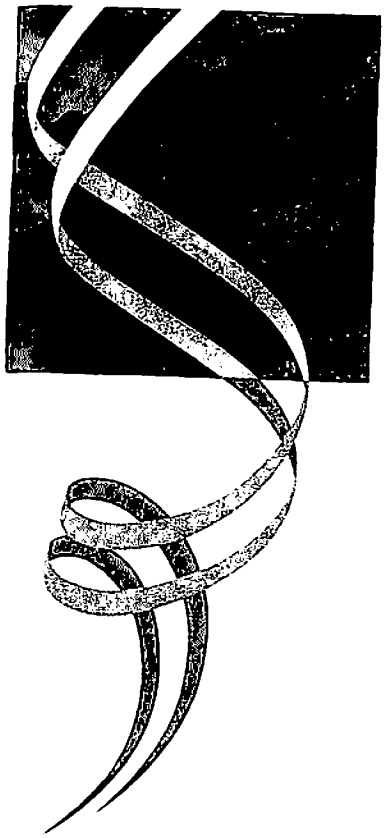


CHAPTER 4

Basic Tools of Recombinant DNA

The deciphering of the genetic code was the crowning achievement of the decade of research following the discovery of the DNA double helix. That research program had been set out clearly in the Central Dogma, and by the mid-1960s, it had been confirmed experimentally. Semiconservative DNA replication as suggested by the double helix had been demonstrated and an enzyme carrying out DNA synthesis had been found. The role of an RNA intermediary—messenger RNA (mRNA)—had been confirmed and its existence demonstrated. Protein synthesis had been studied intensively by biochemists and many of the components involved—mRNA, activated amino acids, transfer RNA (tRNA), and ribosomes—had been isolated and their functions elucidated. With the deciphering of the genetic code, it was possible now in theory to understand the information contained in any DNA or RNA molecule. A new discipline, molecular genetics, had been born.

To many of the biologists who had played a part in that birth, it seemed that molecular genetics was entering a period of consolidation. They felt there would be further growth, but that this would be incremental; new conceptual advances seemed unlikely as the fine details of the old problems were worked out and biochemists dissected in ever more minute detail the pathways involved. At that time, in the 1960s, molecular genetics was almost entirely based on the study of bacteria and phages. New advances would come in working out the functioning of genes in higher

organisms, but the experimental tools were lacking and it seemed as far in the future as putting a man on the moon. So to avoid possibly marking time, several distinguished contributors to our primary knowledge on the storage of genetic information in DNA left molecular genetics to start up new careers in neurobiology. What they could not have foreseen was the very rapid development over the next decade of the enzymological and chemical techniques that gave rise to recombinant DNA, which by now has come to include any technique for manipulating DNA or RNA.

Recombinant DNA techniques are so powerful because they provide the tools to study the genetics of any organism by isolating the DNA of virtually any gene. A particular gene can be isolated and produced in large quantities through *cloning* and its genetic information can be read by *sequencing*. The functions of that gene can then be analyzed by using *in vitro mutagenesis* to make a specific alteration in that information before reintroducing the mutated DNA into the organism to determine the effects of the mutation. By the late 1970s, as it became clear that these tools offered the fastest and surest route to understanding the molecular mechanisms of formerly intractable processes such as development and cell division, they were seized eagerly by biologists in almost every field. The subsequent period of scientific excitement and achievement has seen few, if any, parallels in the history of biological research.

In this and the next two chapters, we examine some of the tools on the molecular biologist's workbench. In this chapter, we look at the basic techniques of recombinant DNA: enzymes that cut, join, and synthesize DNA; electrophoresis for separating DNA molecules; DNA sequencing and chemical synthesis; the polymerase chain reaction; and the use of hybridization to select specific DNA sequences. In Chapter 5, we will turn to what techniques used for isolating genes have taught us about the structures of genomes, and in Chapter 6 we review the current technologies for studying the function of DNA sequences within cells.

Restriction Enzymes Cut DNA at Specific Sequences

Protein chemists had been able to put *proteases*, in particular, those enzymes that cut polypeptide chains at specific sites, to good use in determining the structure of proteins. By using different proteases, cutting at different specific sites, different smaller fragments (peptides) can be obtained and their amino acid sequences determined. The original order of the peptides in the protein can then be established by looking for overlaps in amino acid sequence between the peptides.

Unfortunately for early nucleic acid biochemists, all the known *nucleases*, the enzymes that were first found to break the phosphodiester bonds of nucleic acids, showed very little sequence dependency. The prevailing opinion was that, in contrast to proteases, highly specific nucleases would never be found and therefore the isolation of discrete DNA fragments, even from viral DNA, would not be possible. Doubt was cast on this view with the observations, beginning as early as 1953, that when DNA molecules from one strain of *Escherichia coli* were introduced into a different *E. coli* strain (e.g., *E. coli* strain B and *E. coli* strain C), the DNA was nearly always quickly fragmented into smaller pieces. Occasionally, however, the infecting DNA molecule somehow became modified and remained intact, so that it could now replicate in the new bacterial strain. This phenomenon was recognized as being similar to host restriction when bacteriophages infect bacteria. Most of the bacteriophage DNA is destroyed by bacterial enzymes, but a small amount escapes destruction and phages containing this DNA can efficiently infect that same bacterial strain. It was clear that the bacterial DNA must in some way be modified to escape destruction by its own enzymes, and that occasionally the phage DNA was modified in the same way.

In 1966 chemical analysis of such modified DNA molecules revealed the presence of methylated bases not present in the unmodified DNA. Methylated bases are not inserted as such into growing DNA chains; rather, they arise through the enzymatically catalyzed addition of methyl groups to newly synthesized DNA chains. The stage was thus set. In the 1960s, Stuart Linn and Werner Arber, working in Geneva, found in extracts of cells of *E. coli* strain B both a specific *modification enzyme* that methylated unmethylated DNA and a *restriction nuclease* that broke down unmethylated DNA. The term "restriction" was applied to the nuclease because this enzyme recognized specific unmethylated sequences on the DNA molecule. Over the next several years, other restriction nucleases and their companion modification methylases were identified, suggesting that many site-specific nucleases might exist. None of these early *E. coli* restriction enzymes lived up to these hopes, however; although the enzymes recognized specific

unmethylated sequences, they all cleaved the DNA at random locations far removed from these sites.

The first restriction nuclease that did cleave at a specific site in DNA was discovered in 1970 by Hamilton Smith of Johns Hopkins University, who followed up his accidental finding that the bacterium *Haemophilus influenzae* rapidly broke down foreign phage DNA. This degradative activity was subsequently observed in cell-free extracts and shown to be due to a true restriction nuclease. Thomas Kelly and Smith determined that HindII, as this enzyme is called, bound to the following sequence, in which the arrows indicate the exact cleavage sites on each DNA strand, and "Py" and "Pu" represent any pyrimidine or purine residue:

5′-G T Py ↓ Pu A C-3′
3′-C A Pu ↑ Py T G-5′

Since then, restriction enzymes that cut specific sequences have been isolated from several hundred bacterial strains, and more than 150 different specific cleavage sites have been found (Table 4-1). Nearly all restriction enzymes recognize *palindromic sites*—that is, the sequences from 5′ to 3′ on each strand are the same. We know now that most restriction enzymes are dimers of two identical subunits that fit on the DNA so that their twofold axis of symmetry is the same as that of the sequence. The enzyme moves along the DNA strand, and when it reaches the target sequence, the DNA is distorted and the strands are cut.

Restriction enzymes recognize specific sequences of four to eight base pairs. A given 4-bp site occurs on average every 256 bp, so an enzyme cutting a four-base sequence will produce many more, much smaller DNA fragments than an enzyme cutting an 8-bp sequence, which will occur on average every 65,536 bp. However, predicting the size of DNA fragments produced by a restriction enzyme is complicated by many factors. One factor that affects the frequency with which DNA is cut is its base composition. Take, for example, the enzyme NotI, which has an eight-base recognition sequence that includes two cytidine–guanine dinucleotides (usually abbreviated as CpG). This dinucleotide sequence occurs much less frequently in vertebrate genomes than would be expected statistically—at about 20% of the expected frequency—and thus NotI cuts less frequently than expected. Furthermore, cytidine is

TABLE 4-1. Some restriction enzymes and their cleavage sequences

Microorganism	Enzyme abbreviation	Sequence	Notes
Haemophilus aegyptius	HaeIII	5′ . . . G G|C C . . . 3′ 3′ . . . C C|G G . . . 5′	1
Thermus aquaticus	TaqI	5′ . . . T|C G A . . . 3′ 3′ . . . A G C|T . . . 5′	2
Haemophilus haemolyticus	HhaI	5′ . . . G C G|C . . . 3′ 3′ . . . C|G C G . . . 5′	3
Desulfovibrio desulfuricans	DdeI	5′ . . . C|T *N* A G . . . 3′ 3′ . . . G A *N* T|C . . . 5′	4
Moraxella bovis	MboII	5′ . . . G A A G A $(N)_8$| . . . 3′ 3′ . . . C T T C T $(N)_7$| . . . 5′	5
Escherichia coli	EcoRV	5′ . . . G A T|A T C . . . 3′ 3′ . . . C T A|T A G . . . 5′	1
	EcoRI	5′ . . . G|A A T T C . . . 3′ 3′ . . . C T T A A|G . . . 5′	2
Providencia stuarti	PstI	5′ . . . C T G C A|G . . . 3′ 3′ . . . G A|C G T C . . . 5′	3
Microcoleus	MstII	5′ . . . C C|T *N* A G G . . . 3′ 3′ . . . G G A *N* T|C C . . . 5′	4
Nocardia otitidiscaviarum	NotI	5′ . . . G C|G G C C G C . . . 3′ 3′ . . . C G C C G|G C G . . . 5′	6

Notes
1. Enzyme produces blunt ends.
2. The single strand is the 5′ strand.
3. The single strand is the 3′ strand.
4. The base pair *N* can be any purine or pyrimidine pair.
5. The enzyme does not cut within the recognition sequence, but at whatever sequence lies eight nucleotides 3′ to the recognition site.
6. NotI has an eight-base recognition sequence and cuts mammalian DNA very infrequently.

often chemically modified by methylation and, as NotI cannot cut methylated CpG sites, it will produce even larger fragments. Moreover, the CpG sites are not distributed randomly throughout the genome—they tend to be clustered in the regions 5′ to genes. NotI produces fragments between 1 million and 1.5 million base pairs in size instead of the 65,536 base pairs expected of an "eight-base cutter." Enzymes like NotI that produce very large DNA fragments have proved invaluable for long-range physical mapping of mammalian DNA.

Electrophoresis Is Used to Separate Mixtures of Nucleic Acid Fragments

Proteins and the nucleic acid fragments produced by restriction enzymes can be analyzed by *electrophoresis.* Such molecules, differing in electrical charge, size, and shape, move at different rates in an electrical field and so can be separated under conditions in which they retain their biological activity. It was Arne Tiselius and Theodor Svedberg who, in 1926, developed moving boundary electrophoresis. Subsequently robust solid supports were used, most commonly filter paper and gels made of polymers such as agarose or polyacrylamide.

The rate at which the fragments migrate through the gel is a function of their lengths, charge, and shape, with small fragments generally moving much faster than large fragments (Fig. 4-1). Different supports are used depending on the absolute sizes of the fragments to be separated as well as on the relative differences in the sizes of these fragments. DNA fragments ranging in size from 100 base pairs to several thousand base pairs can be efficiently separated in a gel made of 1% agarose, provided these fragments differ in size from each other by a few hundred base pairs. When it is necessary to resolve fragments of DNA differing from one another by only a single nucleotide (e.g., in DNA sequencing), polyacrylamide gels are needed.

At first, when electrophoresis was used to analyze viral DNA fragments produced by restriction enzymes, the viral DNA was first radioactively labeled and the fragments were detected by autoradiography. A most useful advance came with the realization that DNA-binding dyes such as ethidium bromide could be used to stain DNA fragments directly in the gel. The fragments appear as a series of bands when a stained gel is illuminated with UV light, each band corresponding to a restriction frag-

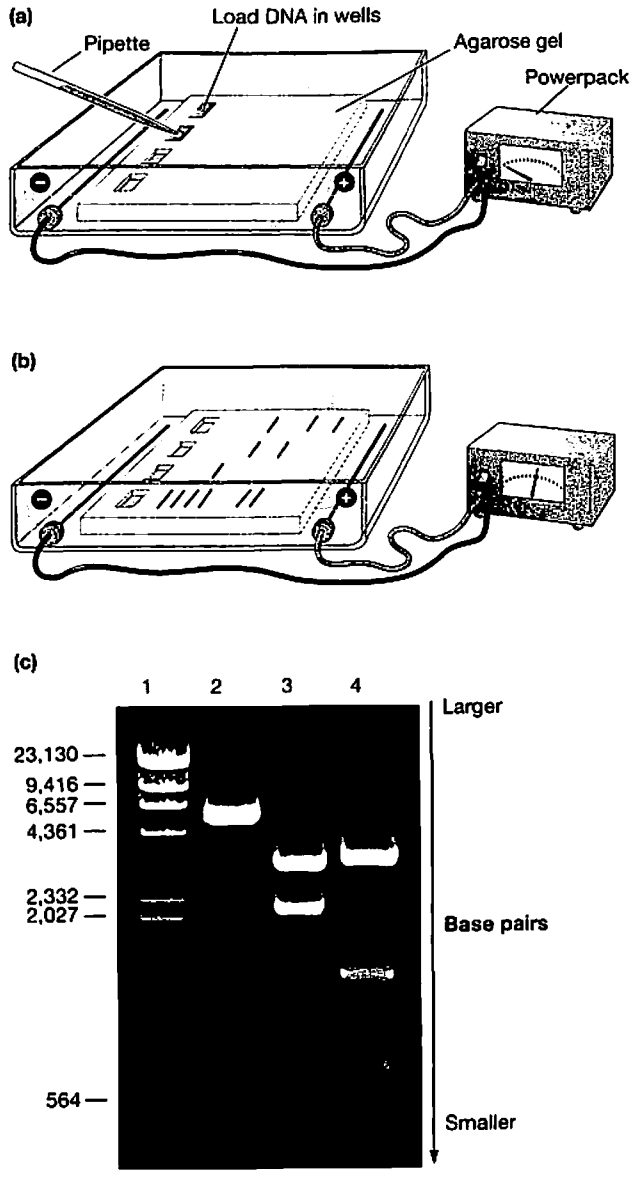


FIGURE 4-1

DNA molecules of different sizes can be separated by electrophoresis. Electrophoresis is a process in which an electrical field is used to move the negatively charged DNA molecules through porous agarose gels. (*a*) A slab of agarose is cast with slots to take the samples. These usually include a blue dye to make it easier to pipette the samples into the slots. (*b*) The current is turned on and the DNA molecules (negatively charged) move toward the positive electrode. Small DNA molecules move faster than larger molecules, and molecules of the same size move at the same speed so that they become separated into bands. (Other factors, including charge and shape, also influence the movement of molecules through the gel.) (*c*) These become visible when the gel is removed from the electrophoresis tank and stained with ethidium bromide. This intercalates between the nucleotides and fluoresces when illuminated with UV light. In the gel shown, the plasmid pCDNA3.1 has been digested with EcoRI (lane *2*), SalI (lane *3*), and NcoI (lane *4*). Lane *1* contains HindIII-digested λ DNA as size markers.

ment of specific size. The sizes can be established by calibration with DNA molecules of known sizes. Bands containing as little as 0.005 μg of DNA can be visualized. Most importantly, DNA molecules move unharmed through these gels and can be recovered (eluted) from the gel as biologically intact double helices for use in experiments.

Restriction Enzyme Sites Are Used to Make Maps of DNA

Different restriction enzymes, having different target sequences, necessarily give different restriction fragments for the same DNA molecule. If restriction enzymes are used singly and in combination, and the lengths of the fragments are carefully determined by electrophoresis, then the fragments can be reassembled to make a *restriction map* of the molecule showing the positions of sites cut by the enzymes. The first such map was obtained in 1971 by Daniel Nathans, a colleague of Hamilton Smith at Johns Hopkins. Nathans used the HindII enzyme to cut ("digest," in the jargon of the field) the circular DNA of the SV40 virus into 11 specific fragments. By measuring the sizes of fragments produced when the DNA was completely cut, and also the overlapping intermediate-sized fragments produced by brief treatments with HindII (resulting therefore in only partial digestion of the DNA), Nathans located the sites on the circular viral DNA that are attacked by the restriction enzyme. With this information, it became possible to determine the regions of biological importance on the circular viral DNA. For example, brief radioactive labeling of replicating viral DNA followed by digestion with HindII allowed Nathans to prove that the replication of SV40 DNA always begins in one specific HindII fragment and proceeds bidirectionally around the circular DNA molecule. Restriction enzymes thus provide a critical tool for molecular biologists to predictably fragment a given DNA molecule and to analyze its structure and features in fine detail. This becomes especially powerful when combined with techniques that detect DNA fragments with specific sequences ("Southern blotting," see Fig. 4-11).

DNA Ligase Joins the Ends of DNA Fragments Produced by Restriction Enzymes

Restriction enzymes together with gel electrophoresis provide a convenient means to purify specific DNA fragments and also, because of the way they cut double-stranded DNA, a way to recombine fragments. Some restriction enzymes, such as HindII, cut DNA at the center of their recognition site to produce blunt-ended fragments; these fragments are fully base-paired out to their ends and have no tendency to stick together (Fig. 4-2). In contrast, many restriction enzymes make staggered or asymmetric cuts on the two strands of the DNA to create short complementary single-stranded tails on the ends of each fragment. The EcoRI recognition sequence is GAATTC and it is cut to generate a four-base-long tail, AATT (Fig. 4-2b). The single-stranded tail sequences left by different restriction enzymes can be similar even though the recognition sites are different. MfeI recognizes the sequence CAATTG and cuts this to produce the same AATT tail as does EcoRI (Fig. 4-2c).

Complementary single-stranded tails tend to associate by base pairing and thus are often called *cohesive*, or *sticky, ends*. For example, the linear molecules that EcoRI generates by cutting circular SV40 DNA often temporarily recyclize because of base pairing between their tails. Base pairing occurs only between complementary base sequences, so the cohesive AATT ends produced by EcoRI will not, for example, pair with the AGCT ends produced by HindIII. However, any two DNA fragments with complementary ends (regardless of the organism from which they originate) can associate by base pairing and temporarily anneal. What was needed was a way to permanently join the associated ends.

In fact, such enzymes had already been discovered in 1967. The *DNA ligases* play important roles in DNA repair and DNA replication, where a ligase links the DNA Okazaki fragments generated on the lagging strand. By combining restriction enzymes and ligases, scientists had the tools to cut DNA and to paste the fragments together. Such experiments were first done at Stanford University in 1972 by Janet Mertz and Ronald Davis, who realized that EcoRI in conjunction with DNA ligase provided a general way to achieve in vitro, site-specific genetic recombination.

Thus, by the early 1970s, techniques for manipulating and isolating genes were available and plans were under way to clone genes and produce large quantities of the cloned genes in bacteria. But even as the first plans were being made to carry out gene cloning experiments, concerns began to be raised about the safety of such experiments.

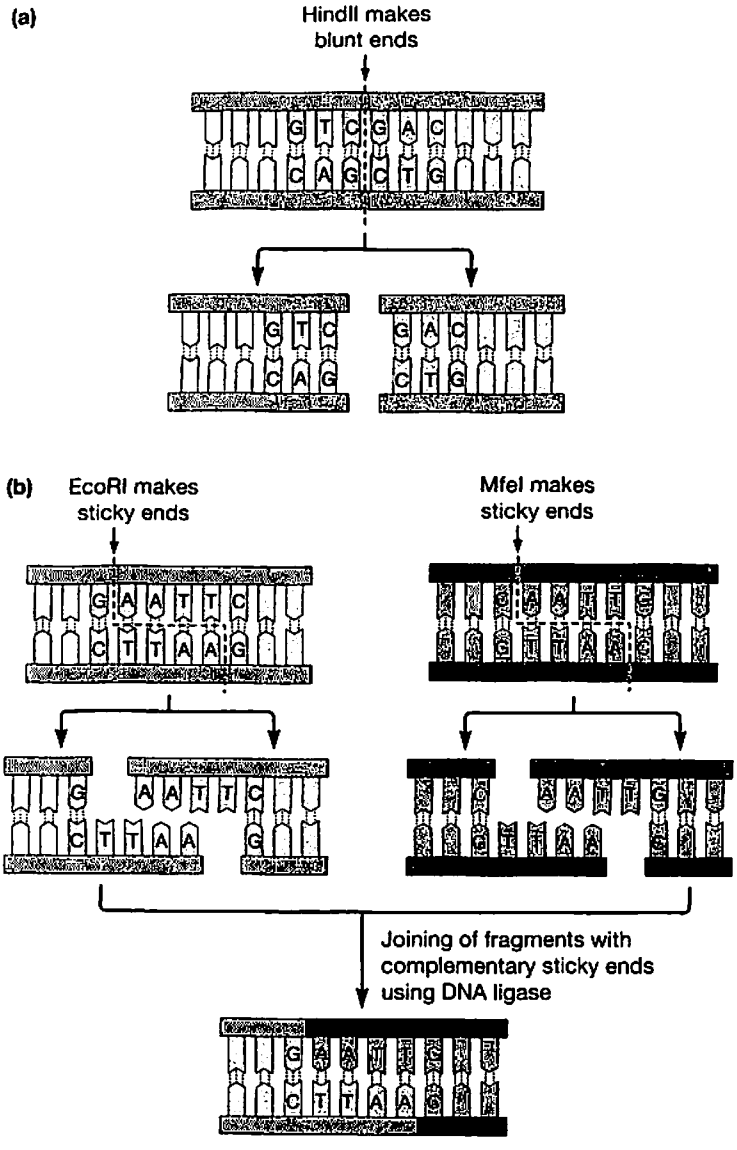


FIGURE 4-2

Restriction enzymes can produce blunt ends or "sticky" ends. (*a*) The HindII restriction enzyme cuts DNA at the center of its recognition site, leaving blunt ends. These blunt ends can be ligated to blunt ends produced by any other process. (*b*) The EcoRI and MfeI restriction enzymes make staggered, symmetrical cuts in DNA away from the center of the recognition site, leaving cohesive, or sticky, ends. A sticky end produced by EcoRI digestion can anneal to any other sticky end produced by EcoRI cleavage or any other sticky end with complementary sequence produced through other methods. For example, the restriction enzyme MfeI produces identical overhangs to EcoRI, although the flanking sequences are different. MfeI sticky ends can be ligated to those from EcoRI, although the ligated product cannot be recut with either enzyme because the resulting sequence is recognized by neither enzyme.

Scientists Voice Concerns about the Dangers of Gene Cloning at the Asilomar Conference

These first concerns—prior to the cloning of any gene—centered on proposals to clone the genomes of DNA tumor viruses and replicate the recombinant molecules in bacterial cells like *E. coli*. The question was: Could *E. coli*, present in the intestinal tract of all human beings, act as a vector to transmit the cancer-causing genes of tumor viruses to human beings? In 1972, it was pointed out that the DNA of mice, and possibly that of human beings, harbored the DNA genomes (proviruses) of latent RNA tumor viruses. Many of the recombinant plasmids that scientists would isolate in their search for specific genes might contain as well these DNA proviruses that might be able to cause human cancer. If so, should all recombinant DNA experiments using vertebrate DNA be regarded as possibly dangerous? The answers were mixed. This was hardly surprising given that the necessary data were not yet available to allow rational assessment of the risks. There was virtually unanimous agreement, however, that researchers should not have unrestricted freedom to do experiments that might have military consequences.

Thus the first discussions of the conjectured hazards of recombinant DNA techniques took place before any experiments had been carried out. However, the announcement of the first successful

cloning experiments brought matters to a head as it became clear that the future of studying the molecular basis of genetics lay in using recombinant DNA techniques. The question then was whether to move ahead as fast as possible or to try to devise methods that would allay worries about the possible risk of these experiments. In July of 1974, a letter appeared in *Science* urging that scientists considering recombinant DNA experiments should pause until there had been further evaluation of the risks involved. This evaluation took place in February of 1975, when a group of more than 100 internationally respected molecular biologists gathered at the Asilomar Conference Center located near Monterey, California. In the absence of clear knowledge whether any danger might exist, a nearly unanimous consensus emerged that some restrictions on DNA cloning were appropriate.

Afterward, the "Asilomar recommendations" were considered by a special committee appointed by the National Institutes of Health (NIH). In its deliberations, the committee recommended guidelines that effectively precluded the use of recombinant DNA techniques for studying the genes of cancer viruses and required the use of genetically disabled bacteria that would not grow well outside the laboratory. These recommendations became codified in official federal regulatory guidelines that took effect in July 1976. These regulations in the United States were paralleled by the establishment of guidelines developed by similar bodies in Europe.

A remarkable feature of the discussions about the safety of recombinant DNA was the unprecedented participation of nonscientists in communities like Cambridge, Massachusetts, in the debates. This led to the appointment of lay members of the Recombinant DNA Advisory Committee and ensured an increasing involvement of nonscientists in discussions about the societal impact of genetics. The most notable example is the Ethical, Legal, and Social Issues (ELSI) Research Program of the Human Genome Project (Chapter 11).

Many scientists thought that the NIH regulations were too restrictive and in some cases scientifically unsound. As more experiments were performed, and more data accumulated, it became increasingly apparent that the actual hazards of recombinant DNA experiments were extremely small. Discussions throughout 1978 led to new, less restrictive NIH regulations that took effect in January 1979 and permitted cloning of viral cancer genes.

Plasmids and Viruses Are Used as Vectors to Carry DNA Sequences

The goal of cloning is not only to isolate a specific gene, but also to produce large quantities of the gene for analysis and further manipulations. The obvious way to do this was to use bacteria to replicate the gene, even if the gene was not bacterial in origin. This requires linking the cloned gene to DNA sequences that have the appropriate signals for directing replication of the hybrid molecule in bacteria. Fortunately, bacteria possess *plasmids*, which have just the right properties to act as *vectors* for carrying foreign DNA fragments into bacteria. Plasmids are tiny, circular DNA molecules—only a few thousand base pairs long—found in bacteria. There may be as few as two or as many as several hundred in each bacterial cell (referred to as *copy number*) and they replicate independently of the host bacterial chromosome. Plasmids were first noticed as genetic elements that were not linked to the main chromosome and that carried genes that made the bacteria resistant to antibiotics such as tetracycline or kanamycin. That these genes were found on plasmids as opposed to main-chromosomal DNA was not a matter of chance. Antibiotic resistance requires relatively large amounts of the enzymes that chemically destroy the antibiotics, and because the genes encoding these enzymes are carried on plasmids, they are present in much higher numbers than would be the case if they were located on the single bacterial chromosome. The small size of plasmids (which makes it easier to manipulate them) and the presence of antibiotic resistance genes (which provides a means to select for bacteria containing the plasmids) made them ideal vectors for cloning.

In the first cloning experiments, Stanley Cohen and Herbert Boyer and their collaborators at Stanford and the University of California, San Francisco, used two plasmids—pSC101 and pSC102—found in *E. coli* (Fig. 4-3). pSC101 was chosen as the vector because it contained only a single EcoRI recognition site that would accept DNA fragments cut with EcoRI, and it could be selected for by using the antibiotic tetracycline because it carried a gene for tetracycline resistance. pSC102 contained a gene for resistance to the drug kanamycin, and this was the gene to be cloned. The two plasmids were mixed and converted into linear molecules with EcoRI sticky ends by treatment with EcoRI. After allowing the DNA fragments to link and using DNA

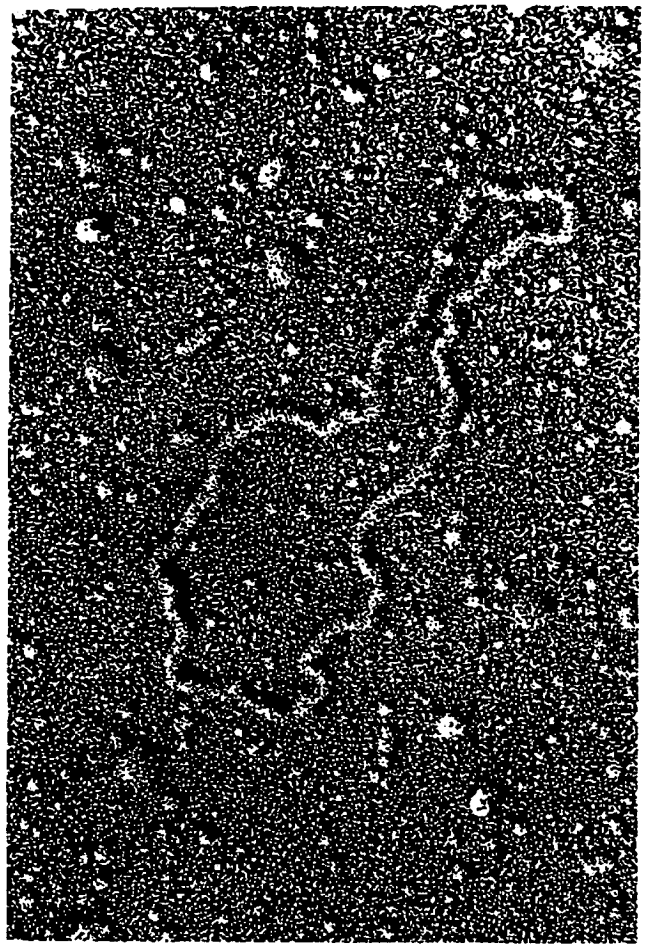


FIGURE 4-3

Electron micrograph of plasmid pSC101. The plasmid is a circle of double-stranded DNA, of 9263 base pairs approximately 1.3 μm long.

ligase to make the links permanent, the recombined DNA molecules were introduced into the bacteria. These were plated on media containing both tetracycline and kanamycin. Those bacteria that survived and grew were resistant to both antibiotics because they contained recombinant plasmids carrying tetracycline resistance from the pSC101 plasmid and kanamycin resistance from pSC102.

Although in the original Cohen–Boyer experiment the two DNAs were both from plasmids, it was clear that all sorts of foreign DNA, from microbes and, most importantly, from higher plants and animals, could be cloned into these plasmids. For example, the *E. coli* chromosome with its 4 million base pairs contains about 500 different recognition sites for EcoRI. By random insertion of individual EcoRI fragments into pSC101, it would be possible to clone all the *E. coli* genes. Plasmids were engineered to be ever more useful and versatile; an early example was pBR322 and elements of this were used to create others, including the popular pUC18 and pUC19. These have additional enzyme cleavage sites for cloning, have different drug-resistance genes, and are present at much higher copy number than other plasmids. All plasmids have one great advantage—they are easily purified from *E. coli* (Fig. 4-4).

Encouraged by their success using plasmids, scientists looked to nature to find other vectors suitable for cloning in bacteria, and bacteriophages were an obvious choice. *Transducing bacteriophages* are naturally occurring vectors that carry parts of bacterial chromosomes from one bacterial cell to another. (Transducing phages integrate into the host bacterial genome to wait for the proper conditions to excise themselves. If this excision process is not precise, the phage may carry a segment of the bacterial genome with it. Such transduction events had long been used for fine-scale mapping of bacterial genes.) Bacteriophage λ was rapidly adapted as a general cloning vector and engineered, by in vitro recombination methods, to carry virtually any fragment of foreign DNA. Much larger fragments of DNA can be cloned in bacteriophages than in plasmids, although there is an upper limit set by the need to package the DNA in the phage protein coat; a maximum of 25 kb of DNA can be cloned in bacteriophage λ. As for plasmids, genetic engineers created more sophisticated phage vectors, including, for example, λgt and λZAP. Subsequently, other vectors with much larger capacity have been developed; bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs) can hold inserts as long as 100–500 kb and 250–1000 kb, respectively (Table 4-2). (Large-scale vectors are discussed in Chapter 11.)

There Are Five Basic Steps in Cloning

There are five steps in the basic process of cloning DNA, described here for cloning in plasmid vectors (Fig. 4-5). The first is to choose the appropriate DNA to be cloned. Sources of DNA for cloning are genomic DNA or a so-called complementary DNA (cDNA), a DNA copy of an mRNA made using the enzyme reverse transcriptase.

The second step is to produce a collection of DNA fragments of a size suitable for inserting into appropriate vectors carrying one or more antibiotic resistance genes. Although these fragments are pre-

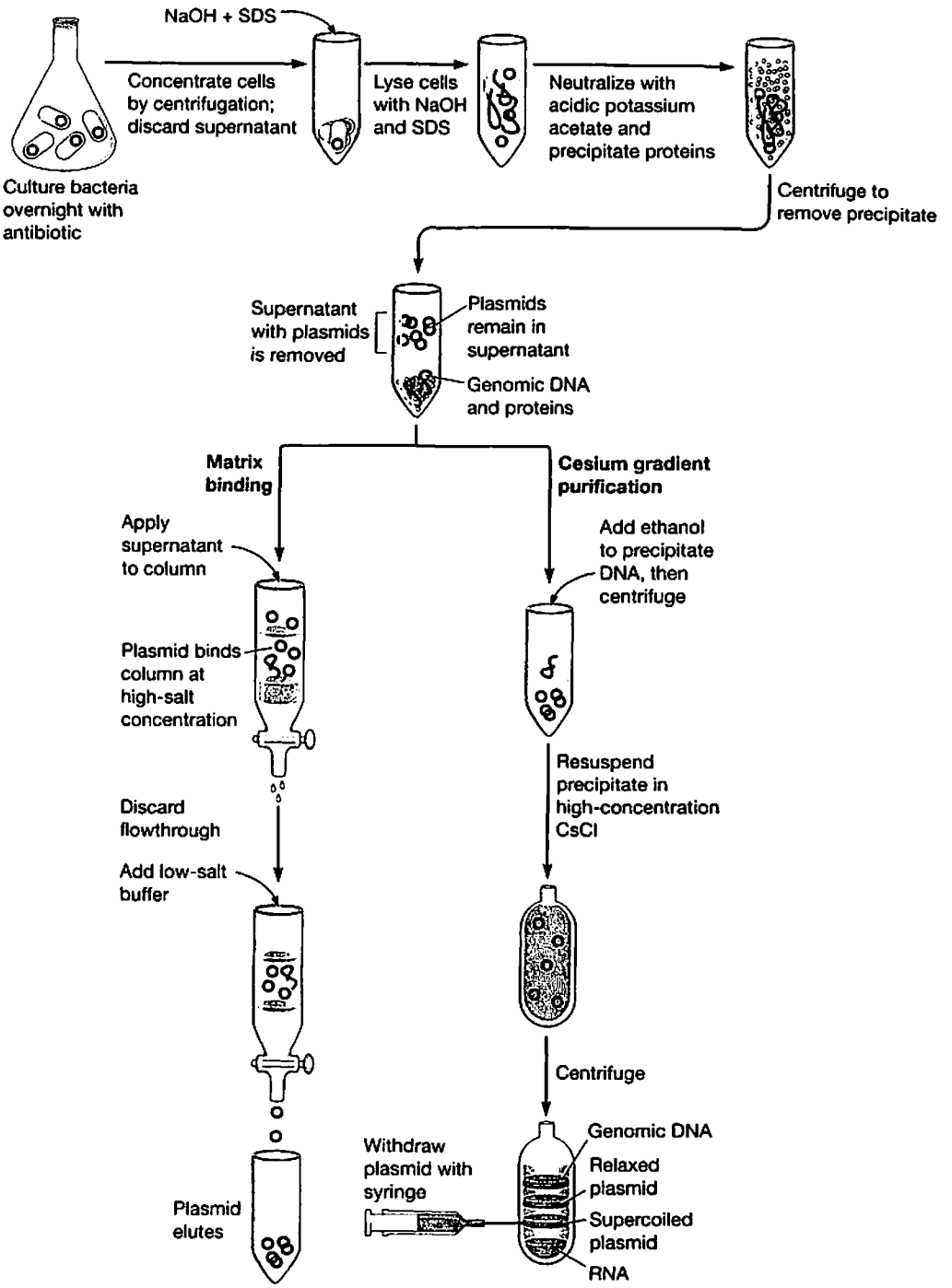


FIGURE 4-4

Purification of plasmid DNA from *Escherichia coli*. Bacteria containing a plasmid with a cloned DNA insert and an antibiotic resistance gene are grown in medium containing the antibiotic to select for bacteria carrying the plasmid. Most plasmids used for molecular biology are *multicopy* (i.e., present in many copies per cell). The medium is centrifuged to pellet the *E. coli* cells and the supernatant is discarded. Sodium hydroxide and the detergent sodium dodecyl sulfate are added to lyse the bacteria, thus freeing the plasmids. Potassium acetate is added, which forms an insoluble precipitate of potassium dodecyl sulfate that contains the *E. coli* genomic DNA and many proteins. This precipitate is removed by centrifugation; the small plasmid DNA remains in the supernatant, which also contains *E. coli* RNA. Two methods are shown for further purification of the plasmids. One method uses density centrifugation in a cell to separate out the plasmid DNA from other nucleic acids. (Plasmid DNA can be very tightly coiled ["supercoiled"], or one strand can be nicked, thus allowing it to adopt a "relaxed shape.") Another approach is to use a resin that specifically binds DNA in buffers of high salt concentration. The column with the bound plasmid DNA is washed several times before eluting the plasmid DNA with a low salt buffer.

TABLE 4-2. Cloning capacity of commonly used vectors

Vector	Insert size range (kb)
Plasmid	<10
Phage	<23
Cosmid	30–46
P1 artificial chromosome (PAC)	130–150
Bacterial artificial chromosome (BAC)	<300
Yeast artificial chromosome (YAC)	200–2000

pared from chromosomal DNA by using a restriction endonuclease, most cDNAs are of a suitable size for cloning without any further manipulation.

The third step is to insert the DNA into the vector and use DNA ligase to covalently link the DNA fragment to the vector DNA. In the fourth step, the vectors with inserted DNA fragments are introduced into a population of bacteria; this step is called *transformation*. The transformed bacteria are plated on agar containing an antibiotic. Only those bacteria containing plasmids will grow because the plasmids confer antibiotic resistance on the bacteria. The bacteria are plated at low density so that as each resistant bacterial cell divides, it gives rise to a colony of bacteria, all derived from a single cell containing a single molecule of recombinant DNA. This is the cloning step proper. Such a collection of cloned DNA fragments propagated in bacteria is called a *library*.

This library is a library without a catalog to tell us which clone contains a particular sequence, so the fifth step is to determine which colonies contain the desired sequence. In the next sections we will review each of these steps in more detail.

Choosing the Right Starting Material Is Essential in Cloning

The first decision to be made is whether to clone a gene starting with chromosomal DNA or a cDNA copy of an mRNA. The choice depends on the particular problem to be tackled. If one is interested in the amino acid sequence of a protein, this information can be obtained most readily from the nucleotide sequence of a cloned cDNA. On the other hand, if one is interested in the regions of a gene that regulate its expression, or in gene sequences not contained within the mRNA, then this information can be obtained only from genes cloned from chromosomal DNA.

The purity of the starting DNA sample is important because if DNA from other organisms is present it will be represented in the final library. This type of

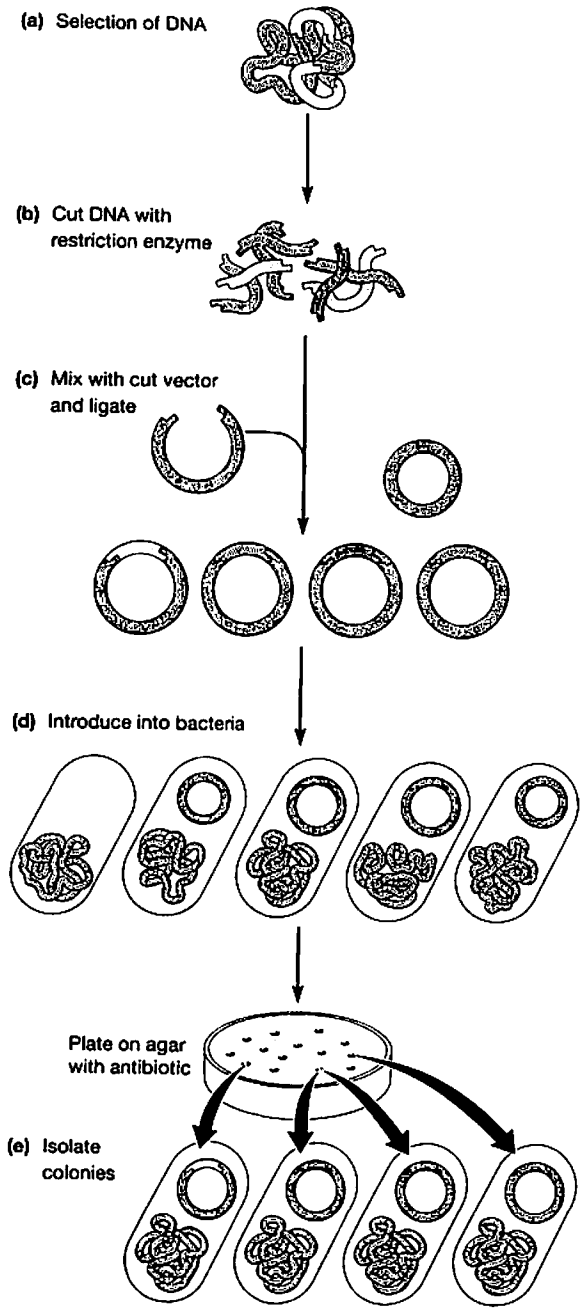


FIGURE 4-5

The five basic steps of cloning DNA in a plasmid. (*a*) DNA containing the gene of interest must be prepared. (*b*) The DNA is treated with a restriction enzyme to cut the DNA into fragments of a size suitable for inserting into a vector that has been cut with the same restriction enzyme. (*c*) Third, the DNA fragments are mixed with vector molecules and DNA ligase is used to covalently link them. (*d*) In the fourth step, the recombinant molecules are introduced into bacteria that are plated on agar containing an antibiotic. This kills any bacteria that have not taken up a plasmid (which carries an antibiotic resistance gene). The bacteria are plated at low density so that each colony arises from a single bacterial cell carrying a plasmid. (*e*) Finally, those very few clones containing the gene of interest are selected from the library.

DNA mixing could occur if, for example, plants harvested for DNA extraction had bacteria and fungi growing on the surface of the leaves and roots. It is also important to ensure that the DNA sample contains the gene of interest. It would not be a good idea to try to clone human Y-chromosome genes beginning with a sample of DNA from a woman.

The mRNA that will serve as the template for the synthesis of a cDNA must be prepared from cells that express the gene of interest. As any given cell type expresses only a subset of its chromosomal genes, proper selection of the starting tissues or cells is especially important. This may not be easy, because many genes are expressed in only a limited number of cell types or under certain growth conditions or at particular stages in development. The first eukaryotic gene to be cloned was the rabbit β-globin gene because there was a readily available and plentiful source of β-globin mRNA—reticulocytes in the blood; 50–90% of the total mRNA in a reticulocyte is globin mRNA. Similarly, pancreatic cells were used as the source of mRNA for cDNA cloning of the insulin gene.

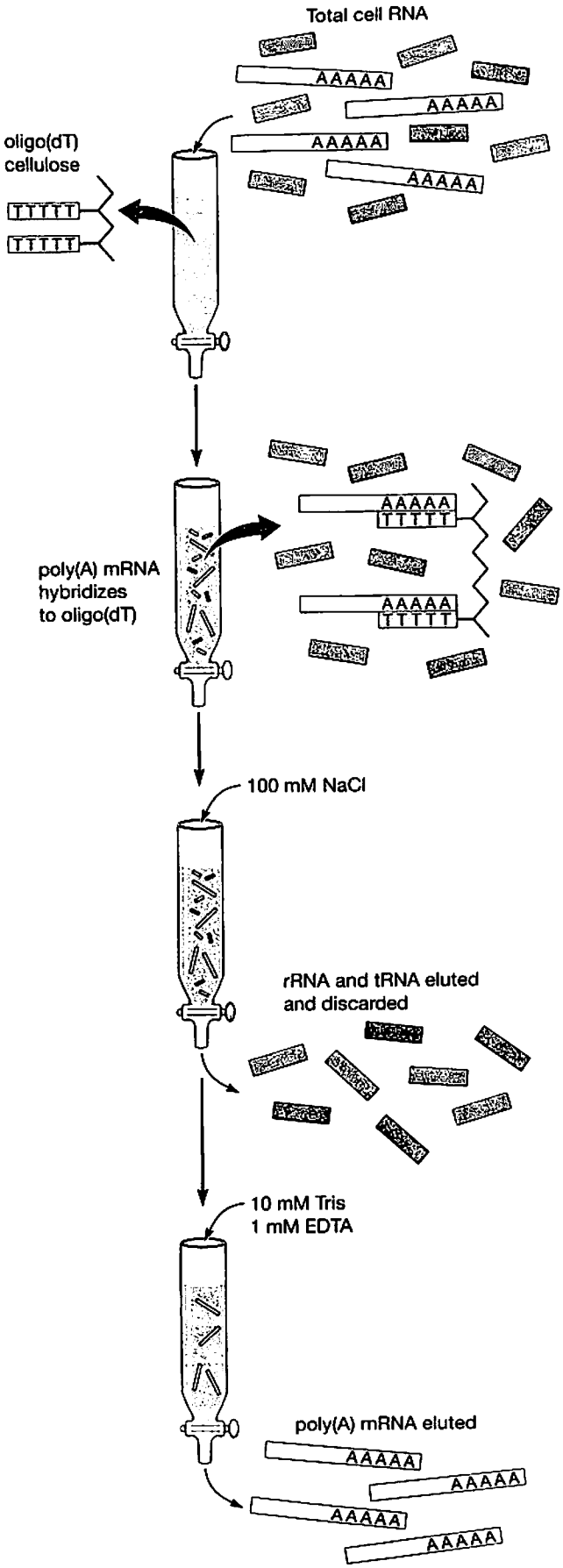
Techniques have improved so that we are now able to clone cDNAs from rare mRNAs, but the relative abundance of mRNAs from different genes affects the relative number of copies of each cDNA present in the library. There will be many more cDNAs derived from an abundant mRNA than from a rare mRNA, so that there will be many more clones of the former than the latter in a library, making it difficult to find clones of a rare transcript. Several techniques have been developed for ***normalization*** of cDNA libraries, which reduce the proportion of clones corresponding to abundant mRNAs.

mRNA Is Converted to cDNA by Enzymatic Reactions

The first step in constructing a cDNA library involves isolating total cellular RNA. A fraction that contains mostly mRNA is then isolated from the total cellular RNA. As we will describe in detail in Chapter 5, the majority of eukaryotic mRNA molecules have a run of adenine nucleotide residues called a ***poly(A) tail*** at their 3′ ends. This tail is important for the function of the mRNA, but also the poly(A) provides a very convenient way to isolate mRNAs from total cellular RNA, the bulk of which is ribosomal RNA and tRNA, which do not have poly(A) tails (Fig. 4-6). (The fraction of poly(A)-containing mRNA is usually only ~1–2% of total cellular RNA.) The selection relies on the use of oligonucleotides composed only of deoxythymidine (oligo(dT)), which can be linked to a solid support such as beads of cellulose. These beads are packed into small columns. When a preparation of total cellular RNA is passed through such a column, the poly(A) tails of the mRNA molecules bind to the oligo(dT), thereby trapping the mRNAs to the support while the rest of the RNA flows through the column. The bound mRNAs are then eluted from the column and collected.

The poly(A) tails of the mRNA molecules are also used for the next step of cDNA cloning, the preparation of a DNA copy of the RNA (Fig. 4-7). Short oligonucleotides containing 12–20 deoxythymidines (oligo(dT)) are mixed with the purified mRNA and hybridized to the poly(A) tails, where they act as primers for reverse transcriptase. This enzyme, which is isolated from certain RNA tumor viruses (see Chapter 5), can use RNA as a template to synthesize a DNA strand. (Its name comes from its ability to reverse the normal first step of gene expression.) The product of the reaction is an RNA–DNA hybrid. Using oligo(dT) as the primer has the disadvantage that, because the reverse transcriptase must begin at the 3′ end of its mRNA, it may not reach the 5′ end of the molecule. This is a particular problem for very long mRNAs. To circumvent this difficulty, a second method, ***randomly primed*** cDNA synthesis, is used. Oligonucleotide fragments, 6–10 nucleotides long and made up of many possible sequences, are used as primers for the cDNA synthesis. In this way, priming of the mRNA occurs from many positions, not only from the 3′ end. Sequences close to the 5′ end of long mRNAs are more readily cloned using this method. Again the product of reverse transcription is an RNA–DNA hybrid. At this point, the RNA–DNA hybrid molecules must be converted into double-stranded DNA molecules that can be cloned into appropriate vectors.

The most common way to synthesize a double-stranded cDNA from an mRNA–cDNA hybrid makes use of an enzyme from *E. coli*, RNase H, that recognizes RNA–DNA hybrid molecules and digests the RNA strand into many short pieces. These RNA pieces remain hybridized to the first cDNA strand and serve as primers for *E. coli* DNA polymerase, which uses the original cDNA as a template to synthesize the complementary strand of DNA. Eventually, this process completely replaces the original RNA with DNA, except for a small piece of RNA at the extreme 5′ end. The new DNA strand is not entirely contiguous but rather contains breaks ("nicks"). These breaks



are joined (ligated) by the action of DNA ligase, thus forming a double-stranded DNA molecule.

cDNA Molecules Are Joined to Vector DNA to Create a Clone Library

The double-stranded cDNA molecules obtained by these procedures are then inserted into a plasmid or a phage vector. Modern plasmid vectors have been extensively engineered to contain a variety of features to aid the cloning process, including, for example, ***polylinker regions*** containing many restriction enzyme sites. For library construction, artificial restriction enzyme sites are added to the ends of the cDNAs (Fig. 4-8). The restriction enzyme sites are contained in 8- to 12-bp oligonucleotides, synthesized chemically, called ***linkers*** or ***adaptors.*** These adaptors are added to the double-stranded cDNAs using DNA ligase and then cut with the appropriate restriction enzyme to produce single-stranded sticky ends.

The cDNA, now carrying sticky ends generated by the restriction enzyme, is combined with a vector that has been cleaved with the same enzyme, and the ends are sealed together using ligase. The recombinant molecules are now ready to be introduced into bacterial cells using a method appropriate for the vector type. Plasmids are introduced by a variety of transformation procedures (detailed in Chapter 6), whereas λ phage vectors are first packaged in vitro to form infectious phage particles that can replicate in bacteria. Although plasmid vectors offer the advantage of ease of manipulation of the inserted cDNA, phage libraries contain a greater number of clones and can be screened in much larger numbers.

FIGURE 4-6

Isolation of poly(A) mRNA. Most eukaryotic mRNAs carry a poly(A) tail, which can be used to purify the mRNA fraction from the bulk of cellular RNA. Cellular RNA is passed over a column consisting of an inert material, often cellulose or agarose, to which oligonucleotides consisting entirely of deoxythymidine (dT) residues have been attached. The poly(A) tails hybridize to this oligo(dT), causing the mRNA to stick to the column, whereas the rest of the RNA runs through. After extensive washing of the column to remove the last traces of contaminating material, the column is washed with a buffer of low ionic strength. Under these conditions the poly(A)–oligo(dT) hybrids dissociate, and the purified mRNA washes off the column.

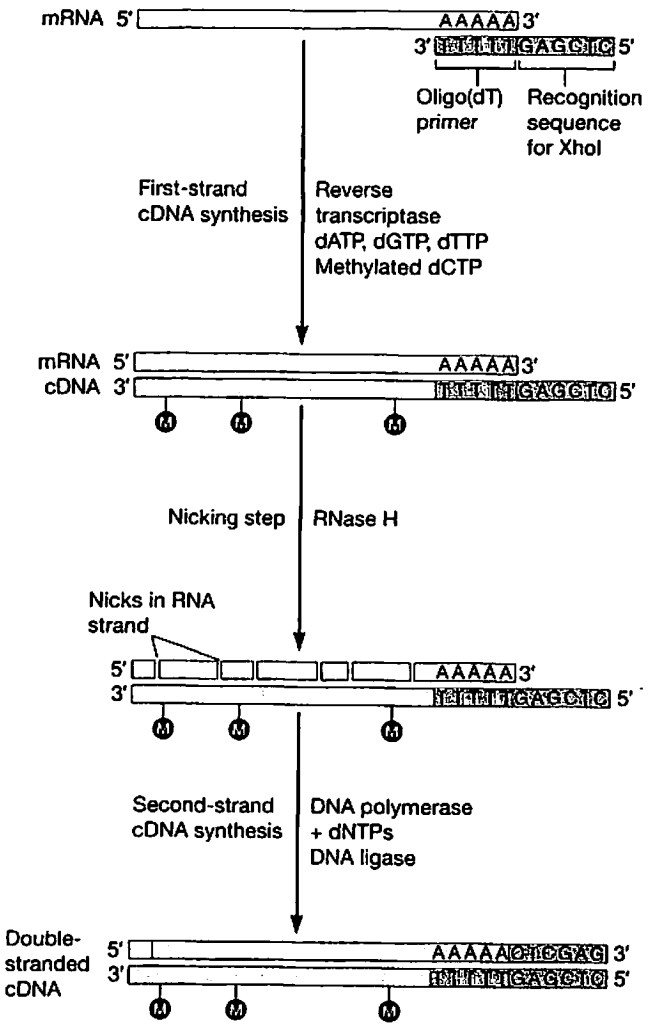


FIGURE 4-7

Synthesis of cDNA. poly(A) mRNA is incubated with deoxythymidine-containing oligonucleotides, which hybridize to the poly(A) tails, forming primed templates for the enzyme reverse transcriptase. When the cDNA will be used to construct a library, the oligonucleotide primers are designed so that they contain a restriction site, such as XhoI, which will be used to create sticky ends for ligation to the vector. Deoxynucleotide triphosphates are added to the mRNA but with methylated deoxycytosine instead of the usual unmethylated deoxycytosine. On addition of reverse transcriptase, cDNA molecules are synthesized on the mRNA templates; this methylated cDNA cannot be cut at internal XhoI sites. The result of the reverse transcriptase reaction is a collection of RNA–DNA hybrids and the RNA strands must be destroyed and replaced with DNA. RNase H nicks the RNA strands and these serve as initiation sites for DNA synthesis by *E. coli* DNA polymerase. This DNA synthesis step is done using unmethylated deoxycytosine to ensure that the XhoI site in the oligonucleotide at the ends of the cDNAs can be digested by XhoI. Eventually, most of the RNA fragments are replaced by DNA. The double-stranded DNA molecules are finally treated with DNA ligase, which seals up any remaining nicks in the new DNA strand.

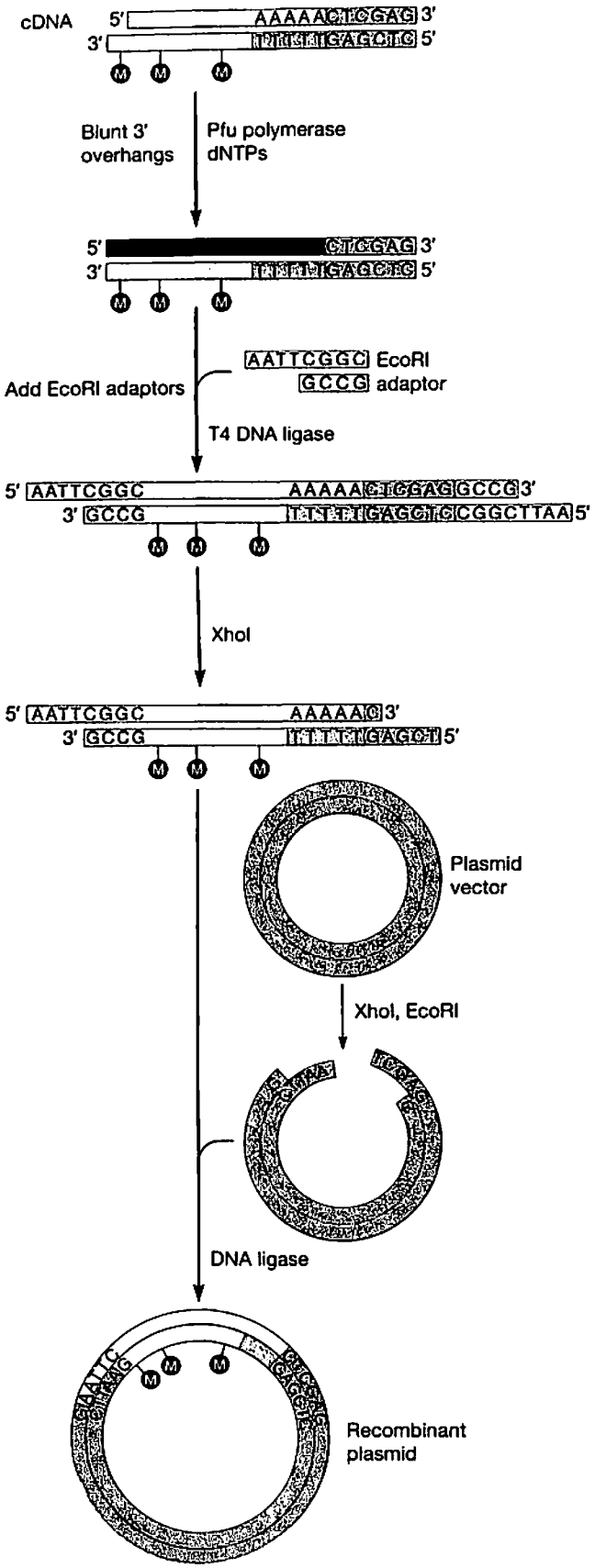
How can the experimenter be certain that the process of cloning to this point has resulted in bacteria containing cloned DNA? One simple approach, called *blue/white screening*, relies on the β-galactosidase gene from the *lac* operon (Chapter 3). The enzyme β-galactosidase hydrolyzes the chemical X-gal to produce an insoluble blue dye. The cloning site of many vectors is within a copy of the β-galactosidase gene, so that the insertion of DNA disrupts the gene and bacterial colonies or phage plaques remain colorless ("white") if X-gal is included in the culture medium (Fig. 4-9). In contrast, colonies or plaques containing vectors without DNA inserts turn blue. Blue/white screening can be used to determine whether recombinant vectors are present at high frequency in the library and picking white colonies or plaques excludes those containing empty vectors.

Libraries of Genomic DNA Represent the Complete Sequence of Organisms

cDNA libraries are useful because they represent the proteins expressed in a given tissue and can be directly used for the expression and study of proteins. However, mRNAs only represent a snapshot of the complete genome of an organism. To study the entire genome (much of which is not expressed as mRNA), it is necessary to create a library containing all segments of the genome of an organism. This is possible by similar techniques to those described above for making cDNA libraries (in fact, it is simpler in that one begins with double-stranded DNA rather than having to make cDNA). One limitation of the plasmid and λ vectors is that they can hold only small DNA fragments. This means not only that many clones are needed to ensure complete coverage of genomic DNA, but also that many large genes cannot be contained on a single clone. Other vectors with much larger capacity, such as BACs and YACs, are indispensable for genome-scale cloning. Genomic libraries have long been used to study specific parts of the genome, and creation of high-quality libraries is essential for genome sequencing, as we shall see in Chapter 11.

Nucleic Acid Probes Are Used to Locate Clones Carrying a Desired DNA Sequence

The net result of plating out a clone library is hundreds of thousands to a million phage plaques or bacterial colonies (for plasmid vectors), each containing



a cloned DNA fragment, distributed on a set of agar plates. Once the library has been plated out, a copy, or *replica*, is prepared on nitrocellulose filters or nylon membranes (Fig. 4-10). This process transfers a portion of each plaque or colony to the nitrocellulose and is done in such a way that the pattern of plaques or colonies on the original plates is maintained on the filters. Screening is carried out by incubating these nitrocellulose replicas with a nucleic acid probe to detect sequences of interest or with an antibody that can detect an expressed protein of interest.

The most direct method of screening is to use nucleic acid hybridization, a very sensitive means for detecting DNA sequences with small pieces of DNA (probes) that are complementary to those sequences. This requires knowledge of the sequences being sought. In some cases, part of the gene may already have been cloned, and this can be used to search for clones that contain additional sequences flanking the starting clone. In other cases, the gene of interest may be selected based on genome sequence information from that organism or from another organism. If the sequence of interest is not precisely known, it is still possible to use hybridization to find partially

FIGURE 4-8

Directional cloning of double-stranded cDNA in a plasmid vector. The cDNA is first manipulated to give it different single-stranded tails, complementary to restriction sites on the vector DNA. The cDNA is treated with *Pfu* DNA polymerase, which has 3′ to 5′ exonuclease activity and 5′ to 3′ polymerase activity. These combined activities fill in 5′ overhangs and remove 3′ overhangs, producing blunt ends on the cDNA product. Next, the EcoRI adaptors are attached to the ends of the cDNA using DNA ligase. Ligation requires a 5′-phosphate and, because the linkers are not phosphorylated, only one linker can be attached at each end of the cDNA, to the free 5′-phosphates. This produces a sticky end compatible with EcoRI at one end of the cDNA. The cDNA is then digested with XhoI. Because methylated cytosines were used for the cDNA synthesis, the only XhoI site that can be cut is that in the reverse transcriptase primer at the 3′ end of the original mRNA. The end product is a cDNA molecule with a sticky EcoRI at one end and a XhoI at the other. Next, the vector DNA is treated with XhoI and EcoRI, and the cut vector is purified. The vector and cDNAs now have compatible ends and they are covalently joined with DNA ligase. The result is a population of circular molecules, with the cDNA directionally cloned with the original 5′ end at the EcoRI site and the 3′ end at the XhoI site.

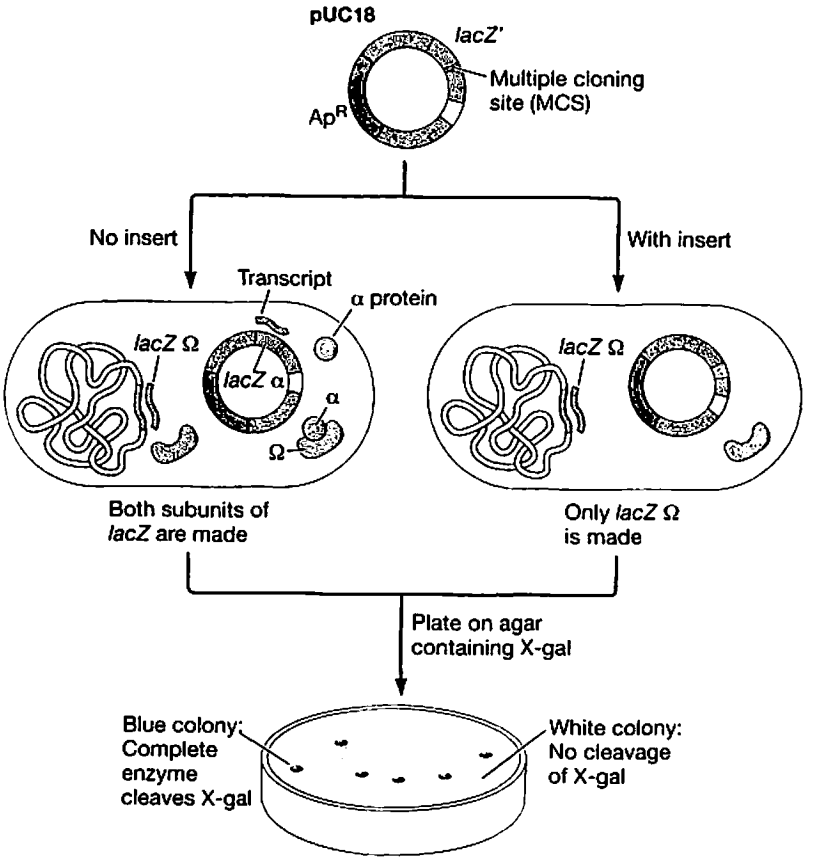


FIGURE 4-9

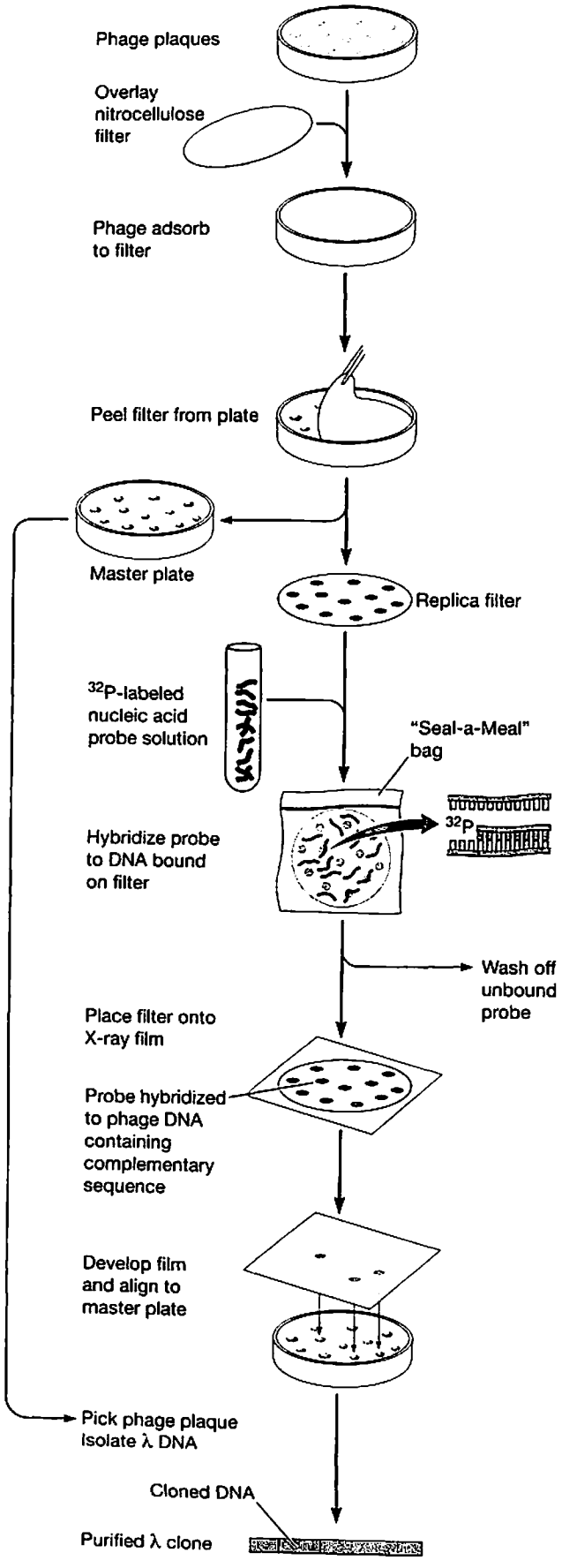
Blue/white screening to recognize vectors containing inserts. The *Escherichia coli lacZ* gene encodes the enzyme β-galactosidase. When bacteria making β-galactosidase are grown on agar containing the colorless chemical X-gal, cleavage of the X-gal by the enzyme produces an insoluble blue product. In many modern vectors, the multiple cloning site is situated within a small 150-nucleotide fragment of the *lacZ* gene (*lacZ'*), which produces a 50-amino-acid-long α fragment. The multiple cloning site does not disrupt the *lacZ'* reading frame. When the α fragment is combined with the *lacZ* Ω fragment, they assemble to form an active enzyme. So, expression of the α fragment from a plasmid in an *E. coli* host producing the Ω fragment leads to the formation of active β-galactosidase, and colonies grown on X-gal containing agar are blue. In contrast, if there is a cDNA insert in the multiple cloning site, the α fragment cannot be synthesized, functional β-galactosidase is not made, and the colonies are colorless ("white"). The blue and white colonies (or plaques if a phage vector has been used) are easily distinguishable, providing a ready means for identifying clones with inserts.

matching sequences. Hybridization between probe and cloned gene sequences containing mismatched bases will occur if screening is performed at lower temperatures (e.g., 42°C instead of 65°C) and higher salt concentrations. The filter with spots showing where the probe hybridized to clones is then compared to the original plate to select clones containing the sequence of interest. These clones are removed from the plates and the phages or plasmids are grown in bulk to produce large quantities of the cloned DNA for further study.

Southern and Northern Blotting Procedures Analyze DNA and RNA by Hybridization

Once a DNA segment of interest has been cloned, it is useful to study its position in the genome and expression in the organism. In 1975 Ed Southern combined gel electrophoresis with hybridization to create an extremely powerful tool for detecting sequences, the so-called *Southern blotting* procedure (Fig. 4-11). Genomic DNA is cut with one or several restriction enzymes. The resultant fragments are denatured to separate the DNA strands and separated by electrophoresis on an agarose gel. The gel is then overlaid with a sheet of nitrocellulose or nylon membrane, and a flow of buffer is set up through the gel, carrying the DNA fragments onto the filter, where they bind. Thus, a replica of the distribution of DNA in the gel is created on the filter. A radioactively labeled nucleic acid probe, complementary to the sequences being sought, is then hybridized to the DNA molecules bound on the filter. Unbound probe is washed off, and the dried filter is exposed to X-ray film. The end product is an *autoradiograph* with a pattern of bands indicating the positions of DNA fragments complementary to the probe. The sizes of the fragments can be estimated by comparisons with size standards run on the same gel.

Southern blotting made possible the production of detailed restriction maps of complex genomes, in

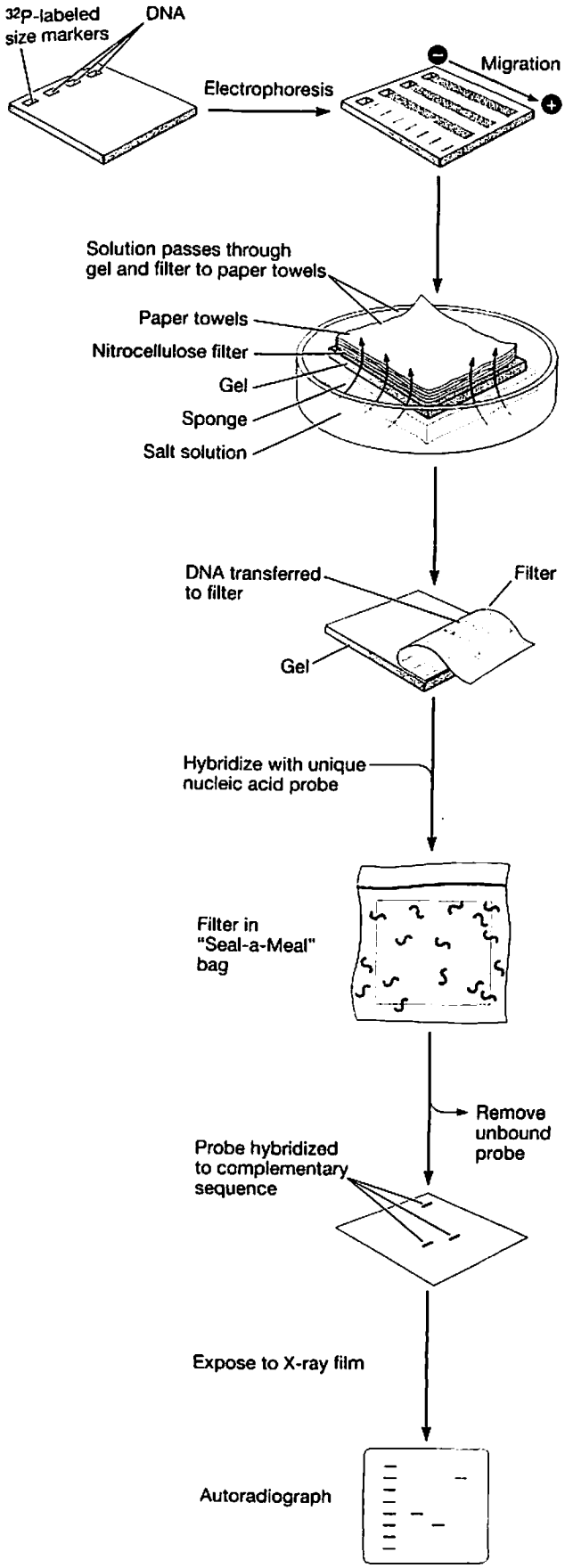


which specific fragments can be detected among the millions of other fragments produced by restriction enzyme digests. It has been used in conjunction with electrophoretic separation of very large DNA molecules to prepare restriction maps covering distances of hundreds of kilobases. In this way genes can be ordered along chromosomes, revealing the clustering of genes into functionally related groups. Southern blotting has also been useful for looking at the distribution of a gene across species. "Zoo blots" are made by hybridizing a gene probe from one species against restriction digests of genomic DNAs from a variety of species. The degree of evolutionary conservation of the gene is indicated by the range of DNAs to which the probe hybridizes.

The analogous technique for analyzing RNA is called *Northern blotting*. Instead of separating DNA on a gel, an RNA sample is electrophoresed, blotted, and hybridized with a labeled probe of interest. Northern blotting is particularly valuable for determining which tissues or cell types express a particular gene, when genes are turned on, and what factors regulate their expression. The availability of extensive sequence information now allows the reverse experiment, called a *microarray*. To construct a microarray, several hundred to hundreds of thousands of known DNA sequences (equivalent to the probes used in

FIGURE 4-10

Screening a library with a nucleic acid probe to find a clone. Libraries are typically screened by spreading several hundred thousand phages on 10–20 large agar plates covered with the host bacterial culture (lawn). After the phage plaques have grown to visible size, nitrocellulose filters are carefully laid onto the surface of the plates. Phage particles from the plate adhere to the filter, creating a precise replica on the filter of the pattern of plaques on the plate. The filters are treated to strip off phage proteins and bind the phage DNA to the filter surface. The filter is incubated in a solution containing a radioactively labeled DNA or RNA probe complementary in sequence to a portion of the gene being sought. This hybridization reaction is often carried out in sealed plastic bags. The filters are carefully washed to remove unbound probe, which leaves behind only the probe molecules tightly bound to complementary sequences within phage DNA. The location of the bound probe is determined by exposing the filters to X-ray film (autoradiography). The position is represented by a spot of exposure on the film. By orienting the film with the original agar plate, the phage plaque carrying the complementary sequence can be identified and the desired clone can be isolated.



Southern blotting) are immobilized in tiny grids and hybridized with labeled DNA or RNA. We shall describe this extremely powerful, recently developed method in more detail in Chapter 13.

Although the basics of Southern and Northern blotting have remained unchanged, the method for detecting DNA and RNA fragments in the gels has changed dramatically. Instead of using photographic film to record the positions of radioactively labeled probes, researchers use phosphorimagers, which have screens containing crystals that respond to radioactive emissions. When these are illuminated by a laser, the crystals emit light, which is recorded by a photomultiplier. This is much more rapid than using X-ray film—exposures are measured in hours rather than days—and much more sensitive. Furthermore, quantitation is more accurate because the response of the imager is linear over a wide range of radioactivity. Alternatively, probes can be labeled using nonradioactive labels. Indirect methods incorporate nucleotides carrying a small molecule such as biotin into a probe and detect this biotin using an antibody carrying alkaline phosphatase. This in turn acts on a

FIGURE 4-11

Southern blotting: analyzing DNA by gel electrophoresis, blotting, and hybridization. DNA cleaved with restriction enzymes is applied to an agarose gel and electrophoretically separated by size. The DNA in the gel is transferred to a nitrocellulose filter to make a precise replica of distribution of DNA as it was in the gel. This is usually done by placing the gel atop a sponge sitting in a tray of buffer. The filter is laid over the gel and covered with a stack of paper towels that acts as a wick, pulling buffer up through the sponge, gel, and filter. DNA fragments from the gel are carried up onto the filter, where they stick tightly. The filter is removed and hybridized with a radioactively labeled probe. Hybridization specifically tags the DNA fragment of interest, even though it may constitute only a minute fraction of the nucleic acids on the filter—this is the basis of the exquisite selectivity of the method. Unbound probe is washed off, and the filter is exposed to X-ray film. The position of a DNA fragment complementary to the probe appears as a band on the film. This procedure is termed Southern blotting when DNA is transferred to nitrocellulose. A very similar procedure called Northern blotting can be performed using RNA. The process is called Western blotting when protein is transferred, usually from an SDS-polyacrylamide gel rather than an agarose gel. In Western blotting, the protein of interest is visualized using an antibody that specifically recognizes it within the background of other cellular proteins.

substrate that emits light when hydrolyzed by the enzyme. Direct methods label a probe with dyes that fluoresce when stimulated by a laser.

Hybridization analogous to Southern and Northern blotting can also be performed directly on cells, a procedure called fluorescence in situ hybridization (FISH; Chapter 14). In FISH, fluorescently labeled probes are hybridized to "fixed" (chemically preserved) cells to determine the localization of a DNA or RNA sequence within the cell.

Only one other point of the compass has come into common usage, although not dealing with nucleic acids. In *Western blotting*, proteins in complex mixtures are separated by electrophoresis, usually on polyacrylamide gels, and then transferred to nitrocellulose or polyvinylidene fluoride (PVDF) filters. Specific proteins are detected by using antibodies.

Powerful Methods Are Used to Sequence DNA

Although by the mid-1970s, genes could be isolated and produced in large quantities for analysis, there was no way to read the nucleotide sequence of cloned DNA. In fact, the first nucleic acid sequences had been determined in the 1960s, but these were of the relatively small tRNA molecules, only 75–80 nucleotides in length. In the late 1960s, Frederick Sanger turned his attention from sequencing proteins to developing fast, simple procedures for sequencing larger RNA molecules. These provided a roundabout way to determine the sequence of DNA by first using RNA polymerase to synthesize a complementary RNA chain, and then sequencing this RNA. A breakthrough came a few years later with the advent of methods that allowed direct sequencing of DNA fragments between 100 and 500 nucleotides in length. The first of these, the *plus–minus* method, was developed by Sanger in 1975, and, in conjunction with using polyacrylamide gel electrophoresis (another Sanger innovation), a large part of the 5386-bp sequence of the small DNA phage ϕX174 was quickly determined (Chapter 10). Two years later, Allan Maxam and Walter Gilbert at Harvard, and Sanger and his colleagues at Cambridge, England, developed new methods that were significant advances on the plus–minus method.

Maxam and Gilbert's method was based on the chemical degradation of DNA chains. In this method, DNA fragments are labeled at one end and then divided into four separate samples. Each sample undergoes a different treatment, producing DNA fragments ending in G, C, a combination of A and G, and a combination of C and T. The DNA fragments in the completed reactions are separated by electrophoresis in polyacrylamide gels and detected by autoradiography. Merging the data from all the samples gives the complete sequence. The sequence of all the 5243 base pairs of SV40 DNA was completed quickly using this method, and that of the small recombinant plasmid pBR322 (4362 bp) was determined in less than a year by Greg Sutcliffe in Gilbert's laboratory.

It is Sanger's second method for sequencing DNA that has become the standard technique (Fig. 4-12). Specific terminators of DNA chain elongation—2′,3′-dideoxynucleoside triphosphates (ddNTPs)—were synthesized. These ddNTPs are incorporated normally into a growing DNA chain through their 5′-triphosphate groups, but they cannot join with the next incoming deoxynucleotide triphosphate (dNTP) because they lack the 3′-OH group needed to make the phosphodiester bond. When a small amount of a specific ddNTP (say, ddATP) is included in the reaction mixture along with the other three dNTPs required for DNA synthesis by DNA polymerase, the products are a series of chains that are specifically terminated wherever a ddATP is incorporated. Four reactions are set up, each with a different ddNTP; the DNA chains of each reaction are separated by electrophoresis; and the sequence is read off.

The next major advance in DNA sequencing came when Michael Hunkapiller and Leroy Hood developed a method for labeling each ddNTP with a differently colored dye, so that all four reactions could be run in a single lane on a gel. Later, semiautomated sequencing machines became available and determining the sequence of a segment of DNA became a standard procedure in every molecular biology laboratory. Continuing developments, including robots for preparing reaction samples and fully automatic capillary-based sequencing machines, have led to sequencing the entire genomes of higher organisms. At the date of this printing, the genomes of more than 1000 viruses and 300 organisms, including human, mouse, and rat, have been determined. Chapters 10, 11, and 12 provide many more details of the strategies and technologies that have enabled us to achieve feats regarded as impossible only a few years ago.

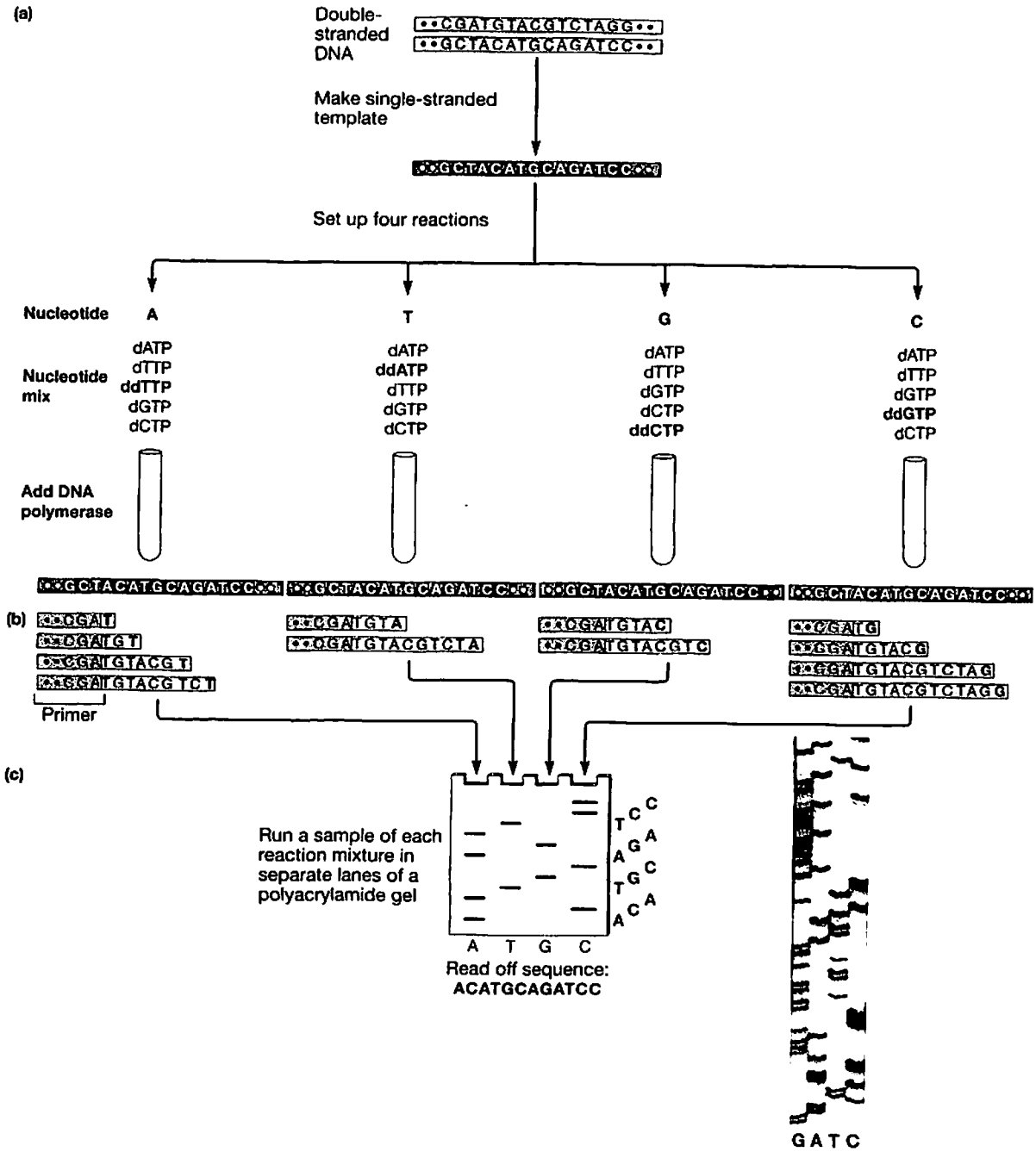


FIGURE 4-12

The Sanger dideoxy DNA sequencing procedure. 2′,3′-dideoxynucleotides of each of the four bases are prepared. These molecules can be incorporated into a growing DNA strand by DNA polymerase because they have a normal 5′-triphosphate; however, once incorporated into a growing DNA strand, the dideoxynucleotide (ddNTP) cannot form a phosphodiester bond with the next incoming dNTP. Growth of that particular DNA chain stops. (*a*) A Sanger sequencing reaction consists of a DNA strand to be sequenced, a short piece of DNA (the primer) that is complementary to the end of that strand, a carefully controlled ratio of one particular ddNTP with its normal dNTP, and the other three dNTPs. A small amount of one or more radioactive dNTPs is also included so that DNA molecules can be visualized later by autoradiography. (*b*) When DNA polymerase is added, normal polymerization will begin from the primer; when a ddNTP is incorporated by chance, the growth of that chain will stop. If the correct ratio of ddNTP:dNTP is chosen, a series of labeled strands will result, the lengths of which are dependent on the location of a particular base relative to the end of the DNA. (*c*) The resulting labeled fragments are separated by size on an acrylamide gel, and autoradiography is performed; the pattern of the fragments gives the DNA sequence. Generally, in modern sequencing methods the products are visualized by detecting fluorescence from dyes conjugated to the chain terminating ddNTPs. By using a different dye for each ddNTP, a single reaction can be done and analyzed on a single lane of a gel.

Oligonucleotides Are Synthesized Chemically

The ready availability of oligonucleotides of defined sequence is critical for many recombinant DNA techniques. Methods for linking nucleotides were worked out during the 1970s and this early phase culminated with the synthesis of a complete alanine tRNA by H. Gobind Khorana's laboratory in 1972.

The emergence of quick, convenient methods for the synthesis of moderately long oligonucleotides with defined sequences depended on three technical advances. The first followed the example of peptide synthesis and attached the growing nucleotide chain to a solid support. The polymerases in cells synthesize DNA and RNA in a 5′ to 3′ direction, whereas it is most frequently used for in vitro oligonucleotide synthesis done in reverse, from 3′ to 5′. Thus, the 3′-hydroxyl of the first base is tethered to the solid support. This leaves just one end of the molecule available for reactions, and adding and washing off reagents is much easier with a tethered molecule.

The second advance was the development of improved chemistries for protecting both the 5′ and the 3′ end of a mono- or oligonucleotide from chemical reactions by attaching *blocking groups* onto the nucleotides used in DNA synthesis. These blocking groups are additional atoms linked to the reactive hydroxyl and amine groups on the nucleotides, which prevent unwanted chemical reactions between the next added base and the growing oligonucleotide chain. After synthesis is complete, the blocking groups are removed, leaving a DNA molecule chemically identical to those synthesized in cells.

The third advance came with the development of programmable machines that synthesize oligonucleotides using phosphoramidite chemistry and solid supports. Until then, synthesis of oligonucleotides was a time-consuming process that was limited to linking together fewer than 20 nucleotides. Now, completely automated machines synthesize oligonucleotides as long as 100 bases in a few hours. The limiting factor is the progressively lower yield of full-length oligonucleotides with increasing length.

Genes Can Be Synthesized from Oligonucleotides

Oligonucleotides are essential for many techniques and they have also been used to synthesize genes. Because the yield of each reaction step of oligonucleotide synthesis is less than 100%, there are limits on the maximum length of synthetic oligonucleotides. Complete genes have to be made by stitching together multiple oligonucleotides and the first gene thus synthesized, in 1970, was for a tRNA. Synthesis of genes encoding small peptide hormones such as the 42-bp human hormone somatostatin and the two chains of human insulin, 63 and 90 bp, soon followed. At that time, chemical synthesis provided a way for researchers to bypass the prohibitions against the insertion of human DNA into a bacterium. Although those prohibitions were soon lifted, chemical synthesis continued to be a practical, although somewhat expensive, way to create genes and, by the 1980s, genes such as tissue plasminogen activator, 1610 bp in length, were being made. Further developments led to the synthesis of the complete genomes of poliovirus (7440 bp) and of the bacteriophage ϕX174 (5386 bp).

The Polymerase Chain Reaction Amplifies Specific Regions of a DNA Target

The *polymerase chain reaction* technique (PCR) was devised by Kary Mullis in the mid-1980s and, like DNA sequencing and synthesis, has revolutionized the practice of molecular genetics; those who learn the craft today find it inconceivable that any research was done prior to 1988 when PCR became a reliable and convenient tool. There are so many variations of the PCR technique, and so many applications, that we can provide only a sampling of them in this chapter. Here, we will go over the basic features of PCR, whereas many of the experiments described in later chapters will provide examples of the applications of PCR.

A major problem in analyzing genes is that they are rare targets in a complex genome—the human genome, for example, has about 23,000 genes. Various procedures were devised to isolate and detect these rare targets, but such techniques were generally complex and time consuming. PCR changed this by enabling us to produce enormous numbers of a specified DNA sequence without resorting to cloning. PCR uses DNA polymerase to synthesize multiple copies of a DNA sequence ("amplify") determined by the primers needed to initiate DNA synthesis.

The starting material for a PCR is a DNA sample containing the sequence to be amplified. DNA preparation for a PCR is relatively easy except for the

most demanding of samples. The target DNA sequences do not have to be isolated from other DNA because they are defined by the oligonucleotide primers used in the reaction. In fact, DNA released by disrupting cells by boiling or by detergents can be used directly without any purification. Because PCR amplifies DNA, the amount of DNA needed to initiate a PCR is very small—even a single DNA molecule will suffice. Remarkable sources of DNA for PCR include biopsies embedded in paraffin for more than 40 years; blood samples, taken by heel prick of newborns, for the neonatal detection of phenylketonuria and stored as dried spots on cards; and even the tooth of a 10- to 11-year-old Neanderthal child dating from 100,000 years ago.

DNA polymerase uses single-stranded DNA as a template for the synthesis of a complementary new strand. These single-stranded templates are produced simply by heating double-stranded DNA to near boiling. DNA polymerase requires a small section of double-stranded DNA to initiate ("prime") synthesis. Therefore, the starting point for DNA synthesis in vitro can be specified by supplying an oligonucleotide primer that anneals to the template at that point (Fig. 4-13).

Both DNA strands of a double helix can serve as templates for synthesis, provided an oligonucleotide primer is supplied for each strand. The primers are chosen to flank the region of DNA that is to be amplified, so that the newly synthesized strands of DNA, starting at each primer, extend beyond the position of the primer on the opposite strand. Therefore new primer binding sites are generated on each newly synthesized DNA strand. The reaction mixture is heated to separate the strands, the primers anneal to the newly available binding sites, and new chains are synthesized. The cycle of heating, primer binding, and extension is repeated (Figs. 4-14 and 4-15), so that at the end of n cycles, the reaction mixture contains a theoretical maximum of 2^n double-stranded DNA molecules that are copies of the sequence between the primers (Table 4-3). In practice, the typical yield is 10–30% of the theoretical maximum. Important factors include the length and

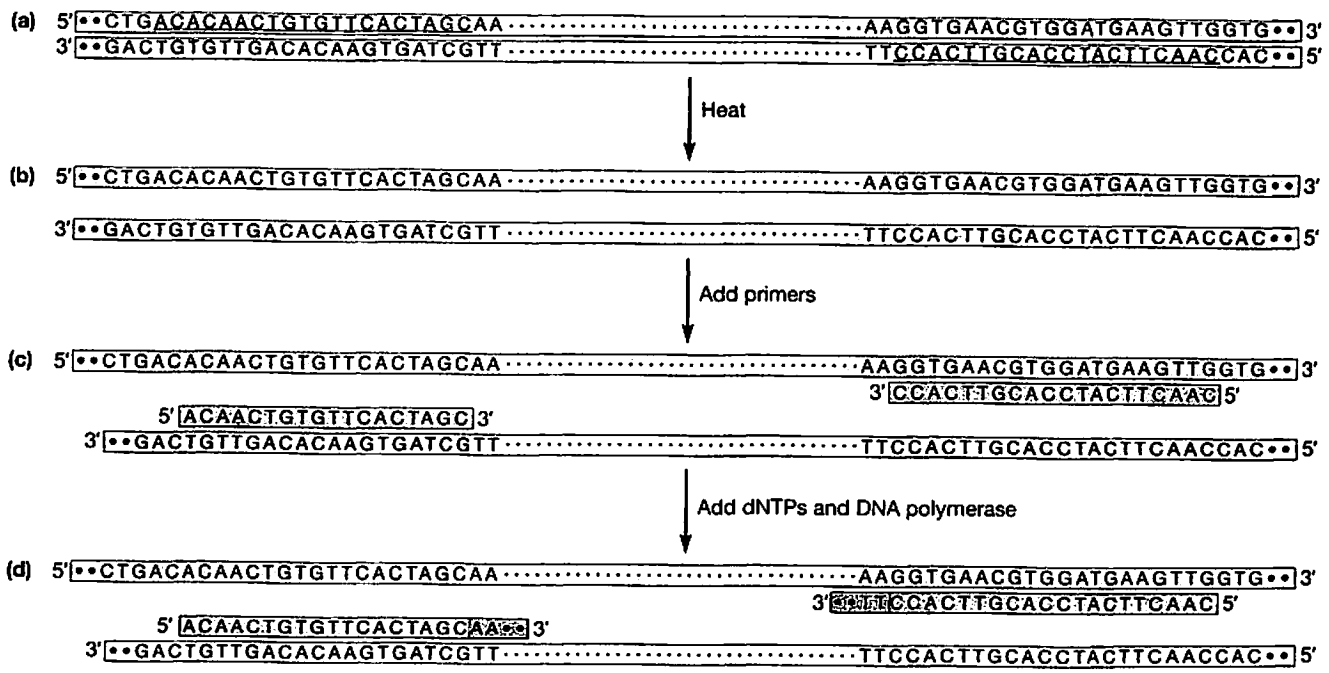


FIGURE 4-13

Primers for DNA polymerase. (a) The target for amplification, a small section covering 110 bp of the β-globin gene, is shown. Two sequences separated by 60 nucleotides are detailed. Oligonucleotide primers for the PCR are made to hybridize to the two sets of 20 nucleotides (*underlined*). (*b*) When the DNA is heated, the strands separate. (c) The oligonucleotide primers (shown in *green*) hybridize specifically to their complementary sequences at the 3′ ends of each strand of the target sequence. (*d*) DNA polymerase uses these primers to begin synthesis of new strands (shown in *pink*) complementary to the target DNA sequences in the 5′ to 3′ directions.

composition of the sequence to be amplified, because DNA polymerase can falter on long sequences or in regions of high GC content. Nevertheless, PCR can produce so many copies that the DNA can be easily visualized on an agarose gel stained with ethidium bromide rather than having to resort to Southern blotting.

Optimizing a PCR requires the adjustment of many parameters, including temperature, salt concentrations, and cycle durations, but the primers are the key factor—it is they that direct where synthesis of new DNA strands will begin. The specificity of primer binding depends on the length of the primers and their nucleotide composition. Oligonucleotides between 18 and 24 nucleotides long tend to be highly specific under standard conditions where the temperature of the primer-binding step (called *annealing*) is close to the temperature at which the primers dissociate from the template DNA (the *melting temperature*, T_m). It is also important to choose primer pairs with similar T_ms, so that both anneal at the chosen temperature. Increasing primer length increases specificity, but at the price of increasing the T_m. If the T_m becomes too high, the required annealing temperature will exceed the temperature for polymerase extension and may risk inaccurate priming.

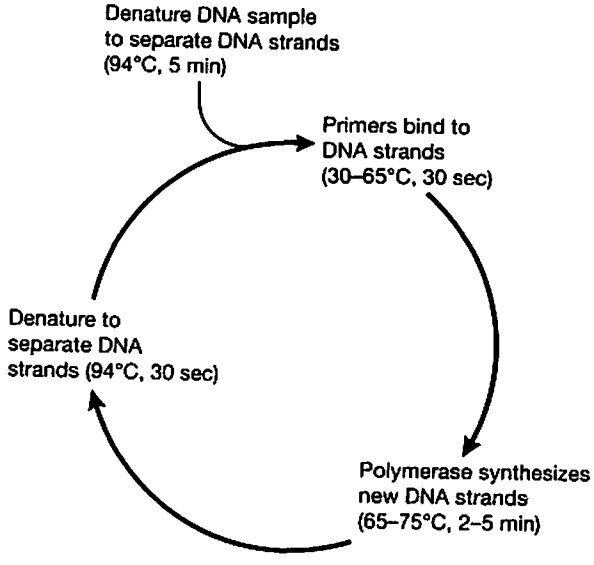


FIGURE 4-14

The PCR cycle. The DNA sample is heated to separate the DNA strands (initial denaturation), and then the reaction mixture goes through repeated cycles of primer annealing, DNA synthesis, and denaturation. The target sequence doubles in concentration for each cycle.

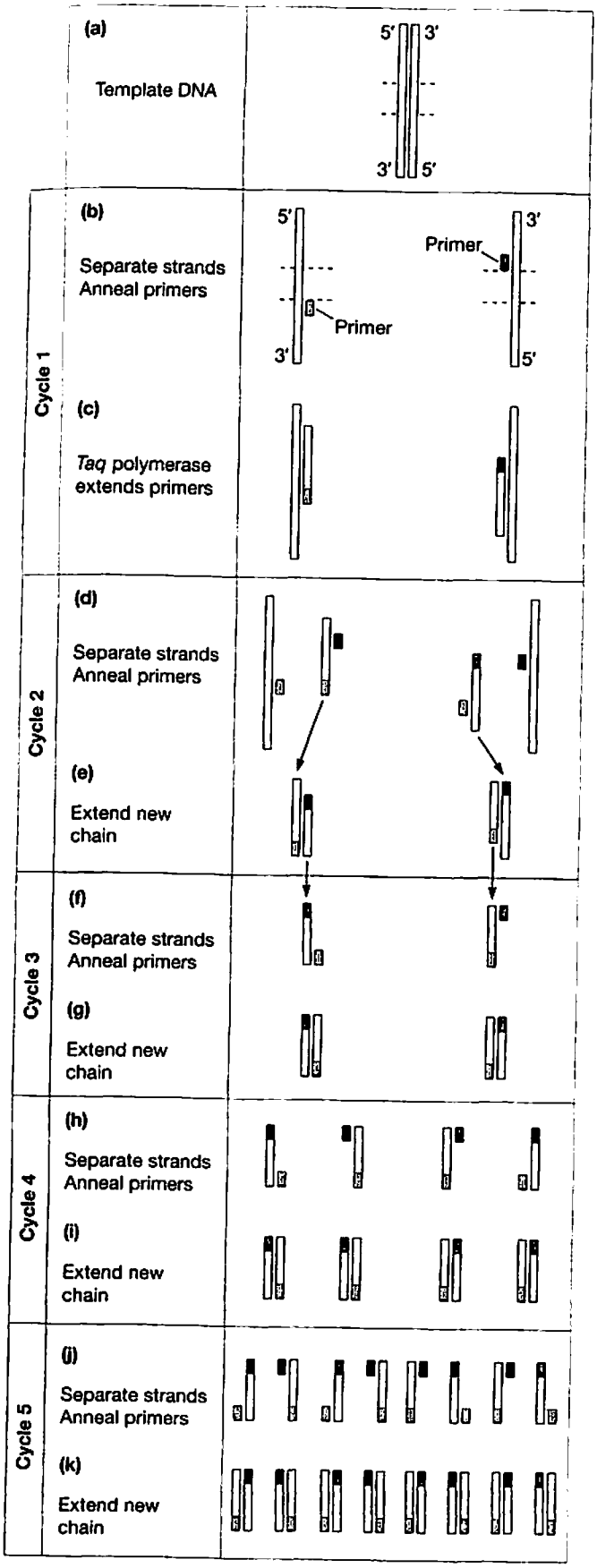
Thermostable Polymerases Simplify and Improve PCR

Once the extraordinary power of PCR had been recognized, many improvements were made in the basic technique. Undoubtedly, the most significant was the use of DNA polymerases from thermophilic bacteria that live at very high temperatures. Because *E. coli* DNA polymerase is destroyed at the temperatures needed to separate double-stranded DNA, fresh enzyme had to be added for each cycle of the reaction. However, the DNA polymerase from the bacterium *Thermus aquaticus*, which lives in water at a temperature of 75°C, has a temperature optimum of 72°C and is reasonably stable even at 94°C. Called Taq polymerase, this enzyme remains active through a complete set of amplification cycles and has enabled automation of PCR using *thermocyclers*, machines designed to carry out the time and temperature cycles for a PCR. For processing very large numbers of samples, the PCRs are set up using robots. This is essential for genome projects where millions of PCRs are carried out without any manual intervention.

Contamination Can Be a Problem in PCR Studies

An unwelcome corollary of the amplification power of the PCR is that minor contamination of the starting material can have serious consequences. This is true especially for critical applications, such as amplifying DNA isolated in forensic cases where the samples taken at a crime scene are rarely in the pristine state of samples used to prepare DNA in the laboratory, as we will see in Chapter 16.

An essential control is to carry out a PCR amplification without template DNA to test for contaminating DNA in the reaction reagents. One source of contamination is the products of previous amplification reactions. A completed PCR mixture may contain as many as 10^{13} amplified fragments, so that even minute volumes, such as droplets in an aerosol from a pipette tip, contain very large numbers of amplifiable molecules. A simple and very effective



way to reduce contamination is to physically separate pre- and postamplification steps so that amplified reactions are never in the same area that is used for setting up the reactions. DNA can also be made inactive as a template for PCR. Often, areas in the lab set aside for PCR are treated with UV light. This cross-links DNA, thus creating intrastrand bridges that stop the polymerase as well as breaking DNA strands.

More sophisticated methods can be used for especially critical studies or to deal with persistent contamination. A clever method is based on the incorporation of deoxyuracil instead of thymidine into PCR products. (Remarkably, this has little effect on the properties of the amplified DNA except that many restriction enzyme sites are lost.) When a new reaction is set up, the enzyme uracil-*N*-glycosylase is included and this degrades any previously amplified DNA contaminating the new reaction by destroying the uracils. dUTP can be used in the new reaction because uracil-*N*-glycosylase enzyme is inactivated by the high temperature in the new PCR.

FIGURE 4-15

The exponential increase of DNA copies during PCR. (*a*) The starting material is a double-stranded DNA molecule. The region to be amplified is marked by *dashed lines*. (*b*) The strands are separated by heating the reaction mixture and then cooling so that the primers anneal to the two primer-binding sites flanking the target region, one on each strand. (*c*) *Taq* polymerase synthesizes new strands of DNA, complementary to the template, that extend a variable distance beyond the position of the primer-binding site on the other template. (*d*) The reaction mixture is heated again; the original and newly synthesized DNA strands separate. Four binding sites are now available to the primers, one on each of the two original strands and the two new DNA strands. (To simplify the diagram, subsequent events involving the original strands are omitted.) (*e*) *Taq* polymerase synthesizes new complementary strands, but the extension of these chains is limited precisely to the target sequence. The two newly synthesized chains thus span exactly the region specified by the primers. (*f*) The process is repeated, and primers anneal to the newly synthesized strands (and also to the variable-length strands, but these are omitted from the figure). (*g*) *Taq* polymerase synthesizes complementary strands, producing two double-stranded DNA fragments that are identical to the target sequence. (*h–k*) The process is repeated and the number of target fragments doubles for each subsequent cycle of the reaction.

TABLE 4-3. Amplification of molecules during polymerase chain reaction

Cycle number	Number of double-stranded target molecules
1	0
2	0
3	2
4	4
5	8
6	16
7	32
8	64
9	128
10	256
11	512
12	1,024
13	2,048
14	4,096
15	8,192
16	16,384
17	32,768
18	65,536
19	131,072
20	262,144
21	524,288
22	1,048,576
23	2,097,152
24	4,194,304
25	8,388,608
26	16,777,216
27	33,554,432
28	67,108,864
29	134,217,728
30	268,435,456
31	536,870,912
32	1,073,741,824

Fidelity of DNA Synthesis Determines the Accuracy of PCR Amplification

Contamination is not the only possible source of error in PCR. Like all other biochemical processes, DNA replication is not a perfect process, and occasionally DNA polymerase will add an incorrect nucleotide to the growing DNA chain. The rate of misincorporation measured in a naturally replicated DNA molecule is approximately 1 in 10^9 nucleotides. Cells achieve such extraordinary accuracy because the DNA replication machinery removes mismatched nucleotides added to the DNA chain. In vitro, *Taq* polymerase does not have this "proofreading" capability, and using the temperatures and salt concentrations typical in a PCR, the enzyme incorporates one incorrect nucleotide for about every 2×10^4 nucleotides incorporated. This is not a serious matter for bulk analysis of PCR products because molecules with the same misincorporated nucleotide will form a minute proportion of the total number of molecules synthesized. But misincorporation is important if PCR fragments are to be used for cloning, where each clone is derived from a single amplified molecule. If this molecule contains one or more misincorporated nucleotides, then all DNA isolated from that clone will carry the identical mutation. This problem can be reduced by beginning the PCR with a large, rather than a small, number of template molecules—fewer cycles of amplification are needed and less total DNA synthesis takes place. Another way to reduce the error rate is to use different DNA polymerases. Polymerases from many thermophilic bacterial species have been isolated for use in PCR, and some of these, such as *Pfu* DNA polymerase, do have proofreading 3′ to 5′ exonuclease activity. The exonuclease activity removes the mismatched nucleotides, which allows the polymerase another opportunity to synthesize the correct sequence.

PCR Will Amplify Sequences from a Single DNA Molecule

One of the most remarkable applications of PCR is in performing linkage analysis using sperm. Human genetic linkage analysis is difficult because it has to be carried out using families where the number of available offspring is often much lower than the geneticist would like for statistical analysis. However, examining 1000 sperm is like studying a family with 1000 children, because each sperm is the result of a meiotic division. Linkage analysis can be done by choosing a male heterozygous for the loci to be tested and determining what proportion of his sperm cells show recombination between the loci. But a sperm cell is haploid, containing only one copy of each chromosome, and each chromosome is a single molecule of DNA. Linkage analysis using sperm requires PCR to work effectively and reliably in a reaction mixture containing a single target DNA molecule!

In the example given here, alleles at three loci were analyzed—two on chromosome 19 (DS19S49 and *APOC2*) and one on chromosome 9 (D9S52) (Fig. 4-16). These loci contain microsatellite repeats, stretches of DNA where a sequence is repeated over and over again. D19S49 has repeats

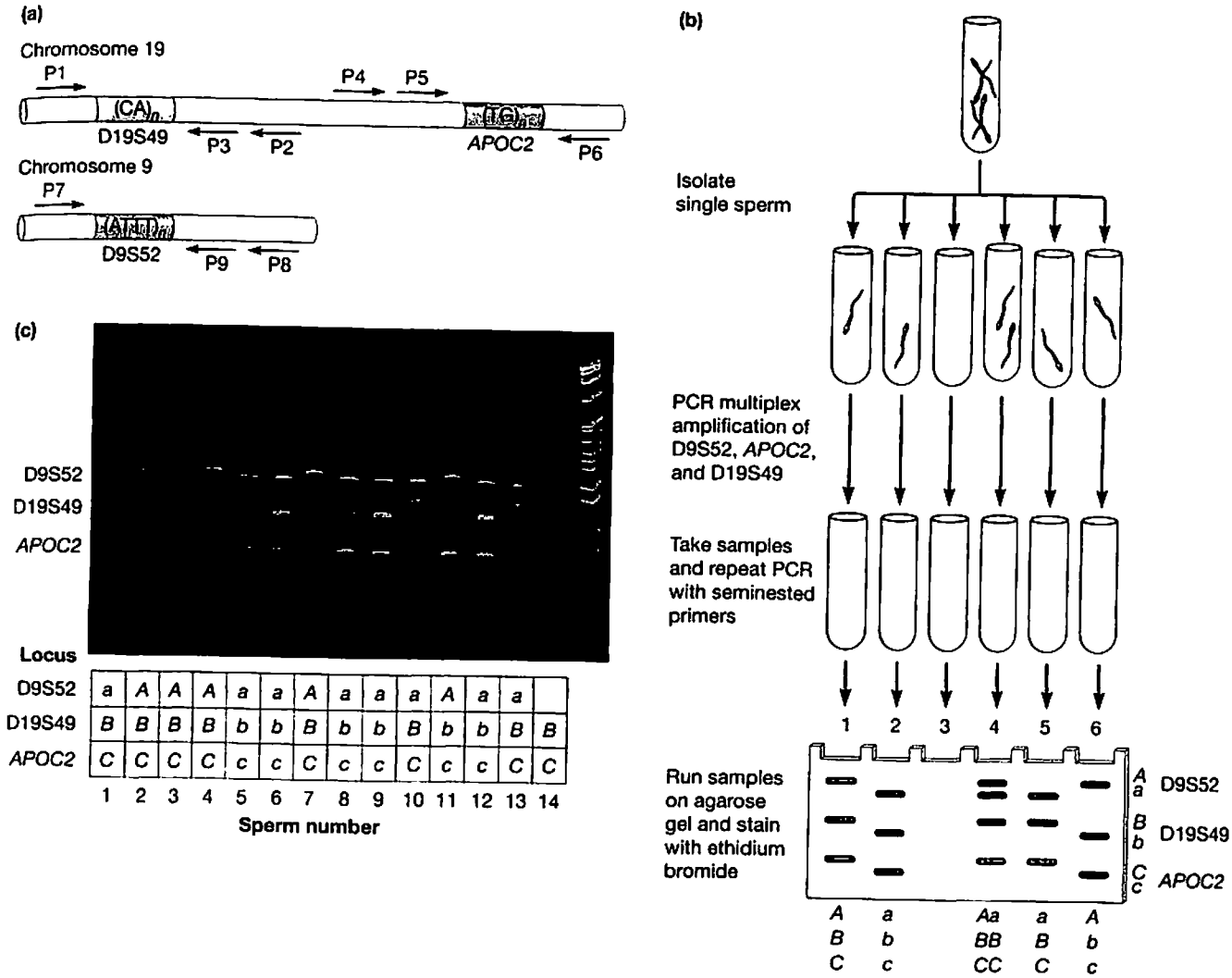


FIGURE 4-16

Amplifying single DNA molecules in linkage analysis using single sperm. (*a*) Three polymorphic regions were used: D19S49 and *APOC2* on chromosome 19 and D9S52 on chromosome 9. The polymorphisms arise from differing numbers of repeat units at each locus: (CA) at D19S49; (TG) at *APOC2*; and (ATTT) at D9S52. Sets of three primers specific for each locus were synthesized. (*b*) A sperm sample was obtained from a man who was heterozygous for each locus—$(CA)_{15}$ and $(CA)_{21}$; $(TG)_{18}$ and $(TG)_{22}$; and $(ATTT)_{12}$ and $(ATTT)_{13}$. Single sperm were transferred to tubes, and amplification of the three loci was carried out simultaneously using the outer primers for each locus (P1 and P2; P4 and P6; P7 and P8). Samples were taken and the amplification repeated using one outer and one inner primer (P1 and P3; P5 and P6; P7 and P9). Samples were run on an acrylamide gel and stained with ethidium bromide. There should be three bands in each lane, corresponding to DNA amplified at each locus, and the haplotypes of each sperm for these loci can be read off from the gel. Thus, the sperm in lane 1 is *ABC*, whereas that in lane 2 is *abc*. However, lane 3 has no bands, presumably because the tube contained no sperm, whereas there are two bands corresponding to both alleles of the D9S52 locus in lane 4. This sample presumably had two sperm in it—*ABC* and *aBC*. (*c*) Data from an experiment. The gel and its analysis are shown. One amplification product—for D9S52—is missing in lane 14. It is not unusual for such "dropout" to occur in multiplex reactions. The data show linkage between D19S49 and *APOC2*. For this man, chromosome 19 has allele *B* of D19S49 associated with allele *C* of *APOC2*, and allele *b* with allele c. There is no linkage between alleles of D9S52 and D19S49-*APOC2* because they are on different chromosomes.

of CAs, *APOC2* has repeats of TG, and D9S52 has repeats of ATTT. The alleles at these sites result from differing numbers of repeats, the minimum difference being one repeat, equivalent to two nucleotides at *APOC2* and D19S49 and four nucleotides at D9S52. Sperm was obtained from men who were heterozygous at all three loci, and single sperm were isolated using a fluorescence-activated cell sorter, more commonly used for preparing different types of lymphocytes. The three loci were simultaneously amplified in what is termed a ***multiplex reaction***. Primers for multiplex PCR must be carefully designed so that each PCR product has a distinct size and so that the different primer sets anneal to the template and not to one another. Samples were electrophoresed using polyacrylamide gels and the alleles were determined. Inspection shows, not surprisingly, that although *APOC2* and D19S49 are linked, D9S52 is not linked because it is on a separate chromosome. Calculation of the recombination frequency between the two linked loci requires statistical treatment of the data because technical complications have to be taken into account. For example, sorting is not perfect, so that some tubes may have more than one sperm, whereas other tubes may have no sperm, and sometimes loci fail to amplify. The frequency of recombination as calculated in this experiment was 0.083, which is in good agreement with the value determined from similar analysis performed on 40 families. These estimates of genetic distances are for *male* chromosomes, which is an important point, given that recombination frequencies for genes differ for male and female meioses.

Real-Time PCR Is Used for the Rapid and Accurate Quantitation of RNA and DNA

It is often necessary to determine the amount of a particular DNA or RNA molecule in a sample. This can be done by measuring the intensities of bands on Southern or Northern blots, but this is not a very sensitive method; PCR provides much more versatile and accurate methods.

In general, the greater the number of target molecules in the starting sample, the more amplified molecules will be present at each step of the reaction. By measuring the final PCR products on a gel, it is possible to get some idea of how much of the target molecule was present at the beginning of the reaction. However, this is not so easy to do reliably. If the quantities of primers and nucleotides are too low, or there are too many cycles of PCR, the primers and nucleotides in the reaction will be exhausted and the reaction will reach a plateau after which no more product can be made. At this stage, the amount of DNA made no longer reflects how much of the target molecule was present initially. A much better strategy is to use quantitative real-time PCR methods that measure the amplification of the target molecule continuously during a PCR. After an initial phase where amplification is undetectable, the PCR enters an exponential phase in which the product is almost doubling at each step. If more target molecules are present at the start, fewer cycles will be required to reach the exponential phase. Comparison of the number of cycles required to reach this point allows the determination of the initial template concentration of the reaction (Fig. 4-17).

FIGURE 4-17

Real-time PCR is used to measure the quantity of target sequences in a sample. (*a,b*) Plots of the theoretical accumulation of DNA molecules during a PCR. In *a*, the scale is linear, and the curve is close to the baseline of cycle 25 because the number of molecules is low compared to later cycles. In *b*, the accumulation of molecules is plotted on a logarithmic scale, producing a linear plot because of the exponential increase. In both cases the curves flatten by cycle 32. (*c,d*) Plots of experimental data. The PCR products were detected by the dye SYBR Green, a nonspecific dye that fluoresces when intercalated into double-stranded DNA. The amount of PCR product is so low in the early cycles that it cannot be detected above background fluorescence. Note that the best fit to the exponential accumulation shown in the theoretical logarithmic graph (*b*) is during cycles 18–24 and it is in this period, when the reaction is behaving truly exponentially, that comparison of template concentrations should be made. Note that the linear portion of the plot in *c*—cycles 26–30—is *not* the correct time to do quantitative determinations! (*e*) Tenfold dilutions of the starting DNA concentrations have been compared using real-time PCR and plotted on a linear scale. The greater the amount of template, the fewer the number of cycles are needed to first be able to detect the product. Note that the reactions reach different steady-state levels that do not reflect the initial differences in template concentration; the reaction corresponding to the *light blue* curve has 1000-fold more template than the *red* reaction, yet produces less final product. This reflects the effects of small, random variations during the initial stages of the reaction, which become amplified during the later stages, and emphasizes the importance of making comparisons at the correct times.

Several different approaches are available for determining the amount of PCR product present at the end of each cycle, but all rely on measuring the amount of a fluorescent tag that is associated with each newly synthesized molecule. Early tags included ethidium bromide and SYBR Green I, both of which insert in the grooves of double-stranded DNA. A more recent approach uses oligonucleotide probes that hybridize to an internal sequence of the target molecule. One type of probe is called a ***molecular beacon***, which is an oligonucleotide carrying a fluorophore at one end and a nonfluorescent quencher at the other. The molecular beacon is designed so that the fluorophore and quencher come together, thus forming a hairpin. When illuminated in this configuration with UV

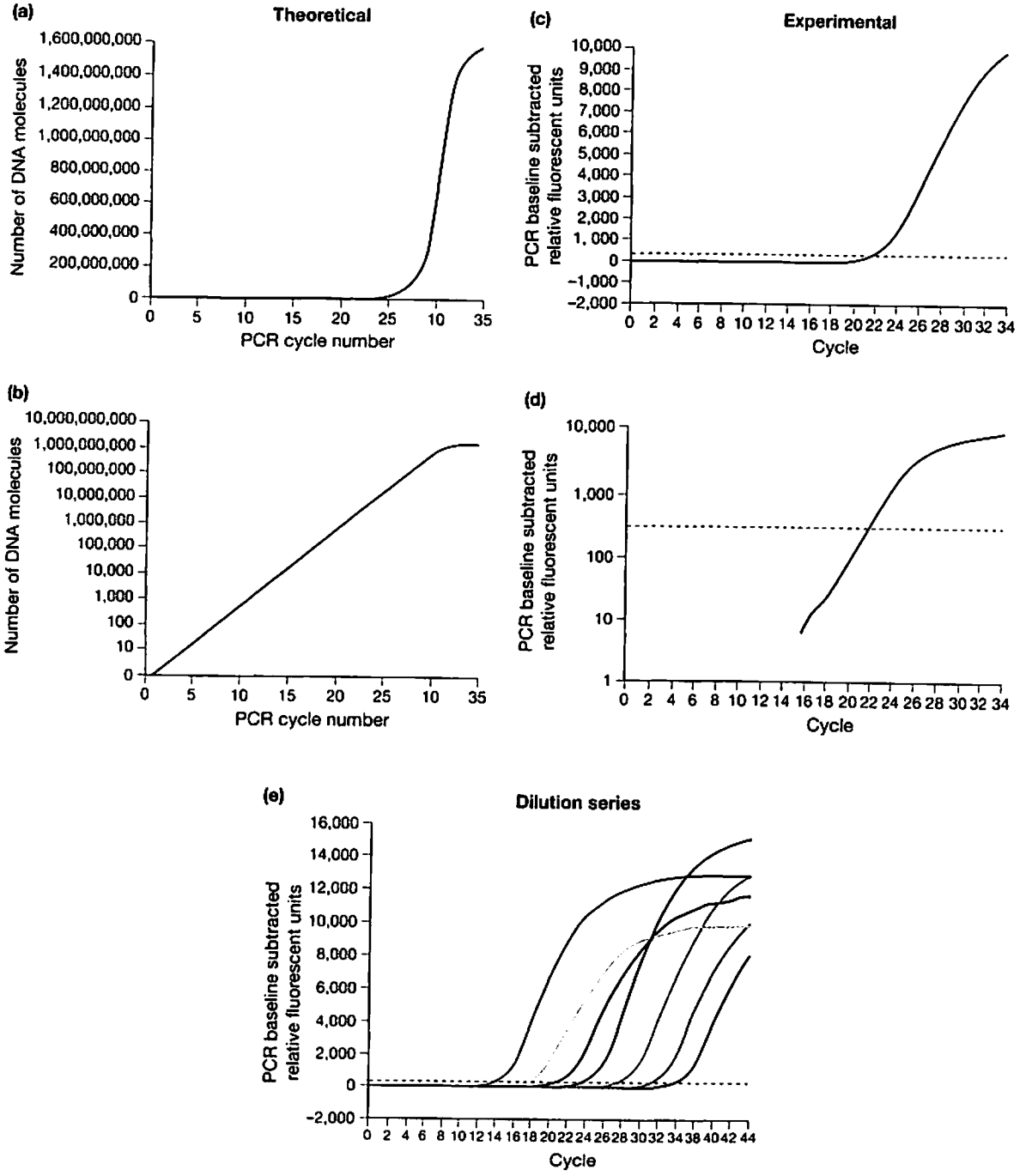


FIGURE 4-17 *(See facing page for legend.)*

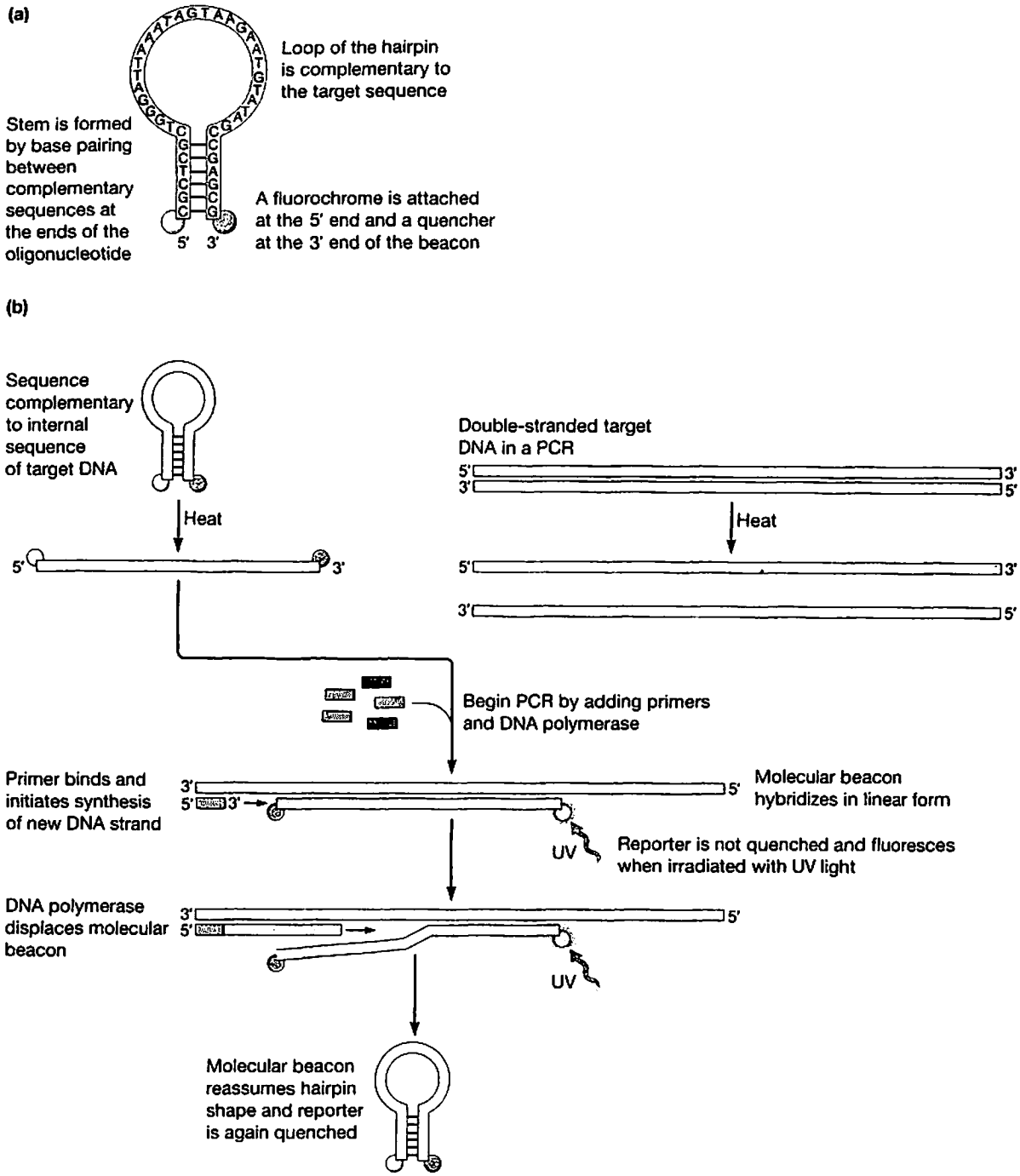


FIGURE 4-18

Molecular beacons for real-time PCR. (*a*) A molecular beacon is a short (30-bp to 40-bp) oligonucleotide synthesized to be complementary to an internal site of the segment to be amplified. The beacon has a fluorescent reporter dye attached to its 5′ nucleotide and a quenching dye attached to its 3′ nucleotide. The first few, usually six, 5′ nucleotides and the last 3′ nucleotides are complementary so that the beacon folds onto itself, forming a hairpin structure that brings the dye and quencher into close proximity. This association permits the transfer of energy from the reporter dye to the quenching dye, so that the quenching dye absorbs the energy emitted by the reporter dye when it is irradiated, preventing any emission of the energy as a fluorescent signal. (*b*) When the PCR is heated to dissociate the target DNA strands, the beacon hairpins also melt, producing a linear molecule. During the annealing step of PCR, when the reaction is cooled to allow hybridization of the primers, the molecular beacon hybridizes to the target strand. The reporter and quencher dyes are now separated, so that the reporter fluoresces when irradiated. Because the reporter fluoresces only when bound to the target sequence, the fluorescence signal is proportional to the quantity of target molecules in the PCR. The beacon does not interfere with the amplification process because it is displaced from the target as the DNA polymerase moves along its template strand. It is then available for the next cycle.

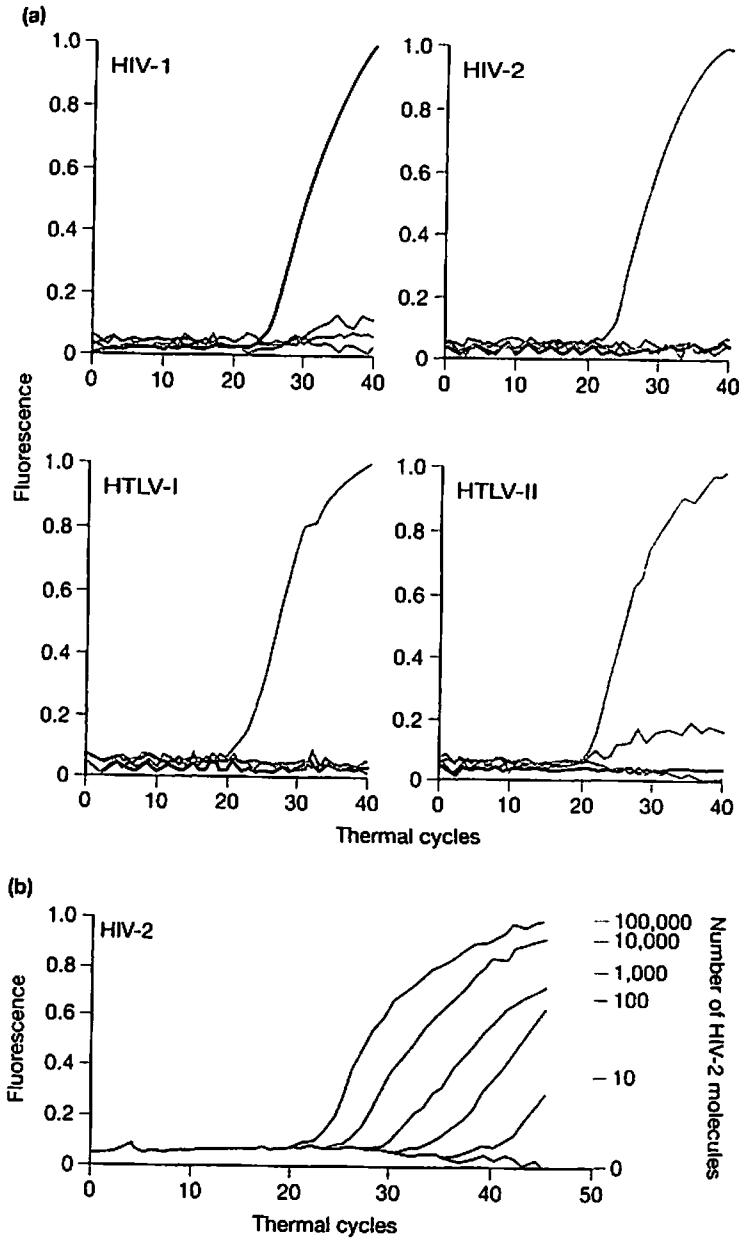


FIGURE 4-19

Detecting different human retroviruses. Primers were designed to amplify the *gag* gene of HIV-1, the *env* gene of HIV-2, the *tax* gene of HTLV-I, and the *pol* gene of HTLV-II. Oligonucleotides specific for conserved internal sequences of each amplified viral sequence were synthesized and each labeled with a different fluorochrome and the same nonfluorescent quencher. (*a*) The four primer–molecular beacon combinations were used in the same reaction with one virus. The specificity of the combinations is clear; the HIV-1 primers (in *purple*) amplify only HIV-1. (*b*) The technique can detect very small numbers of one virus in the presence of large numbers of a second. Here the PCR contained a constant number of HTLV-I molecules (100,000) and tenfold dilutions of HIV-2 molecules, ranging from 100,000 to 10. The number of PCR cycles needed to develop a measurable signal is inversely proportional to the starting number of HIV-2 molecules, and as few as ten HIV-2 molecules can be detected in the presence of 100,000 HTLV-I molecules.

light, energy from the fluorophore is absorbed by the quencher in a process called fluorescence resonance energy transfer (FRET), and no light is emitted. However, when the probe is heated and hybridizes to the amplified DNA sequence, fluorophore and quencher are separated, and now the former emits light when irradiated (Fig. 4-18). The increasing fluorescence measured as the reaction proceeds is a measure of the increase in the number of amplified molecules.

Multiplex PCR has been combined with molecular beacon technology for a very sensitive means of detecting and quantitating human retroviruses—HIV-1, HIV-2, HTLV-I, and HTLV-II. By using different fluorochromes, multiple beacons can be used to report on PCR products amplified from the different viruses in a single reaction. As few as ten molecules of one virus can be detected in the presence of 100,000 molecules of a second virus (Fig. 4-19).

Recombinant DNA Comes of Age

Recombinant DNA technology has fulfilled its promise many times over. What were once esoteric and difficult techniques, limited to a small number

of laboratories, now can be learned readily from one of many manuals and there is a service industry to supply the enzymes, oligonucleotides, and other reagents that once had to be prepared painstakingly by individual researchers. The tools of recombinant DNA are used routinely around the world and biologists have used them to learn many of the essential features of genes and how they function in organisms. In the next chapter we review how these tools have been put to use for cloning genes.

Reading List

The best resource for much of the material covered in this chapter is Sambrook and Russell (2001). Although it is a laboratory manual, the introductions to each chapter contain extensive descriptions of the underlying principles of cloning.

Sambrook J. and Russell D.W. 2001. *Molecular Cloning: A Laboratory Manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Restriction Enzymes and Maps

Linn S. and Arber W. 1968. Host specificity of DNA produced by *Escherichia coli*, X. In vitro restriction of phage fd replicative form. *Proc. Natl. Acad. Sci.* **59:** 1300–1306.

Meselson M. and Yuan R. 1968. DNA restriction enzyme from *E. coli*. *Nature* **217:** 1110–1114.

Kelly T.J., Jr., and H.O. Smith. 1970. A restriction enzyme from *Hemophilus influenzae*, II. Base sequence of the recognition site. *J. Mol. Biol.* **51:** 393–409.

Smith H.O. and Wilox K.W. 1970. A restriction enzyme from *Hemophilus influenzae*, I. Purification and general properties. *J. Mol. Biol.* **51:** 379–391.

Danna K. and Nathans D. 1971. Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proc. Natl. Acad. Sci.* **68:** 2913–2917.

Sharp P.A., Sugden B., and Sambrook J. 1973. Detection of two restriction endonuclease activities in *Hemophilus parainfluenza* using analytical agarose-ethidium bromide electrophoresis. *Biochemistry* **12:** 3055–3062.

Roberts R.J., Vincze T., Posfai J., and Macelis D. 2003. REBASE: Restriction enzymes and methyltransferases. *Nucleic Acids Res.* **31:** 418–420.

The Recombinant DNA Debate

Singer M. and Soll D. 1973. Guidelines for DNA hybrid molecules. *Science* **181:** 1174.

Berg P., Baltimore D., Boyer H.W., Cohen S.N., Davis R.W., et al. 1974. Potential biohazards of recombinant DNA molecules. *Science* **185:** 303.

Berg P., Baltimore D., Brenner S., Roblin R.O., and Singer M.F. 1975. Asilomar conference on recombinant DNA molecules. *Science* **188:** 991–994.

Norman C. 1976. Genetic manipulation: Guidelines issued. *Nature* **262:** 2–4.

Rogers M. 1976. *Biohazard*. Knopf, New York.

Department of Health, Education, and Welfare. 1980. Guidelines for research involving recombinant DNA molecules. *Federal Register*, Tuesday, January 29.

Watson J.D. and Tooze J. 1981. *The DNA Story*. W.H. Freeman, San Francisco. (A sourcebook for documents on the recombinant DNA debate.)

Krimsky S. 1983. *Genetic Alchemy*. MIT Press, Cambridge, Massachusetts.

Zilinskas R.A. and Zimmerman B.K. 1986. *The Gene-splicing Wars: Reflections on the Recombinant DNA Controversy*. Macmillan, New York.

Plasmids and Viruses Are Used as Vectors

Jackson D., Symons R., and Berg P. 1972. Biochemical method for inserting new genetic information into DNA of simian virus 40: Circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci.* **69:** 2904–2909.

Mertz J.E. and Davis R.W. 1972. Cleavage of DNA by RI restriction endonuclease generates cohesive ends. *Proc. Natl. Acad. Sci.* **69:** 3370–3374.

Cohen S., Chang A., Boyer H., and Helfing R. 1973. Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci.* **70:** 3240–3244.

Lobban P. and Kaiser A.D. 1973. Enzymatic end-to-end joining of DNA molecules. *J. Mol. Biol.* **79:** 453–471.

Bolivar F., Rodrigues R.L., Greene P.J., Betlach M.C., Heyneker H.L., et al. 1977. Construction and characterization of new cloning vehicles. II. A multi-purpose cloning system. *Gene* **2:** 95–113.

Yanisch-Peron C., Vieira J., and Messing J. 1985. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* **33:** 103–119.

Hohn B. and Murray K. 1977. Packaging recombinant DNA molecules into bacteriophage particles in vitro. *Proc. Natl. Acad. Sci.* **74:** 3259–3263.

Blattner F.R., Williams B.G., Blechl A.E., Denniston-Thompson K., Faber H.E., et al. 1977. Charon phages: Safer derivatives of bacteriophage lambda for DNA cloning. *Science* **196:** 161–169.

Huynh T.V., Young R.A., and Davis R.W. 1985. Constructing and screening cDNA libraries in lgt10 and lgt11. In *DNA Cloning: A Practical Approach* (ed. D.M. Glover), vol. 1, pp. 49–78. IRL Press, Oxford.

Short J.M., Fernandez J.M., Sorge J.A., and Huse W.D. 1988. Lambda ZAP: A bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Res.* **16:** 7583–7600.

Cloning

Aviv H. and Leder P. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci.* **69:** 1408–1412.

Maniatis T., Kee S.G., Efstratiadis A., and Kafatos F.C. 1976. Amplification and characterization of a β-globin gene synthesized in vitro. *Cell* **8:** 163–182.

Okayama H. and Berg P. 1982. High-efficiency cloning of full length cDNA. *Mol. Cell. Biol.* **2:** 161–170.

Soares M.B., Bonaldo M.D.F., Jelene P., Su L., Lawton L., and Efstratiadis A. 1994. Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci.* **91:** 9228–9232.

Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., et al. 2000. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* **10:** 1617–1630.

Grunstein M. and Hogness D.S. 1975. Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci.* **72:** 3961–3965.

Southern and Northern Blotting

Southern E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:** 503–517.

Southern E.M. 2000. Blotting at 25. *Trends Biochem. Sci.* **25:** 585–588. (Southern provides the inside story of how he devised his eponymous blot.)

Alwine J.C., Kemp D.J., and Stark G.R. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci.* **74:** 5350–5354.

Sequencing DNA

Sanger F. and Dowding M. 1996. *Selected Papers of Frederick Sanger.* World Scientific, Singapore. (Facsimile printings of 48 of Sanger's most important papers, with fascinating commentaries by Sanger.)

Sanger F., Nicklen S., and Coulson A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74:** 5463–5467.

Sanger F. and A.R. Coulson. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94:** 444–448.

Sanger F., Air G.M., Barrel B.G., Brown N.L., Coulson A.R., et al. 1977. Nucleotide sequence of bacteriophage ϕX174. *Nature* **265:** 678–695.

Maxam A.M. and Gilbert W. 1977. A new method of sequencing DNA. *Proc. Natl. Acad. Sci.* **74:** 560–564.

Sutcliffe J.G. 1979. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. *Cold Spring Harbor Symp. Quant. Biol.* **43:** 77–90.

Sutcliffe J.G. 1995. pBR322 and the advent of rapid DNA sequencing. *Trends Biochem. Sci.* **20:** 87–90. (A personal history of the early days of sequencing.)

Oligonucleotide Synthesis and Making Genes

Gait M.J. and Sheppard R.C. 1977. Rapid synthesis of oligodeoxy-ribonucleotides: A new solid-phase method. *Nucleic Acids Res.* **4:** 1135–1158.

Beaucage S.L. and Caruthers M.H. 1981. Deoxynucleotide phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22:** 1859–1862.

Heyneker H.L., Shine J., Goodman H.M., Boyer H., Rosenberg J., et al. 1976. Synthetic *lac* operator is functional in vivo. *Nature* **263:** 748–752.

Khorana H.G. 1979. Total synthesis of a gene. *Science* **203:** 614–625.

Cello J., Pau A.V., and Wimmer E. 2002. Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template. *Science* **297:** 1016–1018.

Polymerase Chain Reaction

Mullis K.B. 1990. The unusual origin of the polymerase chain reaction. *Sci. Am.* **262:** 56–65.

Saiki R.K., Gelfand D.H., Stoffel S., Scharf S.J., Higuchi R., et al. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:** 487–491.

Mullis K. and Faloona F. 1987. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. *Methods Enzymol.* **55:** 335–350.

McCabe E.R.B., Huang S.-Z., Seltzer W.K., and Law M.L. 1987. DNA microextraction from dried blood spots on filter paper blotters: Potential applications to newborn screening. *Hum. Genet.* **75:** 213–216.

Higuchi R., von Beroldingen C.H., Sensabaugh G.F., and Erlich H.A. 1988. DNA typing from single hairs. *Nature* **332:** 543–546.

Krings M., Stone A., Schmitz R.W., Krainitzki H., Stoneking M., and Pääbo S. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* **90:** 19–30.

Chien A., Edgar D.B., and Trela J.M. 1976. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus. J. Bacteriol.* **127:** 1550–1557.

Saiki R.K., Scharf S.J., Faloona F., Mullis K.B., Horn G.T., Erlich H.A., and Arnheim N. 1985. Enzymatic amplification of beta-globin sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230:** 1350–1354.

Kwok S. and Higuchi R. 1989. Avoiding false positives with PCR (review). *Nature* **339:** 237–238.

Longo M.C., Berringer M.S., and Hartley J.L. 1990. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene* **93:** 125–128.

Keohavong P. and Thilly W.G. 1989. Fidelity of DNA polymerases in DNA amplification. *Proc. Natl. Acad. Sci.* **86:** 9253–9257.

Eckert K.A. and Kunkel T.A. 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res.* **18:** 3739–3744.