principles of Biochemistry*, 5th ed. W. H. Freeman.

Voet, D. & Voet, J. G. 2004. *Biochemistry* 4th ed., 2 vols. Wiley.

Additional Topics

Abbott, A. 2008. The plaque plan. *Nature* 456:161–164.

Dill, K. A., et al. 2007. The protein folding problem: when will it be solved? *Curr. Opin. Struct. Biol.* 17:342–346.

Caughey, B. & Baron, G. S. 2006. Prions and their partners in crime. *Nature* 443:803–810.

Dodson, E. J. 2007. Protein predictions. *Nature* 450:176–177.

Dunker, A. K., et al. 2008. Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* 18:756–764.

Fersht, A. R. & Daggett, V., eds. 2007. Folding and binding. *Curr. Opin. Struct. Biol.* 17:#1.

Goloubinoff, P. & De Los Rios, P. 2007. The mechanism of Hsp70 chaperones. *Trends Biochem. Sci.* 32:372–380.

Guarente, L. 2008. Mitochondria—a nexus for aging, calorie restriction, and sirtuins? *Cell* 132:171–176.

Han, J.-H., et al. 2007. The folding and evolution of multidomain proteins. *Nat. Revs. Mol. Cell Biol.* 8:319–330.

Hanash, S. M., et al. 2008. Mining the plasma proteome for cancer biomarkers. *Nature* 452:571–579.

Holtzman, D. M. 2008. Moving towards a vaccine. *Nature* 454:418–420. [alzheimers's disease]

Horwich, A. L., et al. 2007. Two families of chaperonin: physiology and mechanism. *Annu. Rev. Cell Dev. Biol.* 23:115–145.

Mandavilli, A., et al. 2006. Issue on alzheimer's disease. *Nat. Med.* 12:747–784.

Roberts, B. E. & Shorter, J. 2008. Escaping amyloid fate. *Nat. Struct. Mol. Biol.* 15:544–545.

CHAPTER 3 TEXT READINGS

Benkovic, S. J. & Hammes-Schiffer, S. 2003. A perspective on enzyme catalysis. *Science* 301:1196–1202.

Hammes, G. G. 2000. Thermodynamics and Kinetics for the Biological Sciences. Wiley.

Hammes, G. G. 2008. How do enzymes really work? *J. Biol. Chem.* 283:22337–22346.

Harold, F. M. 1986. The Vital Force: A Study of Bioenergetics. Freeman.

Harris, D. A. 1995. *Bioenergetics at a Glance.* Blackwell.

Jencks, W. P. 1997. From chemistry to biochemistry to catalysis to movement. *Annu. Rev. Biochem.* 66:1–18.

Kornberg, A. 1989. *For the Love of Enzymes.* Harvard.

Koshland, D. E., Jr. 2004. Crazy, but correct. *Nature* 432:447. [on postulation of induced fit hypothesis]

Kraut, D. A., et al. 2003. Challenges in enzyme mechanism and energetics. *Annu. Rev. Biochem.* 72:517–571.

Kraut, J. 1988. How do enzymes work? *Science* 242:533–540.

Ringe, D. & Petsko, G. A. 2008. How enzymes work. *Science* 320:1428–1429.

Taubs, G., et al. 2008. Drug resistance. *Science* 321:355–369.

Vrielink, A. & Sampson, N. 2003. Sub-Ångstrom resolution enzyme X-ray structures: is seeing believing? *Curr. Opin. Struct. Biol.* 13:709–715.

Walsh, C., et al. 2001. Reviews on biocatalysis. *Nature* 409:226–268.

Wright, G. D. & Sutherland, A. D. 2007. New strategies for combating multidrug-resistant bacteria. *Trends Mol. Med.* 13:260–267.

CHAPTER 4 TEXT READINGS

Reviews on membranes can be found each year in *Curr. Opin. Struct. Biol.* issue #4.

Bennett, V. & Healy, J. 2008. Organizing the fluid membrane bilayer: diseases linked to spectrin and ankyrin. *Trends Mol. Med.* 14:28–35.

Bezanilla, F. 2008. How membrane proteins sense voltage. *Nat. Revs. Mol. Cell Biol.* 9:323–332.

Boucher, R. C. 2007. Cystic fibrosis: a disease of vulnerability to airway surface dehydration. *Trends Mol. Med.* 13:231–240.

Changeux, J.-P. & Taly, A. 2008. Nicotinic receptors, allosteric proteins and medicine. *Trends Mol. Med.* 14:93–102.

Couzin-Frankel, J. 2009. The promise of a cure: 20 years and counting. *Science* 324:1504–1507. [cystic fibrosis]

Elofsson, A. & von Heijne, G. 2007. Membrane protein structure: prediction versus reality. *Ann. Rev. Biochem.* 76:125–140.

Fleishman, S. J. & Ben-Tal, N. 2006. Progress in structure prediction of alpha-helical membrane proteins. *Curr. Opin. Struct. Biol.* 16:496–504.

Gadsby, D. C. 2007. Ion pumps made crystal clear. *Nature* 450:957–959.

Jacobson, K., et al. 2007. Lipid rafts: at a crossroad between cell biology and physics. *Nat. Cell Biol.* 9:7–14.

Janmey, P. A. & Kinnunen, P. K. J. 2006. Biophysical properties of lipids and dynamic membranes. *Trends Cell Biol.* 16:538–546.

Kozlov, M. M. 2007. Bending over to attract. *Nature* 447:387–389. [lipid curvature]

Riordan, J. R. 2008. CFTR function and prospects for therapy. *Ann. Rev. Biochem.* 77:701–726.

van Meer, G., et al. 2008. Membrane lipids: where they are and how they behave. *Nat. Revs. Mol. Cell Biol.* 9:112–124.

von Heijne, G. 2006. Membrane-protein topology. *Nat. Revs. Mol. Cell Biol.* 7: 909–918.

CHAPTER 5 TEXT READINGS

Bollinger, Jr., J. M. 2008. Electron relay in proteins. *Science* 320:1730–1731.

Brzezinski, P. & Adelroth, P. 2006. Design principles of proton-pumping haem-copper oxidases. *Curr. Opin. Struct. Biol.* 16:465–472.

Chan, D. C. 2006. Mitochondria: dynamic organelles in disease, aging, and development. *Cell* 125:1241–1252.

Detmer, S. A. & Chan, D. C. 2007. Functions and dysfunctions of mitochondrial dynamics. *Nat. Revs. Mol. Cell Biol.* 8:870–879.

Fischer, W. W. 2008. Life before the rise of oxygen. *Nature* 455:1051–1052.

Khrapko, K. & Vijg, J. 2007. Mitochondrial DNA mutations and aging: a case closed? *Nat. Gen.* 39:445–446.

Schrader, M. & Yoon, Y. 2007. Mitochondria and peroxisomes: are the "Big Brother" and the "Little Sister" closer than assumed? *Bioess.* 29:1105–1114.

Senior, A. E. 2007. ATP synthase: motoring to the finish line. *Cell* 130:220–221.

von Ballmoos, C., et al. 2008. Unique rotary ATP synthase and its biological diversity. *Annu. Rev. Biophys.* 37:43–64.

Westermann, B. 2008. Molecular machinery of mitochondrial fusion and fission. *J. Biol. Chem.* 283:13501–13505.

CHAPTER 6 TEXT READINGS

Allen, J. F. & Martin, W. 2007. Out of thin air. *Nature* 445:610–612. [evolution of photosynthesis]

Nelson, N. & Ben-Shem, A. 2004. The complex architecture of oxygenic photosynthesis. *Nature Revs. Mol. Cell Biol.* 5:971–982.

Nelson, N. & Yocum, C. F. 2006. Structure and function of photosystems I and II. *Annu. Rev. Plant Biol.* 57:521–565.

Shikanai, T. 2007. Cyclic electron transport around photosystem I: genetic approaches. *Annu. Rev. Plant Biol.* 58:199–217.

CHAPTER 7 TEXT READINGS

Reviews on cell-to-cell contact and extracellular matrix can be found each year in *Curr. Opin. Cell Biol.* issue #5.

Ainsworth, C. 2008. Stretching the imagination. *Nature* 456:696–699. [mechano sensation]

Balda, M. S. & Matter, K. 2008. Tight junctions at a glance. *J. Cell Sci.* 121: 3677–3682.

Davis, D. M. & Sowinski, S. 2008. Membrane nanotubes: dynamic long-distance connections between animal cells. *Nature Revs. Mol. Cell Biol.* 9:431–436.

Delon, I. & Brown, N. H. 2007. Integrins and the actin cytoskeleton. *Curr. Opin. Cell Biol.* 19:43–50.

Evans, E. A. & Calderwood, D. A. 2007. Forces and bond dynamics in cell adhesion. *Science* 316:1148–1153.

Halbleib, J. M. & Nelson, W. J. 2006. Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. *Genes Develop.* 20:3199–3214.

Kadler, K. E., et al. 2007. Collagens at a glance. *J. Cell Sci.* 120:1955–1958.

Luo, B. H., et al. 2007. Structural basis of integrin regulation and signaling. *Ann. Rev. Immunol.* 25:619–647.

Moser, M., et al. 2009. The tail of integrins, talin, and kindlins. *Science* 324:895–899.

Page-McCaw, A., et al. 2007. Matrix metalloproteinases and the regulation of tissue remodeling. *Nature Revs. Mol. Cell Biol.* 8:221–233.

Pokutta, S. & Weis, W. I. 2007. Structure and mechanism of cadherins and cetenins in cell-cell contacts. *Ann. Rev. Cell Develop. Biol.* 23:237–261.

Schwartz, M. A. 2009. The force is with us. *Science* 323:588–589. [forces on focal adhesions]

Shin, K., et al. 2006. Tight junctions and cell polarity. *Ann. Rev. Cell Develop. Biol.* 22: 207–235.

Shoulders, M. D. & Raines, R. T. 2009. Collagen structure and stability. *Ann. Rev. Biochem.* 78:929–958.

Somerville, C. 2006. Cellulose synthesis in higher plants. *Ann. Rev. Cell Develop. Biol.* 22:53–78.

Sonnenberg, A. & Watt, F. M., eds. 2009. Special issue on integrins. *J. Cell. Sci.* 122:#2.

Steeg, P. S. 2006. Tumor metastases: mechanistic insights and clinical challenges. *Nature Med.* 12:895–904.

Varki, A. 2007. Glycan-based interactions involving vertebrate, sialic-acid-recognizing proteins. *Nature* 446:1023–1029.

Wheelock, M. J., et al. 2008. Cadherin switching. *J. Cell Sci.* 121:727–735.

CHAPTER 8 TEXT READINGS

Reviews on endomembranes and organelles can be found each year in *Curr. Opin. Cell Biol.* issue #4.

Baines, A. C. & Zhang, B. 2007. Receptor-mediated protein transport in the early secretory pathway. *Trends Biochem. Sci.* 32:381–388.

Baker, M. J., et al. 2007. Mitochondrial protein-import machinery: correlating structure with function. *Trends Cell Biol.* 17:456–464.

Collins, R. N. & Zimmerberg, J. 2009. A score for membrane fusion. *Nature* 459:1065–1066.

Couzin, J. 2008. Cholesterol veers off script. *Science* 322:220–223.

De Matteis, M. A. & Luini, A. 2008. Exiting the Golgi complex. *Nat. Revs. Mol. Cell Biol.* 9:273–284.

Di Paolo, G. & De Camilli, P. 2006. Phosphoinositides in cell regulation and membrane dynamics. *Nature* 443:651–657.

Doherty, G. J. & McMahon, H. T. 2009. Mechanisms of endocytosis. *Ann. Rev. Biochem.* 78:857–902.

Fromme, J. C., et al. 2008. Coordination of COPII vesicle trafficking by Sec23. *Trends Cell Biol.* 18:330–336.

Grosshans, B. L., et al. 2006. Rabs and their effectors: achieving specificity in membrane traffic. *PNAS.* 103:11821–11827.

Jahn, R. & Scheller, R. H. 2006. SNAREs—engines for membrane fusion. *Nature Revs. Mol. Cell Biol.* 7:631–643.

Jahn, R., et al. 2008. Reviews on membrane fusion. *Nature Struct. Mol. Cell Biol.* 15: 655–698.

Kutik, S., et al. 2007. Cooperation of translocase complexes in mitochondrial protein import. *J. Cell Biol.* 179:585–591.

Luzio, J. P., et al. 2007. Lysosomes: fusion and function. *Nature Revs. Mol. Cell Biol.* 8: 622–632.

Mayor, S. & Pagano, R. E. 2007. Pathways of clathrin-independent endocytosis. *Nature Revs. Mol. Cell Biol.* 8:603–612.

Mukhopadhyay, D. & Riezman, H. 2007. Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science* 315:201–205.

Neupert, W. & Herrmann, J. M. 2007. Translocation of proteins into mitochondria. *Ann. Rev. Biochem.* 76:723–749.

Ohtsubo, K. & Marth, J. D. 2006. Glycosylation in cellular mechanisms of health and disease. *Cell* 126:855–867.

Pfeffer, S. R. 2007. Unsolved mysteries in membrane traffic. *Ann. Rev. Biochem.* 76:629–645.

Rader, D. J. & Daugherty, A. 2008. Translating molecular discoveries into new therapies for atherosclerosis. *Nature* 451:904–913.

Raiborg, C. & Stenmark, H. 2009. The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature* 458:445–452.

Rapoport, T. A. 2007. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* 450:663–669.

Ron, D. & Walter, P. 2007. Signal integration in the endoplasmic reticulum unfolded protein response. *Nature Revs. Mol. Cell Biol.* 8:519–529.

Vembar, S. S. & Brodsky, J. L. 2008. One step at a time: endoplasmic reticulum-associated degradation. *Nat. Revs. Mol. Cell Biol.* 9: 944–957.

White, S. H. & von Heijne, G. 2008. How translocans select transmembrane helices. *Ann. Rev. Biophys.* 37:23–42.

CHAPTER 9 TEXT READINGS

Reviews on the cytoskeleton and motor proteins can be found each year in *Curr. Opin. Cell Biol.* issue #1.

Akhmanova, A. & Steinmetz, M. O. 2008. Tracking the ends: a dynamic protein network controls the fate of microtubule tips. *Nature Revs. Mol. Cell Biol.* 9:309–328.

Bartolini, F. & Gundersen, G. G. 2006. Generation of noncentrosomal microtubule arrays. *J. Cell Sci.* 119:4155–4163.

Bettencourt–Dias, M. & Glover, D. M. 2007. Centrosome biogenesis and function. *Nature Revs. Mol. Cell Biol.* 8:451–463.

Burbank, K. S. & Mitchison, T. J. 2006. Microtubule dynamic instability. *Curr. Biol.* 16:R516–R517.

Carlier, M.-F. & Pantaloni, D. 2007. Control of actin assembly dynamics in cell motility. *J. Biol. Chem.* 282:23005–23009.

Chhabra, E. S. & Higgs, H. N. 2007. The many faces of actin: matching assembly factors with cellular structures. *Nature Cell Biol.* 9:1110–1120.

Conti, M. A. & Adelstein, R. S. 2008. Nonmuscle myosin II moves in new directions. *J. Cell Sci.* 121:11–18.

Gerdes, J. M., et al. 2009. The vertebrate primary cilium in development, homeostasis, and disease. *Cell* 137:32–45.

Godsel, L. M., et al. 2008. Intermediate filament assembly: dynamics to disease. *Trends Cell Biol.* 18:28–37.

Goley, E. D. & Welch, M. D. 2006. The ARP2/3 complex: an actin nucleator comes of age. *Nature Revs. Mol. Cell Biol.* 7:713–726.

Herrmann, H., et al. 2007. Intermediate filaments: from cell architecture to nanomechanics. *Nature Revs. Mol. Cell Biol.* 8:562–573.

Hirokawa, N. & Noda, Y. 2008. Intracellular transport and kinesin superfamily proteins, KIFs. *Physiol. Rev.* 88:1089–1118.

Kritikou, E., et al. 2008. Milestone papers on the cytoskeleton. *Nature Suppl.* December.

Pollard, T. D. 2008. Regulation of actin filament assembly by Arp2/3 complex and formins. *Ann. Rev. Biophys. Biomol. Struct.* 36:451–477.

Satir, P., et al. 2007. Reviews on mammalian cilia. *Ann. Rev. Physiol.* 69:377–450.

Scholey, J. M. 2008. Intraflagellar transport motors in cilia: moving along the cell's antenna. *J. Cell Biol.* 180:23–29.

Vale, R. D. 2008. Microscopes for fluorimeters: the era of single molecule measurements. *Cell* 135:779–785.

Vallee, R. 2007. An interview discussing many of the important discoveries in the motor protein field. *Curr. Biol.* 17:R903–R905.

van den Heuvel, M. G. L. & Dekker, C. 2007. Motor proteins at work for nanotechnology. *Science* 317:333–336.

CHAPTER 10 TEXT READINGS

Reviews on genomes and evolution can be found each year in *Curr. Opin. Genetics Develop.* #6.

Altshuler, D., et al. 2008. Genetic mapping in human disease. *Science* 322:881–888.

Baker, M. 2008. Genetics by numbers. *Nature* 451:516–518. [genome-wide association]

Cohen, J. 2007. DNA duplications and deletions help determine health. *Science* 317:1315–1317.

Cozzarelli, N. R., et al. 2006. Giant proteins that move DNA. *Nature Revs. Mol. Cell Biol.* 7:580–588. [on topoisomerases]

Frazer, K. A., et al. 2009. Human genetic variation and its contribution to complex traits. *Nature Revs. Gen.* 10:241–251.

Gee, H. 2008. The amphioxus unleashed. *Nature* 453:999–1000. [amphioxus genome]

Goodier, J. L. & Kazazian, Jr., H. H. 2008. Retrotransposons revisited: the restraint and the rehabilitation of parasites. *Cell* 135:23–35.

Hayden, E. C. 2009. The other strand. *Nature* 457:776–779. [genetic basis of being human]

Hurles, M. E., et al. 2008. The functional impact of structural variation in humans. *Trends Gen.* 24:238–245.

Maher, B. 2008. The case of the missing heritability. *Nature* 456:18–21.

Mathew, C. G. 2008. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. *Nature Revs. Gen.* 9:9–14.

Mills, R. E. 2007. Which transposable elements are active in the human genome? *Trends Gen.* 23:183–190.

Nielsen, R., et al. 2007. Recent and ongoing selection in the human genome. *Nature Revs. Gen.* 8:857–868.

Orr, H. T. & Zoghbi, H. Y. 2007. Trinucleotide repeat disorders. *Ann. Rev. Neurosci.* 30: 575–621.

Pennisi, E. 2008. 17q21.31: not your average genomic address. *Science* 322:842–845. [haplotypes, inversions, and disease]

Sebat, J., et al. 2007. Human genomic variation. *Nature Gen.* 39 (Supplement to July issue).

Shurin, S. B. 2008. Pharmacogenomics—ready for prime time? *New Engl. J. Med.* 358: 1061–1063.

Varki, A., et al. 2008. Explaining human uniqueness: genome interactions with environment, behavior and culture. *Nature Revs. Gen.* 9:749–763.

Williams, A. J. & Paulson, H. L. 2008. Polyglutamine neurodegeneration: protein misfolding revisited. *Trends Neurosci.* 31: 521–529.

CHAPTER 11 TEXT READINGS

Reviews on the nucleus and gene expression can be found each year in *Curr. Opin. Cell Biol.* issue #3.

Boisvert, F.-M. 2007. The multifunctional nucleolus, *Nature Revs. Mol. Cell Biol.* 8: 574–585.

Borukhov, S. & Nudler, E. 2008. RNA polymerase: the vehicle of transcription. *Trends Microbiol.* 16:126–133.

Carninci, P. & Hayashizaki, Y. 2007. Noncoding RNA transcription beyond annotated genes. *Curr. Opin. Gen. Develop.* 17:139–144.

Castanotto, D. & Rossi, J. J. 2009. The promises and pitfalls of RNA-interference-based therapeutics. *Nature* 457:426–433.

Chapman, E. J. & Carrington, J. C. 2007. Specialization and evolution of endogenous small RNA pathways. *Nature Revs. Gen.* 8:884–896.

Couzin, J. 2008. MicroRNAs make big impression in disease after disease. *Science* 319:1782–1784.

Darzacq, X., et al. 2009. Imaging transcription in living cells. *Ann. Rev. Biophys.* 38:173–196.

Ghildiyal, M. & Zamore, P. D. 2009. Small silencing RNAs: an expanding universe. *Nature Revs. Gen.* 10:94–108.

Hutvagner, G. & Simard, M. J. 2008. Argonaute proteins: key players in RNA silencing. *Nature Revs. Mol. Cell Biol.* 9:22–32.

Kapranov, P., et al. 2007. Genome-wide transcription and the implications for genomic organization. *Nature Revs. Gen.* 8:413–423.

Kim, V. N., et al. 2009. Biogenesis of small RNAs in animals. *Nature Revs. Mol. Cell Biol.* 10:126–139.

Kornberg, R. D. 2007. The molecular basis of eukaryotic transcription. *PNAS* 104: 12955–12961.

Korostelev, A. & Noller, H. F. 2007. The ribosome in focus: new structures bring new insights. *Trends Biochem. Sci.* 32:434–441.

Müller, F., et al. 2007. New problems in RNA polymerase II transcription initiation. *J. Biol. Chem.* 282:14685–14689.

Nudler, E. 2009. RNA polymerase active center: the molecular engine of transcription. *Ann. Rev. Biochem.* 78:335–361.

O'Donnell, K. A. & Boeke, J. D. 2007. Mighty Piwis defend the germline against genome intruders. *Cell* 129:37–44.

Peters, L. & Meister, G. 2007. Argonaute proteins: mediators of RNA silencing. *Mol. Cell* 26:611–622.

Petherick, A. 2008. The production line. *Nature* 454:1043–1045. [the transcriptome]

Rodnina, M. V., et al. 2007. How ribosomes make peptide bonds. *Trends Biochem. Sci.* 32:20–26.

Sharp, P. A., et al. 2009. Special review issue on RNA. *Cell* 136#4.

Siomi, H. & Siomi, M. C. 2009. RNA silencing—Nature insight. *Nature* 457:396–433.

Steitz, T. A. 2008. A structural understanding of the dynamic ribosome machine. *Nature Revs. Mol. Cell Biol.* 9:242–253.

Svejstrup, J. Q. 2007. Contending with transcriptional arrest during RNAPII transcript elongation. *Trends Biochem. Sci.* 32:165–171.

Wang, G.-S. & Cooper, T. A. 2007. Splicing and disease. *Nature Revs. Gen.* 8:749–761.

CHAPTER 12 TEXT READINGS

Reviews on the nucleus and gene regulation can be found each year in *Curr. Opin. Cell Biol.* #3.

Reviews on chromosomes and gene expression can be found each year in *Curr. Opin. Genetics and Develop.* #2.

Aubert, G. & Lansdorp, P. M. 2008. Telomeres and aging. *Physiol. Revs.* 88:557–579.

Besse, F. & Ephrussi, A. 2008. Translational control of localized mRNAs: restricting protein synthesis in space and time. *Nature Revs. Mol. Cell Biol.* 9:971–980.

Bird, A., et al. 2007. Nature Insight: Epigenetics. *Nature* 447:395–440.

Blackburn, E. H., Greider, C. W., and Szostak, J. W. 2006. Telomeres and telomerase. *Nature Med.* 12:1133–1138.

Cibelli, J. 2007. A decade of cloning mystique. *Science* 316:990–992. [cloning animals]

D'Angelo, M. A. & Hetzer, M. W. 2008. Structure, dynamics and function of nuclear pore complexes. *Trends Cell Biol.* 18:456–466.

Davuluri, R. V., et al. 2008. The functional consequences of alternative promoter use in mammalian genomes. *Trends Gen.* 24:167–176.

Fedor, M. J., et al. 2008. Alternative splicing review series. *J. Biol. Chem.* 283:1209–1233.

Flynt, A. S. & Lai, E. C. 2008. Biological principles of microRNA-mediated regulation. *Nature Revs. Gen.* 9:831–842.

Garneau, N. L., et al. 2007. The highways and byways of mRNA decay. *Nature Revs. Mol. Cell Biol.* 8:113–126.

Goldberg, A. D., et al. 2007. Issue on epigenetics. *Cell* 128:#4.

Henikoff, S. 2008. Nucleosome destabilization in the epigenetic regulation of gene expression. *Nature Revs. Gen.* 9:15–26.

Hurtley, S. M., et al. 2007. Articles on the nucleus. *Science* 318:1399–1416.

Jiang, C. & Pugh, B. F. 2009. Nucleosome positioning and gene regulation. *Nature Revs. Gen.* 10:161–172.

Jirtle, R. J., et al. 2007. Reviews on epigenetics. *Nature Revs. Gen.* 8:#4.

Kim, E., et al. 2008. Alternative splicing: current perspectives. *Bioess.* 30:38–47.

Koch, F., et al. 2008. Genome-wide RNA polymerase II: not genes only. *Trends Biochem. Sci.* 33:265–273.

Kornberg, R. D., et al. 2007. Articles on chromatin. *Nature Struct. Mol. Biol.* 14:#11.

Kumaran, R. I., et al. 2008. Chromatin dynamics and gene positioning. *Cell* 132: 929–934.

Lanctot, C., et al. 2007. Dynamic genome architecture in the nuclear space. *Nature Revs. Gen.* 8:104–115.

Lawrence, J. B. & Clemson, C. M. 2008. Gene associations: true romance or chance meeting in a nuclear neighborhood? *J. Cell Biol.* 182:1035–1038.

Mattick, J. S., et al. 2009. RNA regulation of epigenetic processes. *Bioess.* 31:51–59.

Rando, O. J. & Chang, H. Y. 2009. Genome-wide views of chromatin structure. *Ann. Rev. Biochem.* 78:245–271.

Ruthenburg, A. J. 2007. Multivalent engagement of chromatin modifications by linked binding modules. *Nature Revs. Mol. Cell Biol.* 8:983–994.

Sandelin, A., et al. 2007. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature Revs. Gen.* 8:424–436.

Schones, D. E. & Zhao, K. 2008. Genome-wide approaches to studying chromatin modifications. *Nature Revs. Gen.* 9:179–191.

Sutherland, H. & Bickmore, W. A. 2009. Transcription factories: gene expression in unions? *Nature Revs. Gen.* 10:457–466.

Suzuki, M. M. & Bird, A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature Revs. Gen.* 9: 465–476.

Welstead, G. G., et al. 2008. The reprogramming language of pluripotency. *Curr. Opin. Gen. Develop.* 18:123–129.

CHAPTER 13 TEXT READINGS

Arias, E. E. & Walter, J. C. 2007. Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes Develop.* 21:497–518.

Corpet, A. & Almouzni, G. 2009. Making copies of chromatin: the challenge of nucleosomal organization and epigenetic information. *Trends Cell Biol.* 19:29–40.

Garinis, G. A., et al. 2008. DNA damage and ageing. *Nature Cell Biol.* 10:1241–1247.

Groth, A., et al. 2007. Chromatin challenges during DNA replication and repair. *Cell* 128:721–733.

Hamdan, S. M. & Richardson, C. C. 2009. Motors, switches, and contacts in the replisome. *Ann. Rev. Biochem.* 78:205–243.

Hanawalt, P. C. 2007. Paradigms for the three Rs: DNA replication, recombination, and repair. *Mol. Cell* 28:702–707.

Loeb, L. A. & Monnat, Jr., R. J. 2008. DNA polymerases and human disease. *Nature Revs. Gen.* 9:594–604.

Pomerantz, R. T. & O'Donnell, M. 2007. Replisome mechanics: insights into a twin DNA polymerase machine. *Trends Microbiol.* 15:156–164.

Sclafani, R. A. & Holzen, T. M. 2007. Cell cycle regulation of DNA replication. *Ann. Rev. Gen.* 41:237–280.

Stillman, B. 2008. DNA polymerases at the replication fork in eukaryotes. *Mol. Cell* 30:259–260.

Wickelgren, I. 2007. A healthy tan? *Science* 315:1214–1216.

Yang, W. & Woodgate, R. 2007. What a difference a decade makes: insights into translesion DNA synthesis. *PNAS* 104:15591–15598.

CHAPTER 14 TEXT READINGS

Reviews on cell division can be found each year in *Curr. Opin. Cell Biol.* #6.

Barr, F. A. & Gruneberg, U. 2007. Cytokinesis: placing and making the final cut. *Cell* 131:847–860.

Bartek, J. & Lukas, J. 2007. DNA damage checkpoints: from initiation to recovery or adaptation. *Curr. Opin. Cell Biol.* 19:238–245.

Bloom, K. 2008. Kinetochores and microtubules wed without a ring. *Cell* 135:211–213.

Cheeseman, I. M. & Desai, A. 2008. Molecular architecture of the kinetochore-microtubule interface. *Nature Revs. Mol. Cell Biol.* 9:33–46.

Cimprich, K. A. & Cortez, D. 2008. ATR: an essential regulator of genome integrity. *Nature Revs. Mol. Cell Biol.* 9:616–627.

Davis, T. N. & Wordeman, L. 2007. Rings, bracelets, sleeves, and chevrons: new structures of kinetochore proteins. *Trends Cell Biol.* 17:377–382.

Güttinger, S., et al. 2009. Orchestrating nuclear envelope disassembly and reassembly during mitosis. *Nature Revs. Mol. Cell Biol.* 10:178–191.

Hochegger, H., et al. 2008. Cyclin-dependent kinases and cell-cycle transitions: does one fit all? *Nature Revs. Mol. Cell Biol.* 9:910–916.

Hochwagen, A. 2008. Meiosis. *Curr. Biol.* 18:R641–R645.

Hunt, P. A. & Hassold, T. J. 2007. Human female meiosis: what makes a good egg go bad? *Trends Gen.* 24:86–93.

Kwok, B. H. & Kapoor, T. M. 2007. Microtubule flux: drivers wanted. *Curr. Opin. Cell Biol.* 19:36–42.

Lowe, M. & Barr, F. A. 2007. Inheritance and biogenesis of organelles in the secretory pathway. *Nature Revs. Mol. Cell Biol.* 8:429–439.

Musacchio, A. & Salmon, E. D. 2007. The spindle assembly checkpoint in space and time. *Nature Revs. Mol. Cell Biol.* 8:379–393.

O'Connell, C. B. & Khodjakov, A. L. 2007. Cooperative mechanisms of mitotic spindle formation. *J. Cell Sci.* 120:1717–1722.

Peters, J.-M., et al. 2008. The cohesin complex and its roles in chromosome biology. *Genes Develop.* 22:3089–3114.

Peters, J.-M. 2006. The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nature Revs. Mol. Cell Biol.* 7:644–656.

Sullivan, M. & Morgan, D. O. 2007. Finishing mitosis, one step at a time. *Nature Revs. Mol. Cell Biol.* 8:894–903.

Yu, H. 2007. Cdc20: a WD40 activator for a cell cycle degradation machine. *Mol. Cell* 27:3–16.

CHAPTER 15 TEXT READINGS

Reviews on cell regulation can be found each year in *Curr. Opin. Cell Biol.* #2.

Bos, J. L., et al. 2007. GEFs and GAPs: critical elements in the control of small G proteins. *Cell* 129:865–877.

Chou, I., et al. 2006. Reviews on taste and smell. *Nature* 444:287–321.

Clapham, D. E. 2007. Calcium signaling. *Cell* 131:1047–1058.

Cohen, P. 2006. The twentieth century struggle to decipher insulin signalling. *Nature Revs. Mol. Cell Biol.* 7:867–873.

De Meyts, P. 2008. The insulin receptor: a prototype for dimeric, allosteric membrane receptors? *Trends Biochem. Sci.* 33:376–384.

Kobilka, B. K., et al. 2007. Special Issue on GPCRs. *Trends Pharmacol. Sci.* 28:#8.

Lewis, R. S. 2007. The molecular choreography of a store-operated calcium channel. *Nature* 446:284–287.

Manning, B. D. & Cantley, L. C. 2007. AKT/PKB signaling: navigating downstream. *Cell* 129:1261–1274.

Muoio, D. M. & Newgard, C. B. 2008. Molecular and metabolic mechanisms of insulin resistance and β-cell failure in type 2 diabetes. *Nature Revs. Mol. Cell Biol.* 9:193–205.

Oldham, W. M. & Hamm, H. E. 2008. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nature Revs. Mol. Cell Biol.* 9:60–71.

Riedl, S. J. & Salvesen, G. S. 2007. The apoptosome: signalling platform of cell death. *Nature Revs. Mol. Cell Biol.* 8:405–413.

Santner, A. & Estelle, M. 2009. Recent advances and emerging trends in plant hormone signalling. *Nature* 459:1071–1078.

Schlessinger, J. & Lemmon, M. A. 2006. Nuclear signaling by receptor tyrosine kinases. *Cell* 127:45–48.

Schwartz, T. W. & Hubbell, W. L. 2008. A moving story of receptors. *Nature* 455: 473–474. [GPCR crystal structure]

Sheridan, C. & Martin, S. J. 2008. Commitment in apoptosis: slightly dead but mostly alive. *Trends Cell Biol.* 18:353–357.

Taguchi, A. & White, M. F. 2008. Insulin-like signaling, nutrient homostasis, and life span. *Ann. Rev. Physiol.* 70:191–212.

Taylor, R. C., et al. 2008. Apoptosis: controlled demolition at the cellular level. *Nature Revs. Mol. Cell Biol.* 9:231–241.

Taylor, S. S. & Ghosh, G., eds. 2006. Catalysis and regulation [of protein kinases]. *Curr. Opin. Struct. Biol.* 16:#6.

Ward, C. W., et al. 2007. The insulin and EGF receptor structures: new insights into ligand-induced receptor activation. *Trends Biochem. Sci.* 32:129–137.

Youle, R. J. & Strasser, A. 2008. The BCL-2 protein family: opposing activities that mediate cell death. *Nature Revs. Mol. Cell Biol.* 9:47–59.

CHAPTER 16 TEXT READINGS

Reviews on the genetic and cellular mechanisms of oncogenesis can be found each year in *Curr. Opin. Gen. Develop.* Issue #1.

Burkhart, D. L. & Sage, J. 2008. Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nature Revs. Cancer* 8:671–681.

Campisi, J. & d'Adda di Fagagna, F. 2007. Cellular senescence: when bad things happen to good cells. *Nature Revs. Mol. Cell Biol.* 8:729–740.

Di Micco, R., et al. 2007. Breaking news: high-speed race ends in arrest—how oncogenes induce senescence. *Trends Cell Biol.* 17:529–535.

Eaves, C. J. 2008. Here, there, everywhere? *Nature* 456:581–582. [cancer stem cells]

Gray-Schopfer, V., et al. 2007. Melanoma biology and new targeted therapy. *Nature* 445:851–857.

Harley, C. B. 2008. Telomerase and cancer therapeutics. *Nature Revs. Cancer* 8:167–179.

Jones, P. A. & Baylin, S. B. 2007. The epigenomics of cancer. *Cell* 128:683–692.

Kastan, M. B. 2007. Wild-type p53: tumors can't stand it. *Cell* 128:837–840.

Kruse, J.-P. & Gu, W. 2009. Modes of p53 regulation. *Cell* 137:609–622.

Kutzler, M. A. & Weiner, D. B. 2008. DNA vaccines: ready for prime time? *Nature Revs. Gen.* 9:776–788.

Landis, M. W., et al. 2007. Mouse models of cancer. *Cell* 129:Supplement to #4.

Leen, A. M., et al. 2007. Improving T cell therapy for cancer. *Ann. Rev. Immunol.* 25:243–265.

Livingston, D. M. 2009. Complicated supercomplexes. *Science* 324:602–603. [BRCA mechanism]

Luo, J., et al. 2009. Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell* 136:823–837.

Marte, B., et al. 2008. Nature insight on molecular cancer diagnostics. *Nature* 452:547–589.

Nevins, J. R. & Potti, A. 2007. Mining gene expression profiles: expression signatures as cancer phenotypes. *Nature Revs. Gen.* 8: 601–609.

Nguyen, D. X. & Massagué, J. 2007. Genetic determinants of cancer metastasis. *Nature Revs. Gen.* 8:341–352.

Rice, J., et al. 2008. DNA vaccines: precision tools for activating effective immunity against cancer. *Nature Revs. Cancer* 8:108–118.

Rosen, J. M. & Jordan, C. T. 2009. The increasing complexity of the cancer stem cell paradigm. *Science* 324:1670–1673.

Sotiriou, C. & Piccart, M. J. 2007. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nature Revs. Cancer* 7:545–555.

Stratton, M. R., et al. 2009. The cancer genome. *Nature* 458:719–724.

Stratton, M.R. & Rahman, N. 2008. The emerging landscape of breast cancer susceptibility. *Nature Gen.* 40:17–22.

Trent, J. M. & Touchman, J. W. 2007. The gene topography of cancer. *Science* 318:1079–1080.

Vousden, K. H. & Prives, C. 2009. Blinded by the light: the growing complexity of p53. *Cell* 137:413–431.

CHAPTER 17 TEXT READINGS

The following consist entirely of reviews in Immunology:

Advances in Immunology
Annual Review of Immunology
Critical Reviews in Immunology
Current Opinion in Immunology
Immunological Reviews
Nature Reviews Immunology
Trends in Immunology

Bousso, P. 2008. T-cell activation by dendritic cells in the lymph node: lessons from the movies. *Nature Revs. Immunol.* 8:675–684.

Chatenoud, L. 2006. Immune therapies of autoimmune diseases: are we approaching a real cure? *Curr. Opin. Immunol.* 18: 710–717.

Cohen, J., et al. 2007. Challenges in immunology. *Science* 317:614–629.

Deng, L. & Mariuzza, R. A. 2007. Recognition of self-peptide—MHC complexes by autoimmune T-cell receptors. *Trends Biochem. Sci.* 32:500–508.

Jensen, P. E. 2007. Recent advances in antigen processing and presentation. *Nature Immunol.* 8:1041–1048.

Medzhitov, R. 2007. Recognition of microorganisms and activation of the immune response. *Nature* 449:819–826.

Medzhitov, R. 2008. Origin and physiological roles of inflammation. *Nature* 454:428–435.

Murphy, K. M., et al., 2007. Janeway's Immunobiology, 7th ed. Garland.

Nossal, G. J. V., et al. 2007. Articles on the development of the clonal selection theory. *Nature Immunol.* 8:1015–1025.

Paul, W. E. 2007. Dendritic cells bask in the limelight. *Cell* 130:967–970.

Sakaguchi, S., et al. 2008. Regulatory T cells and immune tolerance. *Cell* 133:775–787.

Steinman, R. M. 2007. Dendritic cells: versatile controllers of the immune system. *Nature Med.* 13:1155–1159.

Vignali, D. A. A., et al. 2008. How regulatory T cells work. *Nature Revs. Immunol.* 8: 523–531.

Vyas, J. M., et al. 2008. The known unknowns of antigen processing and presentation. *Nature Revs. Immunol.* 8:607–618.

INDEX

Topics of Human Interest