

# 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

**N**ucleotides 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

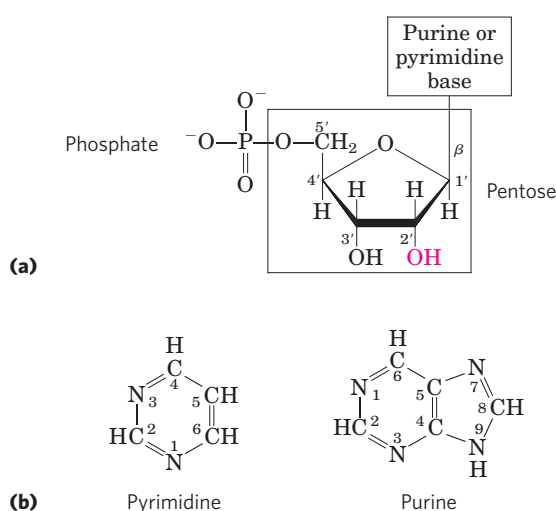
**Nucleotides, Building Blocks of Nucleic Acids** The amino acid sequence of every protein in a cell, and the nucleotide sequence of every RNA, is specified by a nucleotide sequence in the cell's DNA. A segment of a DNA molecule that contains the information required for the synthesis of a functional 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.

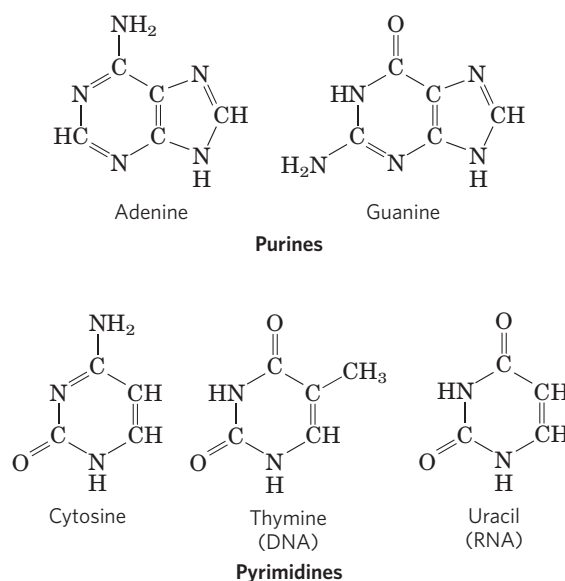
compounds, **pyrimidine** and **purine**. The bases and pentoses of the common nucleotides are heterocyclic compounds.

**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.

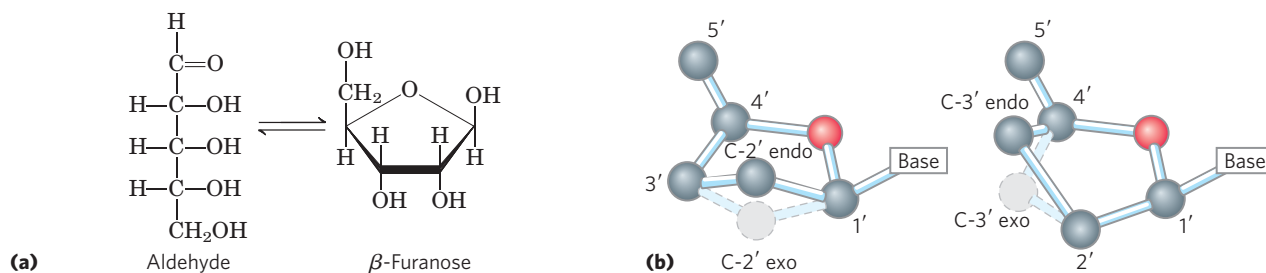
form. As **Figure 8–3** shows, the pentose ring is not planar but occurs in one of a variety of conformations generally described as “puckered.”

**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.

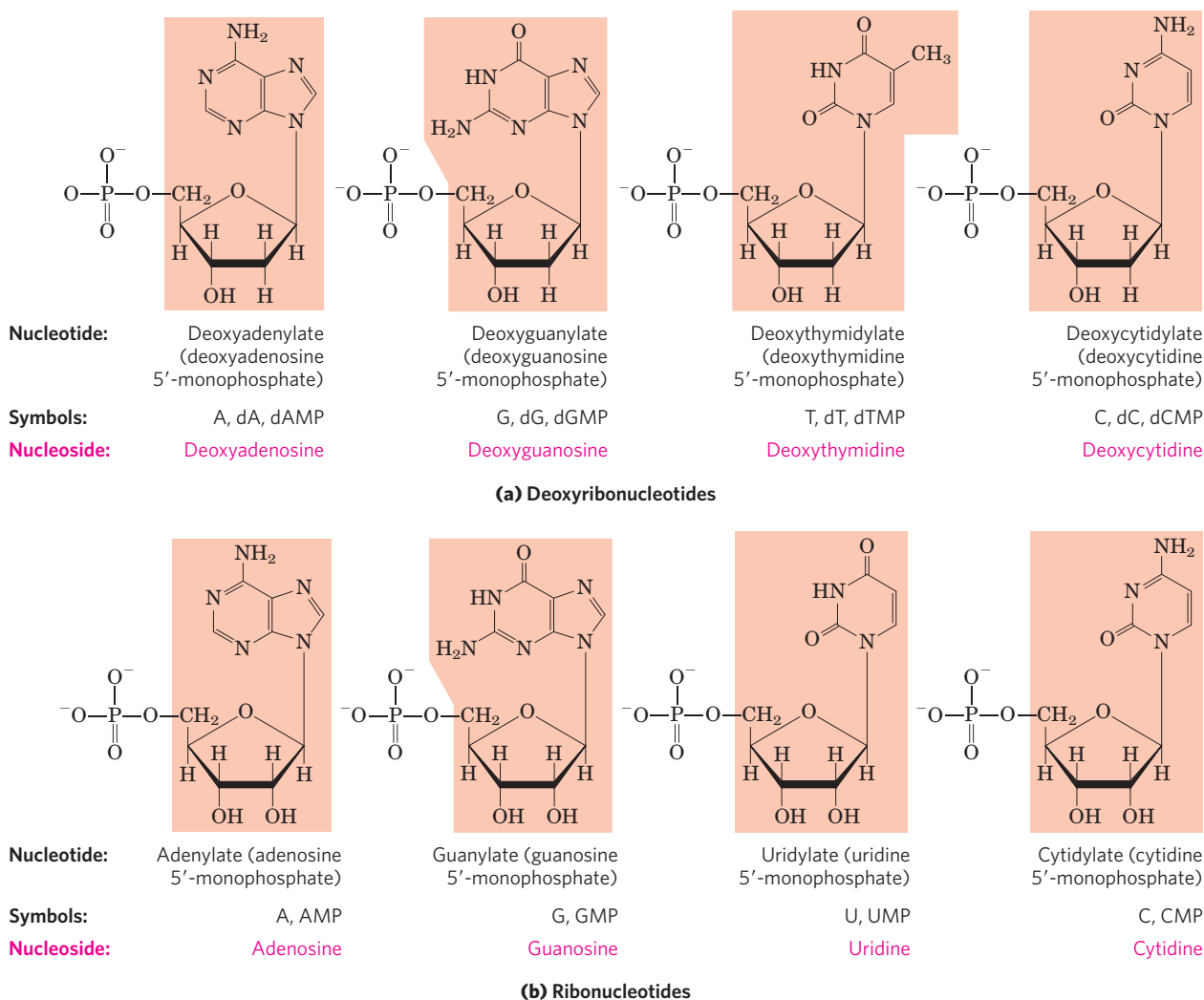


**FIGURE 8-3 Conformations of ribose.** (a) In solution, the straight-chain (aldehyde) and ring ( $\beta$ -furanose) forms of free ribose are in equilibrium. RNA contains only the ring form,  $\beta$ -D-ribofuranose. Deoxyribose undergoes a similar interconversion in solution, but in DNA exists solely as  $\beta$ -2'-deoxy-D-ribofuranose. (b) Ribofuranose rings in nucleotides can

exist in four different puckered conformations. In all cases, four of the five atoms are nearly in a single plane. The fifth atom (C-2' or C-3') is on either the same (endo) or the opposite (exo) side of the plane relative to the C-5' atom.

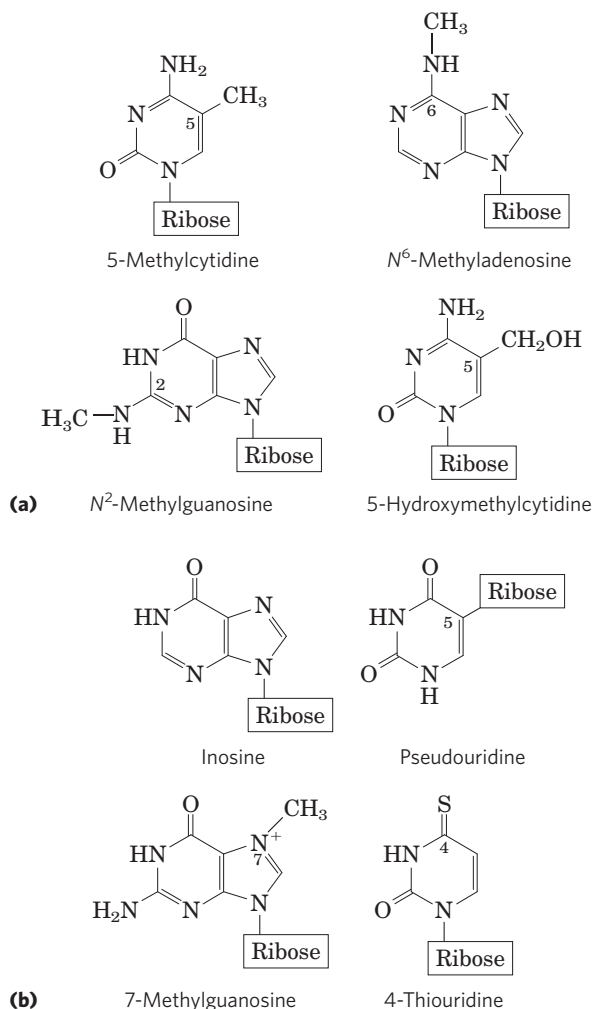
**Figure 8-4** gives the structures and names of the four major **deoxyribonucleotides** (deoxyribonucleoside 5'-monophosphates), the structural units of

DNAs, and the four major **ribonucleotides** (ribonucleoside 5'-monophosphates), the structural units of RNAs.



**FIGURE 8-4 Deoxyribonucleotides and ribonucleotides of nucleic acids.** All nucleotides are shown in their free form at pH 7.0. The nucleotide units of DNA (a) are usually symbolized as A, G, T, and C, sometimes as dA, dG, dT, and dC; those of RNA (b) as A, G, U, and C. In their free form the deoxyribonucleotides are commonly abbreviated dAMP, dGMP, dTMP, and dCMP; the ribonucleotides, AMP, GMP,

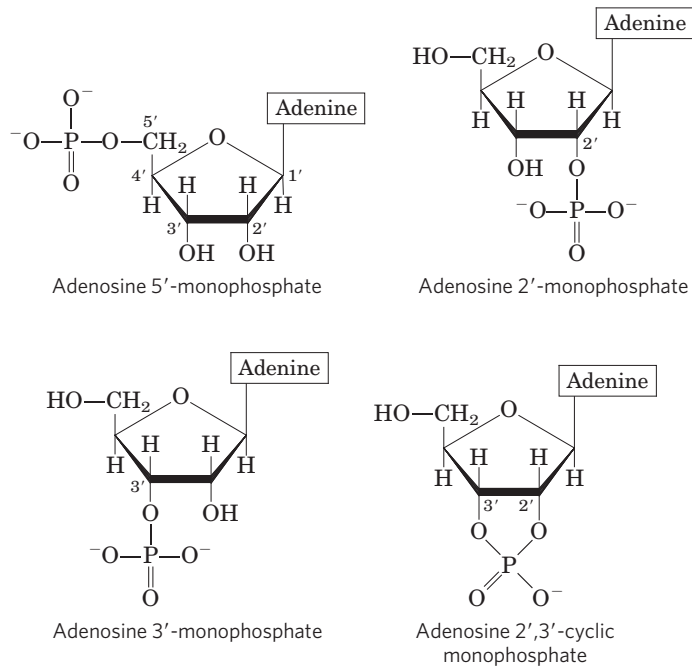
UMP, and CMP. For each nucleotide in the figure, the more common name is followed by the complete name in parentheses. All abbreviations assume that the phosphate group is at the 5' position. The nucleoside portion of each molecule is shaded in light red. In this and the following illustrations, the ring carbons are not shown.



**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,  $N^6$ -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  $N^6$ ; similarly, the carbonyl oxygen and amino nitrogen at C-6 and C-2 of guanine are  $O^6$  and  $N^2$ , respectively. Examples of this nomenclature are  $N^6$ -methyladenosine and  $N^2$ -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

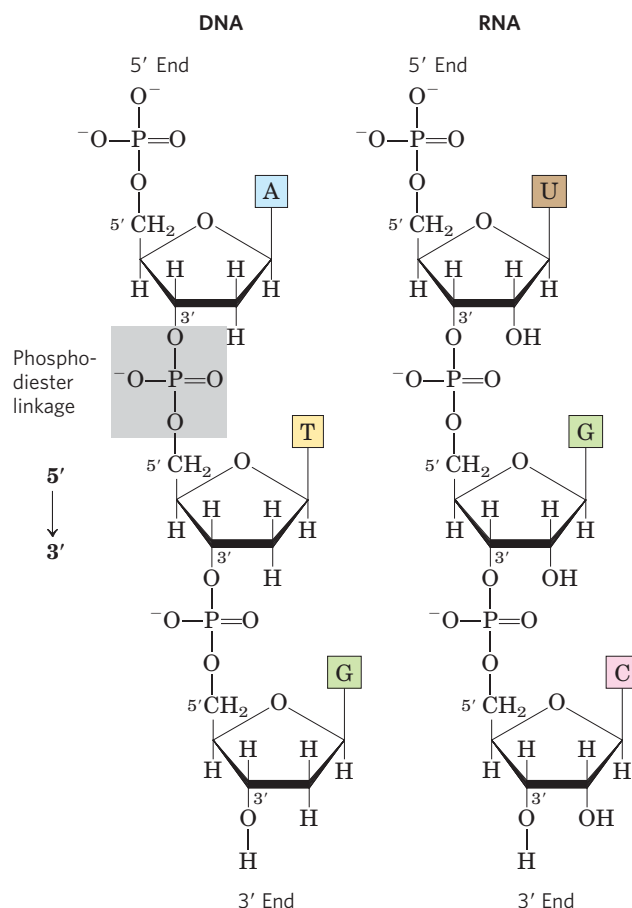
The successive nucleotides of both DNA and RNA are covalently linked through phosphate-group “bridges,” in which the 5'-phosphate group of one nucleotide unit is joined to the 3'-hydroxyl group of the next nucleotide,

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.

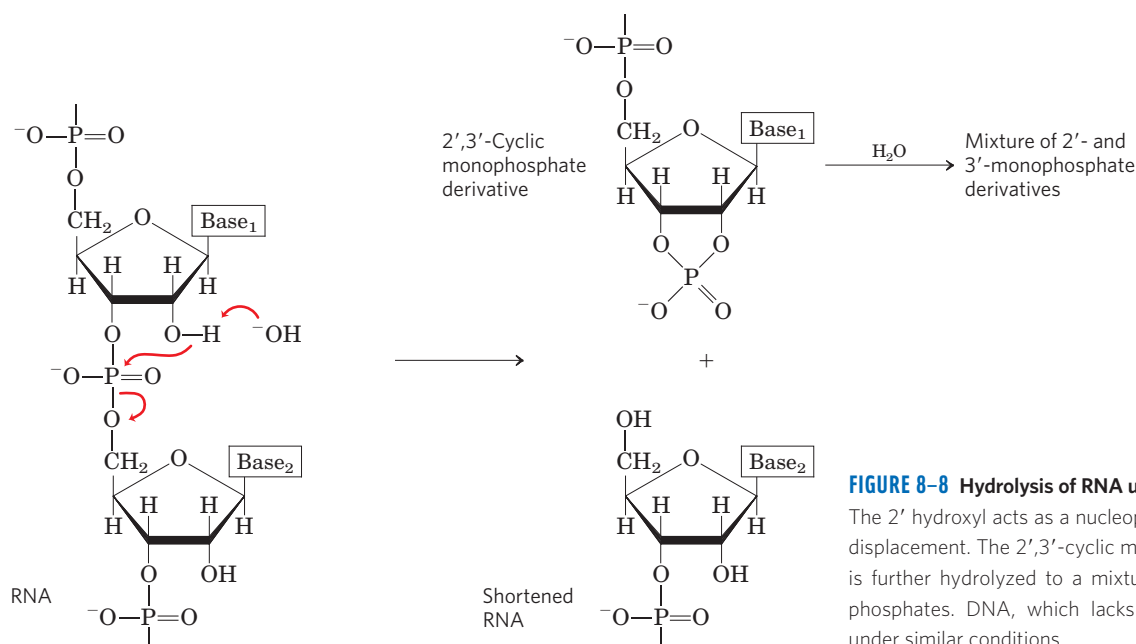
**KEY CONVENTION:** All the phosphodiester linkages in DNA and RNA have the same orientation along the chain (Fig. 8-7), giving each linear nucleic acid strand a specific polarity and distinct 5' and 3' ends. By definition, the **5' end** lacks a nucleotide at the 5' position and the **3' end** lacks 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

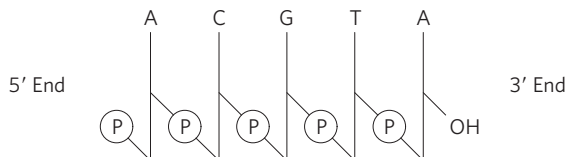


**FIGURE 8-7** Phosphodiester linkages in the covalent backbone of DNA and RNA. The phosphodiester bonds (one of which is shaded in the DNA) link successive nucleotide units. The backbone of alternating pentose and phosphate groups in both types of nucleic acid is highly polar. The 5' and 3' ends of 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  $\textcircled{\text{P}}$ , 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  $\textcircled{\text{P}}$ ) are drawn diagonally from the middle (C-3') of the deoxyribose of one nucleotide to the bottom (C-5') of the next.



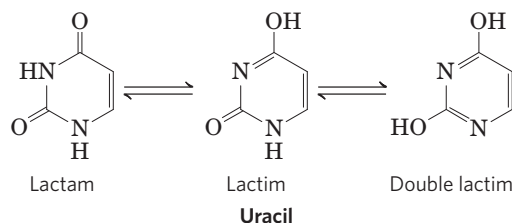
Some simpler representations of this pentadeoxyribonucleotide are pA-C-G-T-A<sub>OH</sub>, pApCpGpTpA, and pACGTA.

**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



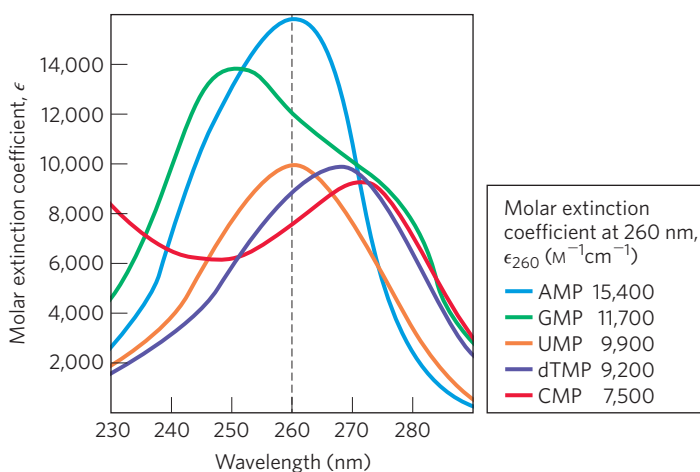
**FIGURE 8–9 Tautomeric forms of uracil.** The lactam form predominates at pH 7.0; the other forms become more prominent as pH decreases. The other free pyrimidines and the free purines also have tautomeric forms, but they are more rarely encountered.

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).

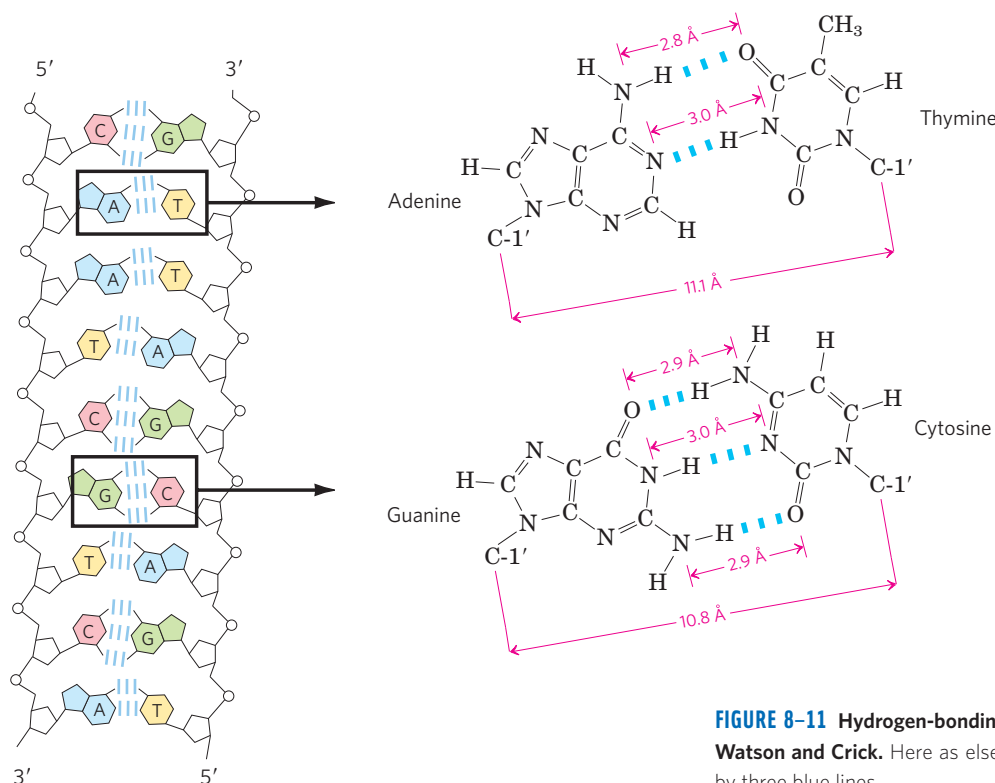
The purine and pyrimidine bases are hydrophobic and relatively insoluble in water at the near-neutral pH of the cell. At acidic or alkaline pH the bases become charged and their solubility in water increases. Hydrophobic stacking interactions in which two or more bases are positioned with the planes 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

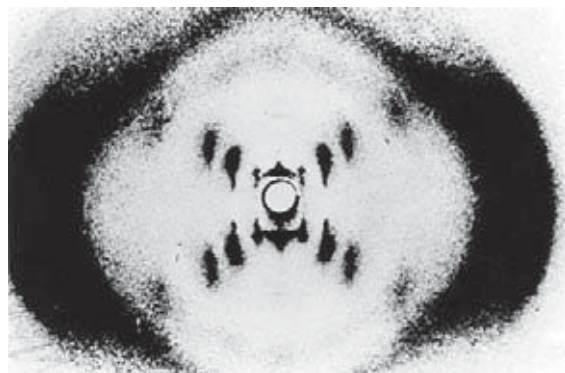
DNA was first isolated and characterized by Friedrich Miescher in 1868. He called the phosphorus-containing substance “nuclein.” Not until the 1940s, with the work of Oswald T. Avery, Colin MacLeod, and Maclyn McCarty, was there any compelling evidence that DNA was the genetic material. Avery and his 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 cytidine 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.



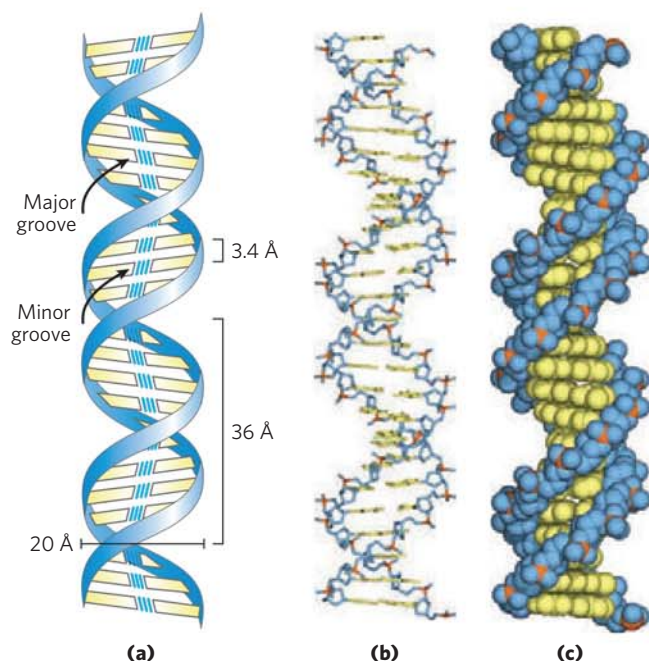
Rosalind Franklin,  
1920-1958



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 hydrophobic and nearly planar





**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.

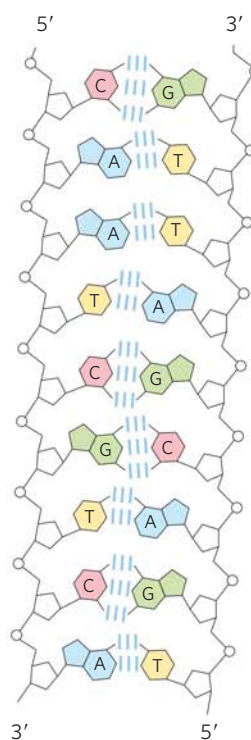
To account for the periodicities observed in the x-ray diffraction patterns of DNA fibers, Watson and Crick manipulated molecular models to arrive at a structure in which the vertically stacked bases inside the double helix would be 3.4 Å apart; the secondary repeat

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**).

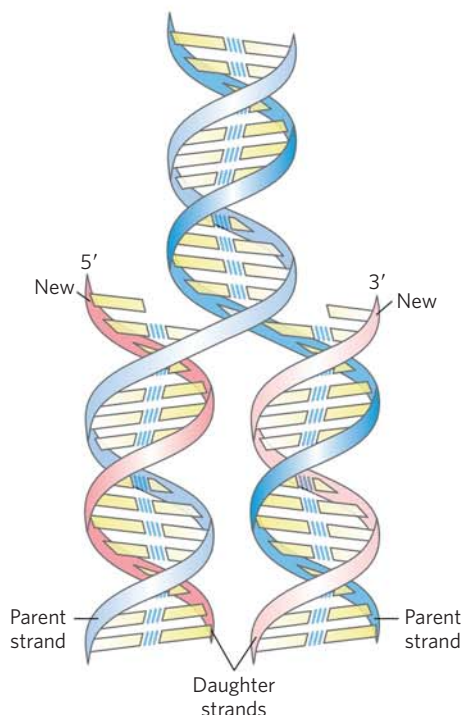
As **Figure 8-14** shows, the two antiparallel polynucleotide chains of double-helical DNA are not identical in either base sequence or composition. Instead they are **complementary** to each other. Wherever adenine occurs in one chain, thymine is found in the other; similarly, wherever guanine occurs 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  $A_3T_2G_1C_3$ ; the right,  $A_2T_3G_3C_1$ . 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.



**FIGURE 8-15** Replication of DNA as suggested by Watson and Crick. The preexisting or “parent” strands become separated, and each is the template for biosynthesis of a complementary “daughter” strand (in pink).

strand functions as a template to guide the synthesis of one complementary strand (Fig. 8-15). These expectations were experimentally confirmed, inaugurating a revolution in our understanding of biological inheritance.

### WORKED EXAMPLE 8-1 Base Pairing in DNA

In samples of DNA isolated from two unidentified species of bacteria, X and Y, adenine makes up 32% and 17%, respectively, of the total bases. What relative proportions of adenine, guanine, thymine, and cytosine would you expect to find in the two DNA samples? What assumptions have you made? One of these species was isolated from a hot spring (64°C). Which species is most likely the thermophilic bacterium, and why?

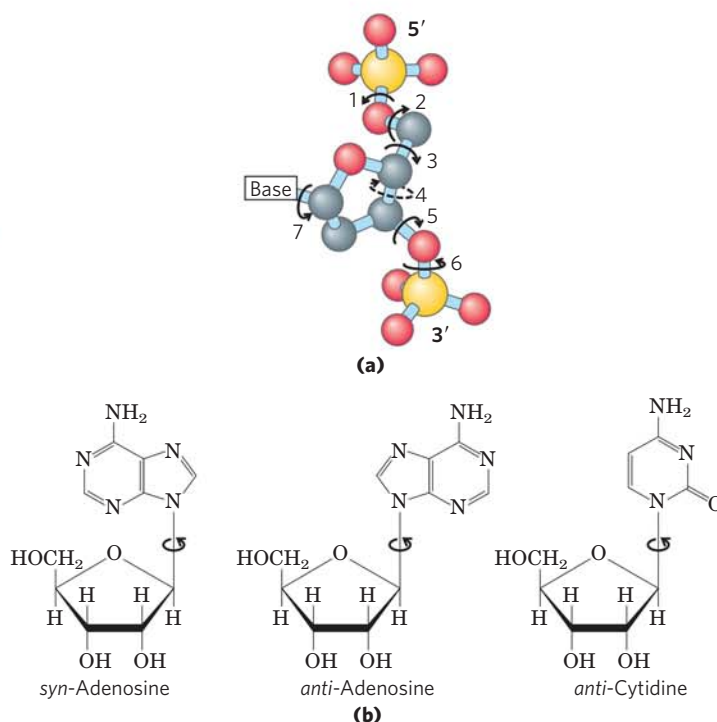
**Solution:** For any double-helical DNA,  $A = T$  and  $G = C$ . The DNA from species X has 32% A and therefore must contain 32% T. This accounts for 64% of the bases and leaves 36% as  $G \equiv C$  pairs: 18% G and 18% C. The sample from species Y, with 17% A, must contain 17% T, accounting for 34% of the base pairs. The remaining 66% of the bases are thus equally distributed as 33% G and 33% C. This calculation is based on the assumption that both DNA molecules are double-stranded.

The higher the  $G + C$  content of a DNA molecule, the higher the melting temperature. Species Y, having the DNA with the higher  $G + C$  content (66%), most likely is the thermophilic bacterium; its DNA has a higher melting temperature and thus is more stable at the temperature of the hot spring.

### DNA Can Occur in Different Three-Dimensional Forms

DNA is a remarkably flexible molecule. Considerable rotation is possible around several types of bonds in the sugar–phosphate (phosphodeoxyribose) backbone, and thermal fluctuation can produce bending, stretching, and unpairing (melting) of the strands. Many significant deviations from the Watson-Crick DNA structure are found in cellular DNA, some or all of which may be important in DNA metabolism. These structural variations generally do not affect the key properties of DNA defined by Watson and Crick: strand complementarity, antiparallel strands, and the requirement for  $A \equiv T$  and  $G \equiv C$  base pairs.

Structural variation in DNA reflects three things: the different possible conformations of the deoxyribose, rotation about the contiguous bonds that make up the phosphodeoxyribose backbone (Fig. 8-16a), and free rotation about the C-1'–N-glycosyl bond (Fig. 8-16b). Because of steric constraints, purines in purine nucleotides are restricted to two stable conformations with respect to deoxyribose, called *syn* and *anti* (Fig. 8-16b). Pyrimidines are generally restricted to the *anti* conformation because of steric interference between the sugar and the carbonyl oxygen at C-2 of the pyrimidine.



**FIGURE 8-16** Structural variation in DNA. (a) The conformation of a nucleotide in DNA is affected by rotation about seven different bonds. Six of the bonds rotate freely. The limited rotation about bond 4 gives rise to ring pucker. This conformation is *endo* or *exo*, depending on whether the atom is displaced to the same side of the plane as C-5' or to the opposite side (see Fig. 8-3b). (b) For purine bases in nucleotides, only two conformations with respect to the attached ribose units are sterically permitted, *anti* or *syn*. Pyrimidines occur in the *anti* conformation.

The Watson-Crick structure is also referred to as **B-form DNA**, or B-DNA. The B form is the most stable structure for a random-sequence DNA molecule under physiological conditions and is therefore the standard point of reference in any study of the properties of DNA. Two structural variants that have been well characterized in crystal structures are the **A** and **Z forms**. These three DNA conformations are shown in **Figure 8–17**, with a summary of their properties. The A form is favored in many solutions that are relatively devoid of water. The DNA is still arranged in a right-handed double helix, but the helix is wider and the number of base pairs per helical turn is 11, rather than 10.5 as in B-DNA. The plane of the base pairs in A-DNA is tilted about 20° relative to B-DNA base pairs, thus the base pairs in A-DNA are not perfectly perpendicular to the helix axis. These structural changes deepen the major groove while making the minor groove shallower. The reagents used to promote crystallization of DNA tend to dehydrate it, and thus most short DNA molecules tend to crystallize in the A form.

Z-form DNA is a more radical departure from the B structure; the most obvious distinction is the left-handed helical rotation. There are 12 base pairs per helical turn, and the structure appears more slender and elongated. The DNA backbone takes on a zigzag appearance. Certain nucleotide sequences fold into left-handed Z helices much more readily than others. Prominent examples are sequences in which pyrimidines alternate with

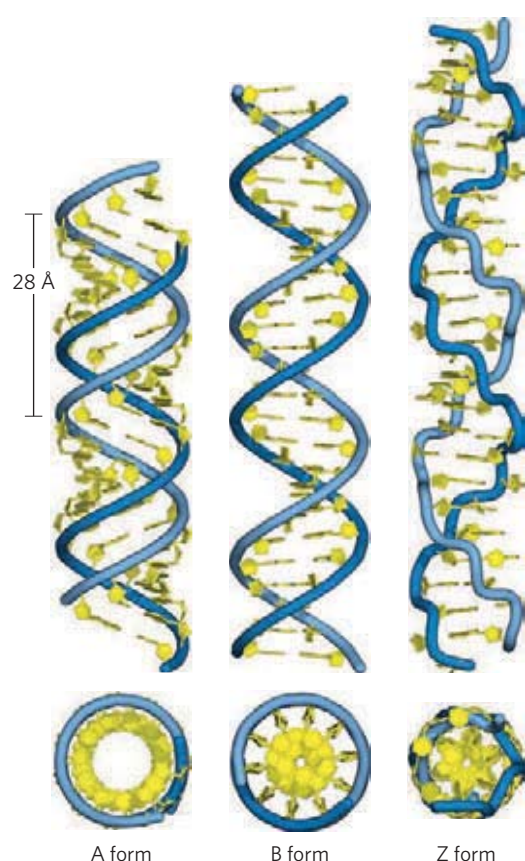
purines, especially alternating C and G or 5-methyl-C and G residues. To form the left-handed helix in Z-DNA, the purine residues flip to the syn conformation, alternating with pyrimidines in the anti conformation. The major groove is barely apparent in Z-DNA, and the minor groove is narrow and deep.

Whether A-DNA occurs in cells is uncertain, but there is evidence for some short stretches (tracts) of Z-DNA in both bacteria and eukaryotes. These Z-DNA tracts may play a role (as yet undefined) in regulating the expression of some genes or in genetic recombination.

### Certain DNA Sequences Adopt Unusual Structures

Other sequence-dependent structural variations found in larger chromosomes may affect the function and metabolism of the DNA segments in their immediate vicinity. For example, bends occur in the DNA helix wherever four or more adenosine residues appear sequentially in one strand. Six adenosines in a row produce a bend of about 18°. The bending observed with this and other sequences may be important in the binding of some proteins to DNA.

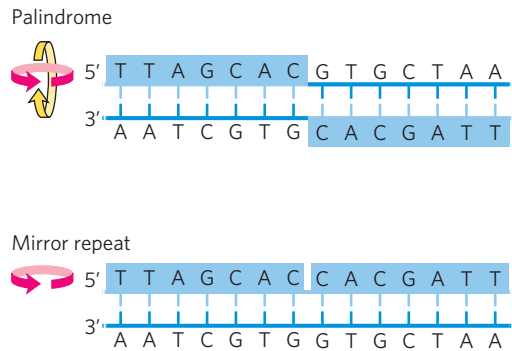
A rather common type of DNA sequence is a **palindrome**. A palindrome is a word, phrase, or sentence that is spelled identically read either forward or backward; two examples are ROTATOR and NURSES RUN. The term is applied to regions of DNA with **inverted repeats** of base sequence having twofold symmetry



**FIGURE 8–17** Comparison of A, B, and Z forms of DNA. Each structure shown here has 36 base pairs. The riboses and bases are shown in yellow. The phosphodiester backbone is represented as a blue rope. Blue is the color used to represent DNA strands in later chapters. The table summarizes some properties of the three forms of DNA.

	A form	B form	Z form
Helical sense	Right handed	Right handed	Left handed
Diameter	~26 Å	~20 Å	~18 Å
Base pairs per helical turn	11	10.5	12
Helix rise per base pair	2.6 Å	3.4 Å	3.7 Å
Base tilt normal to the helix axis	20°	6°	7°
Sugar pucker conformation	C-3' endo	C-2' endo	C-2' endo for pyrimidines; C-3' endo for purines
Glycosyl bond conformation	Anti	Anti	Anti for pyrimidines; syn for purines

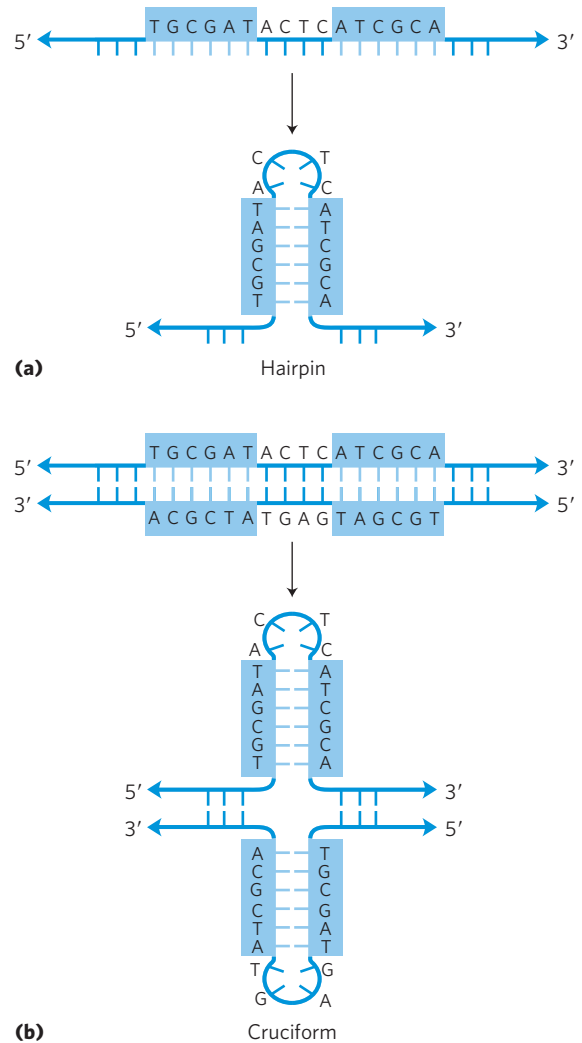




**FIGURE 8-18 Palindromes and mirror repeats.** Palindromes are sequences of double-stranded nucleic acids with twofold symmetry. In order to superimpose one repeat (shaded sequence) on the other, it must be rotated 180° about the horizontal axis then 180° about the vertical axis, as shown by the colored arrows. A mirror repeat, on the other hand, has a symmetric sequence within each strand. Superimposing one repeat on the other requires only a single 180° rotation about the vertical axis.

over two strands of DNA (**Fig. 8-18**). Such sequences are self-complementary within each strand and therefore have the potential to form **hairpin** or **cruciform** (cross-shaped) structures (**Fig. 8-19**). When the inverted repeat occurs within each individual strand of the DNA, the sequence is called a **mirror repeat**. Mirror repeats do not have complementary sequences within the same strand and cannot form hairpin or cruciform structures. Sequences of these types are found in virtually every large DNA molecule and can encompass a few base pairs or thousands. The extent to which palindromes occur as cruciforms in cells is not known, although some cruciform structures have been demonstrated in vivo in *Escherichia coli*. Self-complementary sequences cause isolated single strands of DNA (or RNA) in solution to fold into complex structures containing multiple hairpins.

Several unusual DNA structures involve three or even four DNA strands. Nucleotides participating in a Watson-Crick base pair (**Fig. 8-11**) can form additional hydrogen bonds, particularly with functional groups arrayed in the major groove. For example, a cytidine residue (if protonated) can pair with the guanosine residue of a G≡C nucleotide pair (**Fig. 8-20**); a thymidine can pair with the adenosine of an A=T pair. The N-7, O<sup>6</sup>, and N<sup>6</sup> of purines, the atoms that participate in the hydrogen bonding of triplex DNA, are often referred to as **Hoogsteen positions**, and the non-Watson-Crick pairing is called **Hoogsteen pairing**, after Karst Hoogsteen, who in 1963 first recognized the potential for these unusual pairings. Hoogsteen pairing allows the formation of **triplex DNAs**. The triplexes shown in **Figure 8-20** (a, b) are most stable at low pH because the C≡G·C<sup>+</sup> triplet requires a protonated cytosine. In the triplex, the pK<sub>a</sub> of this cytosine is >7.5, altered from its normal value of 4.2. The triplexes also form most readily within long sequences containing only pyrimidines or only purines in a given strand. Some triplex DNAs contain two pyrimidine

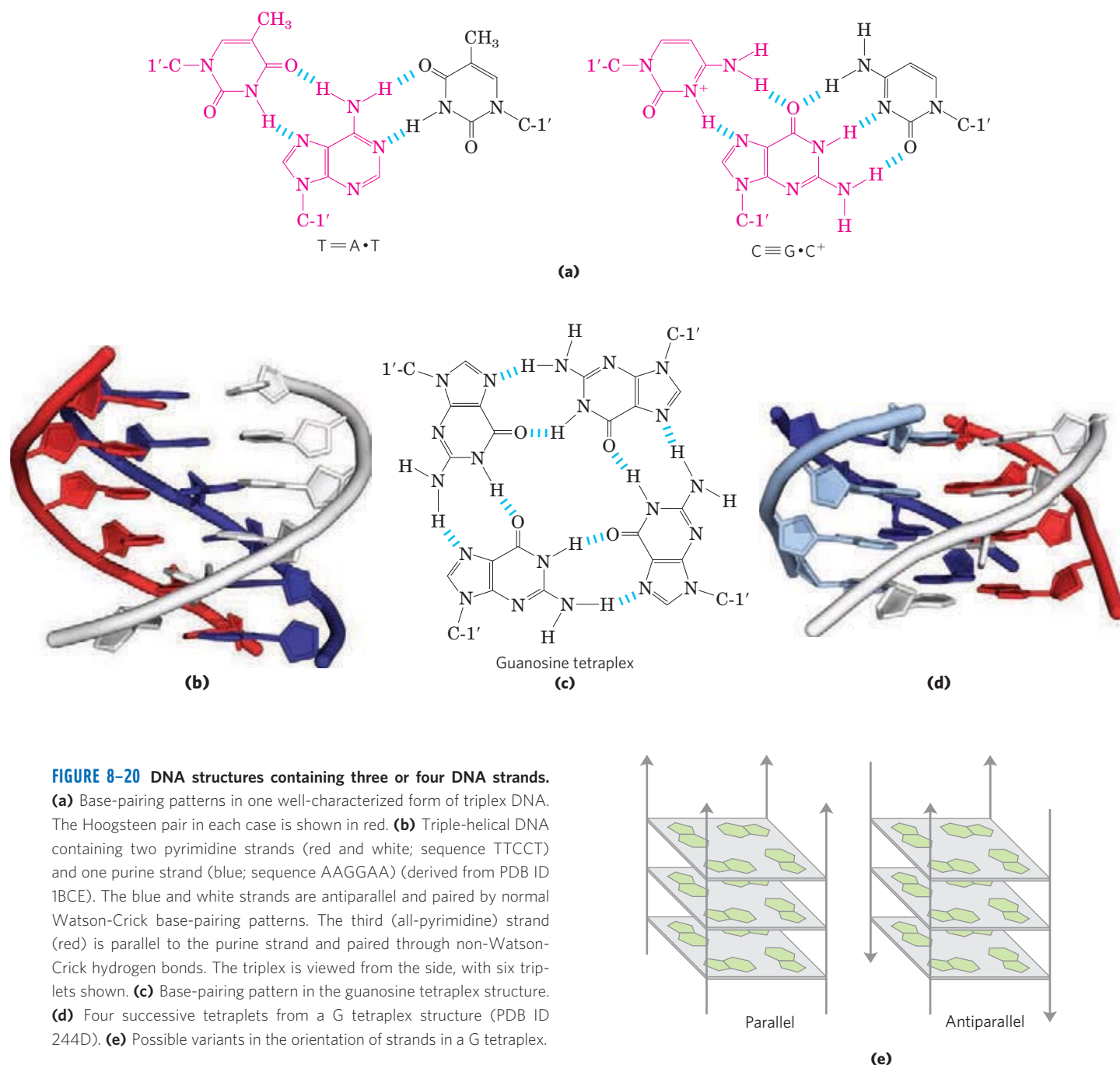


**FIGURE 8-19 Hairpins and cruciforms.** Palindromic DNA (or RNA) sequences can form alternative structures with intrastrand base pairing. **(a)** When only a single DNA (or RNA) strand is involved, the structure is called a hairpin. **(b)** When both strands of a duplex DNA are involved, it is called a cruciform. Blue shading highlights asymmetric sequences that can pair with the complementary sequence either in the same strand or in the complementary strand.

strands and one purine strand; others contain two purine strands and one pyrimidine strand.

Four DNA strands can also pair to form a tetraplex (quadruplex), but this occurs readily only for DNA sequences with a very high proportion of guanosine residues (**Fig. 8-20c, d**). The guanosine tetraplex, or **G tetraplex**, is quite stable over a wide range of conditions. The orientation of strands in the tetraplex can vary as shown in **Figure 8-20e**.

In the DNA of living cells, sites recognized by many sequence-specific DNA-binding proteins (**Chapter 28**) are arranged as palindromes, and polypyrimidine or polypurine sequences that can form triple helices are found within regions involved in the regulation of expression of some eukaryotic genes. In principle, synthetic DNA strands designed to pair with these sequences to form triplex DNA could disrupt gene



**FIGURE 8-20** DNA structures containing three or four DNA strands.

**(a)** Base-pairing patterns in one well-characterized form of triplex DNA. The Hoogsteen pair in each case is shown in red. **(b)** Triple-helical DNA containing two pyrimidine strands (red and white; sequence TTCCT) and one purine strand (blue; sequence AAGGAA) (derived from PDB ID 1BCE). The blue and white strands are antiparallel and paired by normal Watson-Crick base-pairing patterns. The third (all-pyrimidine) strand (red) is parallel to the purine strand and paired through non-Watson-Crick hydrogen bonds. The triplex is viewed from the side, with six triplets shown. **(c)** Base-pairing pattern in the guanosine tetraplex structure. **(d)** Four successive tetraplets from a G tetraplex structure (PDB ID 244D). **(e)** Possible variants in the orientation of strands in a G tetraplex.

expression. This approach to controlling cellular metabolism is of commercial interest for its potential application in medicine and agriculture.

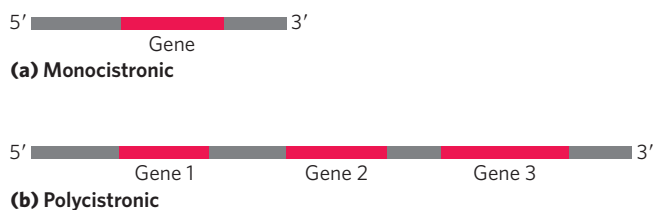
### Messenger RNAs Code for Polypeptide Chains

We now turn our attention to the expression of the genetic information that DNA contains. RNA, the second major form of nucleic acid in cells, has many functions. In gene expression, RNA acts as an intermediary by using the information encoded in DNA to specify the amino acid sequence of a functional protein.

Given that the DNA of eukaryotes is largely confined to the nucleus whereas protein synthesis occurs

on ribosomes in the cytoplasm, some molecule other than DNA must carry the genetic message from the nucleus to the cytoplasm. As early as the 1950s, RNA was considered the logical candidate: RNA is found in both the nucleus and the cytoplasm, and an increase in protein synthesis is accompanied by an increase in the amount of cytoplasmic RNA and an increase in its rate of turnover. These and other observations led several researchers to suggest that RNA carries genetic information from DNA to the protein biosynthetic machinery of the ribosome. In 1961 François Jacob and Jacques Monod presented a unified (and essentially correct) picture of many aspects of this process. They proposed





**FIGURE 8–21 Bacterial mRNA.** Schematic diagrams show **(a)** monocistronic and **(b)** polycistronic mRNAs of bacteria. Red segments represent RNA coding for a gene product; gray segments represent noncoding RNA. In the polycistronic transcript, noncoding RNA separates the three genes.

the name “messenger RNA” (mRNA) for that portion of the total cellular RNA carrying the genetic information from DNA to the ribosomes, where the messengers provide the templates that specify amino acid sequences in polypeptide chains. Although mRNAs from different genes can vary greatly in length, the mRNAs from a particular gene generally have a defined size. The process of forming mRNA on a DNA template is known as **transcription**.

In bacteria and archaea, a single mRNA molecule may code for one or several polypeptide chains. If it carries the code for only one polypeptide, the mRNA is **monocistronic**; if it codes for two or more different polypeptides, the mRNA is **polycistronic**. In eukaryotes, most mRNAs are monocistronic. (For the purposes of this discussion, “cistron” refers to a gene. The term itself has historical roots in the science of genetics, and its formal genetic definition is beyond the scope of this text.) The minimum length of an mRNA is set by the length of the polypeptide chain for which it codes. For example, a polypeptide chain of 100 amino acid residues requires an RNA coding sequence of at least 300 nucleotides, because each amino acid is coded by a nucleotide triplet (this and other details of protein synthesis are discussed in Chapter 27). However, mRNAs transcribed from DNA are always somewhat longer than the length needed simply to code for a polypeptide sequence (or sequences). The additional, noncoding RNA includes sequences that regulate protein synthesis. **Figure 8–21** summarizes the general structure of bacterial mRNAs.

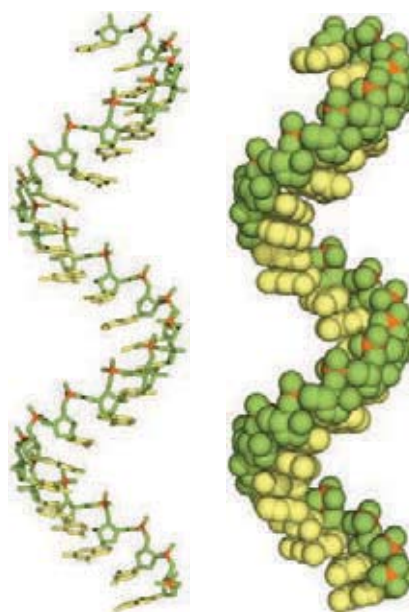
### Many RNAs Have More Complex Three-Dimensional Structures

Messenger RNA is only one of several classes of cellular RNA. Transfer RNAs are adapter molecules in protein synthesis; covalently linked to an amino acid at one end, they pair with the mRNA in such a way that amino acids are joined to a growing polypeptide in the correct sequence. Ribosomal RNAs are components of ribosomes. There is also a wide variety of special-function RNAs, including some (called ribozymes) that have enzymatic activity. All the RNAs are considered in detail

in Chapter 26. The diverse and often complex functions of these RNAs reflect a diversity of structure much richer than that observed in DNA molecules.

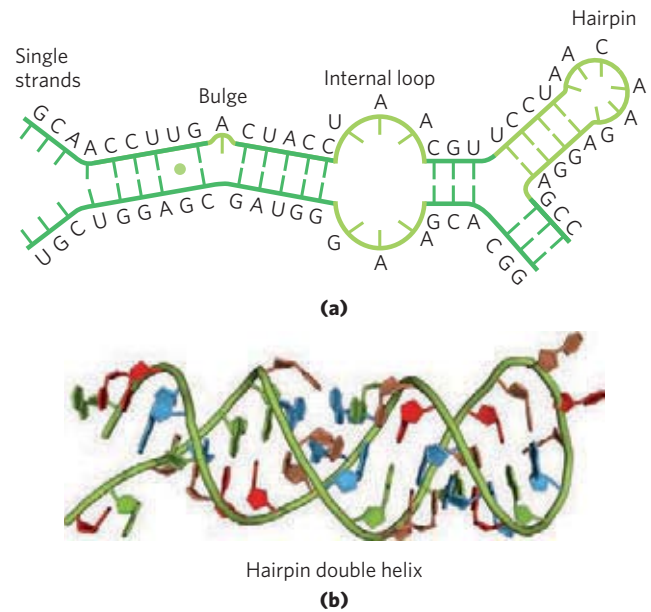
The product of transcription of DNA is always single-stranded RNA. The single strand tends to assume a right-handed helical conformation dominated by base-stacking interactions (**Fig. 8–22**), which are stronger between two purines than between a purine and pyrimidine or between two pyrimidines. The purine-purine interaction is so strong that a pyrimidine separating two purines is often displaced from the stacking pattern so that the purines can interact. Any self-complementary sequences in the molecule produce more complex structures. RNA can base-pair with complementary regions of either RNA or DNA. Base pairing matches the pattern for DNA: G pairs with C and A pairs with U (or with the occasional T residue in some RNAs). One difference is that base pairing between G and U residues—unusual in DNA—is allowed in RNA (see Fig. 8–24) when complementary sequences in two single strands of RNA pair with each other. The paired strands in RNA or RNA-DNA duplexes are antiparallel, as in DNA.

When two strands of RNA with perfectly complementary sequences are paired, the predominant double-stranded structure is an A-form right-handed double helix. However, strands of RNA that are perfectly paired over long regions of sequence are uncommon. The three-dimensional structures of many RNAs, like those of proteins, are complex and unique. Weak



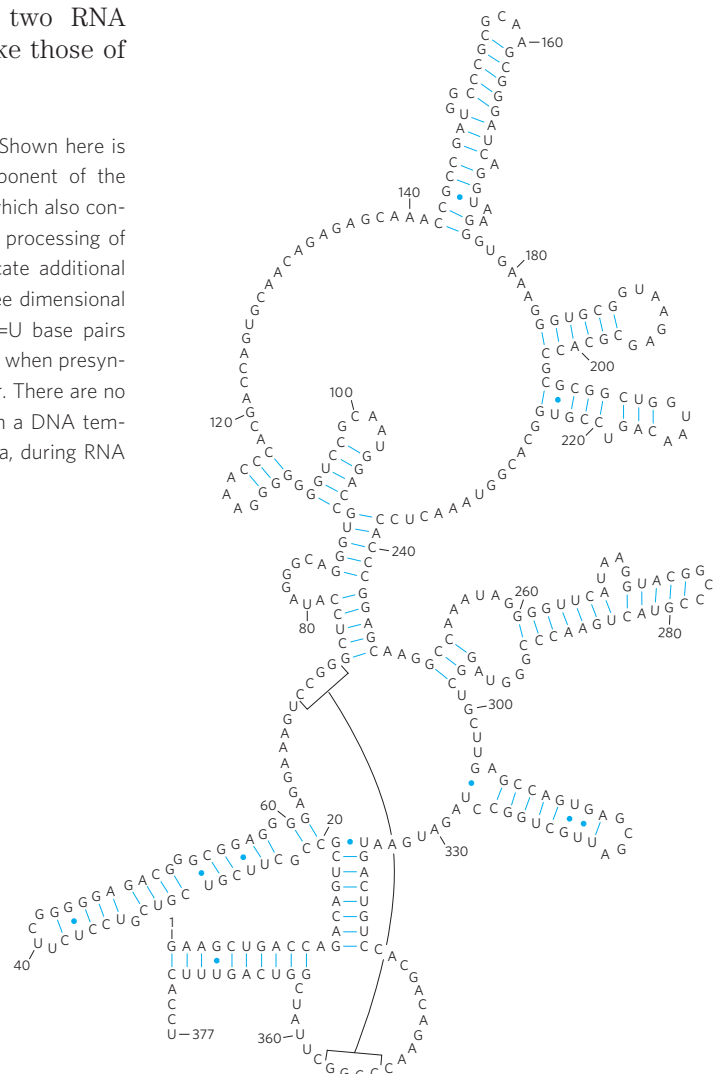
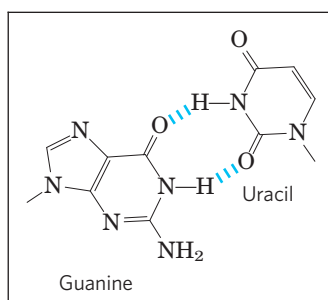
**FIGURE 8–22 Typical right-handed stacking pattern of single-stranded RNA.** The bases are shown in yellow, the phosphorus atoms in orange, and the riboses and phosphate oxygens in green. Green is used to represent RNA strands in succeeding chapters, just as blue is used for DNA.

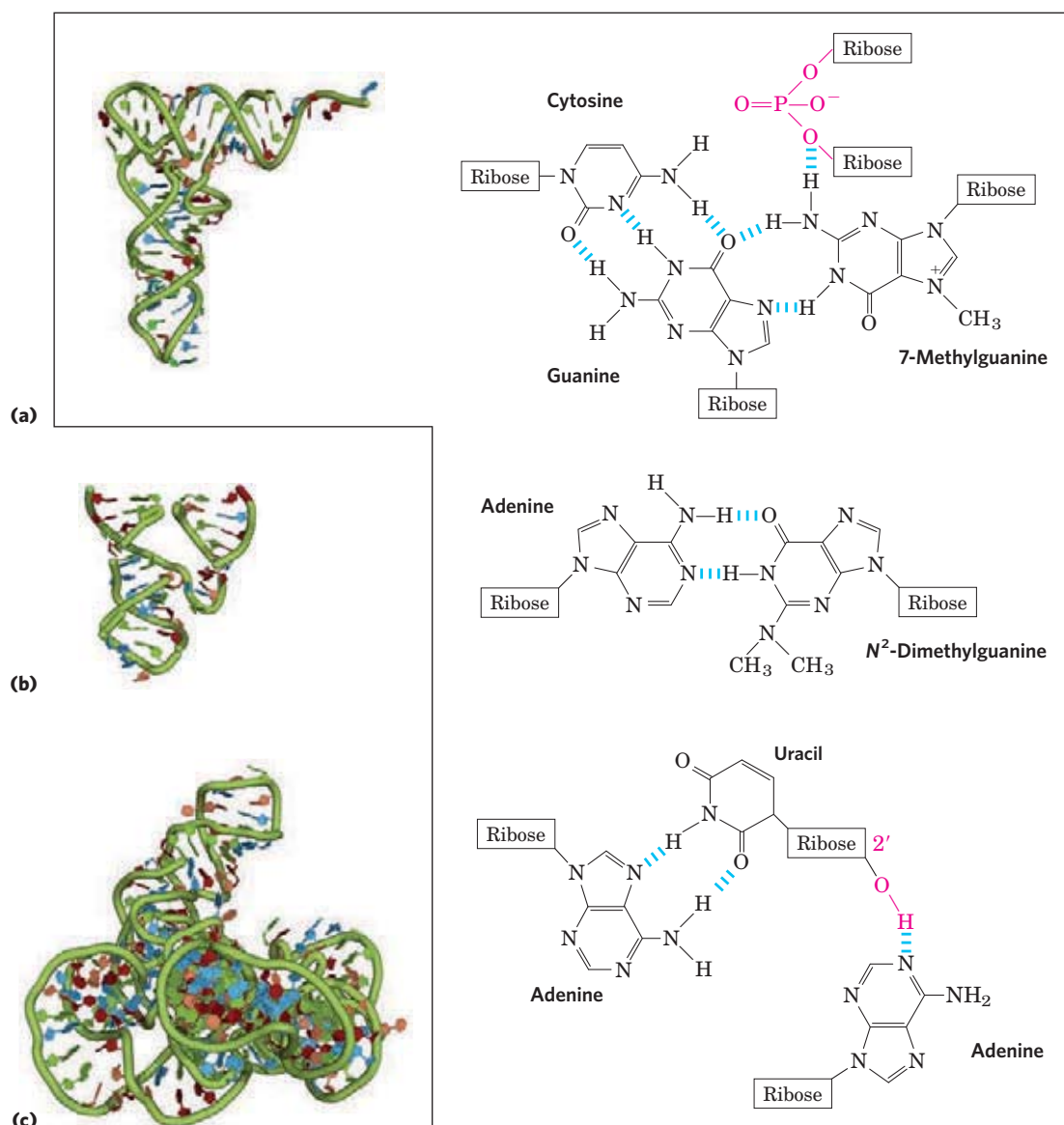
interactions, especially base-stacking interactions, help stabilize RNA structures, just as they do in DNA. Z-form helices have been made in the laboratory (under very high-salt or high-temperature conditions). The B form of RNA has not been observed. Breaks in the regular A-form helix caused by mismatched or unmatched bases in one or both strands are common and result in bulges or internal loops (Fig. 8-23). Hairpin loops form between nearby self-complementary (palindromic) sequences. The potential for base-paired helical segments in many RNAs is extensive (Fig. 8-24), and the resulting hairpins are the most common type of secondary structure in RNA. Specific short base sequences (such as UUCG) are often found at the ends of RNA hairpins and are known to form particularly tight and stable loops. Such sequences may act as starting points for the folding of an RNA molecule into its precise three-dimensional structure. Other contributions are made by hydrogen bonds that are not part of standard Watson-Crick base pairs. For example, the 2'-hydroxyl group of ribose can hydrogen-bond with other groups. Some of these properties are evident in the structure of the phenylalanine transfer RNA of yeast—the tRNA responsible for inserting Phe residues into polypeptides—and in two RNA enzymes, or ribozymes, whose functions, like those of



**FIGURE 8-23** Secondary structure of RNAs. (a) Bulge, internal loop, and hairpin loop. (b) The paired regions generally have an A-form right-handed helix, as shown for a hairpin (derived from PDB ID 1GID).

**FIGURE 8-24** Base-paired helical structures in an RNA. Shown here is the possible secondary structure of the M1 RNA component of the enzyme RNase P of *E. coli*, with many hairpins. RNase P, which also contains a protein component (not shown), functions in the processing of transfer RNAs (see Fig. 26-26). The two brackets indicate additional complementary sequences that may be paired in the three dimensional structure. The blue dots indicate non-Watson-Crick G=U base pairs (boxed inset). Note that G=U base pairs are allowed only when presynthesized strands of RNA fold up or anneal with each other. There are no RNA polymerases (the enzymes that synthesize RNAs on a DNA template) that insert a U opposite a template G, or vice versa, during RNA synthesis.





**FIGURE 8-25 Three-dimensional structure in RNA.** (a) Three-dimensional structure of phenylalanine tRNA of yeast (PDB ID 1TRA). Some unusual base-pairing patterns found in this tRNA are shown. Note also the involvement of the oxygen of a ribose phosphodiester bond in one hydrogen-bonding arrangement, and a ribose 2'-hydroxyl group in another (both in red). (b) A hammerhead ribozyme (so named because the secondary structure at the active site looks like the head of a hammer), derived from certain plant viruses (derived from PDB ID 1MME).

Ribozymes, or RNA enzymes, catalyze a variety of reactions, primarily in RNA metabolism and protein synthesis. The complex three-dimensional structures of these RNAs reflect the complexity inherent in catalysis, as described for protein enzymes in Chapter 6. (c) A segment of mRNA known as an intron, from the ciliated protozoan *Tetrahymena thermophila* (derived from PDB ID 1GRZ). This intron (a ribozyme) catalyzes its own excision from between exons in an mRNA strand (discussed in Chapter 26).

protein enzymes, depend on their three-dimensional structures (**Fig. 8-25**).

The analysis of RNA structure and the relationship between its structure and its function is an emerging field of inquiry that has many of the same complexities as the analysis of protein structure. The importance of understanding RNA structure grows as we become increasingly aware of the large number of functional roles for RNA molecules.

## SUMMARY 8.2 Nucleic Acid Structure

- ▶ Many lines of evidence show that DNA bears genetic information. Some of the earliest evidence came from the Avery-MacLeod-McCarty experiment, which showed that DNA isolated from one bacterial strain can enter and transform the cells of another strain, endowing it with some of the inheritable characteristics of the donor. The

Hershey-Chase experiment showed that the DNA of a bacterial virus, but not its protein coat, carries the genetic message for replication of the virus in a host cell.

- ▶ Putting together the available data, Watson and Crick postulated that native DNA consists of two antiparallel chains in a right-handed double-helical arrangement. Complementary base pairs, A=T and G=C, are formed by hydrogen bonding within the helix. The base pairs are stacked perpendicular to the long axis of the double helix, 3.4 Å apart, with 10.5 base pairs per turn.
- ▶ DNA can exist in several structural forms. Two variations of the Watson-Crick form, or B-DNA, are A- and Z-DNA. Some sequence-dependent structural variations cause bends in the DNA molecule. DNA strands with appropriate sequences can form hairpin or cruciform structures or triplex or tetraplex DNA.
- ▶ Messenger RNA transfers genetic information from DNA to ribosomes for protein synthesis. Transfer RNA and ribosomal RNA are also involved in protein synthesis. RNA can be structurally complex; single RNA strands can fold into hairpins, double-stranded regions, or complex loops.

### 8.3 Nucleic Acid Chemistry

The role of DNA as a repository of genetic information depends in part on its inherent stability. The chemical transformations that do occur are generally very slow in the absence of an enzyme catalyst. The long-term storage of information without alteration is so important to a cell, however, that even very slow reactions that alter DNA structure can be physiologically significant. Processes such as carcinogenesis and aging may be intimately linked to slowly accumulating, irreversible alterations of DNA. Other, nondestructive alterations also occur and are essential to function, such as the strand separation that must precede DNA replication or transcription. In addition to providing insights into physiological processes, our understanding of nucleic acid chemistry has given us a powerful array of technologies that have applications in molecular biology, medicine, and forensic science. We now examine the chemical properties of DNA and some of these technologies.

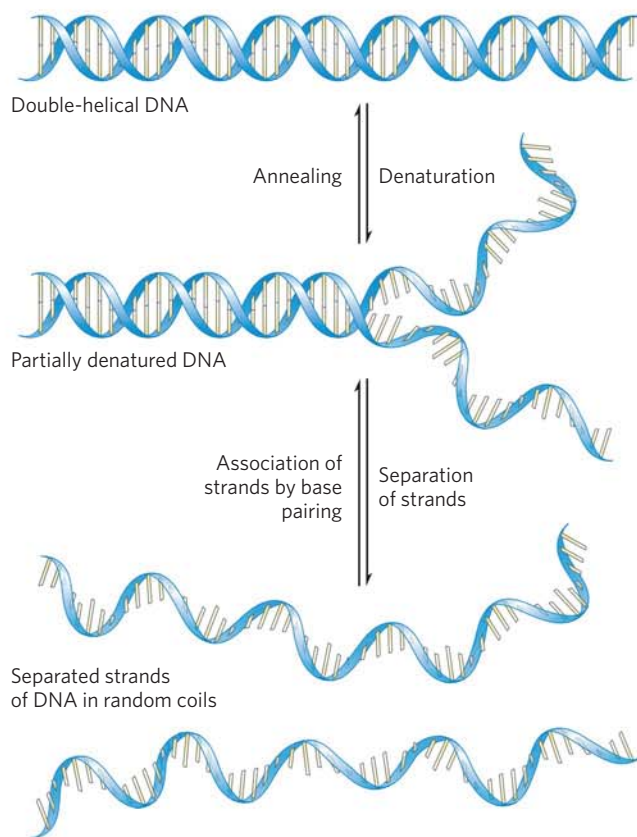
#### Double-Helical DNA and RNA Can Be Denatured

Solutions of carefully isolated, native DNA are highly viscous at pH 7.0 and room temperature (25 °C). When such a solution is subjected to extremes of pH or to temperatures above 80 °C, its viscosity decreases sharply, indicating that the DNA has undergone a physical change. Just as heat and extremes of pH denature globular proteins, they also cause denaturation, or melting, of double-helical DNA. Disruption of the hydrogen

bonds between paired bases and of base stacking causes unwinding of the double helix to form two single strands, completely separate from each other along the entire length or part of the length (partial denaturation) of the molecule. No covalent bonds in the DNA are broken (**Fig. 8-26**).

Renaturation of a DNA molecule is a rapid one-step process, as long as a double-helical segment of a dozen or more residues still unites the two strands. When the temperature or pH is returned to the range in which most organisms live, the unwound segments of the two strands spontaneously rewind, or **anneal**, to yield the intact duplex (**Fig. 8-26**). However, if the two strands are completely separated, renaturation occurs in two steps. In the first, relatively slow step, the two strands “find” each other by random collisions and form a short segment of complementary double helix. The second step is much faster: the remaining unpaired bases successively come into register as base pairs, and the two strands “zipper” themselves together to form the double helix.

The close interaction between stacked bases in a nucleic acid has the effect of decreasing its absorption of UV light relative to that of a solution with the same concentration of free nucleotides, and the absorption is decreased further when two complementary nucleic acid strands are paired. This is called the hypochromic

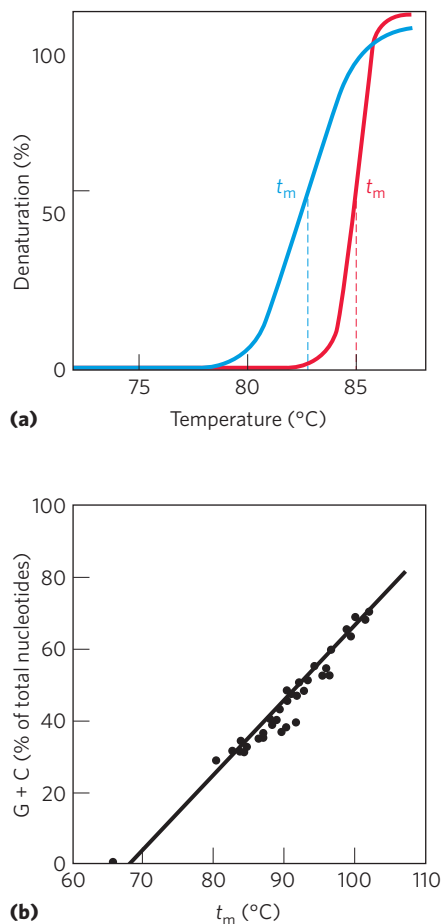


**FIGURE 8-26** Reversible denaturation and annealing (renaturation) of DNA.

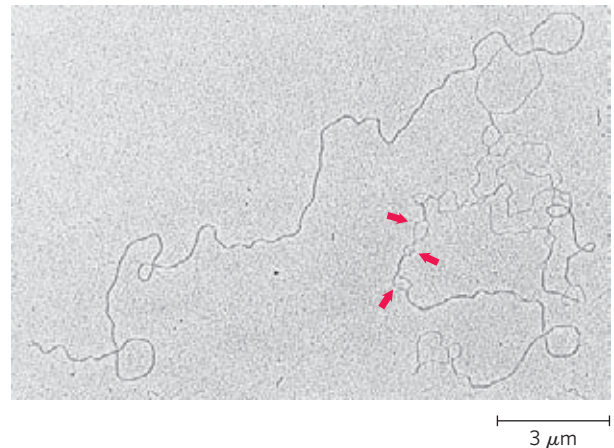


effect. Denaturation of a double-stranded nucleic acid produces the opposite result: an increase in absorption called the hyperchromic effect. The transition from double-stranded DNA to the single-stranded, denatured form can thus be detected by monitoring UV absorption at 260 nm.

Viral or bacterial DNA molecules in solution denature when they are heated slowly (**Fig. 8-27**). Each species of DNA has a characteristic denaturation temperature, or melting point ( $t_m$ ; formally, the temperature at which half the DNA is present as separated single strands): the higher its content of G≡C base pairs, the higher the melting point of the DNA. This is because G≡C base pairs, with three hydrogen bonds, require more heat energy to dissociate than A=T base pairs. Thus the melting point of a DNA molecule, determined under fixed conditions of pH and ionic strength, can yield an estimate of its base composition. If denaturation conditions are carefully controlled, regions that are rich in A=T base pairs will specifically denature while most of the DNA remains double-stranded. Such



**FIGURE 8-27 Heat denaturation of DNA.** (a) The denaturation, or melting, curves of two DNA specimens. The temperature at the mid-point of the transition ( $t_m$ ) is the melting point; it depends on pH and ionic strength and on the size and base composition of the DNA. (b) Relationship between  $t_m$  and the G+C content of a DNA.



**FIGURE 8-28 Partially denatured DNA.** This DNA was partially denatured, then fixed to prevent renaturation during sample preparation. The shadowing method used to visualize the DNA in this electron micrograph increases its diameter approximately fivefold and obliterates most details of the helix. However, length measurements can be obtained, and single-stranded regions are readily distinguishable from double-stranded regions. The arrows point to some single-stranded bubbles where denaturation has occurred. The regions that denature are highly reproducible and are rich in A=T base pairs.

denatured regions (called bubbles) can be visualized with electron microscopy (**Fig. 8-28**). Note that in the strand separation of DNA that occurs *in vivo* during processes such as DNA replication and transcription, the sites where these processes are initiated are often rich in A=T base pairs, as we shall see.

Duplexes of two RNA strands or one RNA strand and one DNA strand (RNA-DNA hybrids) can also be denatured. Notably, RNA duplexes are more stable to heat denaturation than DNA duplexes. At neutral pH, denaturation of a double-helical RNA often requires temperatures 20°C or more higher than those required for denaturation of a DNA molecule with a comparable sequence, assuming the strands in each molecule are perfectly complementary. The stability of an RNA-DNA hybrid is generally intermediate between that of RNA and DNA duplexes. The physical basis for these differences in thermal stability is not known.

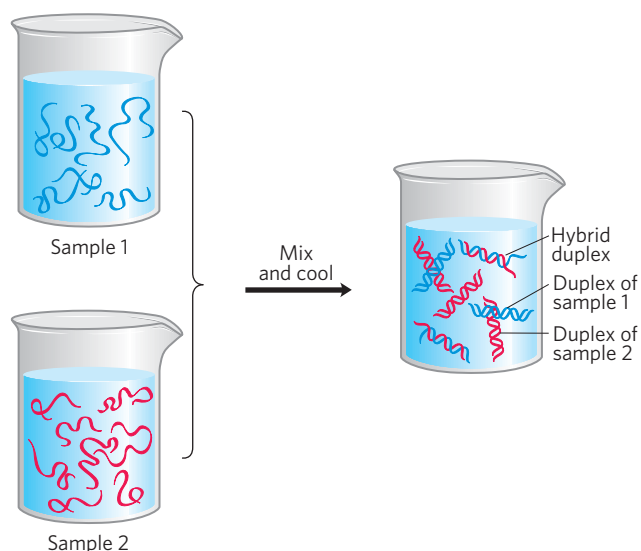
### Nucleic Acids from Different Species Can Form Hybrids

The ability of two complementary DNA strands to pair with one another can be used to detect similar DNA sequences in two different species or within the genome of a single species. If duplex DNAs isolated from human cells and from mouse cells are completely denatured by heating, then mixed and kept at about 25°C below their  $t_m$  for many hours, much of the DNA will anneal. The rate of DNA annealing is affected by temperature, the length and concentration of the DNA fragments being annealed, the concentration of salts in the reaction



mixture, and properties of the sequence itself (e.g., complexity and  $G \equiv C$  content). Temperature is especially important. If the temperature is too low, short sequences with coincidental similarity from distant, heterologous parts of the DNA molecules will anneal unproductively and interfere with the more general alignment of complementary DNA strands. Temperatures that are too high will favor denaturation. Most of the reannealing occurs between complementary mouse DNA strands to form mouse duplex DNA; similarly, most human DNA strands anneal with complementary human DNA strands. However, some strands of the mouse DNA will associate with human DNA strands to yield **hybrid duplexes**, in which segments of a mouse DNA strand form base-paired regions with segments of a human DNA strand (**Fig. 8–29**). This reflects a common evolutionary heritage; different organisms generally have many proteins and RNAs with similar functions and, often, similar structures. In many cases, the DNAs encoding these proteins and RNAs have similar sequences. The closer the evolutionary relationship between two species, the more extensively their DNAs will hybridize. For example, human DNA hybridizes much more extensively with mouse DNA than with DNA from yeast.

The hybridization of DNA strands from different sources forms the basis for a powerful set of techniques



**FIGURE 8–29 DNA hybridization.** Two DNA samples to be compared are completely denatured by heating. When the two solutions are mixed and slowly cooled, DNA strands of each sample associate with their normal complementary partner and anneal to form duplexes. If the two DNAs have significant sequence similarity, they also tend to form partial duplexes or hybrids with each other: the greater the sequence similarity between the two DNAs, the greater the number of hybrids formed. Hybrid formation can be measured in several ways. One of the DNAs is usually labeled with a radioactive isotope to simplify their detection and measurement.

essential to the practice of modern molecular genetics. A specific DNA sequence or gene can be detected in the presence of many other sequences if one already has an appropriate complementary DNA strand (usually labeled in some way) to hybridize with it (Chapter 9). The complementary DNA can be from a different species or from the same species, or it can be synthesized chemically in the laboratory using techniques described later in this chapter. Hybridization techniques can be varied to detect a specific RNA rather than DNA. The isolation and identification of specific genes and RNAs rely on these hybridization techniques. Applications of this technology make possible the identification of an individual on the basis of a single hair left at the scene of a crime or the prediction of the onset of a disease decades before symptoms appear (see Box 9–1).

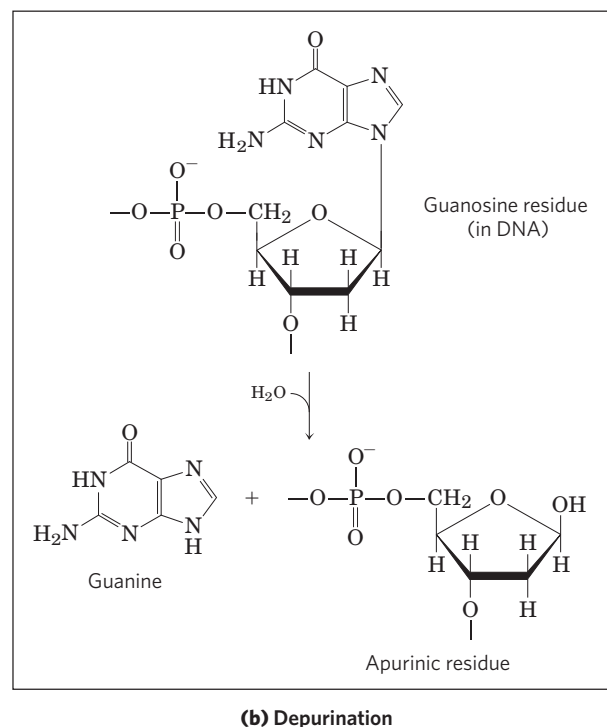
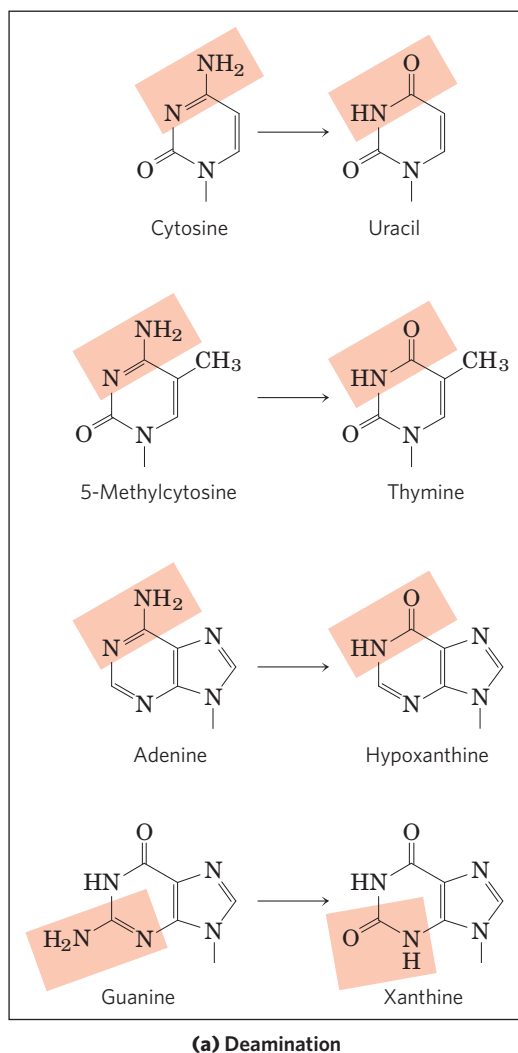
### Nucleotides and Nucleic Acids Undergo Nonenzymatic Transformations



Purines and pyrimidines, along with the nucleotides of which they are a part, undergo spontaneous alterations in their covalent structure. The rate of these reactions is generally *very slow*, but they are physiologically significant because of the cell's very low tolerance for alterations in its genetic information. Alterations in DNA structure that produce permanent changes in the genetic information encoded therein are called **mutations**, and much evidence suggests an intimate link between the accumulation of mutations in an individual organism and the process of aging and carcinogenesis.

Several nucleotide bases undergo spontaneous loss of their exocyclic amino groups (deamination) (**Fig. 8–30a**). For example, under typical cellular conditions, deamination of cytosine (in DNA) to uracil occurs in about one of every  $10^7$  cytidine residues in 24 hours. This corresponds to about 100 spontaneous events per day, on average, in a mammalian cell. Deamination of adenine and guanine occurs at about 1/100th this rate.

The slow cytosine deamination reaction seems innocuous enough, but is almost certainly the reason why DNA contains thymine rather than uracil. The product of cytosine deamination (uracil) is readily recognized as foreign in DNA and is removed by a repair system (Chapter 25). If DNA normally contained uracil, recognition of uracils resulting from cytosine deamination would be more difficult, and unrepaired uracils would lead to permanent sequence changes as they were paired with adenines during replication. Cytosine deamination would gradually lead to a decrease in  $G \equiv C$  base pairs and an increase in  $A = U$  base pairs in the DNA of all cells. Over the millennia, cytosine deamination could eliminate  $G \equiv C$  base pairs and the genetic code that depends on them. Establishing thymine as one of the four bases in DNA may well



**FIGURE 8-30** Some well-characterized nonenzymatic reactions of nucleotides. **(a)** Deamination reactions. Only the base is shown. **(b)** Depurination, in which a purine is lost by hydrolysis of the *N*- $\beta$ -glycosyl bond. Loss of pyrimidines via a similar reaction occurs, but much more slowly. The resulting lesion, in which the deoxyribose is present but the base is not, is called an abasic site or an AP site (apurinic site or, rarely, apyrimidinic site). The deoxyribose remaining after depurination is readily converted from the  $\beta$ -furanose to the aldehyde form (see Fig. 8-3), further destabilizing the DNA at this position. More nonenzymatic reactions are illustrated in Figures 8-31 and 8-32.

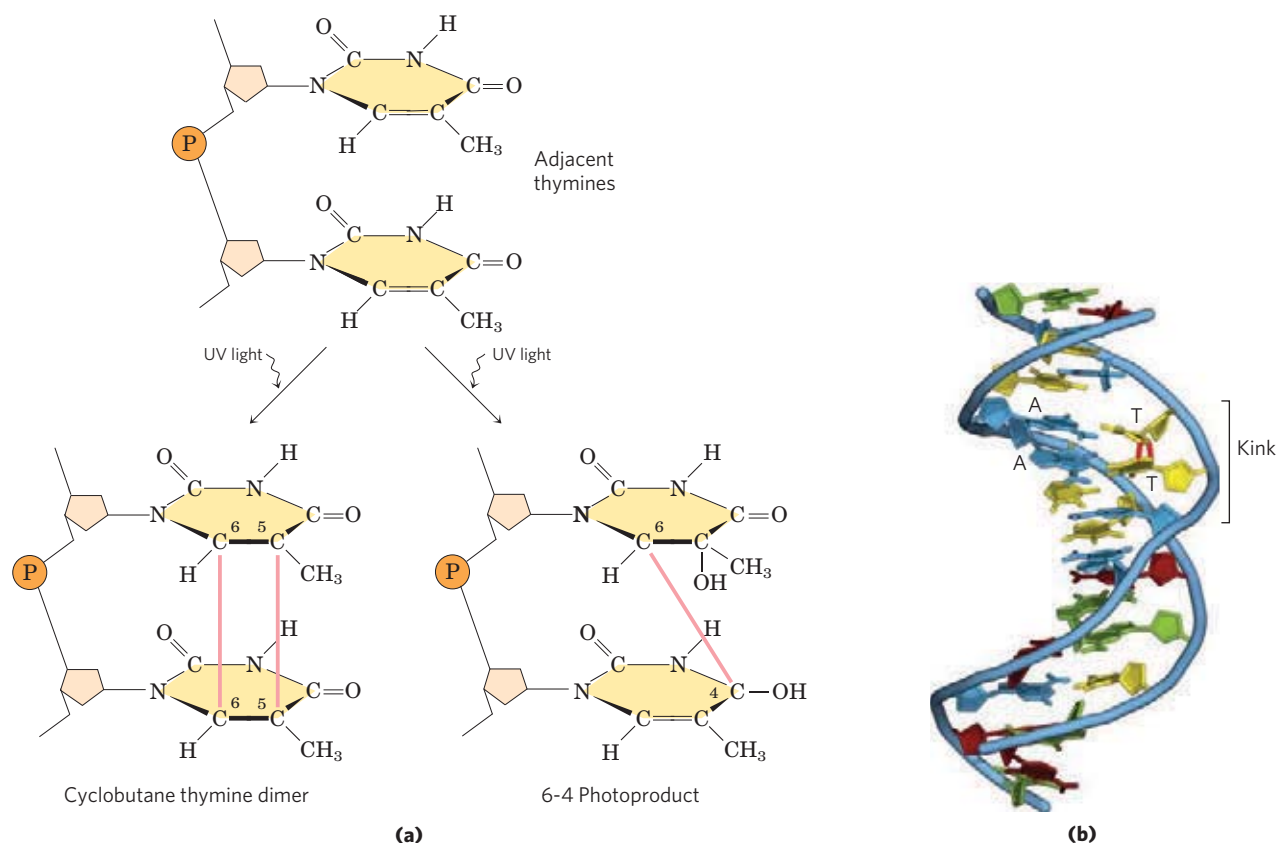
have been one of the crucial turning points in evolution, making the long-term storage of genetic information possible.

Another important reaction in deoxyribonucleotides is the hydrolysis of the *N*- $\beta$ -glycosyl bond between the base and the pentose, to create a DNA lesion called an AP (apurinic, apyrimidinic) site or abasic site (Fig. 8-30b). This occurs at a higher rate for purines than for pyrimidines. As many as one in  $10^5$  purines (10,000 per mammalian cell) are lost from DNA every 24 hours under typical cellular conditions. Depurination of ribonucleotides and RNA is much slower and generally is not considered physiologically significant. In the test tube, loss of purines can be accelerated by dilute acid. Incubation of DNA at pH 3 causes selective removal of the purine bases, resulting in a derivative called apurinic acid.

Other reactions are promoted by radiation. UV light induces the condensation of two ethylene groups to form a cyclobutane ring. In the cell, the same reaction between adjacent pyrimidine bases in nucleic acids forms cyclobutane pyrimidine dimers. This happens most frequently between adjacent thymidine residues

on the same DNA strand (**Fig. 8-31**). A second type of pyrimidine dimer, called a 6-4 photoproduct, is also formed during UV irradiation. Ionizing radiation (x rays and gamma rays) can cause ring opening and fragmentation of bases as well as breaks in the covalent backbone of nucleic acids.

Virtually all forms of life are exposed to energy-rich radiation capable of causing chemical changes in DNA. Near-UV radiation (with wavelengths of 200 to 400 nm), which makes up a significant portion of the solar spectrum, is known to cause pyrimidine dimer formation and other chemical changes in the DNA of bacteria and of human skin cells. We are subject to a constant field of ionizing radiation in the form of cosmic rays, which can penetrate deep into the earth, as well as radiation emitted from radioactive elements, such as radium, plutonium, uranium, radon,  $^{14}\text{C}$ , and  $^3\text{H}$ . X rays used in medical and dental examinations and in radiation therapy of cancer and other diseases are another form of ionizing radiation. It is estimated that UV and ionizing radiations are responsible for about 10% of all DNA damage caused by environmental agents.



**FIGURE 8-31** Formation of pyrimidine dimers induced by UV light.

(a) One type of reaction (on the left) results in the formation of a cyclobutyl ring involving C-5 and C-6 of adjacent pyrimidine residues. An alternative reaction (on the right) results in a 6-4 photoproduct, with a

linkage between C-6 of one pyrimidine and C-4 of its neighbor.

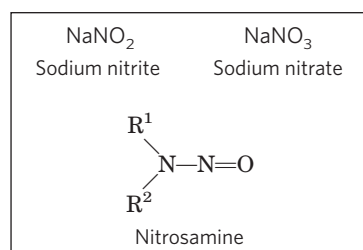
(b) Formation of a cyclobutane pyrimidine dimer introduces a bend or kink into the DNA (PDB ID 1TTD).

DNA also may be damaged by reactive chemicals introduced into the environment as products of industrial activity. Such products may not be injurious per se but may be metabolized by cells into forms that are. There are two prominent classes of such agents (**Fig. 8-32**): (1) deaminating agents, particularly nitrous acid ( $\text{HNO}_2$ ) or compounds that can be metabolized to nitrous acid or nitrites, and (2) alkylating agents.

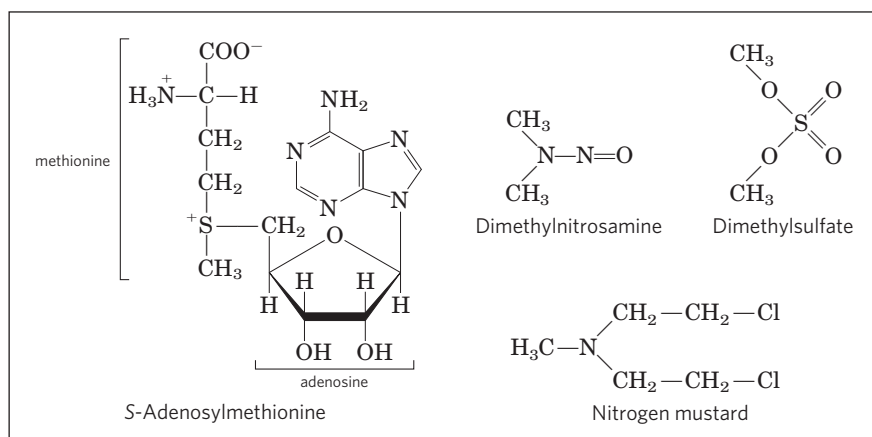
Nitrous acid, formed from organic precursors such as nitrosamines and from nitrite and nitrate

salts, is a potent accelerator of the deamination of bases. Bisulfite has similar effects. Both agents are used as preservatives in processed foods to prevent the growth of toxic bacteria. They do not seem to increase cancer risks significantly when used in this way, perhaps because they are used in small amounts and make only a minor contribution to the overall levels of DNA damage. (The potential health risk from food spoilage if these preservatives were not used is much greater.)

**FIGURE 8-32** Chemical agents that cause DNA damage. (a) Precursors of nitrous acid, which promotes deamination reactions. (b) Alkylating agents. Only S-adenosylmethionine acts enzymatically.

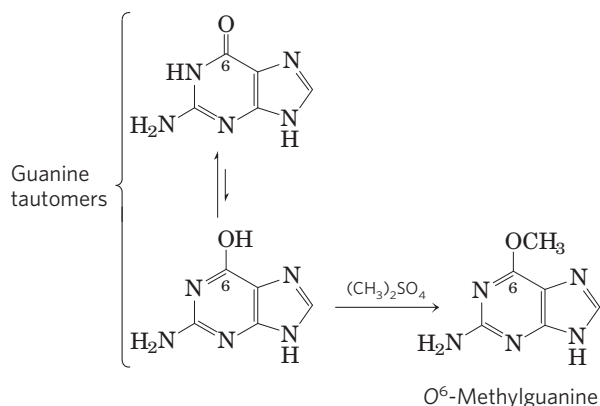


(a) Nitrous acid precursors



(b) Alkylating agents

Alkylating agents can alter certain bases of DNA. For example, the highly reactive chemical dimethylsulfate (Fig. 8–32b) can methylate a guanine to yield  $O^6$ -methylguanine, which cannot base-pair with cytosine.



Many similar reactions are brought about by alkylating agents normally present in cells, such as *S*-adenosyl methionine.

The most important source of mutagenic alterations in DNA is oxidative damage. Excited-oxygen species such as hydrogen peroxide, hydroxyl radicals, and superoxide radicals arise during irradiation or as a byproduct of aerobic metabolism. Of these species, the hydroxyl radicals are responsible for most oxidative DNA damage. Cells have an elaborate defense system to destroy reactive oxygen species, including enzymes such as catalase and superoxide dismutase that convert reactive oxygen species to harmless products. A fraction of these oxidants inevitably escape cellular defenses, however, and damage to DNA occurs through any of a large, complex group of reactions ranging from oxidation of deoxyribose and base moieties to strand breaks. Accurate estimates for the extent of this damage are not yet available, but every day the DNA of each human cell is subjected to thousands of damaging oxidative reactions.

This is merely a sampling of the best-understood reactions that damage DNA. Many carcinogenic compounds in food, water, or air exert their cancer-causing effects by modifying bases in DNA. Nevertheless, the integrity of DNA as a polymer is better maintained than that of either RNA or protein, because DNA is the only macromolecule that has the benefit of extensive biochemical repair systems. These repair processes (described in Chapter 25) greatly lessen the impact of damage to DNA. ■

### Some Bases of DNA Are Methylated

Certain nucleotide bases in DNA molecules are enzymatically methylated. Adenine and cytosine are methylated more often than guanine and thymine. Methylation is generally confined to certain sequences or regions of a DNA molecule. In some cases the function of methylation is well understood; in others the function remains unclear. All known DNA methylases use *S*-adenosylmethionine as a methyl group donor (Fig. 8–32b). *E. coli*

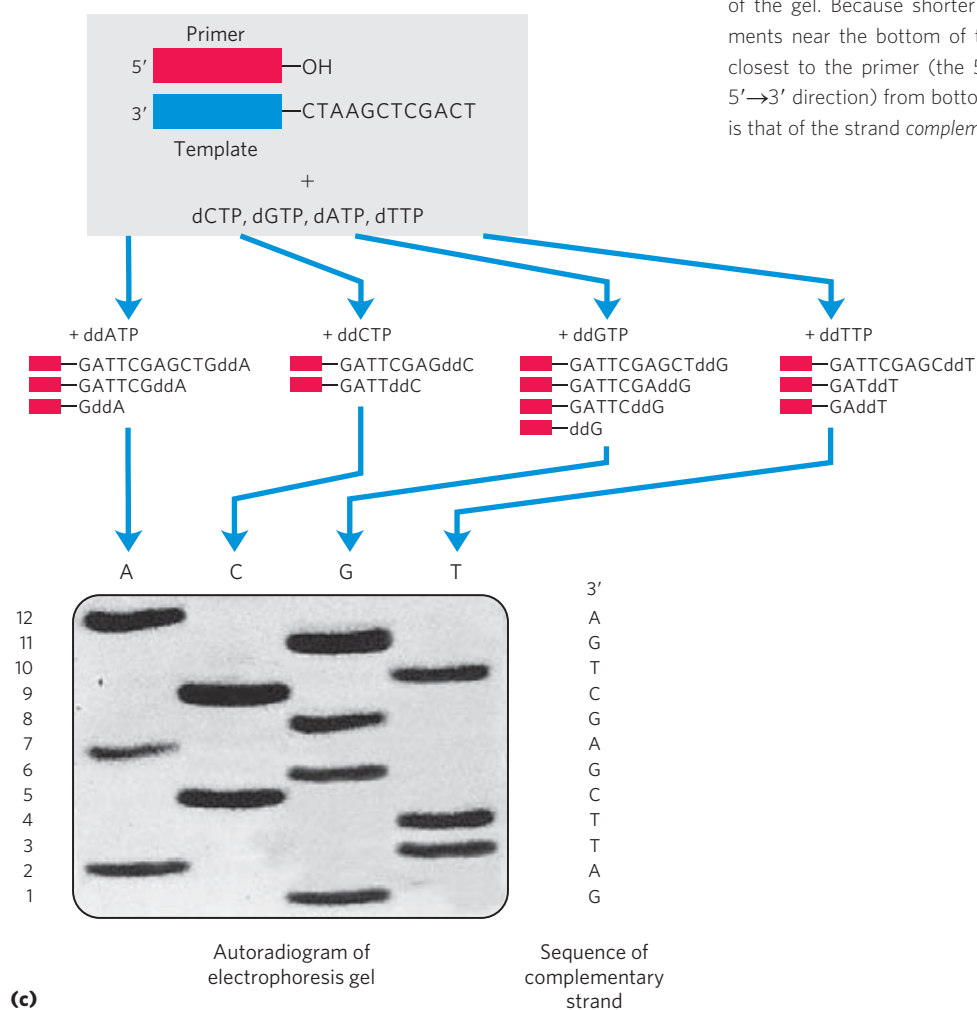
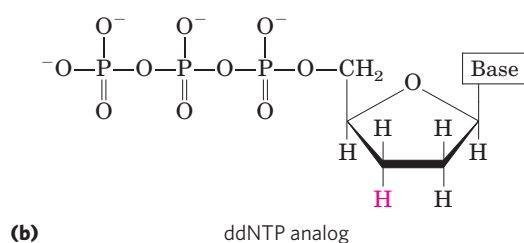
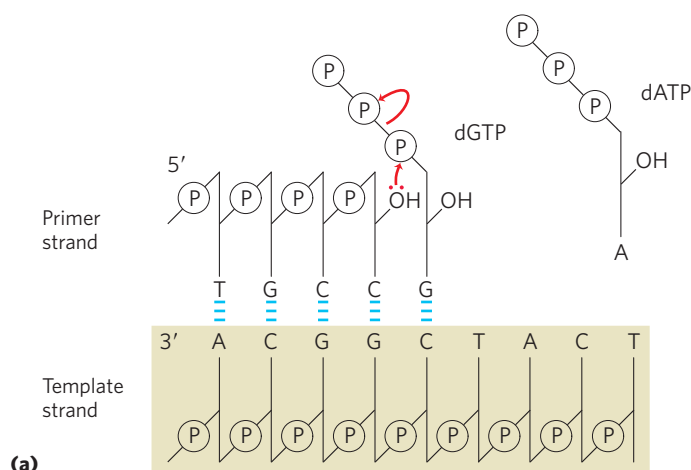
has two prominent methylation systems. One serves as part of a defense mechanism that helps the cell to distinguish its DNA from foreign DNA by marking its own DNA with methyl groups and destroying (foreign) DNA without the methyl groups (this is known as a restriction-modification system; see p. 314). The other system methylates adenosine residues within the sequence (5')GATC(3') to *N*<sup>6</sup>-methyladenosine (Fig. 8–5a). This is mediated by the Dam (*DNA adenine methylation*) methylase, a component of a system that repairs mismatched base pairs formed occasionally during DNA replication (see Fig. 25–21).

In eukaryotic cells, about 5% of cytidine residues in DNA are methylated to 5-methylcytidine (Fig. 8–5a). Methylation is most common at CpG sequences, producing methyl-CpG symmetrically on both strands of the DNA. The extent of methylation of CpG sequences varies by molecular region in large eukaryotic DNA molecules.

### The Sequences of Long DNA Strands Can Be Determined

In its capacity as a repository of information, a DNA molecule's most important property is its nucleotide sequence. Until the late 1970s, determining the sequence of a nucleic acid containing even five or ten nucleotides was very laborious. The development of two new techniques in 1977, one by Alan Maxam and Walter Gilbert and the other by Frederick Sanger, made possible the sequencing of larger DNA molecules with an ease unimagined just a few years before. The techniques depend on an improved understanding of nucleotide chemistry and DNA metabolism, and on electrophoretic methods for separating DNA strands differing in size by only one nucleotide. Electrophoresis of DNA is similar to that of proteins (see Fig. 3–18). Polyacrylamide is often used as the gel matrix in work with short DNA molecules (up to a few hundred nucleotides); agarose is generally used for longer pieces of DNA.

In both Sanger and Maxam-Gilbert sequencing, the general principle is to reduce the DNA to four sets of labeled fragments. The reaction producing each set is base-specific, so the lengths of the fragments correspond to positions in the DNA sequence where a certain base occurs. For example, for an oligonucleotide with the sequence pAATCGACT, labeled at the 5' end (the left end), a reaction that breaks the DNA after each C residue will generate two labeled fragments: a four-nucleotide and a seven-nucleotide fragment; a reaction that breaks the DNA after each G will produce only one labeled, five-nucleotide fragment. Because the fragments are radioactively labeled at their 5' ends, only the fragment to the 5' side of the break is visualized. The fragment sizes correspond to the relative positions of C and G residues in the sequence. When the sets of fragments corresponding to each of the four bases are electrophoretically separated side by side, they produce a ladder of bands from which the sequence can be read directly (Fig. 8–33). We illustrate only the Sanger



**FIGURE 8-33 DNA sequencing by the Sanger method.** This method makes use of the mechanism of DNA synthesis by DNA polymerases (Chapter 25). **(a)** DNA polymerases require both a primer (a short oligonucleotide strand), to which nucleotides are added, and a template strand to guide selection of each new nucleotide. In cells, the 3'-hydroxyl group of the primer reacts with an incoming deoxynucleoside triphosphate (dNTP) to form a new phosphodiester bond. **(b)** The Sanger sequencing procedure uses dideoxynucleoside triphosphate (ddNTP) analogs to interrupt DNA synthesis. (The Sanger method is also known as the dideoxy method.) When a ddNTP is inserted in place of a dNTP, strand elongation is halted after the analog is added, because it lacks the 3'-hydroxyl group needed for the next step. **(c)** The DNA to be sequenced is used as the template strand, and a short primer, radioactively or fluorescently labeled, is annealed to it. By addition of small amounts of a single ddNTP, for example ddCTP, to an otherwise normal reaction system, the synthesized strands will be prematurely terminated at some locations where dC normally occurs. Given the excess of dCTP over ddCTP, the chance that the analog will be incorporated whenever a dC is to be added is small. However, ddCTP is present in sufficient amounts to ensure that each new strand has a high probability of acquiring at least one ddC at some point during synthesis. The result is a solution containing a mixture of labeled fragments, each ending with a C residue. Each C residue in the sequence generates a set of fragments of a particular length, such that the different-sized fragments, separated by electrophoresis, reveal the location of C residues. This procedure is repeated separately for each of the four ddNTPs, and the sequence can be read directly from an autoradiogram of the gel. Because shorter DNA fragments migrate faster, the fragments near the bottom of the gel represent the nucleotide positions closest to the primer (the 5' end), and the sequence is read (in the 5'→3' direction) from bottom to top. Note that the sequence obtained is that of the strand *complementary* to the strand being analyzed.

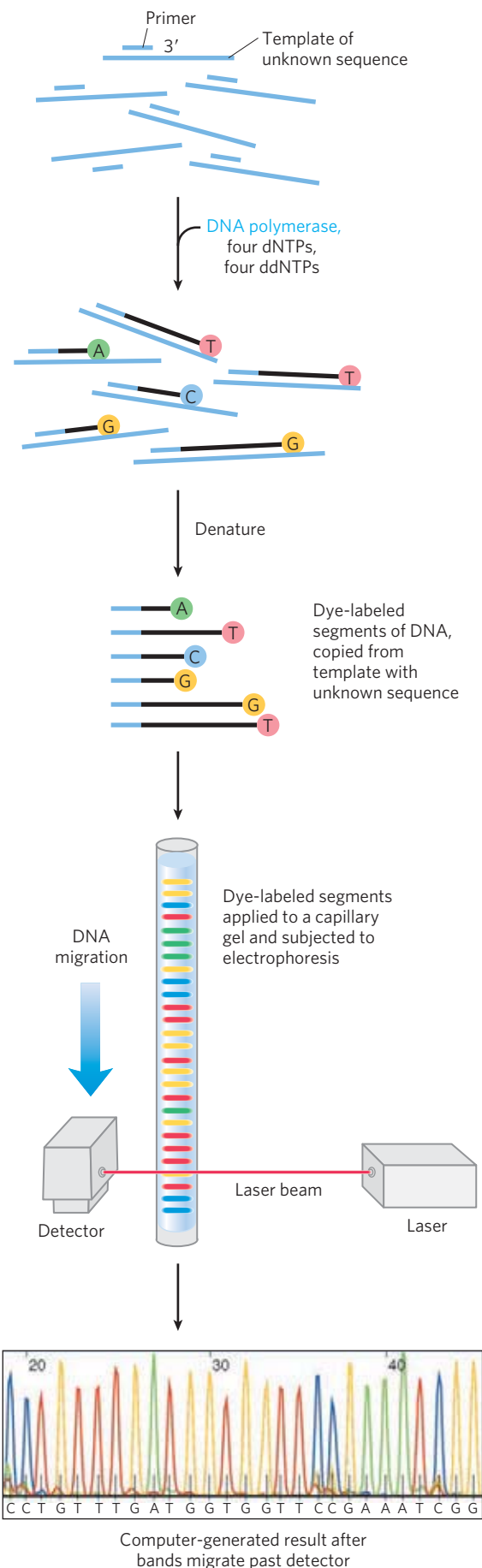


method, because it has proved to be technically easier and is in more widespread use. It requires the enzymatic synthesis of a DNA strand complementary to the strand under analysis, using a radioactively labeled “primer” and dideoxynucleotides.

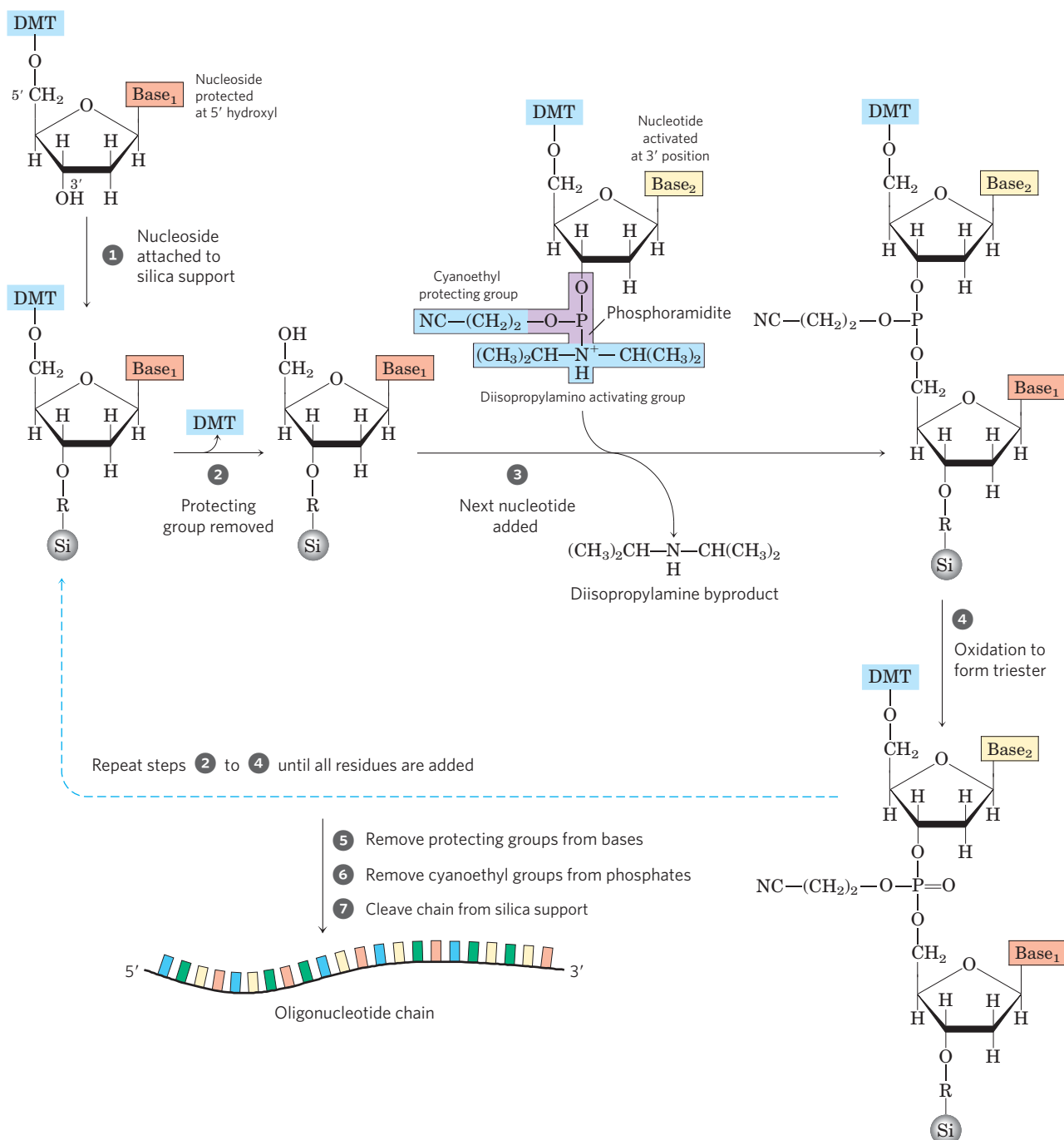
Since these first practical DNA-sequencing methods appeared, methodology has improved rapidly. Much of the advance has been fueled by the Human Genome Project, described in Chapter 9. A variation of Sanger’s sequencing method, in which the dideoxynucleotides used for each reaction are labeled with a differently colored fluorescent tag (**Fig. 8-34**), was used in early efforts to automate large DNA-sequencing efforts. With this technology, researchers can sequence DNA molecules containing thousands of nucleotides in a few hours. This approach was heavily used in the initial efforts to sequence entire organism genomes, and is still used for routine sequencing of genes or DNA segments. However, modern genomic sequencing now makes use of vastly more efficient methods, sometimes referred to as next-generation or **next-gen sequencing**. These are described in Chapter 9. 🧪 **Dideoxy Sequencing of DNA**

### The Chemical Synthesis of DNA Has Been Automated

An important practical advance in nucleic acid chemistry was the rapid and accurate synthesis of short oligonucleotides of known sequence. The methods were pioneered by H. Gobind Khorana and his colleagues in the 1970s. Refinements by Robert Letsinger and Marvin Caruthers led to the chemistry now in widest use, called the phosphoramidite method (**Fig. 8-35**). The synthesis is carried out with the growing strand attached to a solid support, using principles similar to those used by Merrifield for peptide synthesis (see Fig. 3-32), and is readily automated. The efficiency of each addition step is very high, allowing the routine synthesis of polymers containing 70 or 80 nucleotides and, in some laboratories, much longer strands. The availability of relatively inexpensive DNA polymers with predesigned sequences is having a powerful impact on all areas of biochemistry (Chapter 9).



**FIGURE 8-34 Strategy for automating DNA-sequencing reactions.** Each dideoxynucleotide used in the Sanger method can be linked to a fluorescent molecule that gives all the fragments terminating in that nucleotide a particular color. All four labeled ddNTPs are added to a single tube. The resulting colored DNA fragments are then separated by size in a single electrophoretic gel contained in a capillary tube (a refinement of gel electrophoresis that allows for faster separations). All fragments of a given length migrate through the capillary gel in a single peak, and the color associated with each peak is detected using a laser beam. The DNA sequence is read by determining the sequence of colors in the peaks as they pass the detector. This information is fed directly to a computer, which determines the sequence.



**FIGURE 8-35 Chemical synthesis of DNA by the phosphoramidite method.**

Automated DNA synthesis is conceptually similar to the synthesis of polypeptides on a solid support. The oligonucleotide is built up on the solid support (silica), one nucleotide at a time, in a repeated series of chemical reactions with suitably protected nucleotide precursors. **1** The first nucleoside (which will be the 3' end) is attached to the silica support at the 3' hydroxyl (through a linking group, R) and is protected at the 5' hydroxyl with an acid-labile dimethoxytrityl group (DMT). The reactive groups on all bases are also chemically protected. **2** The protecting DMT group is removed by washing the column with acid (the DMT group is colored, so this reaction can be followed spectrophotometrically). **3** The next nucleotide has a reactive phosphoramidite at its 3' position: a trivalent phosphite (as opposed to the more oxidized pentavalent phosphate normally present in nucleic acids) with one linked oxygen

replaced by an amino group or substituted amine. In the common variant shown, one of the phosphoramidite oxygens is bonded to the deoxyribose, the other is protected by a cyanoethyl group, and the third position is occupied by a readily displaced diisopropylamino group. Reaction with the immobilized nucleotide forms a 5',3' linkage, and the diisopropylamino group is eliminated. In step **4**, the phosphite linkage is oxidized with iodine to produce a phosphotriester linkage. Reactions **2** through **4** are repeated until all nucleotides are added. At each step, excess nucleotide is removed before addition of the next nucleotide. In steps **5** and **6** the remaining protecting groups on the bases and the phosphates are removed, and in **7** the oligonucleotide is separated from the solid support and purified. The chemical synthesis of RNA is somewhat more complicated because of the need to protect the 2' hydroxyl of ribose without adversely affecting the reactivity of the 3' hydroxyl.

**SUMMARY 8.3 Nucleic Acid Chemistry**

- ▶ Native DNA undergoes reversible unwinding and separation of strands (melting) on heating or at extremes of pH. DNAs rich in G≡C pairs have higher melting points than DNAs rich in A=T pairs.
- ▶ Denatured single-stranded DNAs from two species can form a hybrid duplex, the degree of hybridization depending on the extent of sequence similarity. Hybridization is the basis for important techniques used to study and isolate specific genes and RNAs.
- ▶ DNA is a relatively stable polymer. Spontaneous reactions such as deamination of certain bases, hydrolysis of base-sugar *N*-glycosyl bonds, radiation-induced formation of pyrimidine dimers, and oxidative damage occur at very low rates, yet are important because of a cell's very low tolerance for changes in genetic material.
- ▶ DNA sequences can be determined with a range of modern methods.
- ▶ Oligonucleotides of known sequence can be synthesized rapidly and accurately.

**8.4 Other Functions of Nucleotides**

In addition to their roles as the subunits of nucleic acids, nucleotides have a variety of other functions in every cell: as energy carriers, components of enzyme cofactors, and chemical messengers.

**Nucleotides Carry Chemical Energy in Cells**

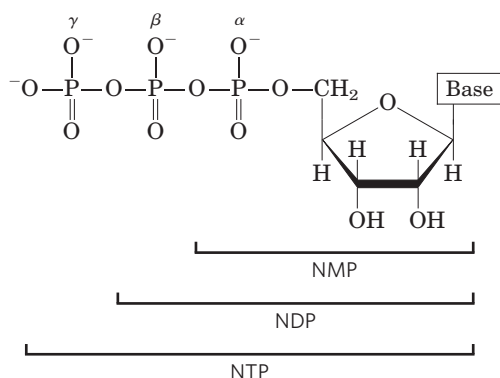
The phosphate group covalently linked at the 5' hydroxyl of a ribonucleotide may have one or two additional phosphates attached. The resulting molecules are referred to as nucleoside mono-, di-, and triphosphates (**Fig. 8–36**). Starting from the ribose, the three phosphates are generally labeled  $\alpha$ ,  $\beta$ , and  $\gamma$ . Hydrolysis of

nucleoside triphosphates provides the chemical energy to drive many cellular reactions. Adenosine 5'-triphosphate, ATP, is by far the most widely used for this purpose, but UTP, GTP, and CTP are also used in some reactions. Nucleoside triphosphates also serve as the activated precursors of DNA and RNA synthesis, as described in Chapters 25 and 26.

The energy released by hydrolysis of ATP and the other nucleoside triphosphates is accounted for by the structure of the triphosphate group. The bond between the ribose and the  $\alpha$  phosphate is an ester linkage. The  $\alpha,\beta$  and  $\beta,\gamma$  linkages are phosphoanhydrides (**Fig. 8–37**). Hydrolysis of the ester linkage yields about 14 kJ/mol under standard conditions, whereas hydrolysis of each anhydride bond yields about 30 kJ/mol. ATP hydrolysis often plays an important thermodynamic role in biosynthesis. When coupled to a reaction with a positive free-energy change, ATP hydrolysis shifts the equilibrium of the overall process to favor product formation (recall the relationship between equilibrium constant and free-energy change described by Eqn 6–3 on p. 194).

**Adenine Nucleotides Are Components of Many Enzyme Cofactors**

A variety of enzyme cofactors serving a wide range of chemical functions include adenosine as part of their structure (**Fig. 8–38**). They are unrelated structurally except for the presence of adenosine. In none of these cofactors does the adenosine portion participate directly in the primary function, but removal of adenosine generally results in a drastic reduction of cofactor activities. For example, removal of the adenine nucleotide (3'-phosphoadenosine diphosphate) from acetoacetyl-CoA, the coenzyme A derivative of acetoacetate, reduces its reactivity as a substrate for  $\beta$ -ketoacyl-CoA transferase (an enzyme of lipid metabolism) by a factor of  $10^6$ . Although this requirement for adenosine has not been investigated in detail, it must involve the binding energy between

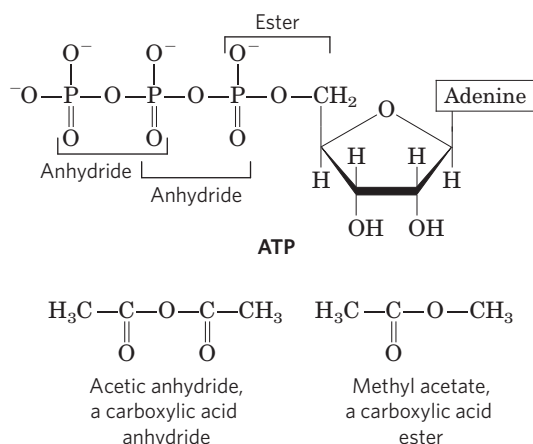


Abbreviations of ribonucleoside 5'-phosphates			
Base	Mono-	Di-	Tri-
Adenine	AMP	ADP	ATP
Guanine	GMP	GDP	GTP
Cytosine	CMP	CDP	CTP
Uracil	UMP	UDP	UTP

Abbreviations of deoxyribonucleoside 5'-phosphates			
Base	Mono-	Di-	Tri-
Adenine	dAMP	dADP	dATP
Guanine	dGMP	dGDP	dGTP
Cytosine	dCMP	dCDP	dCTP
Thymine	dTMP	dTDP	dTTP

**FIGURE 8–36 Nucleoside phosphates.** General structure of the nucleoside 5'-mono-, di-, and triphosphates (NMPs, NDPs, and NTPs) and

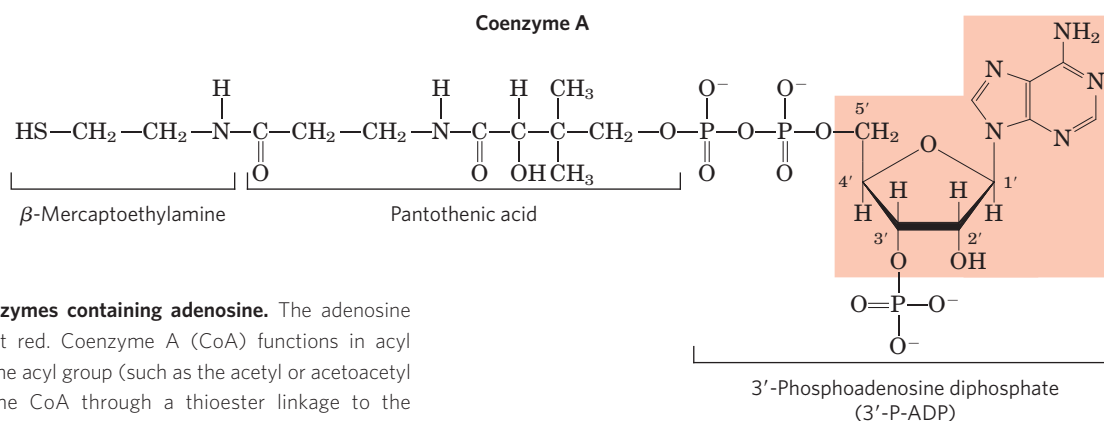
their standard abbreviations. In the deoxyribonucleoside phosphates (dNMPs, dNDPs, and dNTPs), the pentose is 2'-deoxy-D-ribose.



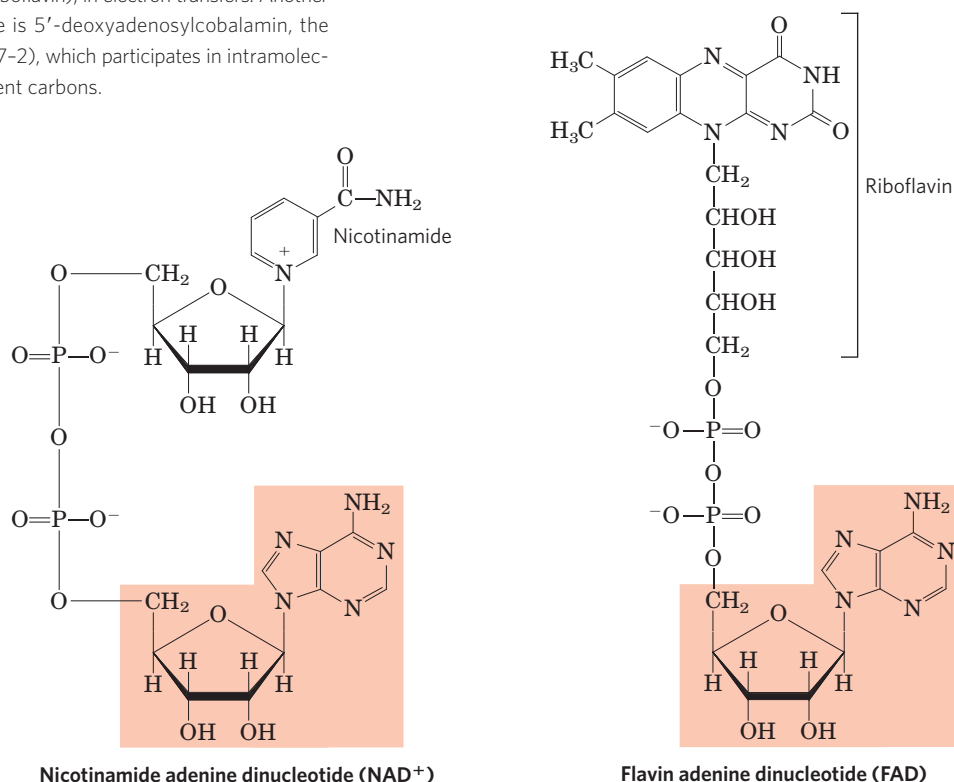
**FIGURE 8-37** The phosphate ester and phosphoanhydride bonds of ATP. Hydrolysis of an anhydride bond yields more energy than hydrolysis of the ester. A carboxylic acid anhydride and carboxylic acid ester are shown for comparison.

enzyme and substrate (or cofactor) that is used both in catalysis and in stabilizing the initial enzyme-substrate complex (Chapter 6). In the case of  $\beta$ -ketoacyl-CoA transferase, the nucleotide moiety of coenzyme A seems to be a binding “handle” that helps to pull the substrate (acetoacetyl-CoA) into the active site. Similar roles may be found for the nucleoside portion of other nucleotide cofactors.

Why is adenosine, rather than some other large molecule, used in these structures? The answer here may involve a form of evolutionary economy. Adenosine is certainly not unique in the amount of potential binding energy it can contribute. The importance of adenosine probably lies not so much in some special chemical characteristic as in the evolutionary advantage of using one compound for multiple roles. Once ATP became the universal source of chemical energy, systems developed



**FIGURE 8-38** Some coenzymes containing adenosine. The adenosine portion is shaded in light red. Coenzyme A (CoA) functions in acyl group transfer reactions; the acyl group (such as the acetyl or acetoacetyl group) is attached to the CoA through a thioester linkage to the  $\beta$ -mercaptoethylamine moiety. NAD<sup>+</sup> functions in hydride transfers, and FAD, the active form of vitamin B<sub>2</sub> (riboflavin), in electron transfers. Another coenzyme incorporating adenosine is 5'-deoxyadenosylcobalamin, the active form of vitamin B<sub>12</sub> (see Box 17-2), which participates in intramolecular group transfers between adjacent carbons.





to synthesize ATP in greater abundance than the other nucleotides; because it is abundant, it becomes the logical choice for incorporation into a wide variety of structures. The economy extends to protein structure. A single protein domain that binds adenosine can be used in different enzymes. Such a domain, called a **nucleotide-binding fold**, is found in many enzymes that bind ATP and nucleotide cofactors.

### Some Nucleotides Are Regulatory Molecules

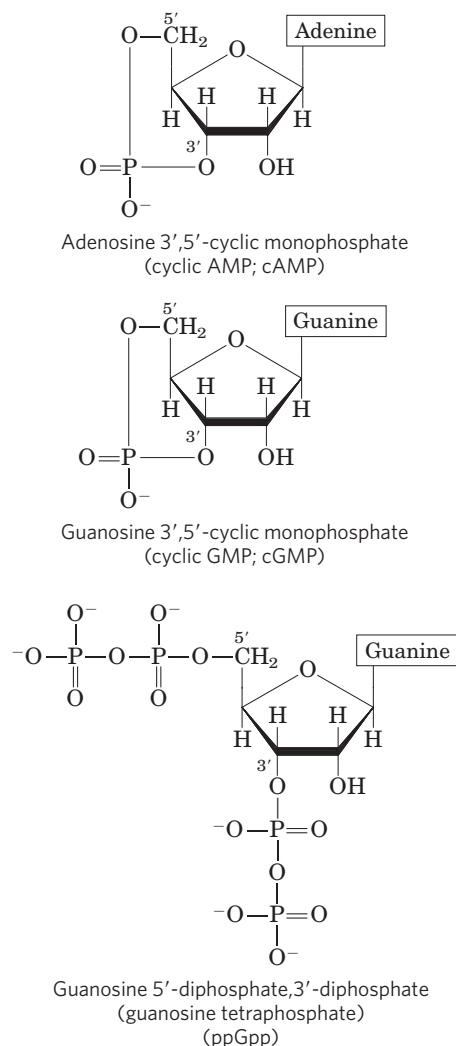
Cells respond to their environment by taking cues from hormones or other external chemical signals. The interaction of these extracellular chemical signals (“first messengers”) with receptors on the cell surface often leads to the production of **second messengers** inside the cell, which in turn leads to adaptive changes in the cell interior (Chapter 12). Often, the second messenger is a nucleotide (**Fig. 8–39**). One of the most common is **adenosine 3′,5′-cyclic monophosphate (cyclic AMP, or cAMP)**, formed from ATP in a reaction catalyzed by adenylyl cyclase, an enzyme associated with

the inner face of the plasma membrane. Cyclic AMP serves regulatory functions in virtually every cell outside the plant kingdom. Guanosine 3′,5′-cyclic monophosphate (cGMP) occurs in many cells and also has regulatory functions.

Another regulatory nucleotide, ppGpp (**Fig. 8–39**), is produced in bacteria in response to a slowdown in protein synthesis during amino acid starvation. This nucleotide inhibits the synthesis of the rRNA and tRNA molecules (see **Fig. 28–22**) needed for protein synthesis, preventing the unnecessary production of nucleic acids.

### SUMMARY 8.4 Other Functions of Nucleotides

- ▶ ATP is the central carrier of chemical energy in cells. The presence of an adenosine moiety in a variety of enzyme cofactors may be related to binding-energy requirements.
- ▶ Cyclic AMP, formed from ATP in a reaction catalyzed by adenylyl cyclase, is a common second messenger produced in response to hormones and other chemical signals.



**FIGURE 8–39** Three regulatory nucleotides.

### Key Terms

*Terms in bold are defined in the glossary.*

<b>gene</b>	281	major groove	289
<b>ribosomal RNA</b>		minor groove	289
<b>(rRNA)</b>	281	B-form DNA	291
<b>messenger RNA</b>		A-form DNA	291
<b>(mRNA)</b>	281	Z-form DNA	291
<b>transfer RNA</b>		<b>palindrome</b>	291
<b>(tRNA)</b>	281	<b>hairpin</b>	292
<b>nucleotide</b>	281	<b>cruciform</b>	292
<b>nucleoside</b>	281	triplex DNA	292
<b>pyrimidine</b>	282	G tetraplex	292
<b>purine</b>	282	<b>transcription</b>	294
<b>deoxyribonucleotides</b>	283	<b>monocistronic</b>	
<b>ribonucleotide</b>	283	<b>mRNA</b>	294
<b>phosphodiester</b>		<b>polycistronic mRNA</b>	294
<b>linkage</b>	285	<b>mutation</b>	299
<b>5′ end</b>	285	<b>second messenger</b>	308
<b>3′ end</b>	285	<b>adenosine 3′,5′-cyclic</b>	
<b>oligonucleotide</b>	286	<b>monophosphate (cyclic</b>	
<b>polynucleotide</b>	286	<b>AMP, cAMP)</b>	308
<b>base pair</b>	287		

### Further Reading

#### General

**Cox, M.M., Doudna, J.A., & O'Donnell, M.** (2012) *Molecular Biology: Principles and Practice*, W. H. Freeman and Company, New York.

The best place to start to learn more about nucleic acid structure and function.

**Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., & Ellenberger, T.** (2006) *DNA Repair and Mutagenesis*, 2nd edn, ASM Press, Washington, DC.

A good source for more information on the chemistry of nucleotides and nucleic acids.

### Historical

**Judson, H.F.** (1996) *The Eighth Day of Creation: Makers of the Revolution in Biology*, expanded edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

**Olby, R.C.** (1994) *The Path to the Double Helix: The Discovery of DNA*, Dover Publications, Inc., New York.

**Sayre, A.** (1978) *Rosalind Franklin and DNA*, W. W. Norton & Co., Inc., New York.

**Watson, J.D.** (1968) *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*, Atheneum, New York. [Paperback edition, Touchstone Books, 2001.]

### Nucleic Acid Structure

**Frank-Kamenetskii, M.D. & Mirkin, S.M.** (1995) Triplex DNA structures. *Annu. Rev. Biochem.* **64**, 65–95.

**Holbrook, S.R.** (2008) Structural principles from large RNAs. *Annu. Rev. Biophys.* **37**, 445–464.

**Keniry, M.A.** (2000) Quadruplex structures in nucleic acids. *Biopolymers* **56**, 123–146.

Good summary of the structural properties of quadruplexes.

### Nucleic Acid Chemistry

**Bonetta, L.** (2006) Genome sequencing in the fast lane. *Nat. Methods* **3**, 141–147.

This paper introduces a newer generation of sequencing methods that are described in Chapter 9.

**Collins, A.R.** (1999) Oxidative DNA damage, antioxidants, and cancer. *Bioessays* **21**, 238–246.

**Cooke, M.S., Evans, M.D., Dizdaroglu, M., & Lunt J.** (2003) Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **17**, 1195–1214.

**Imlay, J.A.** (2008) Cellular defenses against superoxide and hydrogen peroxide. *Annu. Rev. Biochem.* **77**, 755–776.

**Marnett, L.J. & Plataras, J.P.** (2001) Endogenous DNA damage and mutation. *Trends Genet.* **17**, 214–221.

### ATP as Energy Carrier

**Jencks, W.P.** (1987) Economics of enzyme catalysis. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 65–73.

A relatively short article, full of insights.

the moon (~320,000 km). The DNA double helix weighs about  $1 \times 10^{-18}$  g per 1,000 nucleotide pairs; each base pair extends 3.4 Å. For an interesting comparison, your body contains about 0.5 g of DNA!

**4. DNA Bending** Assume that a poly(A) tract five base pairs long produces a 20° bend in a DNA strand. Calculate the total (net) bend produced in a DNA if the center base pairs (the third of five) of two successive (dA)<sub>5</sub> tracts are located (a) 10 base pairs apart; (b) 15 base pairs apart. Assume 10 base pairs per turn in the DNA double helix.

**5. Distinction between DNA Structure and RNA Structure** Hairpins may form at palindromic sequences in single strands of either RNA or DNA. How is the helical structure of a long and fully base-paired (except at the end) hairpin in RNA different from that of a similar hairpin in DNA?

**6. Nucleotide Chemistry** The cells of many eukaryotic organisms have highly specialized systems that specifically repair G–T mismatches in DNA. The mismatch is repaired to form a G≡C (not A=T) base pair. This G–T mismatch repair mechanism occurs in addition to a more general system that repairs virtually all mismatches. Suggest why cells might require a specialized system to repair G–T mismatches.

**7. Denaturation of Nucleic Acids** A duplex DNA oligonucleotide in which one of the strands has the sequence TAATACGACTCACTATAGGG has a melting temperature ( $t_m$ ) of 59 °C. If an RNA duplex oligonucleotide of identical sequence (substituting U for T) is constructed, will its melting temperature be higher or lower?

**8. Spontaneous DNA Damage** Hydrolysis of the N-glycosyl bond between deoxyribose and a purine in DNA creates an AP site. An AP site generates a thermodynamic destabilization greater than that created by any DNA mismatched base pair. This effect is not completely understood. Examine the structure of an AP site (see Fig. 8–30b) and describe some chemical consequences of base loss.

**9. Prediction of Nucleic Acid Structure from Its Sequence** A part of a sequenced chromosome has the sequence (on one strand) ATTGCATCCGCGCGTGC GCGCGCG-ATCCCGTTACTTTCCG. Which part of this sequence is most likely to take up the Z conformation?

**10. Nucleic Acid Structure** Explain why the absorption of UV light by double-stranded DNA increases (the hyperchromic effect) when the DNA is denatured.

**11. Determination of Protein Concentration in a Solution Containing Proteins and Nucleic Acids** The concentration of protein or nucleic acid in a solution containing both can be estimated by using their different light absorption properties: proteins absorb most strongly at 280 nm and nucleic acids at 260 nm. Estimates of their respective concentrations in a mixture can be made by measuring the absorbance ( $A$ ) of the solution at 280 and 260 nm and using the table on the next page, which gives  $R_{280/260}$ , the ratio of absorbances at 280 and 260 nm; the percentage of total mass that is nucleic acid; and a factor,  $F$ , that corrects the  $A_{280}$  reading and gives a more accurate protein

## Problems

**1. Nucleotide Structure** Which positions in the purine ring of a purine nucleotide in DNA have the potential to form hydrogen bonds but are not involved in Watson-Crick base pairing?

**2. Base Sequence of Complementary DNA Strands** One strand of a double-helical DNA has the sequence (5')GCGCAATATTTCTCAAAATATTGCGC(3'). Write the base sequence of the complementary strand. What special type of sequence is contained in this DNA segment? Does the double-stranded DNA have the potential to form any alternative structures?

**3. DNA of the Human Body** Calculate the weight in grams of a double-helical DNA molecule stretching from the Earth to

estimate. The protein concentration (in mg/mL) =  $F \times A_{280}$  (assuming the cuvette is 1 cm wide). Calculate the protein concentration in a solution of  $A_{280} = 0.69$  and  $A_{260} = 0.94$ .

$R_{280/260}$	Proportion of nucleic acid (%)	$F$
1.75	0.00	1.116
1.63	0.25	1.081
1.52	0.50	1.054
1.40	0.75	1.023
1.36	1.00	0.994
1.30	1.25	0.970
1.25	1.50	0.944
1.16	2.00	0.899
1.09	2.50	0.852
1.03	3.00	0.814
0.979	3.50	0.776
0.939	4.00	0.743
0.874	5.00	0.682
0.846	5.50	0.656
0.822	6.00	0.632
0.804	6.50	0.607
0.784	7.00	0.585
0.767	7.50	0.565
0.753	8.00	0.545
0.730	9.00	0.508
0.705	10.00	0.478
0.671	12.00	0.422
0.644	14.00	0.377
0.615	17.00	0.322
0.595	20.00	0.278

**12. Solubility of the Components of DNA** Draw the following structures and rate their relative solubilities in water (most soluble to least soluble): deoxyribose, guanine, phosphate. How are these solubilities consistent with the three-dimensional structure of double-stranded DNA?

**13. Sanger Sequencing Logic** In the Sanger (dideoxy) method for DNA sequencing, a small amount of a dideoxynucleotide triphosphate—say, ddCTP—is added to the sequencing reaction along with a larger amount of the corresponding dCTP. What result would be observed if the dCTP were omitted?

**14. DNA Sequencing** The following DNA fragment was sequenced by the Sanger method. The red asterisk indicates a fluorescent label.



A sample of the DNA was reacted with DNA polymerase and each of the nucleotide mixtures (in an appropriate buffer) listed below. Dideoxynucleotides (ddNTPs) were added in relatively small amounts.

1. dATP, dTTP, dCTP, dGTP, ddTTP
2. dATP, dTTP, dCTP, dGTP, ddGTP

3. dATP, dCTP, dGTP, ddTTP

4. dATP, dTTP, dCTP, dGTP

The resulting DNA was separated by electrophoresis on an agarose gel, and the fluorescent bands on the gel were located. The band pattern resulting from nucleotide mixture 1 is shown below. Assuming that all mixtures were run on the same gel, what did the remaining lanes of the gel look like?



**15. Snake Venom Phosphodiesterase** An exonuclease is an enzyme that sequentially cleaves nucleotides from the end of a polynucleotide strand. Snake venom phosphodiesterase, which hydrolyzes nucleotides from the 3' end of any oligonucleotide with a free 3'-hydroxyl group, cleaves between the 3' hydroxyl of the ribose or deoxyribose and the phosphoryl group of the next nucleotide. It acts on single-stranded DNA or RNA and has no base specificity. This enzyme was used in sequence determination experiments before the development of modern nucleic acid sequencing techniques. What are the products of partial digestion by snake venom phosphodiesterase of an oligonucleotide with the following sequence?



**16. Preserving DNA in Bacterial Endospores** Bacterial endospores form when the environment is no longer conducive to active cell metabolism. The soil bacterium *Bacillus subtilis*, for example, begins the process of sporulation when one or more nutrients are depleted. The end product is a small, metabolically dormant structure that can survive almost indefinitely with no detectable metabolism. Spores have mechanisms to prevent accumulation of potentially lethal mutations in their DNA over periods of dormancy that can exceed 1,000 years. *B. subtilis* spores are much more resistant than are the organism's growing cells to heat, UV radiation, and oxidizing agents, all of which promote mutations.

(a) One factor that prevents potential DNA damage in spores is their greatly decreased water content. How would this affect some types of mutations?

(b) Endospores have a category of proteins called small acid-soluble proteins (SASPs) that bind to their DNA, preventing formation of cyclobutane-type dimers. What causes cyclobutane dimers, and why do bacterial endospores need mechanisms to prevent their formation?

**17. Oligonucleotide Synthesis** In the scheme of Figure 8–35, each new base to be added to the growing oligonucleotide is modified so that its 3' hydroxyl is activated and the 5' hydroxyl has a dimethoxytrityl (DMT) group attached. What is the function of the DMT group on the incoming base?

## Using the Web

**18. The Structure of DNA** Elucidation of the three-dimensional structure of DNA helped researchers understand how this molecule conveys information that can be faithfully replicated from one generation to the next. To see the secondary structure of double-stranded DNA, go to the Protein Data Bank website ([www.pdb.org](http://www.pdb.org)). Use the PDB identifiers listed below to retrieve the structure summaries for the two forms of DNA. Open the structures using Jmol, and use the controls in the Jmol menu (accessed with a control-click or by clicking on the Jmol logo in the lower right corner of the image screen) to complete the following exercises. Refer to the Jmol help links as needed.

(a) Obtain the file for 141D, a highly conserved, repeated DNA sequence from the end of the HIV-1 (the virus that causes AIDS) genome. Display the molecule as a ball-and-stick structure and color by element. Identify the sugar-phosphate backbone for each strand of the DNA duplex. Locate and identify individual bases. Identify the 5' end of each strand. Locate the major and minor grooves. Is this a right- or left-handed helix?

(b) Obtain the file for 145D, a DNA with the Z conformation. Display the molecule as a ball-and-stick structure and color by element. Identify the sugar-phosphate backbone for each strand of the DNA duplex. Is this a right- or left-handed helix?

(c) To fully appreciate the secondary structure of DNA, view the molecules in stereo. On the control menu, Select > All, then Style > Stereographic > Cross-eyed viewing or Wall-eyed viewing. (If you have stereographic glasses available, select the appropriate option.) You will see two images of the DNA molecule. Sit with your nose approximately 10 inches from the monitor and focus on the tip of your nose (cross-eyed) or the opposite edges of the screen (wall-eyed). In the background you should see three images of the DNA helix. Shift your focus to the middle image, which should appear three-dimensional. (Note that only one of the two authors can make this work.)

## Data Analysis Problem

**19. Chargaff's Studies of DNA Structure** The chapter section "DNA Is a Double Helix That Stores Genetic Information" includes a summary of the main findings of Erwin Chargaff and his coworkers, listed as four conclusions ("Chargaff's rules"; p. 288). In this problem, you will examine the data Chargaff collected in support of these conclusions.

In one paper, Chargaff (1950) described his analytical methods and some early results. Briefly, he treated DNA samples with acid to remove the bases, separated the bases by

paper chromatography, and measured the amount of each base with UV spectroscopy. His results are shown in the three tables below. The *molar ratio* is the ratio of the number of moles of each base in the sample to the number of moles of phosphate in the sample—this gives the fraction of the total number of bases represented by each particular base. The *recovery* is the sum of all four bases (the sum of the molar ratios); full recovery of all bases in the DNA would give a recovery of 1.0.

**Molar ratios in ox DNA**

Base	Thymus			Spleen		Liver
	Prep. 1	Prep. 2	Prep. 3	Prep. 1	Prep. 2	Prep. 1
Adenine	0.26	0.28	0.30	0.25	0.26	0.26
Guanine	0.21	0.24	0.22	0.20	0.21	0.20
Cytosine	0.16	0.18	0.17	0.15	0.17	
Thymine	0.25	0.24	0.25	0.24	0.24	
Recovery	0.88	0.94	0.94	0.84	0.88	

**Molar ratios in human DNA**

Base	Sperm		Thymus		Liver
	Prep. 1	Prep. 2	Prep. 1	Normal	Carcinoma
Adenine	0.29	0.27	0.28	0.27	0.27
Guanine	0.18	0.17	0.19	0.19	0.18
Cytosine	0.18	0.18	0.16		0.15
Thymine	0.31	0.30	0.28		0.27
Recovery	0.96	0.92	0.91		0.87

**Molar ratios in DNA of microorganisms**

Base	Yeast		Avian tubercle bacilli
	Prep. 1	Prep. 2	Prep. 1
Adenine	0.24	0.30	0.12
Guanine	0.14	0.18	0.28
Cytosine	0.13	0.15	0.26
Thymine	0.25	0.29	0.11
Recovery	0.76	0.92	0.77

(a) Based on these data, Chargaff concluded that "no differences in composition have so far been found in DNA from different tissues of the same species." This corresponds to conclusion 2 in this chapter. However, a skeptic looking at the data above might say, "They certainly look different to me!" If you were Chargaff, how would you use the data to convince the skeptic to change her mind?

(b) The base composition of DNA from normal and cancerous liver cells (hepatocarcinoma) was not distinguishably different. Would you expect Chargaff's technique to be capable of detecting a difference between the DNA of normal and cancerous cells? Explain your reasoning.

As you might expect, Chargaff's data were not completely convincing. He went on to improve his techniques, as described



in his 1951 paper, in which he reported molar ratios of bases in DNA from a variety of organisms:

Source	A:G	T:C	A:T	G:C	Purine:pyrimidine
Ox	1.29	1.43	1.04	1.00	1.1
Human	1.56	1.75	1.00	1.00	1.0
Hen	1.45	1.29	1.06	0.91	0.99
Salmon	1.43	1.43	1.02	1.02	1.02
Wheat	1.22	1.18	1.00	0.97	0.99
Yeast	1.67	1.92	1.03	1.20	1.0
<i>Haemophilus influenzae</i> type c	1.74	1.54	1.07	0.91	1.0
<i>E. coli</i> K-12	1.05	0.95	1.09	0.99	1.0
Avian tubercle bacillus	0.4	0.4	1.09	1.08	1.1
<i>Serratia marcescens</i>	0.7	0.7	0.95	0.86	0.9
<i>Bacillus schatz</i>	0.7	0.6	1.12	0.89	1.0

(c) According to Chargaff, as stated in conclusion 1 in this chapter, “The base composition of DNA generally varies from one species to another.” Provide an argument, based on the data presented so far, that supports this conclusion.

(d) According to conclusion 4, “In all cellular DNAs, regardless of the species . . .  $A + G = T + C$ .” Provide an argument, based on the data presented so far, that supports this conclusion.

Part of Chargaff’s intent was to disprove the “tetranucleotide hypothesis”; this was the idea that DNA was a monotonous tetranucleotide polymer  $(AGCT)_n$  and therefore not

capable of containing sequence information. Although the data presented above show that DNA cannot be simply a tetranucleotide—if so, all samples would have molar ratios of 0.25 for each base—it was still possible that the DNA from different organisms was a slightly more complex, but still monotonous, repeating sequence.

To address this issue, Chargaff took DNA from wheat germ and treated it with the enzyme deoxyribonuclease for different time intervals. At each time interval, some of the DNA was converted to small fragments; the remaining, larger fragments he called the “core.” In the table below, the “19% core” corresponds to the larger fragments left behind when 81% of the DNA was degraded; the “8% core” corresponds to the larger fragments left after 92% degradation.

Base	Intact DNA	19% Core	8% Core
Adenine	0.27	0.33	0.35
Guanine	0.22	0.20	0.20
Cytosine	0.22	0.16	0.14
Thymine	0.27	0.26	0.23
Recovery	0.98	0.95	0.92

(e) How would you use these data to argue that wheat germ DNA is not a monotonous repeating sequence?

## References

- Chargaff, E.** (1950) Chemical specificity of nucleic acids and mechanism of their enzymic degradation. *Experientia* **6**, 201–209.
- Chargaff, E.** (1951) Structure and function of nucleic acids as cell constituents. *Fed. Proc.* **10**, 654–659.