

# Nucleotides and Nucleic Acids

8.1 Some Basics 281  
 8.2 Nucleic Acid Structure 287  
 8.3 Nucleic Acid Chemistry 297  
 8.4 Other Functions of Nucleotides 306

Nucleotides have a variety of roles in cellular metabolism. They are the energy currency in metabolic transactions, the essential chemical links in the response of cells to hormones and other extracellular stimuli, and the structural components of an array of enzyme cofactors and metabolic intermediates. And, last but certainly not least, they are the constituents of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), the molecular repositories of genetic information. The structure of every protein, and ultimately of every biomolecule and cellular component, is a product of information programmed into the nucleotide sequence of a cell's nucleic acids. The ability to store and transmit genetic information from one generation to the next is a fundamental condition for life.

This chapter provides an overview of the chemical nature of the nucleotides and nucleic acids found in most cells; a more detailed examination of the function of nucleic acids is the focus of Part III of this text.

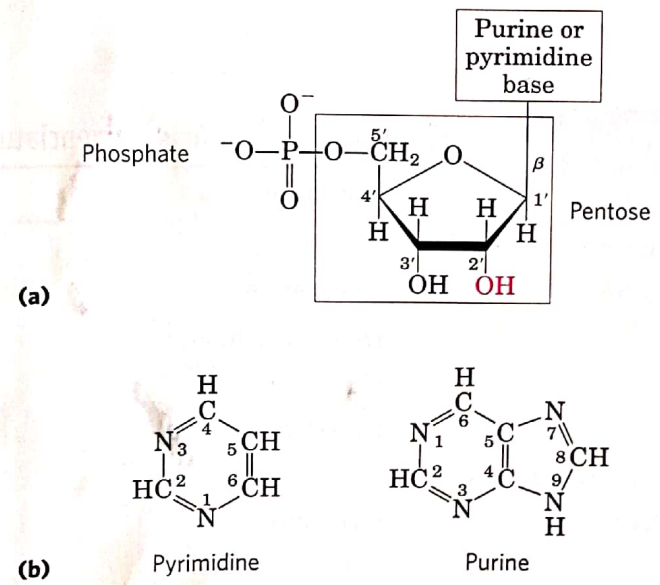
## 8.1 Some Basics

**Nucleotides, Building Blocks of Nucleic Acids** The amino acid sequence of every protein in a cell, and the nucleotide sequence of every RNA, is specified by a nucleotide sequence in the cell's DNA. A segment of a DNA molecule that contains the information required for the synthesis of a functional biological product, whether protein or RNA, is referred to as a **gene**. A cell typically has many thousands of genes, and DNA molecules, not surprisingly, tend to be very large. The storage and transmission of biological information are the only known functions of DNA. RNAs have a broader range of functions, and several classes are found in cells. **Ribosomal RNAs**

(**rRNAs**) are components of ribosomes, the complexes that carry out the synthesis of proteins. **Messenger RNAs (mRNAs)** are intermediaries, carrying genetic information from one or a few genes to a ribosome, where the corresponding proteins can be synthesized. **Transfer RNAs (tRNAs)** are adapter molecules that faithfully translate the information in mRNA into a specific sequence of amino acids. In addition to these major classes there is a wide variety of RNAs with special functions, described in depth in Part III.

## Nucleotides and Nucleic Acids Have Characteristic Bases and Pentoses

**Nucleotides** have three characteristic components: (1) a nitrogenous (nitrogen-containing) base, (2) a pentose, and (3) one or more phosphates (**Fig. 8-1**). The molecule without a phosphate group is called a **nucleoside**. The nitrogenous bases are derivatives of two parent



**FIGURE 8-1 Structure of nucleotides.** (a) General structure showing the numbering convention for the pentose ring. This is a ribonucleotide. In deoxyribonucleotides the —OH group on the 2' carbon (in red) is replaced with H. (b) The parent compounds of the pyrimidine and purine bases of nucleotides and nucleic acids, showing the numbering conventions.

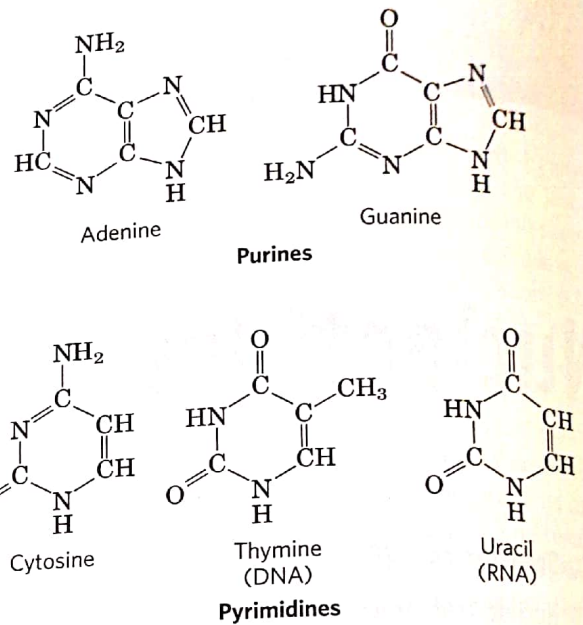
compounds, **pyrimidine** and **purine**. The bases and pentoses of the common nucleotides are heterocyclic compounds.

**KEY CONVENTION:** The carbon and nitrogen atoms in the parent structures are conventionally numbered to facilitate the naming and identification of the many derivative compounds. The convention for the pentose ring follows rules outlined in Chapter 7, but in the pentoses of nucleotides and nucleosides the carbon numbers are given a prime (') designation to distinguish them from the numbered atoms of the nitrogenous bases. ■

The base of a nucleotide is joined covalently (at N-1 of pyrimidines and N-9 of purines) in an *N*- $\beta$ -glycosyl bond to the 1' carbon of the pentose, and the phosphate is esterified to the 5' carbon. The *N*- $\beta$ -glycosyl bond is formed by removal of the elements of water (a hydroxyl group from the pentose and hydrogen from the base), as in *O*-glycosidic bond formation (see Fig. 7-30).

Both DNA and RNA contain two major purine bases, **adenine** (A) and **guanine** (G), and two major pyrimidines. In both DNA and RNA one of the pyrimidines is **cytosine** (C), but the second common pyrimidine is not the same in both: it is **thymine** (T) in DNA and **uracil** (U) in RNA. Only occasionally does thymine occur in RNA or uracil in DNA. The structures of the five major bases are shown in **Figure 8-2**, and the nomenclature of their corresponding nucleotides and nucleosides is summarized in Table 8-1.

Nucleic acids have two kinds of pentoses. The recurring deoxyribonucleotide units of DNA contain 2'-deoxy-D-ribose, and the ribonucleotide units of RNA contain D-ribose. In nucleotides, both types of pentoses are in their  $\beta$ -furanose (closed five-membered ring)



**FIGURE 8-2 Major purine and pyrimidine bases of nucleic acids.** Some of the common names of these bases reflect the circumstances of their discovery. Guanine, for example, was first isolated from guano (bird manure), and thymine was first isolated from thymus tissue.

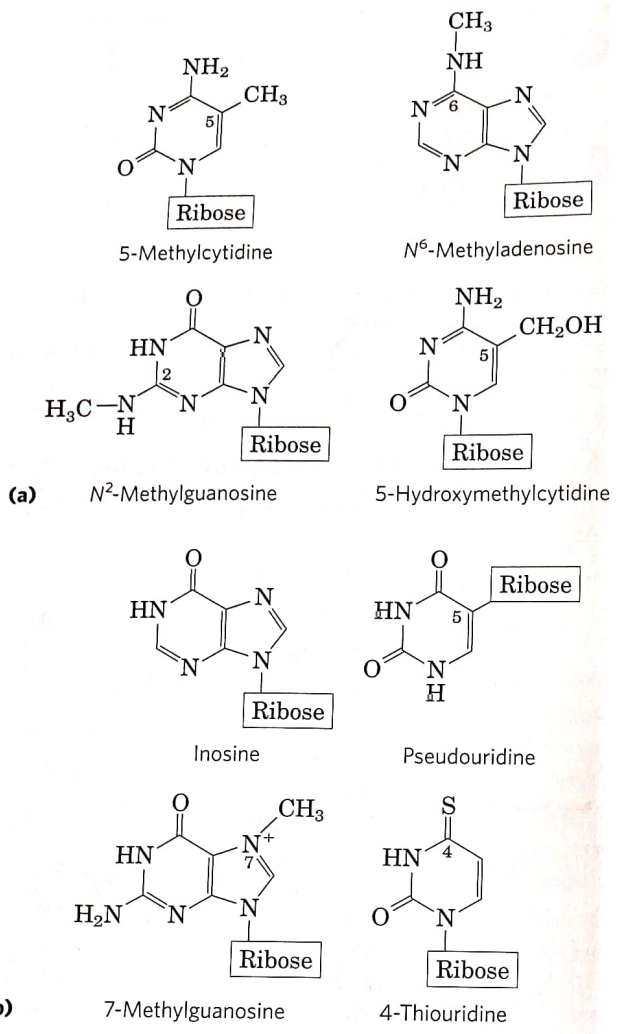
form. As **Figure 8-3** shows, the pentose ring is not planar but occurs in one of a variety of conformations generally described as “puckered.”

**KEY CONVENTION:** Although DNA and RNA seem to have two distinctions—different pentoses and the presence of uracil in RNA and thymine in DNA—it is the pentoses that define the identity of a nucleic acid. If the nucleic acid contains 2'-deoxy-D-ribose, it is DNA by definition even though it may contain uracil. Similarly, if the nucleic acid contains D-ribose, it is RNA regardless of its base composition. ■

**TABLE 8-1 Nucleotide and Nucleic Acid Nomenclature**

Base	Nucleoside	Nucleotide	Nucleic acid
<b>Purines</b>			
Adenine	Adenosine	Adenylate	RNA
	Deoxyadenosine	Deoxyadenylate	DNA
Guanine	Guanosine	Guanylate	RNA
	Deoxyguanosine	Deoxyguanylate	DNA
<b>Pyrimidines</b>			
Cytosine	Cytidine	Cytidylate	RNA
	Deoxycytidine	Deoxycytidylate	DNA
Thymine	Thymidine or deoxythymidine	Thymidylate or deoxythymidylate	DNA
Uracil	Uridine	Uridylate	RNA

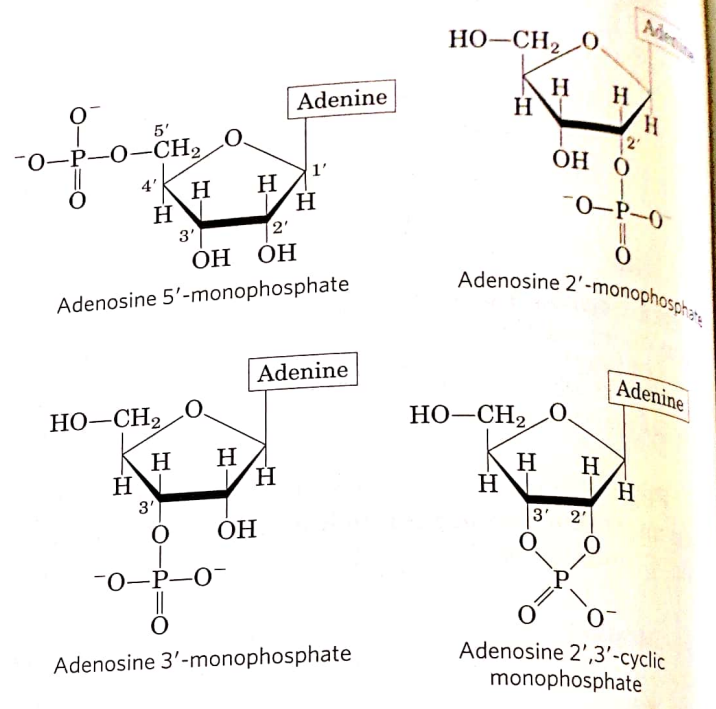
**Note:** “Nucleoside” and “nucleotide” are generic terms that include both ribo- and deoxyribo- forms. Also, ribonucleosides and ribonucleotides are here designated simply as nucleosides and nucleotides (e.g., riboadenosine as adenosine), and deoxyribonucleosides and deoxyribonucleotides as deoxynucleosides and deoxynucleotides (e.g., deoxyriboadenosine as deoxyadenosine). Both forms of naming are acceptable, but the shortened names are more commonly used. Thymine is an exception; “ribothymidine” is used to describe its unusual occurrence in RNA.



**FIGURE 8-5** Some minor purine and pyrimidine bases, shown as the nucleosides. (a) Minor bases of DNA. 5-Methylcytidine occurs in the DNA of animals and higher plants, *N*<sup>6</sup>-methyladenosine in bacterial DNA, and 5-hydroxymethylcytidine in the DNA of animals and of bacteria infected with certain bacteriophages. (b) Some minor bases of tRNAs. Inosine contains the base hypoxanthine. Note that pseudouridine, like uridine, contains uracil; they are distinct in the point of attachment to the ribose—in uridine, uracil is attached through N-1, the usual attachment point for pyrimidines; in pseudouridine, through C-5.

Although nucleotides bearing the major purines and pyrimidines are most common, both DNA and RNA also contain some minor bases (Fig. 8-5). In DNA the most common of these are methylated forms of the major bases; in some viral DNAs, certain bases may be hydroxymethylated or glucosylated. Altered or unusual bases in DNA molecules often have roles in regulating or protecting the genetic information. Minor bases of many types are also found in RNAs, especially in tRNAs (see Fig. 8-25 and Fig. 26-22).

**KEY CONVENTION:** The nomenclature for the minor bases can be confusing. Like the major bases, many have common names—hypoxanthine, for example, shown as its nucleoside inosine in Figure 8-5. When an atom in the purine or pyrimidine ring is substituted, the usual



**FIGURE 8-6** Some adenosine monophosphates. Adenosine 2'-monophosphate, 3'-monophosphate, and 2',3'-cyclic monophosphate are formed by enzymatic and alkaline hydrolysis of RNA.

convention (used here) is simply to indicate the ring position of the substituent by its number—for example, 5-methylcytosine, 7-methylguanine, and 5-hydroxymethylcytosine (shown as the nucleosides in Fig. 8-5). The element to which the substituent is attached (N, C, O) is not identified. The convention changes when the substituted atom is exocyclic (not within the ring structure), in which case the type of atom is identified and the ring position to which it is attached is denoted with a superscript. The amino nitrogen attached to C-6 of adenine is *N*<sup>6</sup>; similarly, the carbonyl oxygen and amino nitrogen at C-6 and C-2 of guanine are *O*<sup>6</sup> and *N*<sup>2</sup>, respectively. Examples of this nomenclature are *N*<sup>6</sup>-methyladenosine and *N*<sup>2</sup>-methylguanosine (Fig. 8-5). ■

Cells also contain nucleotides with phosphate groups in positions other than on the 5' carbon (Fig. 8-6). **Ribonucleoside 2',3'-cyclic monophosphates** are isolatable intermediates, and **ribonucleoside 3'-monophosphates** are end products of the hydrolysis of RNA by certain ribonucleases. Other variations are adenosine 3',5'-cyclic monophosphate (cAMP) and guanosine 3',5'-cyclic monophosphate (cGMP), considered at the end of this chapter.

### Phosphodiester Bonds Link Successive Nucleotides in Nucleic Acids

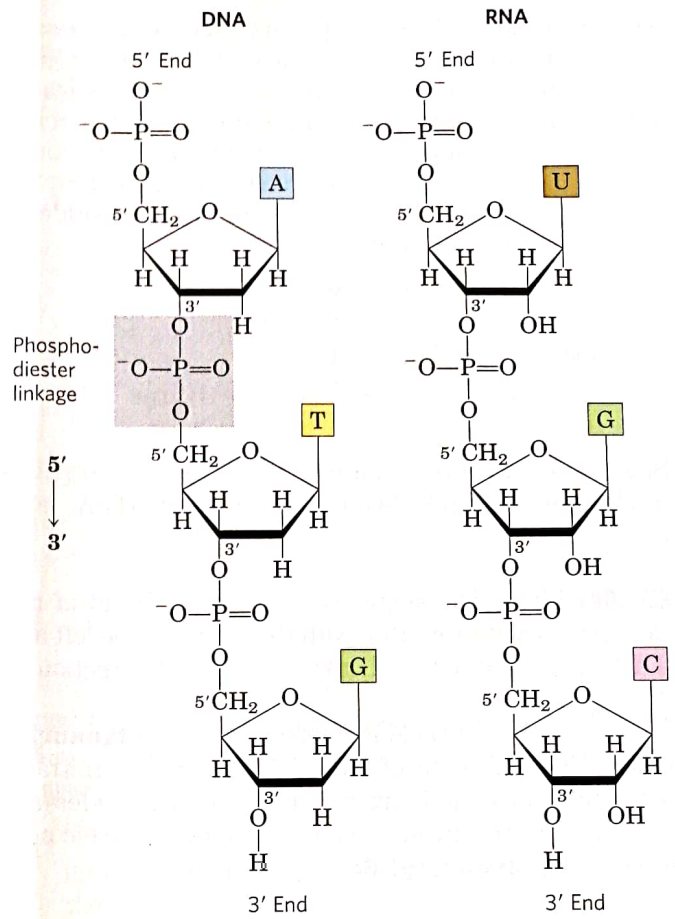
The successive nucleotides of both DNA and RNA are covalently linked through phosphate-group "bridges," in which the 5'-phosphate group of one nucleotide unit is joined to the 3'-hydroxyl group of the next nucleotide.

creating a **phosphodiester linkage (Fig. 8-7)**. Thus the covalent backbones of nucleic acids consist of alternating phosphate and pentose residues, and the nitrogenous bases may be regarded as side groups joined to the backbone at regular intervals. The backbones of both DNA and RNA are hydrophilic. The hydroxyl groups of the sugar residues form hydrogen bonds with water. The phosphate groups, with a  $pK_a$  near 0, are completely ionized and negatively charged at pH 7, and the negative charges are generally neutralized by ionic interactions with positive charges on proteins, metal ions, and polyamines.

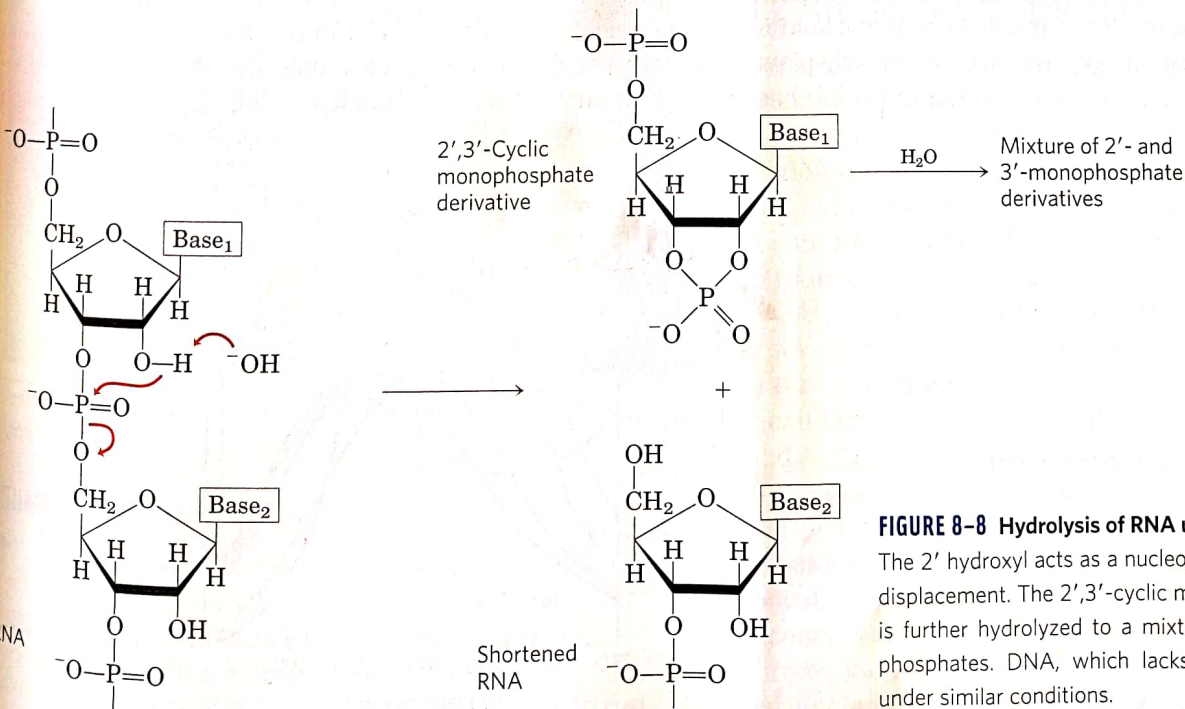
**KEY CONVENTION:** All the phosphodiester linkages in DNA and RNA have the same orientation along the chain (Fig. 8-7), giving each linear nucleic acid strand a specific polarity and distinct 5' and 3' ends. By definition, the **5' end** lacks a nucleotide at the 5' position and the **3' end** lacks a nucleotide at the 3' position. Other groups (most often one or more phosphates) may be present on one or both ends. The 5' to 3' orientation of a strand of nucleic acid refers to the *ends* of the strand, not the orientation of the individual phosphodiester bonds linking its constituent nucleotides. ■

The covalent backbone of DNA and RNA is subject to slow, nonenzymatic hydrolysis of the phosphodiester bonds. In the test tube, RNA is hydrolyzed rapidly under alkaline conditions, but DNA is not; the 2'-hydroxyl groups in RNA (absent in DNA) are directly involved in the process. Cyclic 2',3'-monophosphate nucleotides are the first products of the action of alkali on RNA and are rapidly hydrolyzed further to yield a mixture of 2'- and 3'-nucleoside monophosphates (Fig. 8-8).

The nucleotide sequences of nucleic acids can be represented schematically, as illustrated below by a segment of DNA with five nucleotide units. The phosphate

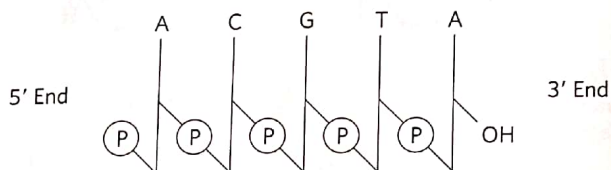


**FIGURE 8-7 Phosphodiester linkages in the covalent backbone of DNA and RNA.** The phosphodiester bonds (one of which is shaded in the DNA) link successive nucleotide units. The backbone of alternating pentose and phosphate groups in both types of nucleic acid is highly polar. The 5' and 3' ends of the macromolecule may be free or may have an attached phosphoryl group.



**FIGURE 8-8 Hydrolysis of RNA under alkaline conditions.** The 2' hydroxyl acts as a nucleophile in an intramolecular displacement. The 2',3'-cyclic monophosphate derivative is further hydrolyzed to a mixture of 2'- and 3'-monophosphates. DNA, which lacks 2' hydroxyls, is stable under similar conditions.

groups are symbolized by  $\text{P}$ , and each deoxyribose is symbolized by a vertical line, from C-1' at the top to C-5' at the bottom (but keep in mind that the sugar is always in its closed-ring  $\beta$ -furanose form in nucleic acids). The connecting lines between nucleotides (which pass through  $\text{P}$ ) are drawn diagonally from the middle (C-3') of the deoxyribose of one nucleotide to the bottom (C-5') of the next.



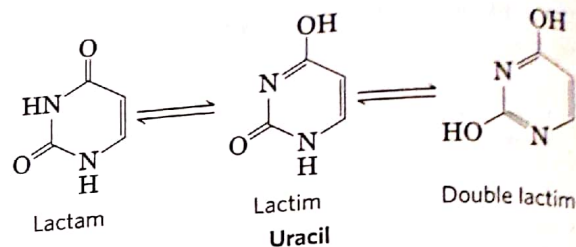
Some simpler representations of this pentadeoxyribonucleotide are pA-C-G-T-A<sub>OH</sub>, pApCpGpTpA, and pACGTA.

**KEY CONVENTION:** The sequence of a single strand of nucleic acid is always written with the 5' end at the left and the 3' end at the right—that is, in the 5'→3' direction. ■

A short nucleic acid is referred to as an **oligonucleotide**. The definition of “short” is somewhat arbitrary, but polymers containing 50 or fewer nucleotides are generally called oligonucleotides. A longer nucleic acid is called a **polynucleotide**.

### The Properties of Nucleotide Bases Affect the Three-Dimensional Structure of Nucleic Acids

Free pyrimidines and purines are weakly basic compounds and thus are called bases. The purines and pyrimidines common in DNA and RNA are aromatic molecules (Fig. 8-2), a property with important consequences for the structure, electron distribution, and light absorption of nucleic acids. Electron delocalization among atoms in the ring gives most of the bonds partial double-bond character. One result is that pyrimidines are planar molecules and purines are very nearly planar, with a slight pucker. Free pyrimidine and purine bases



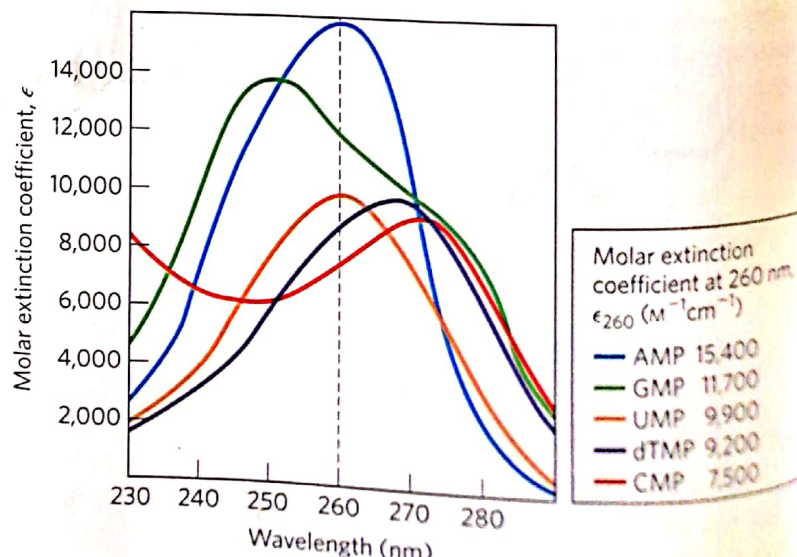
**FIGURE 8-9 Tautomeric forms of uracil.** The lactam form predominates at pH 7.0; the other forms become more prominent as pH decreases. The other free pyrimidines and the free purines also have tautomeric forms, but they are more rarely encountered.

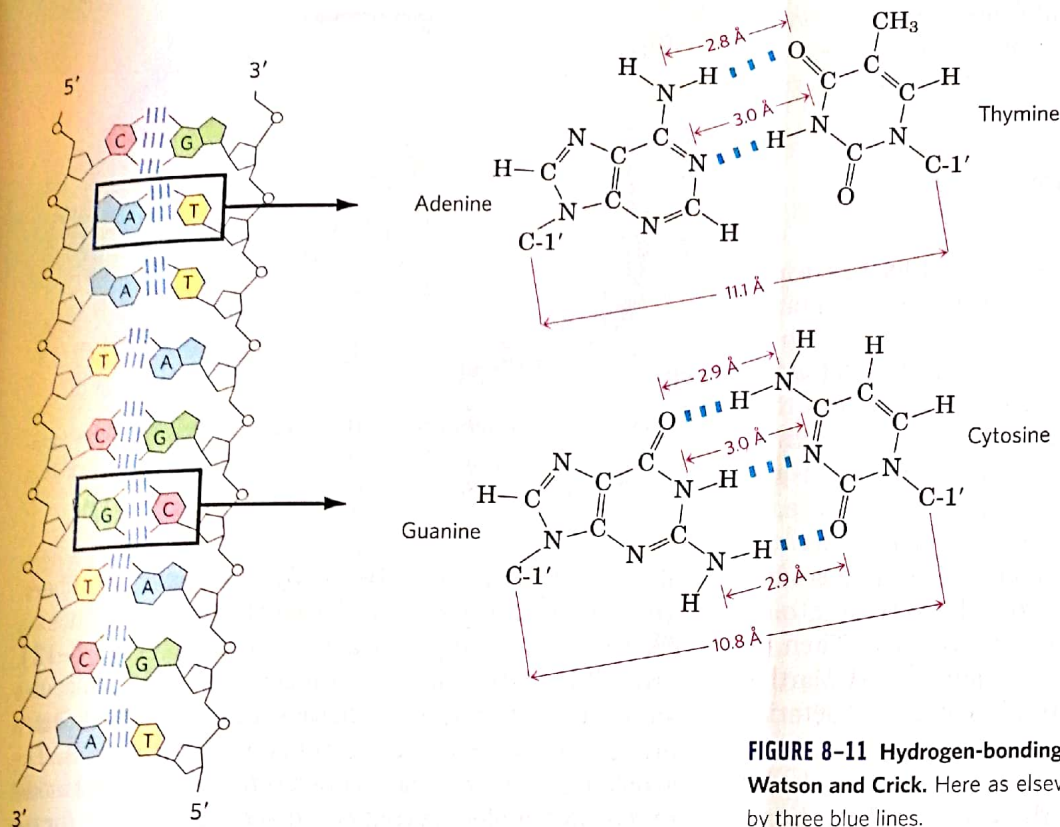
may exist in two or more tautomeric forms depending on the pH. Uracil, for example, occurs in lactam, lactim, and double lactim forms (Fig. 8-9). The structures shown in Figure 8-2 are the tautomers that predominate at pH 7.0. All nucleotide bases absorb UV light, and nucleic acids are characterized by a strong absorption at wavelengths near 260 nm (Fig. 8-10).

The purine and pyrimidine bases are hydrophobic and relatively insoluble in water at the near-neutral pH of the cell. At acidic or alkaline pH the bases become charged and their solubility in water increases. Hydrophobic stacking interactions in which two or more bases are positioned with the planes of their rings parallel (like a stack of coins) are one of two important modes of interaction between bases in nucleic acids. The stacking also involves a combination of van der Waals and dipole-dipole interactions between the bases. Base stacking helps to minimize contact of the bases with water, and base-stacking interactions are very important in stabilizing the three-dimensional structure of nucleic acids, as described later.

The functional groups of pyrimidines and purines are ring nitrogens, carbonyl groups, and exocyclic amino groups. Hydrogen bonds involving the amino and carbonyl groups are the most important mode of interaction between two (and occasionally three or four) complementary strands of nucleic acid. The most common

**FIGURE 8-10 Absorption spectra of the common nucleotides.** The spectra are shown as the variation in molar extinction coefficient with wavelength. The molar extinction coefficients at 260 nm and pH 7.0 ( $\epsilon_{260}$ ) are listed in the table. The spectra of corresponding ribonucleotides and deoxyribonucleotides, as well as the nucleosides, are essentially identical. For mixtures of nucleotides, a wavelength of 260 nm (dashed vertical line) is used for absorption measurements.





**FIGURE 8-11** Hydrogen-bonding patterns in the base pairs defined by Watson and Crick. Here as elsewhere, hydrogen bonds are represented by three blue lines.

hydrogen-bonding patterns are those defined by James D. Watson and Francis Crick in 1953, in which A bonds specifically to T (or U) and G bonds to C (**Fig. 8-11**). These two types of **base pairs** predominate in double-stranded DNA and RNA, and the tautomers shown in Figure 8-2 are responsible for these patterns. It is this specific pairing of bases that permits the duplication of genetic information, as we shall discuss later in this chapter.

group of one pentose and the 3'-hydroxyl group of the next.

- ▶ There are two types of nucleic acid: RNA and DNA. The nucleotides in RNA contain ribose, and the common pyrimidine bases are uracil and cytosine. In DNA, the nucleotides contain 2'-deoxyribose, and the common pyrimidine bases are thymine and cytosine. The primary purines are adenine and guanine in both RNA and DNA.



James D. Watson



Francis Crick, 1916-2004

### SUMMARY 8.1 Some Basics

A nucleotide consists of a nitrogenous base (purine or pyrimidine), a pentose sugar, and one or more phosphate groups. Nucleic acids are polymers of nucleotides, joined together by phosphodiester linkages between the 5'-hydroxyl

## 8.2 Nucleic Acid Structure

The discovery of the structure of DNA by Watson and Crick in 1953 gave rise to entirely new disciplines and influenced the course of many established ones. In this section we focus on DNA structure, some of the events that led to its discovery, and more recent refinements in our understanding of DNA. RNA structure is also introduced.

As in the case of protein structure (Chapter 4), it is sometimes useful to describe nucleic acid structure in terms of hierarchical levels of complexity (primary, secondary, tertiary). The primary structure of a nucleic acid is its covalent structure and nucleotide sequence. Any regular, stable structure taken up by some or all of the nucleotides in a nucleic acid can be referred to as secondary structure. All structures considered in the remainder of this chapter fall under the heading of secondary structure. The complex folding of large chromosomes within eukaryotic chromatin and bacterial nucleoids, or the elaborate folding of large tRNA

or rRNA molecules, is generally considered tertiary structure. DNA tertiary structure is discussed in Chapter 24, and RNA tertiary structure is considered in Chapter 26.

### DNA Is a Double Helix That Stores Genetic Information

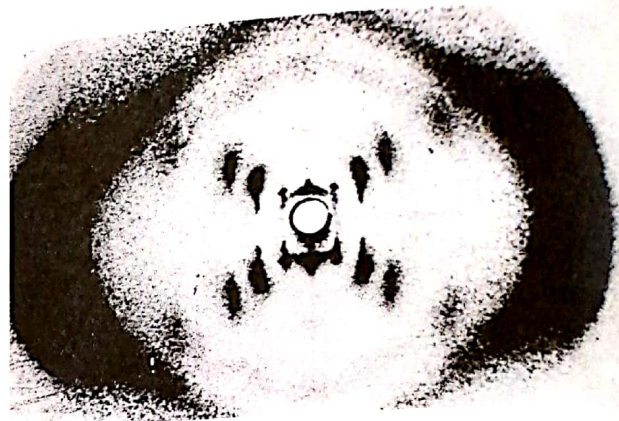
DNA was first isolated and characterized by Friedrich Miescher in 1868. He called the phosphorus-containing substance "nuclein." Not until the 1940s, with the work of Oswald T. Avery, Colin MacLeod, and Maclyn McCarty, was there any compelling evidence that DNA was the genetic material. Avery and his colleagues found that DNA extracted from a virulent (disease-causing) strain of the bacterium *Streptococcus pneumoniae* and injected into a nonvirulent strain of the same bacterium transformed the nonvirulent strain into a virulent strain. They concluded that the DNA from the virulent strain carried the genetic information for virulence. Then in 1952, experiments by Alfred D. Hershey and Martha Chase, in which they studied the infection of bacterial cells by a virus (bacteriophage) with radioactively labeled DNA or protein, removed any remaining doubt that DNA, not protein, carried the genetic information.

Another important clue to the structure of DNA came from the work of Erwin Chargaff and his colleagues in the late 1940s. They found that the four nucleotide bases of DNA occur in different ratios in the DNAs of different organisms and that the amounts of certain bases are closely related. These data, collected from DNAs of a great many different species, led Chargaff to the following conclusions:

1. The base composition of DNA generally varies from one species to another.
2. DNA specimens isolated from different tissues of the same species have the same base composition.
3. The base composition of DNA in a given species does not change with an organism's age, nutritional state, or changing environment.
4. In all cellular DNAs, regardless of the species, the number of adenosine residues is equal to the number of thymidine residues (that is,  $A = T$ ), and the number of guanosine residues is equal to the number of cytosine residues ( $G = C$ ). From these relationships it follows that the sum of the purine residues equals the sum of the pyrimidine residues; that is,  $A + G = T + C$ .

These quantitative relationships, sometimes called "Chargaff's rules," were confirmed by many subsequent researchers. They were a key to establishing the three-dimensional structure of DNA and yielded clues to how genetic information is encoded in DNA and passed from one generation to the next.

To shed more light on the structure of DNA, Rosalind Franklin and Maurice Wilkins used the powerful method



**FIGURE 8-12** X-ray diffraction pattern of DNA fibers. The spots forming a cross in the center denote a helical structure. The heavy bands at the left and right arise from the recurring bases.

of x-ray diffraction (see Box 4-5) to analyze DNA fibers. They showed in the early 1950s that DNA produces a characteristic x-ray diffraction pattern (**Fig. 8-12**). From this pattern it was deduced that DNA molecules are helical with two periodicities along their long axis, a primary one of 3.4 Å and a secondary one of 34 Å. The problem then was to formulate a three-dimensional model of the DNA molecule that could account not only for the x-ray diffraction data but also for the specific  $A = T$  and  $G = C$  base equivalences discovered by Chargaff and for the other chemical properties of DNA.

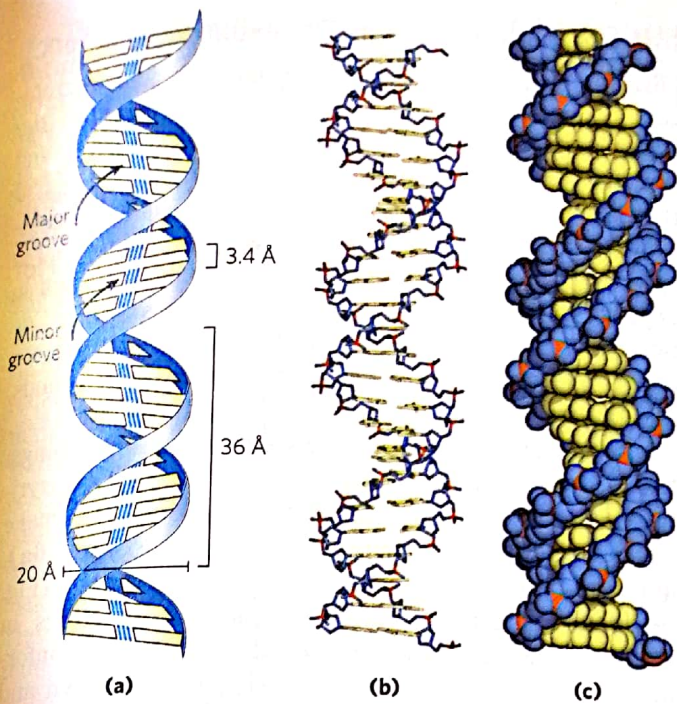


Rosalind Franklin,  
1920-1958



Maurice Wilkins,  
1916-2004

James Watson and Francis Crick relied on this accumulated information about DNA to set about deducing its structure. In 1953 they postulated a three-dimensional model of DNA structure that accounted for all the available data. It consists of two helical DNA chains wound around the same axis to form a right-handed double helix (see Box 4-1 for an explanation of the right- or left-handed sense of a helical structure). The hydrophilic backbones of alternating deoxyribose and phosphate groups are on the outside of the double helix, facing the surrounding water. The furanose ring of each deoxyribose is in the C-2' endo conformation. The purine and pyrimidine bases of both strands are stacked inside the double helix, with their



**FIGURE 8-13** Watson-Crick model for the structure of DNA. The original model proposed by Watson and Crick had 10 base pairs, or 34 Å (3.4 nm), per turn of the helix; subsequent measurements revealed 10.5 base pairs, or 36 Å (3.6 nm), per turn. (a) Schematic representation, showing dimensions of the helix. (b) Stick representation showing the backbone and stacking of the bases. (c) Space-filling model.

ring structures very close together and perpendicular to the long axis. The offset pairing of the two strands creates a **major groove** and **minor groove** on the surface of the duplex (Fig. 8-13). Each nucleotide base of one strand is paired in the same plane with a base of the other strand. Watson and Crick found that the hydrogen-bonded base pairs illustrated in Figure 8-11, G with C and A with T, are those that fit best within the structure, providing a rationale for Chargaff's rule that in any DNA,  $G = C$  and  $A = T$ . It is important to note that three hydrogen bonds can form between G and C, symbolized  $G \equiv C$ , but only two can form between A and T, symbolized  $A = T$ . This is one reason for the finding that separation of paired DNA strands is more difficult the higher the ratio of  $G \equiv C$  to  $A = T$  base pairs. Other pairings of bases tend (to varying degrees) to destabilize the double-helical structure.

When Watson and Crick constructed their model, they had to decide at the outset whether the strands of DNA should be **parallel** or **antiparallel**—whether their 3',5'-phosphodiester bonds should run in the same or opposite directions. An antiparallel orientation produced the most convincing model, and later work with DNA polymerases (Chapter 25) provided experimental evidence that the strands are indeed antiparallel, a finding ultimately confirmed by x-ray analysis.

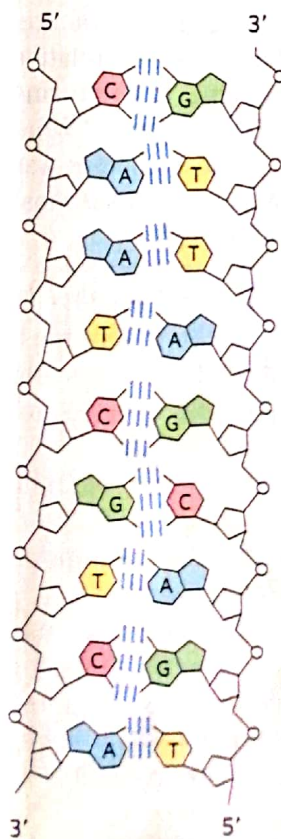
To account for the periodicities observed in the x-ray diffraction patterns of DNA fibers, Watson and Crick manipulated molecular models to arrive at a structure in which the vertically stacked bases inside the double helix would be 3.4 Å apart; the secondary repeat

distance of about 34 Å was accounted for by the presence of 10 base pairs in each complete turn of the double helix. In aqueous solution the structure differs slightly from that in fibers, having 10.5 base pairs per helical turn (Fig. 8-13).

As **Figure 8-14** shows, the two antiparallel polynucleotide chains of double-helical DNA are not identical in either base sequence or composition. Instead they are **complementary** to each other. Wherever adenine occurs in one chain, thymine is found in the other; similarly, wherever guanine occurs in one chain, cytosine is found in the other.

The DNA double helix, or duplex, is held together by two forces, as described earlier: hydrogen bonding between complementary base pairs (Fig. 8-11) and base-stacking interactions. The complementarity between the DNA strands is attributable to the hydrogen bonding between base pairs. The base-stacking interactions, which are largely nonspecific with respect to the identity of the stacked bases, make the major contribution to the stability of the double helix.

The important features of the double-helical model of DNA structure are supported by much chemical and biological evidence. Moreover, the model immediately suggested a mechanism for the transmission of genetic information. The essential feature of the model is the complementarity of the two DNA strands. As Watson and Crick were able to see, well before confirmatory data became available, this structure could logically be replicated by (1) separating the two strands and (2) synthesizing a complementary strand for each. Because nucleotides in each new strand are joined in a sequence specified by the base-pairing rules stated above, each preexisting



**FIGURE 8-14** Complementarity of strands in the DNA double helix. The complementary antiparallel strands of DNA follow the pairing rules proposed by Watson and Crick. The base-paired antiparallel strands differ in base composition: the left strand has the composition  $A_3T_2G_3C_3$ , the right,  $A_2T_3G_3C_3$ . They also differ in sequence when each chain is read in the 5'→3' direction. Note the base equivalences:  $A = T$  and  $G = C$  in the duplex.