Лекция 5.
Структура нуклеиновых кислот

CHAPTER

Nucleotides and Nucleic Acids

The final classes of biomolecules to be considered, the nucleotides and molecules derived from them, represent a clear case in which last is not least. Nucleotides themselves participate in a plethora of crucial supporting roles in cell metabolism, and polymers of nucleotides, the nucleic acids, provide the script for everything that occurs in a cell.

Nucleotides are energy-rich compounds that drive metabolic processes (primarily biosyntheses) in all cells. They also serve as chemical signals, key links in cellular systems that respond to hormones and other extracellular stimuli, and are structural components of a number of enzyme cofactors and metabolic intermediates.

The nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), are the molecular repositories for genetic information. The structure of every protein, and ultimately of every cell constituent, is a product of information programmed into the nucleotide sequence of a cell's nucleic acids.

This chapter provides an overview of the nucleotides and nucleic acids found in most cells. The metabolism of nucleotides is discussed in Chapter 21, and a more detailed examination of the function of nucleic acids is the focus of Part IV of this text.

Some Basics

The amino acid sequence of every protein and the nucleotide sequence of every RNA molecule in a cell are specified by that cell's DNA. The necessary protein or RNA sequence information is found in corresponding nucleotide sequences in the DNA. A segment of DNA that contains the information required for the synthesis of a functional biological product (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 of biological information is the only known function of DNA.

Several classes of RNAs are found in cells, each with a distinct function. **Ribosomal RNAs** (rRNA) are structural components of ribosomes, the large complexes that carry out the synthesis of proteins. **Messenger RNAs** (mRNA) are nucleic acids that carry the information from one or a few genes to the ribosome, where the corresponding proteins can be synthesized. **Transfer RNAs** (tRNA) are adapter molecules that faithfully translate the information in mRNA into a specific sequence of amino acids. In addition to these major classes there are a

wide variety of special-function RNAs, described in depth in Part IV. We introduce here the chemical structures of nucleotides and nucleic acids.

Nucleotides Have Characteristic Bases and Pentoses

Nucleotides have three characteristic components: (1) a nitrogenous base, (2) a pentose, and (3) a phosphate (Fig. 12–1a). The nitrogenous bases are derivatives of two parent compounds, **pyrimidine** and **purine** (Fig. 12–1b). The bases and pentoses found in the common nucleotides are heterocyclic compounds. The carbon and nitrogen atoms in the parent structures are conventionally numbered to facilitate naming and identification of the many derivative compounds. The convention for the pentose ring follows rules outlined in Chapter 11, but in the pentoses of nucleotides the carbon numbers are given a prime (′) designation (Fig. 12–1a) to distinguish them from the numbered atoms of the nitrogenous bases.

The base is joined covalently (at N-1 of pyrimidines and N-9 of purines) in an *N*-glycosidic linkage to the 1′ carbon of the pentose, and the phosphate is esterified to the 5′ carbon. The *N*-glycosidic 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. 11–11). Without the phosphate group, the molecule is called a **nucleoside.**

DNA and RNA both contain two major purine bases, **adenine** (A) and **guanine** (G). DNA and RNA also contain two major pyrimidines; in both types of nucleic acid one of these is **cytosine** (C). The single important difference between the bases of DNA and those of RNA is the nature of the second major pyrimidine: **thymine** (T) in DNA and **uracil** (U) in RNA. Only rarely does thymine occur in RNA or uracil in DNA. The structures of the five major bases are shown in Figure 12–2, and the nomenclature of their corresponding nucleotides and nucleosides is summarized in Table 12–1.

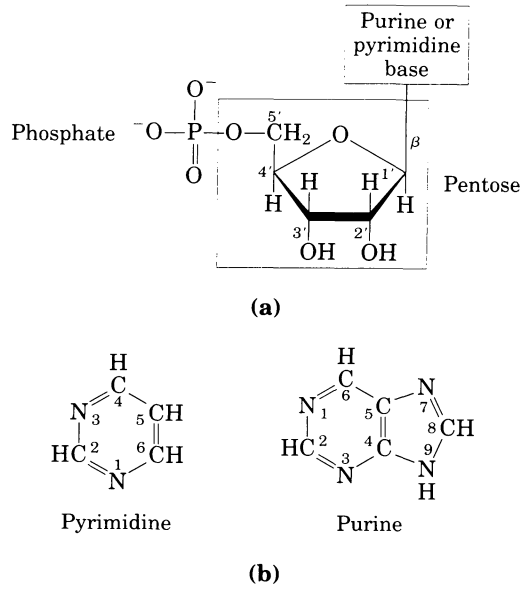


Figure 12–1 (a) The general structure of nucleotides, showing the numbering convention for the pentose. The structure shown is that of a ribonucleotide. In deoxyribonucleotides the —OH group on the 2′ carbon (in red) is replaced with —H. **(b)** The parent compounds of the pyrimidine and purine bases of nucleotides and nucleic acids, showing the numbering conventions for the ring structures.

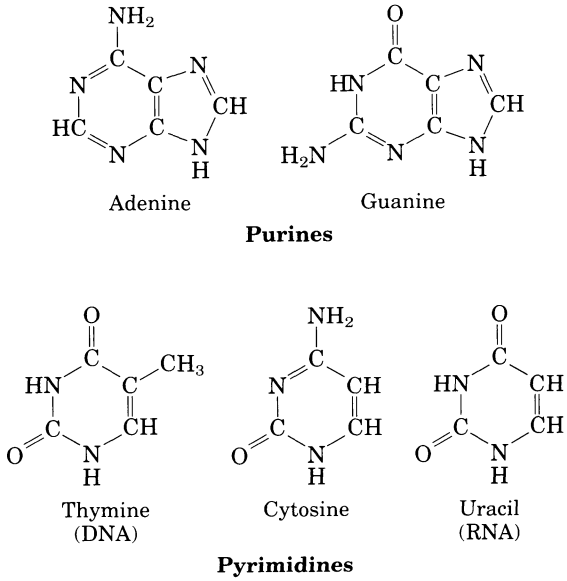


Figure 12–2 The major purine and pyrimidine bases of nucleic acids. Some of the common names of these bases reflect the circumstances of their discovery. Guanine, for example, was first isolated from guano (bird manure), and thymine was first isolated from thymus tissue.

Table 12–1 Nucleotide and nucleic acid nomenclature

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

* *Nucleoside* and *nucleotide* are generic terms that include both ribo- and deoxyribo- forms. Note that here ribonucleosides and ribonucleotides are 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.

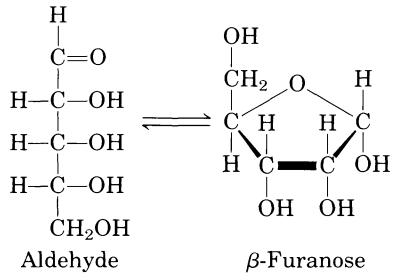
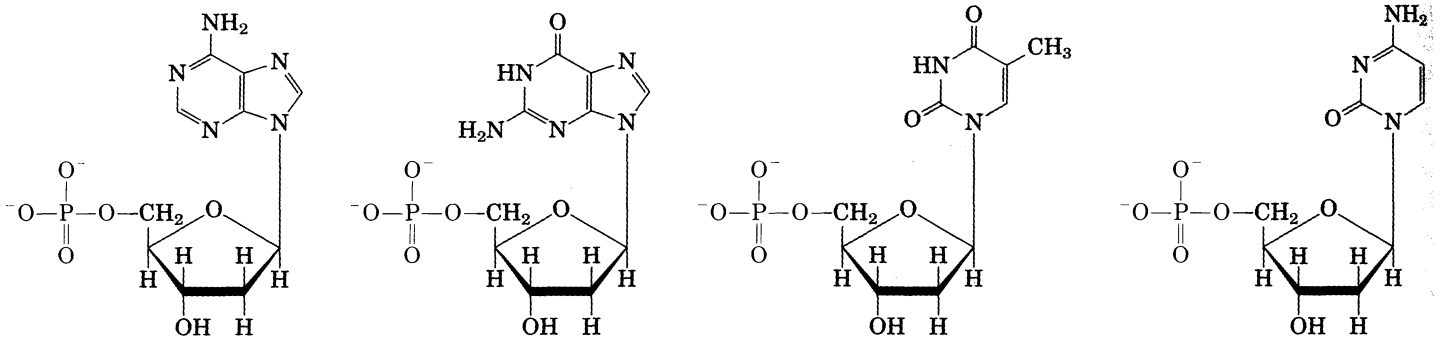


Figure 12–3 The straight-chain (aldehyde) and ring (β-furanose) forms of ribose. When ribose is free in solution, the two forms are in equilibrium. RNA contains only the ring form, β-D-ribofuranose. Deoxyribose undergoes a similar interconversion in solution, but in DNA exists solely as β-2′-deoxy-D-ribofuranose.

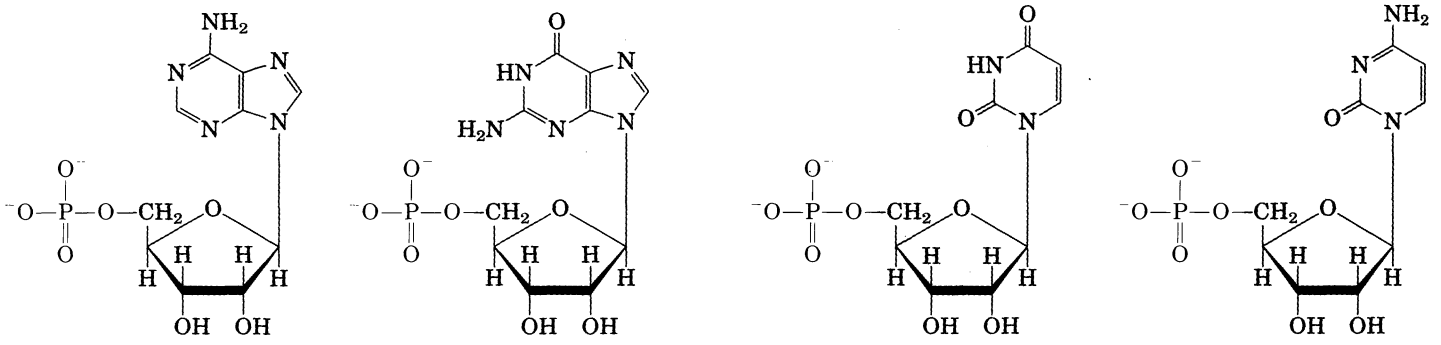
Two kinds of pentoses are found in nucleic acids. The recurring deoxyribonucleotide units of DNA contain 2′-deoxy-D-ribose, and the ribonucleotide units of RNA contain D-ribose. In nucleotides, both types of pentoses are in their β-furanose (closed five-member ring) form (Fig. 12–3).

Figure 12–4 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. Specific long sequences of A, T, G, and C nucleotides in DNA encode the genetic information. Although nucleotides bearing one of these major bases are most common, both DNA and RNA also contain some minor bases



Nucleotide: Deoxyadenylate, deoxyadenosine 5′-monophosphate	Deoxyguanylate, deoxyguanosine 5′-monophosphate	Deoxythymidylate, deoxythymidine 5′-monophosphate	Deoxycytidylate, deoxycytidine 5′-monophosphate
Symbols: A, dA, dAMP	G, dG, dGMP	T, dT, dTMP	C, dC, dCMP
Nucleoside: Deoxyadenosine	Deoxyguanosine	Deoxythymidine	Deoxycytidine

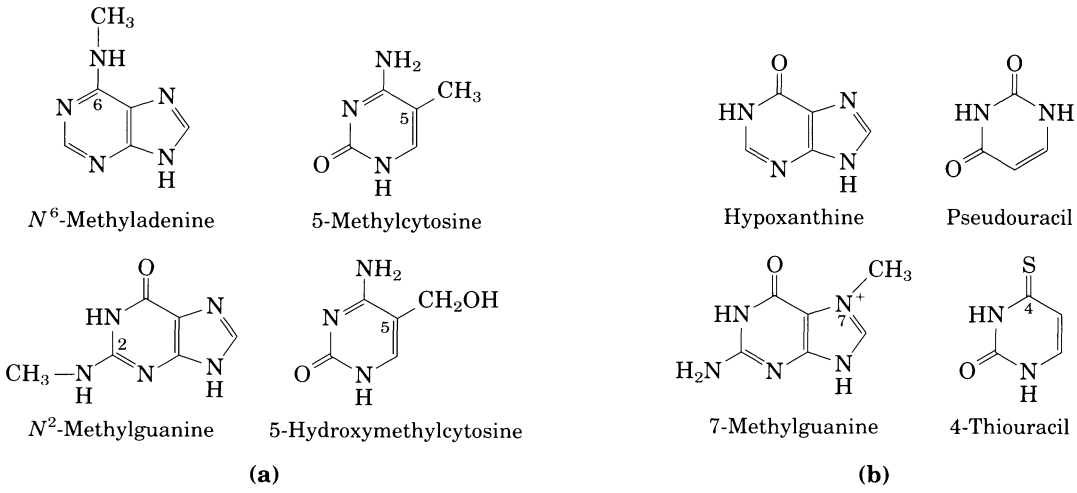
(a)



Nucleotide: Adenylate, adenosine 5′-monophosphate	Guanylate, guanosine 5′-monophosphate	Uridylate, uridine 5′-monophosphate	Cytidylate, cytidine 5′-monophosphate
Symbols: A, AMP	G, GMP	U, UMP	C, CMP
Nucleoside: Adenosine	Guanosine	Uridine	Cytidine

(b)

Figure 12–4 (a) The deoxyribonucleotide units of DNA in free form at pH 7.0. In DNA they are usually symbolized as A, G, T, and C, and sometimes as dA, dG, dT, and dC. In their free form these nucleotides are commonly abbreviated dAMP, dGMP, dTMP, and dCMP. **(b)** The ribonucleotide units of RNAs. All abbreviations assume that the phosphate group is at the 5′ position. The nucleoside portion of each molecule is shaded in red. In this and the following illustrations, the ring carbons are not shown in the purine and pyrimidine bases, as is also the convention for the pentoses.



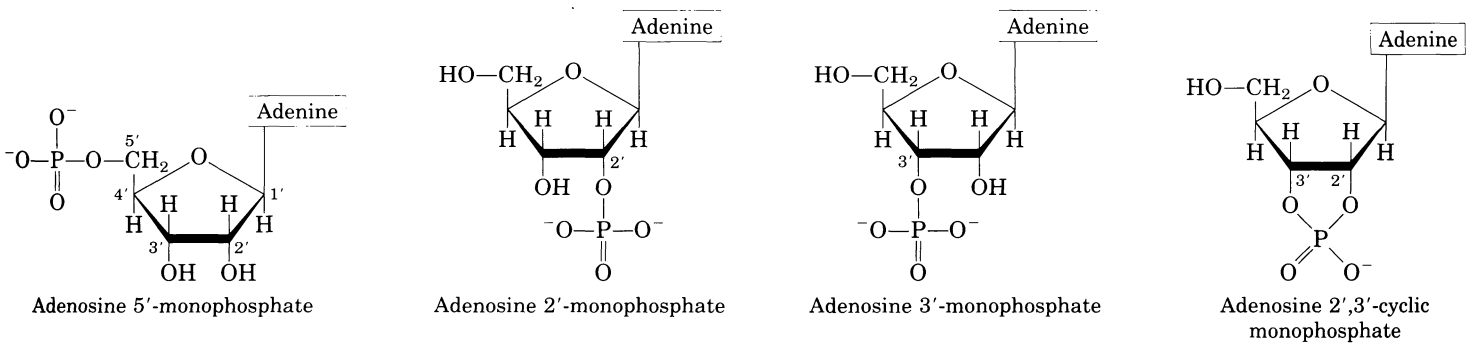
(Fig. 12–5). In DNA the most common of these are methylated forms of the major bases, but in some viral DNAs certain bases may be hydroxymethylated or glucosylated. Such altered or unusual bases in DNA molecules are in many cases specific signals for regulating or protecting the genetic information. Minor bases of many types are also found in RNAs, especially in tRNA.

The nomenclature used for the minor bases can be confusing. As indicated in Figures 12–4 and 12–5, many of the minor bases (such as hypoxanthine) have common names, just as the major bases do. For substituted forms of these bases, when the substitution involves an atom in the purine or pyrimidine rings, the usual convention (used here) is simply to indicate the ring position of the substitution by its number (e.g., 5-methylcytosine, 7-methylguanine, and 5-hydroxymethylcytosine in Fig. 12–5). The type of atom to which the substituent is attached (N, C, O, etc.) is not identified. The convention changes when the substituted atom is exocyclic, 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 in adenine becomes N^6; similarly, the carbonyl oxygen and the amino group at C-6 and C-2 of guanine become O^6 and N^2, respectively. Examples of bases substituted on exocyclic atoms are N^6-methyladenine, and N^2-methylguanine, as shown in Figure 12–5.

Cells also contain nucleotides with phosphate groups in positions other than on the 5′ carbon (Fig. 12–6). **Ribonucleoside 2′,3′-cyclic phosphates** are intermediates and **ribonucleoside 3′-phosphates** are end products of the hydrolysis of RNA by certain ribonucleases. Another variation is represented by adenosine 3′,5′-cyclic monophosphate (cAMP) and guanosine 3′,5′-cyclic monophosphate (cGMP), considered at the end of this chapter.

Figure 12–5 Some minor purine and pyrimidine bases. **(a)** Minor bases found in DNA. 5-Methylcytosine occurs in the DNA of animals and higher plants, N^6-methyladenine in bacterial DNA, and 5-hydroxymethylcytosine in bacteria infected with certain bacteriophages. **(b)** Some minor bases of tRNAs. Note that pseudouracil is identical to uracil; the distinction is the point of attachment to the ribose—uracil is attached through N-1, the normal attachment point for pyrimidines, and pseudouracil is attached through C-5.

Figure 12–6 Some adenosine monophosphates. Adenosine 2′-monophosphate, 3′-monophosphate, and 2′,3′-cyclic monophosphate are intermediates in the alkaline hydrolysis of RNA.



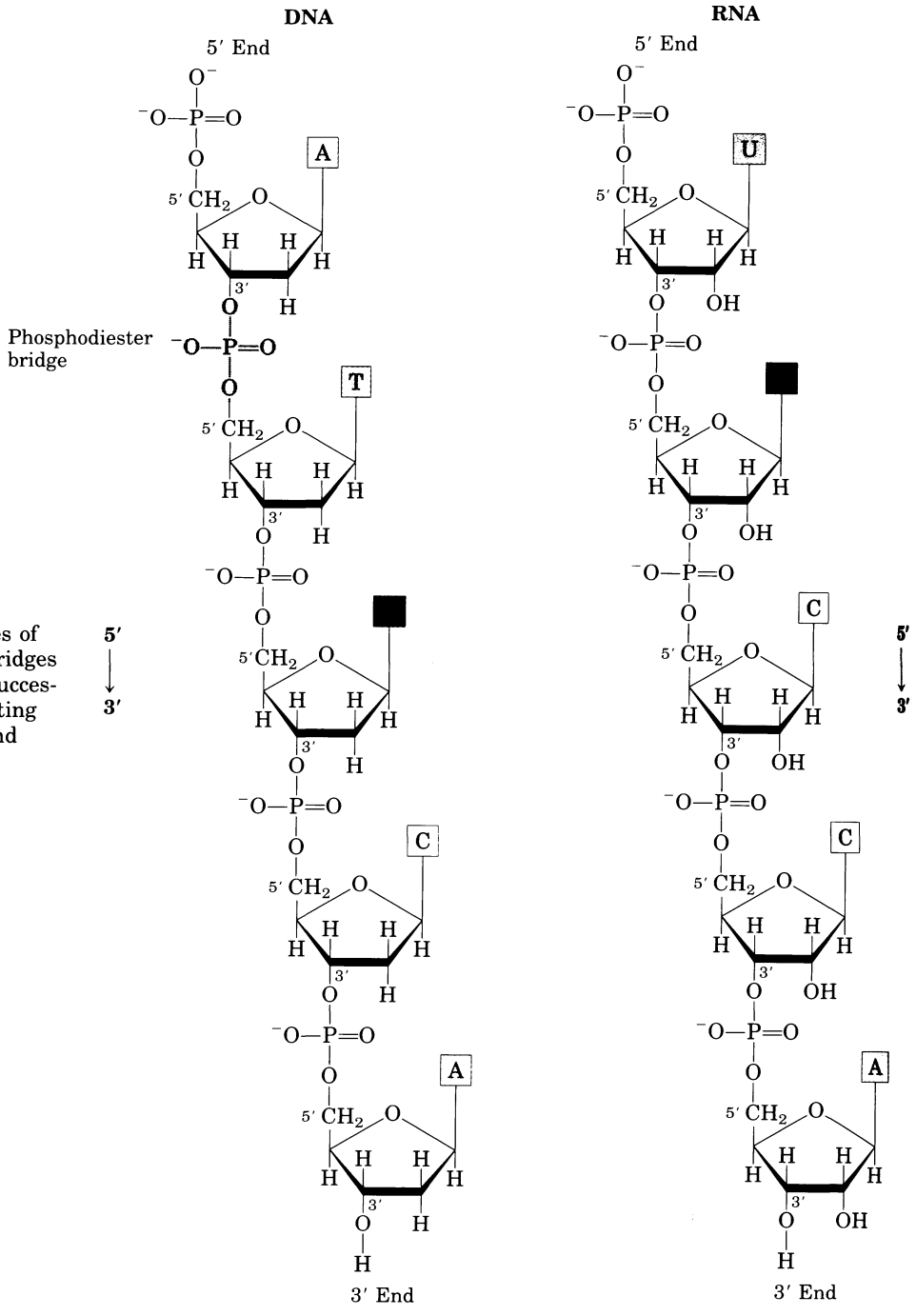


Figure 12–7 The covalent backbone structures of DNA and RNA, showing the phosphodiester bridges (one of which is shaded in the DNA) linking successive nucleotide units. The backbone of alternating pentose and phosphate groups of both DNA and RNA is highly polar.

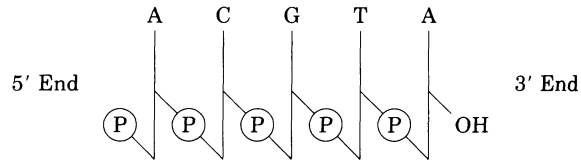
Phosphodiester Bonds Link Successive Nucleotides in Nucleic Acids

The successive nucleotides of both DNA and RNA are covalently linked through phosphate-group "bridges." Specifically, the 5′-hydroxyl group of one nucleotide unit is joined to the 3′-hydroxyl group of the next nucleotide by a **phosphodiester linkage** (Fig. 12–7). Thus the covalent backbones of nucleic acids consist of alternating phosphate and pentose residues, and the characteristic bases may be regarded as side groups joined to the backbone at regular intervals. Also note that the backbones of both DNA and RNA are hydrophilic. The hydroxyl groups of the sugar residues form hydrogen bonds with water. The phosphate groups in the polar backbone have a pK near 0 and are completely ionized and negatively charged at pH 7; thus DNA is an acid. These negative charges are generally neutralized by ionic interactions with positive charges on proteins, metal ions, and polyamines.

All the phosphodiester linkages in DNA and RNA strands have the same orientation along the chain (Fig. 12–7), 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 (Fig. 12–7). Other groups (most often one or more phosphates) may be present on one or both ends.

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′-monophosphates 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. 12–8).

The nucleotide sequences of nucleic acids can be represented schematically, as illustrated (at right) by a segment of DNA having five nucleotide units. The phosphate groups are symbolized by Ⓟ and each deoxyribose by a vertical line. The carbons in the deoxyribose are represented from 1′ at the top to 5′ at the bottom of the vertical line (even though the sugar is always in its closed-ring β-furanose form in nucleic acids). The connecting lines between nucleotides (through Ⓟ) are drawn diagonally from the middle (3′) of the deoxyribose of one nucleotide to the bottom (5′) of the next. By convention, the structure of a single strand of nucleic acid is always written with the 5′ end at the left and the 3′ end at the right; i.e., in the 5′→3′ direction. Some simpler representations of the pentadeoxyribonucleotide illustrated are pA-C-G-T-A_{OH}, pApCpGpTpA, and pACGTA. A short nucleic acid is referred to as an **oligonucleotide.** The definition of "short" is somewhat arbitrary, but the term oligonucleotide is often used for polymers containing 50 or fewer nucleotides. A longer nucleic acid is called a **polynucleotide.**



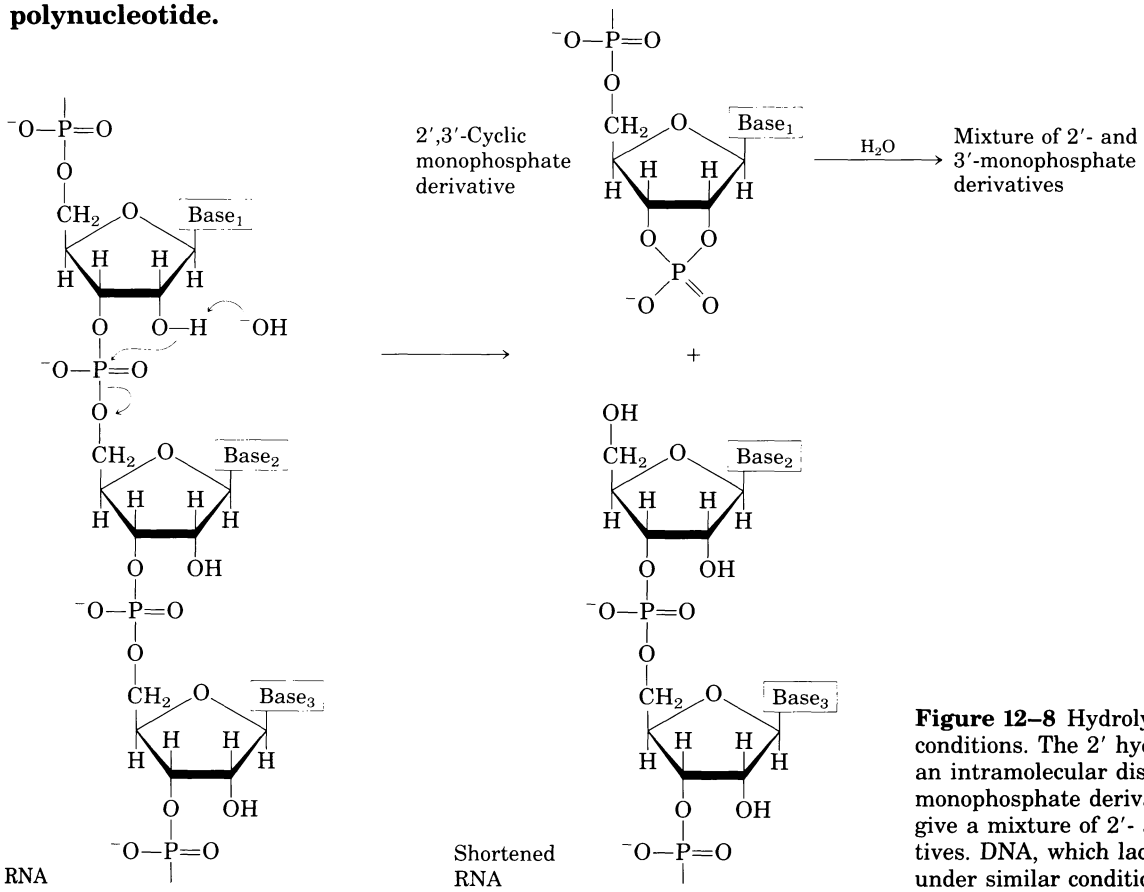


Figure 12–8 Hydrolysis of RNA under alkaline conditions. The 2′ hydroxyl acts as a nucleophile in an intramolecular displacement, the 2′,3′-cyclic monophosphate derivative is further hydrolyzed to give a mixture of 2′- and 3′-monophosphate derivatives. DNA, which lacks 2′ hydroxyls, is stable under similar conditions.

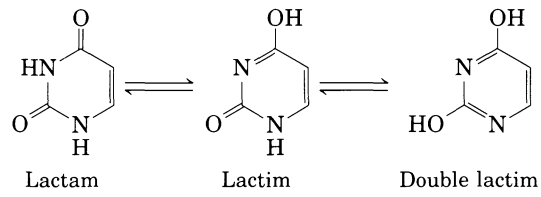


Figure 12–9 Tautomeric forms of uracil. At pH 7.0, the lactam form predominates; the other forms become more prominent as pH decreases.

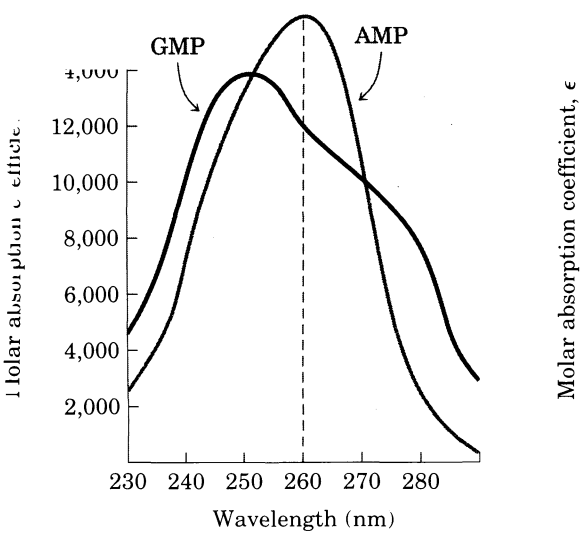
The Properties of Nucleotide Bases Affect the Structure of Nucleic Acids

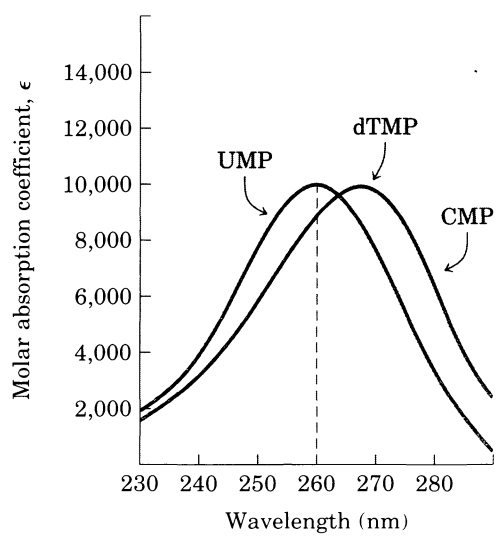
The bases have a variety of chemical properties that affect the structure, and ultimately the function, of nucleic acids. Free pyrimidines and purines are weakly basic compounds, and are thus called bases. The purines and pyrimidines common in DNA and RNA are highly conjugated molecules (see Fig. 12–2). This property has important effects on the structure, electron distribution, and light absorption of nucleic acids. Resonance involving many atoms in the ring gives most of the bonds a partially double-bonded character. One result is that pyrimidines are planar molecules; purines are very nearly planar, with a slight pucker. Free pyrimidine and purine bases may exist in two or more tautomeric forms depending upon the pH. Uracil, for example, occurs in lactam, lactim, and double lactim forms (Fig. 12–9). The structures of the purines and pyrimidines shown in Figure 12–2 are the tautomers predominating at pH 7.0. Again as a result of resonance, all of the bases absorb UV light, and nucleic acids are characterized by a strong absorption at wavelengths near 260 nm (Fig. 12–10).

The purines and pyrimidines are also hydrophobic and relatively insoluble in water at the near neutral pH of the cell. At acidic or alkaline pH the purines and pyrimidines 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 (similar to a stack of coins) represent one of two important modes of interaction between two bases. The stacking involves a combination of van der Waals and dipole–dipole interactions between the bases. These base-stacking interactions help to minimize contact with water and are very important in stabilizing the three-dimensional structure of nucleic acids, as described later. The close interaction between stacked bases in a nucleic acid has the effect of decreasing the absorption of UV light relative to a solution with the same concentration of free nucleotides. This is called the **hypochromic effect.**

The most important 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 second important mode of interaction between bases. Hydrogen bonds be-

Figure 12–10 The absorption spectra of the common nucleotides and their molar absorption coefficients at 260 nm and pH 7.0 (ϵ_{260}). The spectra of the corresponding ribonucleotides and deoxyribonucleotides, as well as the nucleosides, are essentially identical. When mixtures of nucleotides are present, the wavelength at 260 nm (dashed vertical lines) is used for measurements.





Molar absorption coefficient of nucleotides, ϵ_{260} ($M^{-1}cm^{-1}$)	
AMP	15,400
GMP	11,700
CMP	7,500
UMP	9,900
dTMP	9,200

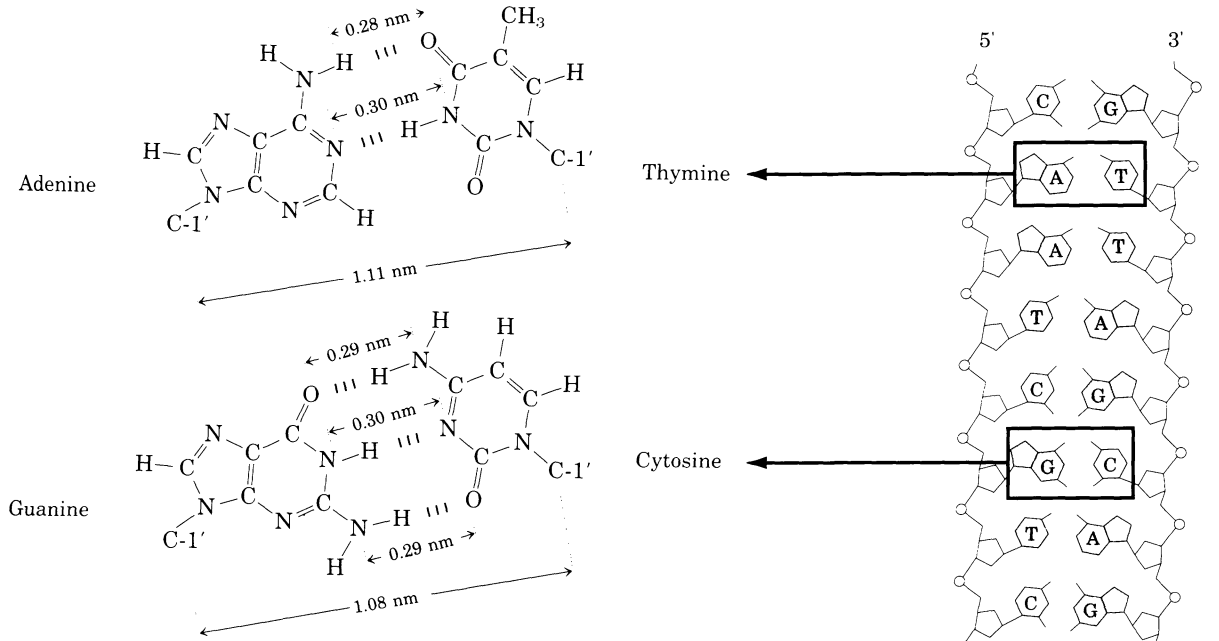


Figure 12–11 Hydrogen-bonding patterns in the base pairs defined by Watson and Crick.

tween bases permit a complementary association of two and occasionally three strands of nucleic acid. The most important hydrogen-bonding patterns are those defined by James Watson and Francis Crick in 1953, in which A bonds specifically to T (or U) and G bonds to C (Fig. 12–11). These two types of base pairs predominate in double-stranded DNA and RNA, and the tautomers shown in Figure 12–2 are responsible for these patterns. This specific pairing of bases permits the duplication of genetic information by the synthesis of nucleic acid strands that are complementary to existing strands, as we shall discuss later in this chapter.

Nucleic Acid Structure

The discovery of the structure of DNA by Watson and Crick in 1953 was a momentous event in science, an event that gave rise to entirely new disciplines and influenced the course of many others. Our present understanding of the storage and utilization of a cell's genetic information is based on work made possible by this discovery. Although information pathways are not treated in detail until Part IV of this book, the outline of these pathways presented in Chapters 1 and 3 is now a prerequisite for discussion of any area of biochemistry. Here, we concern ourselves with DNA structure itself, events that led to its discovery, and more recent refinements in our understanding. RNA structure will also be introduced.

As in the case of protein structure (Chapters 6 and 7), 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 of the structures considered in the following pages of this chapter fall under the heading of secondary structure. The complex folding of large chromosomes within the bacterial nucleoid and eukaryotic chromatin is generally considered tertiary structure; this is considered in Chapter 23.

James Watson

Francis Crick

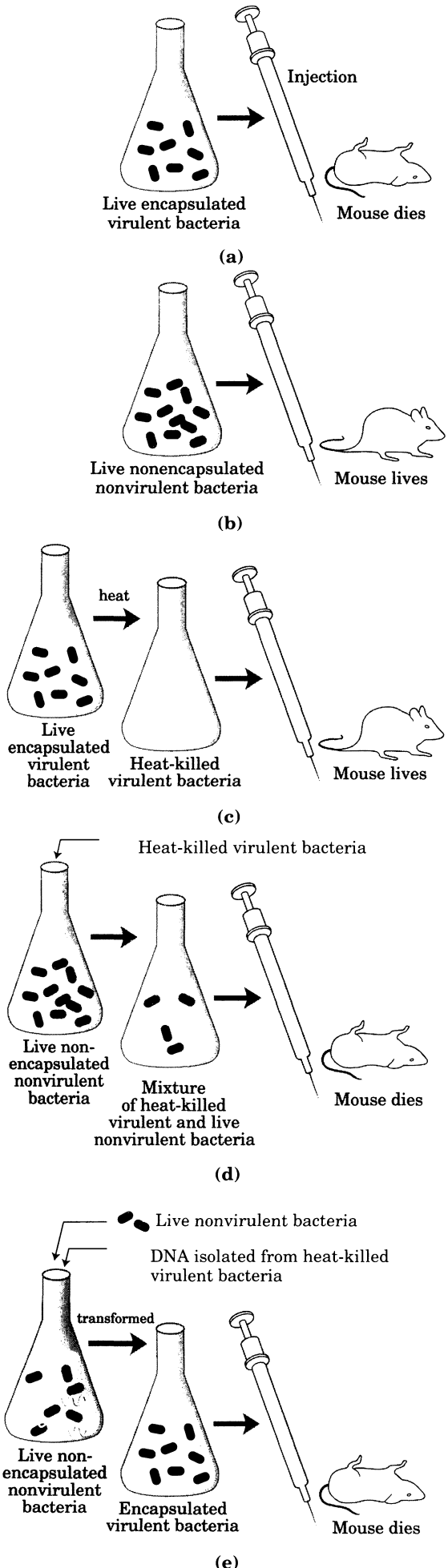


Figure 12–12 The Avery–MacLeod–McCarty experiment. When injected into mice, the encapsulated strain of pneumococcus **(a)** is lethal, whereas the nonencapsulated strain **(b)** is harmless, as is the heat-killed encapsulated strain **(c).** Earlier research by the bacteriologist Frederick Griffith had shown that adding heat-killed virulent bacteria (which alone are harmless to mice) to a live nonvirulent strain permanently transformed the latter into lethal, virulent, encapsulated bacteria **(d).** He concluded that a transforming factor in the heat-killed virulent bacteria had gained entrance into the live nonvirulent bacteria and rendered them virulent and encapsulated.

Avery and his colleagues identified the Griffith transforming factor as DNA. **(e)** They extracted the DNA from heat-killed virulent pneumococci, removing the protein as completely as possible, and added this DNA to nonvirulent bacteria. The nonvirulent pneumococci were permanently transformed into a virulent strain. The DNA evidently gained entrance into the nonvirulent bacteria, and the genes for virulence and capsule formation became incorporated into the chromosomes of the nonvirulent bacteria. All subsequent generations of these bacteria were therefore virulent and encapsulated.

DNA Stores Genetic Information

The biochemical investigation of DNA began with Friedrich Miescher, who carried out the first systematic chemical studies of cell nuclei. In 1868 Miescher isolated a phosphorus-containing substance, which he called "nuclein," from the nuclei of pus cells (leukocytes) obtained from discarded surgical bandages. He found nuclein to consist of an acidic portion, which we know today as DNA, and a basic portion, protein. Miescher later found a similar acidic substance in the heads of salmon sperm cells. Although he partially purified the nucleic acid and studied its properties, the covalent (primary) structure of DNA (as shown in Fig. 12–7) did not become known with certainty until the late 1940s.

Miescher and many others suspected that nuclein or nucleic acid was associated in some way with cell inheritance, but the first direct evidence that DNA is the bearer of genetic information came in 1944 through a discovery made by Oswald T. Avery, Colin MacLeod, and Maclyn McCarty. These investigators found that DNA extracted from a virulent (disease-causing) strain of the bacterium *Streptococcus pneumoniae,* also known as pneumococcus, genetically transformed a nonvirulent strain of this organism into a virulent form (Fig. 12–12). Avery and his colleagues concluded that the DNA extracted from the virulent strain carried the inheritable genetic message for virulence. Not everyone accepted these conclusions, because traces of protein impurities present in the DNA could have been the actual carrier of the genetic information. This possibility was soon eliminated by the finding that treatment of the DNA with proteolytic enzymes did not destroy the transforming activity, but treatment with deoxyribonucleases (DNA-hydrolyzing enzymes) did.

A second important experiment provided independent evidence that DNA carries genetic information. In 1952 Alfred D. Hershey and Martha Chase used radioactive phosphorus (^{32}P) and radioactive sulfur (^{35}S) tracers to show that when the bacterial virus (bacteriophage) T2 infects its host cell, *E. coli,* it is the phosphorus-containing DNA of

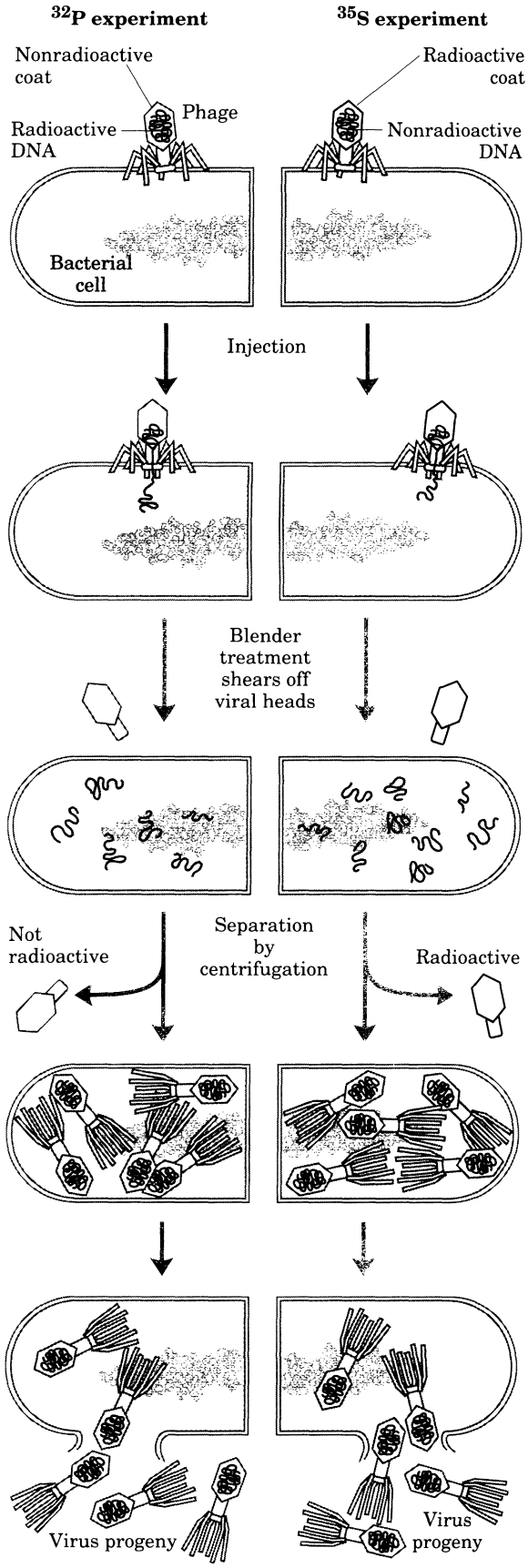


Figure 12–13 Summary of the Hershey–Chase experiment. Two batches of isotopically labeled bacteriophage particles were prepared. One was labeled with ^{32}P in the phosphate groups of the DNA and the other with ^{35}S in the sulfur-containing amino acids of the protein coats (capsids). (Note that DNA contains no sulfur, and viral protein no phosphorus.) The two batches of labeled phage were then added to separate suspensions of unlabeled bacteria. Each suspension of phage-infected cells was agitated in a blender to shear the viral capsids from the bacteria. The bacteria and empty viral coats (ghosts) were then separated by centrifugation. The cells infected with the ^{32}P-labeled phage were found to contain ^{32}P, indicating that the labeled viral DNA had entered the cells, and the viral ghosts contained no radioactivity. The cells infected with ^{35}S-labeled phage were found to have no radioactivity after blender treatment, but the viral ghosts contained ^{35}S. Progeny virus particles were produced in both batches of bacteria some time after the viral coats were removed, thus the genetic message for their replication had been introduced by viral DNA, not by viral protein.

the viral particle, not the sulfur-containing protein of the viral coat, that actually enters the host cell and furnishes the genetic information for viral replication (Fig. 12–13).

These important early experiments and many other lines of evidence have shown that DNA is definitely the exclusive chromosomal component bearing the genetic information of living cells.

DNAs Have Distinctive Base Compositions

A most 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 in 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 the organism's age, nutritional state, or changing environment.
4. In *all* DNAs, regardless of the species, the number of adenine residues is equal to the number of thymine residues (that is, A = T), and the number of guanine residues is equal to the number of cytosine residues (G = C). From these relationships it follows that the sum of the purine residues equals the sum of the pyrimidine residues; that is, A + G = T + C.

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

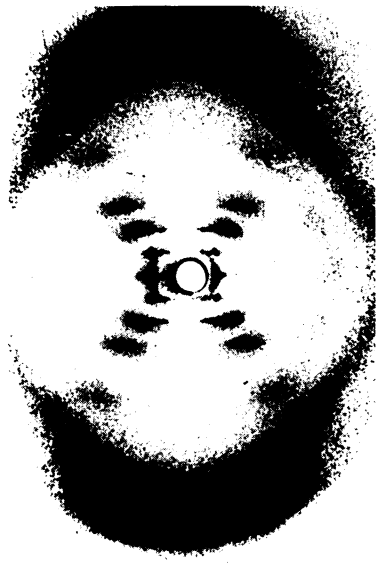


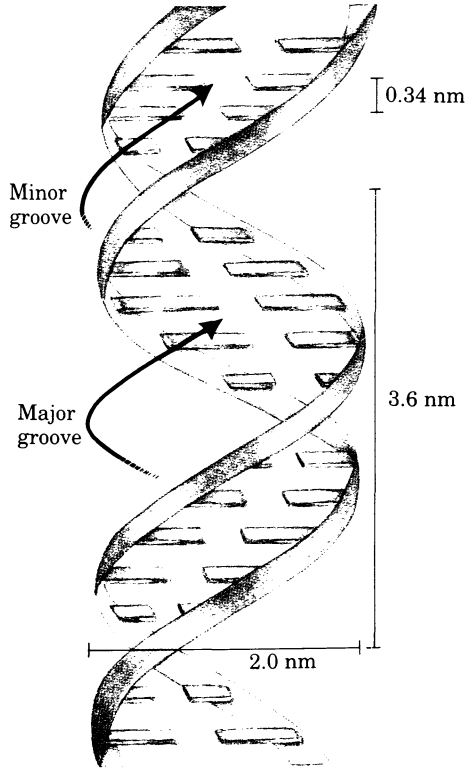
Figure 12–14 The x-ray diffraction pattern of DNA. The spots forming a cross in the center denote a helical structure. The heavy bands at the top and bottom correspond to the recurring bases.

DNA Is a Double Helix

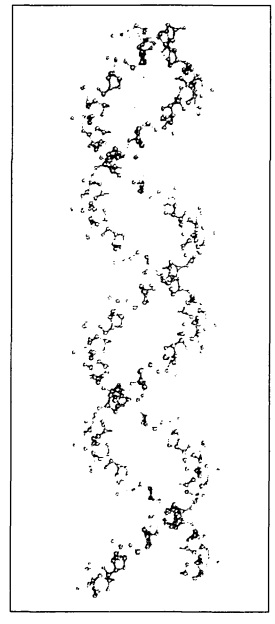
To shed more light on the structure of DNA, Rosalind Franklin and Maurice Wilkins used the powerful method of x-ray diffraction (see Box 7–3) to analyze DNA crystals. They showed in the early 1950s that DNA produces a characteristic x-ray diffraction pattern (Fig. 12–14). From this pattern it was deduced that DNA polymers are helical with two periodicities along their long axis, a primary one of 0.34 nm and a secondary one of 3.4 nm. The pattern also indicated that the molecule contains two strands, a clue that was crucial to determining the structure. 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.

In 1953 Watson and Crick postulated a three-dimensional model of DNA structure that accounted for all of the available data (Fig. 12–15). It consists of two helical DNA chains coiled around the same axis to form a right-handed double helix (see Box 7–1 for an explanation of the right- or left-handed sense of a helical structure). The hydrophilic backbones of alternating deoxyribose and negatively charged phosphate groups are on the outside of the double helix, facing the surrounding water. The purine and pyrimidine bases of both strands are stacked inside the double helix, with their hydrophobic and nearly planar ring structures very close together and perpendicular to the long axis of the helix. The spatial relationship between these strands creates a **major groove** and **minor groove** between the two strands. Each 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 12–11 are those that fit best within the structure, providing a rationale for Chargaff's rules. It is important to note that three hydrogen bonds can form between G and C, symbolized G≡C, but only two can form between A and T, symbolized A=T. Other pairings of bases tend (to varying degrees) to destabilize the double-helical structure.

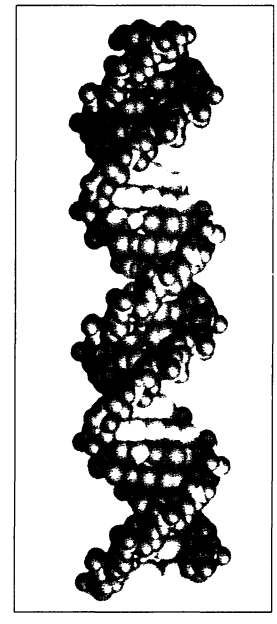
Figure 12–15 The Watson–Crick model for the structure of DNA. The original model proposed that there are 10 base pairs or 3.4 nm per turn of the helix. Subsequent measurements have shown that there are 10.5 base pairs or 3.6 nm per turn. **(a)** Schematic representation, showing dimensions of the helix. **(b)** Line model showing the backbone and stacking of the bases. **(c)** Space-filling model.



(a)



(b)



(c)

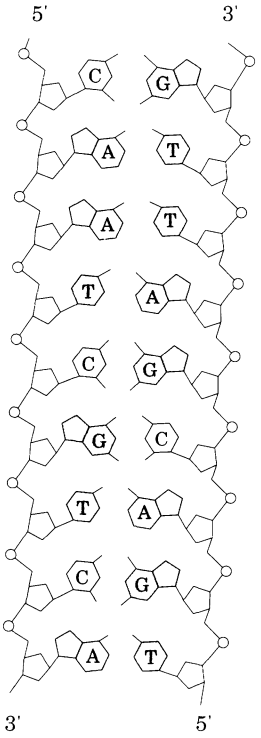


Figure 12–16 Schematic drawing of complementary antiparallel strands of DNA following the pairing rules proposed by Watson and Crick. The base-paired antiparallel strands differ in base composition: the left strand has the composition A_3 T_2 G_1 C_3; the right, A_2 T_3 G_3 C_1. They also differ in sequence when each chain is read in the 5′ → 3′ direction. Note the base equivalences: A = T and G = C. In this and following illustrations, the hydrogen bonds between base pairs are often represented by sets of blue lines.

In the Watson–Crick structure, the two chains or strands of the helix are **antiparallel;** their 5′,3′-phosphodiester bonds run in opposite directions. Later work with DNA polymerases (Chapter 24) provided experimental evidence, confirmed by x-ray crystallography, that the strands are indeed antiparallel.

To account for the periodicities observed in the x-ray diffraction pattern, Watson and Crick used molecular models to show that the vertically stacked bases inside the double helix would be 0.34 nm apart and that the secondary repeat distance of about 3.4 nm could be accounted for by the presence of 10 (now 10.5) nucleotide residues in each complete turn of the double helix (Fig. 12–15a). As can be seen in Figure 12–16, 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 appears in one chain, thymine is found in the other; similarly, wherever guanine is found in one chain, cytosine is found in the other.

The DNA double helix or duplex is held together by two sets of forces, as described earlier: hydrogen bonding between complementary base pairs (Fig. 12–11) and base-stacking interactions. The specificity that maintains a given base sequence in each DNA strand is contributed entirely by 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. Making a copy of this structure (replication) could logically proceed by (1) separating the two strands and (2) synthesizing a complementary strand for each by joining nucleotides in a sequence specified by the base-paring rules stated above. Each preexisting strand could function as a template to guide the synthesis of the complementary strand (Fig. 12–17). These expectations have been experimentally confirmed, and this discovery was a revolution in our understanding of DNA metabolism.

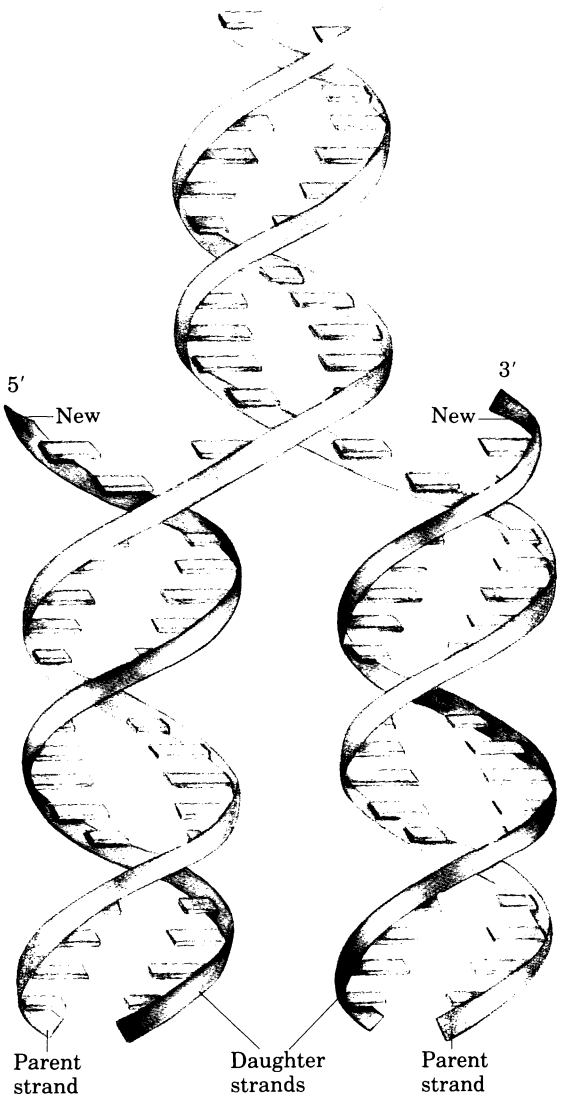


Figure 12–17 Replication of DNA as suggested by Watson and Crick. The parent strands become separated, and each forms the template for biosynthesis of a complementary daughter strand (in red).

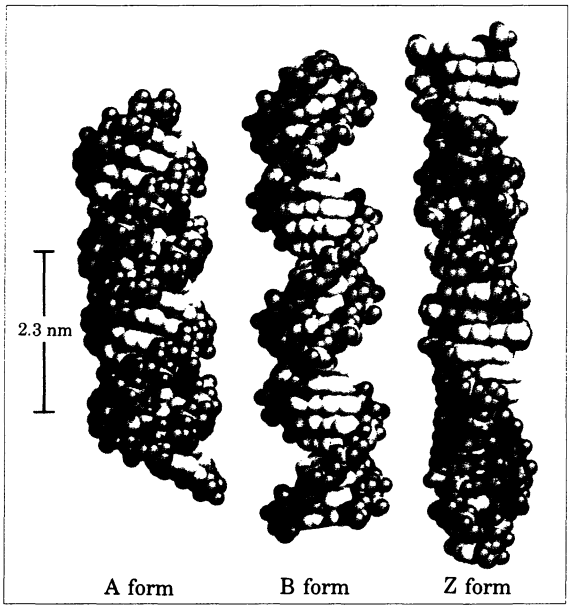


Figure 12–18 Comparison of the A, B, and Z forms of DNA. There are 24 base pairs in each of the structures shown.

DNA Can Occur in Different Structural Forms

DNA is a remarkably flexible molecule. Considerable rotation is possible around a number of bonds in the sugar–phosphate backbone, and thermal fluctuation can produce bending, stretching, and unpairing (melting) in the structure. Many significant deviations from the Watson–Crick DNA structure are found in cellular DNA, and some or all of these may play important roles 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=T and G≡C base pairs.

The Watson–Crick structure is also referred to as B-form 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 DNA structural variants that have been well characterized in crystal structures are the A and Z forms (Fig. 12–18). 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 rise per base pair is 0.23 nm and the number of base pairs per helical turn is 11, relative to the 0.34 nm rise and 10.5 base pairs per turn found in B-DNA. For a given DNA molecule, the A form will be shorter and have a greater diameter than the B form. The reagents used to promote crystallization of DNA tend to dehydrate it, and this leads to a tendency for many DNAs 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, with a rise of 0.38 nm per base pair. The DNA backbone takes on a zig-zag appearance. Certain nucleotide sequences fold up into left-handed Z helices more readily than do others. Prominent examples are sequences in which pyrimidines alternate with purines, especially alternating C and G or 5-methyl-C and G. Whether A-form DNA actually occurs in cells is uncertain, but there is evidence for some short stretches (tracts) of Z-DNA in both prokaryotes and eukaryotes. These Z-DNA tracts may play an as yet undefined role in the regulation of the expression of some genes or in genetic recombination.

Certain DNA Sequences Adopt Unusual Structures

A number of other sequence-dependent structural variations have been detected that may serve locally important functions in DNA metabolism. For example, some sequences cause bends in the DNA helix. Bends are produced whenever four or more adenine residues appear sequentially in one of the two strands (Fig. 12–19). Six adenines 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.

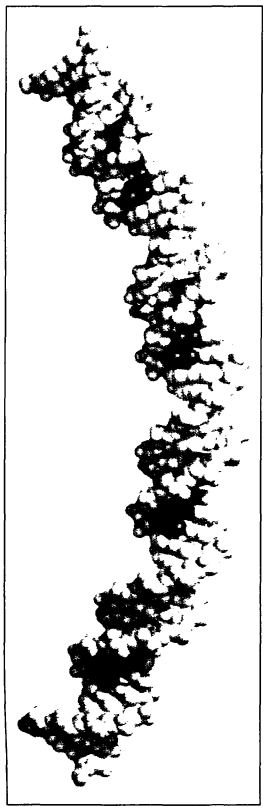


Figure 12–19 A model for the bending of DNA produced by poly(A) tracts. The bend here is produced by four $(dA)_5$ tracts, separated by five base pairs. The adenine bases are shown in red.

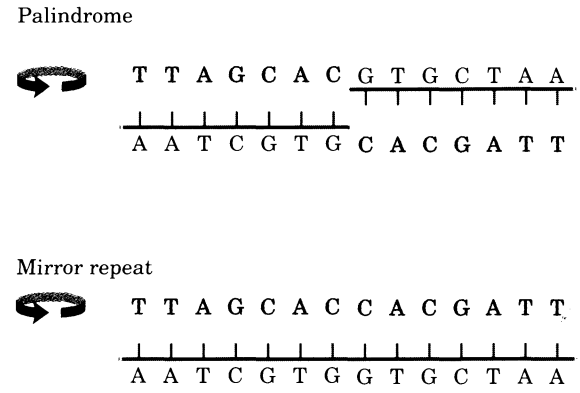


Figure 12–20 Palindromes and mirror repeats. Palindromes are defined in nucleic acids as sequences with twofold symmetry. In order to superimpose one repeat (shaded sequence) on the other, it must be rotated 180° around the horizontal axis and then again about the vertical axis, as shown by the colored arrows. A mirror repeat, on the other hand, has a symmetric sequence on each strand. Superimposing one repeat on the other requires only a single 180° rotation about the vertical axis.

A rather common type of sequence found in DNA is a **palindrome.** A palindrome is a word, phrase, or sentence that is spelled identically reading forward or backward; two examples are ROTATOR and NURSES RUN. The term is applied to regions of DNA in which there are **inverted repetitions** of base sequence with twofold symmetry occurring over two strands of DNA (Fig. 12–20). Such sequences are self-complementary within each of the strands and therefore have the potential to form hairpin or cruciform (cross-shaped) structures (Fig. 12–21). When the inverted sequence 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 involve a few or up to thousands of base pairs. It is not known how many palindromes actually occur as cruciforms in cells, although the existence of at least some cruciform structures has been demonstrated in vivo in *E. coli*. Self-complementary sequences cause isolated single strands of DNA to fold up in solution into complex structures containing multiple hairpins.

A particularly unusual DNA structure, known as H-DNA, is found in polypyrimidine/polypurine tracts that also incorporate a mirror repeat within the sequence. One simple example is a long stretch of alternating T and C residues, as shown in Figure 12–22. A novel feature of H-DNA is the pairing and interwinding of *three* strands of DNA to form a triple helix. Triple-helical DNA forms spontaneously only within long sequences containing only pyrimidines (or only purines) in one strand. Two of the three strands in the H-DNA triple helix (Fig. 12–22c, d) contain pyrimidines and the third contains purines.

These structural variations are interesting because there is a tendency for many of them to appear at sites where important events in DNA metabolism (replication, recombination, transcription) are initi-

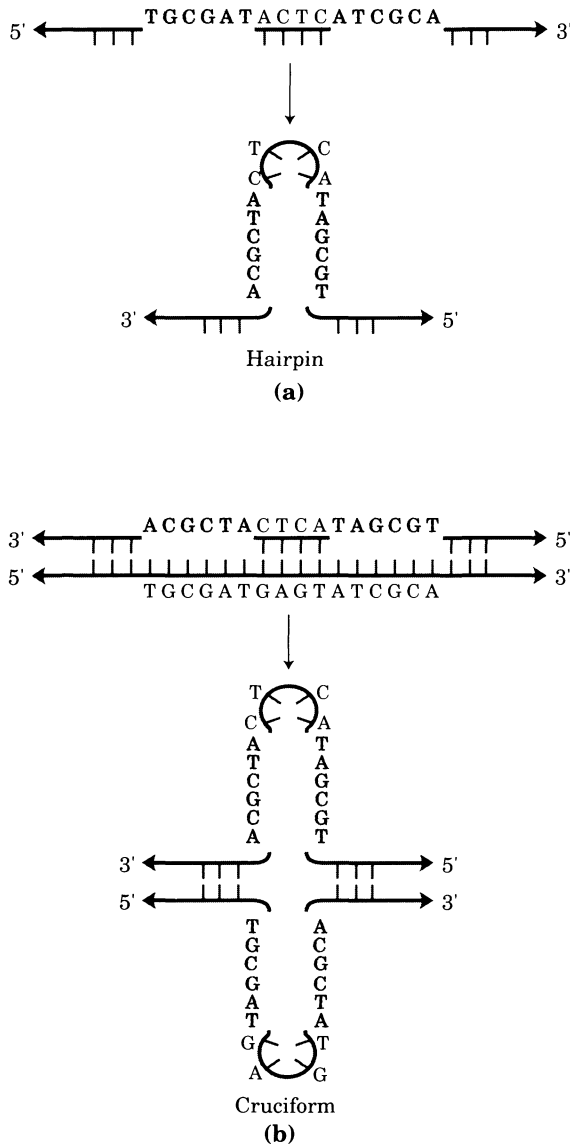
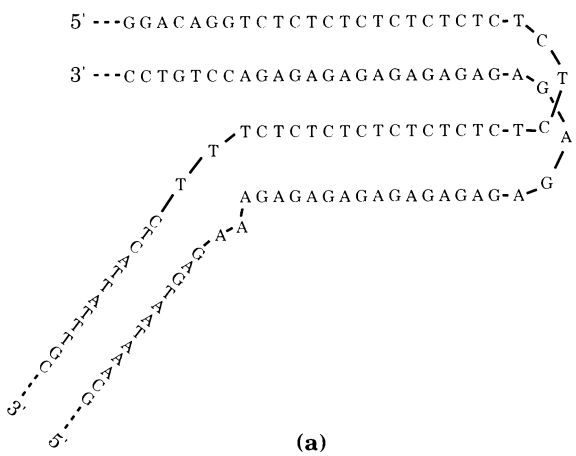
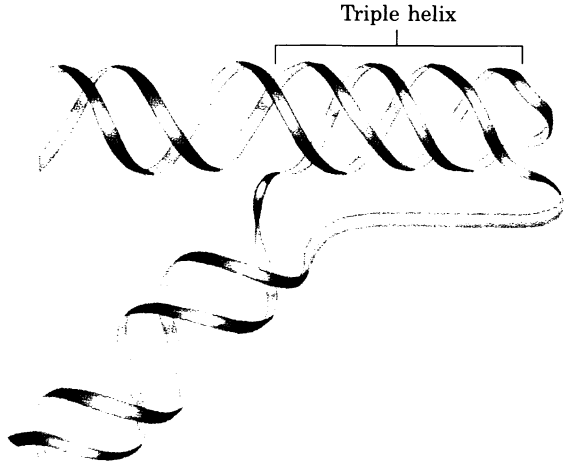


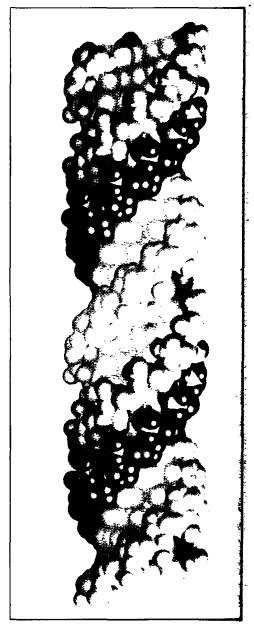
Figure 12–21 Hairpins and cruciforms. Palindromic DNA (or RNA) sequences can form alternative structures with intrastrand base pairing. When only a single DNA (or RNA) strand is involved it is called a hairpin **(a).** When both strands of a duplex DNA are involved, the structure is called a cruciform **(b).** Blue shading highlights asymmetric sequences that can pair alternatively with a complementary sequence in the same or opposite strand.



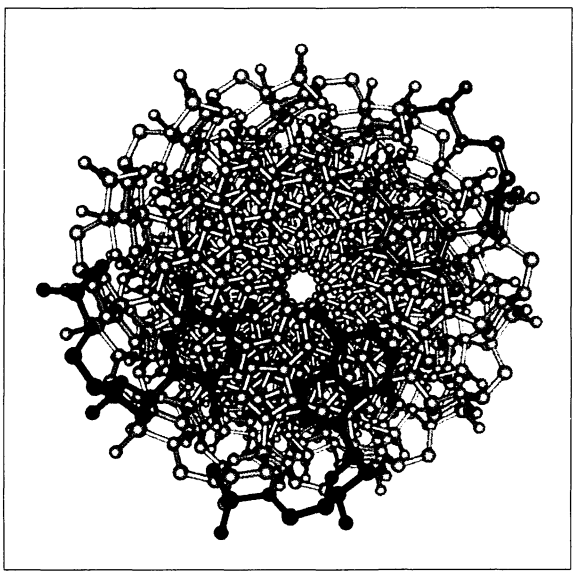
(a)



(b)



(c)



(d)

Figure 12–22 H-DNA. A sequence of alternating T and C residues can be considered a mirror repeat centered about one of the central T or C residues **(a).** These sequences form an unusual structure in which the strands in one half of the mirror repeat are separated, and the pyrimidine-containing strand folds back on the other half of the repeat to form a triple helix **(b).** The purine strand (alternating A and G residues) is left unpaired. This structure produces a sharp bend in the DNA. **(c)** A triple-helical DNA formed from two pyrimidine strands (polydeoxythymidine, shown with gray and light blue backbones) and one purine strand (polydeoxyadenine, with a dark blue backbone). Phosphorus atoms are shown in yellow. In this structure the light blue and dark blue strands are antiparallel and paired via normal Watson–Crick base pairing patterns. The third (gray) strand is parallel to the dark blue (purine) strand and paired through non–Watson–Crick hydrogen bonds, including one between the C-4 carbonyl group of thymine and the N-7 of adenine. **(d)** An end view of the triple-helical DNA shown in **(c),** with the base triplet at one end.

ated or regulated. For example, the sites recognized by many sequence-specific DNA-binding proteins (Chapter 27) are arranged as palindromes, and sequences that can form H-DNA are found within regions involved in the regulation of expression of a number of genes in eukaryotes. Much work is still required to define these structures and determine their functional significance.

Messenger RNAs Code for Polypeptide Chains

We now turn our attention briefly from DNA structure to the expression of the genetic information contained in DNA. RNA, the second major form of nucleic acid in cells, plays the role of intermediary in converting this information into a functional protein.

In eukaryotes DNA is largely confined to the nucleus, whereas protein synthesis occurs on ribosomes in the cytoplasm. Therefore some molecule other than DNA must carry the genetic message for protein synthesis 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 cytoplasm, and the onset of protein synthesis is accompanied by an increase in the amount of RNA in the cytoplasm 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, Francois Jacob and Jacques Monod presented a unified (and essentially correct) picture of many aspects of this process. They proposed the name messenger RNA (mRNA) for that portion of the total cell RNA carrying the genetic information from DNA to the ribosomes, where the messengers provide the templates for specifying amino acid sequences in polypeptide chains. Although mRNAs from different genes can vary greatly in length, the mRNAs from a particular gene will generally have a defined size. The process of forming mRNA on a DNA template is known as transcription.

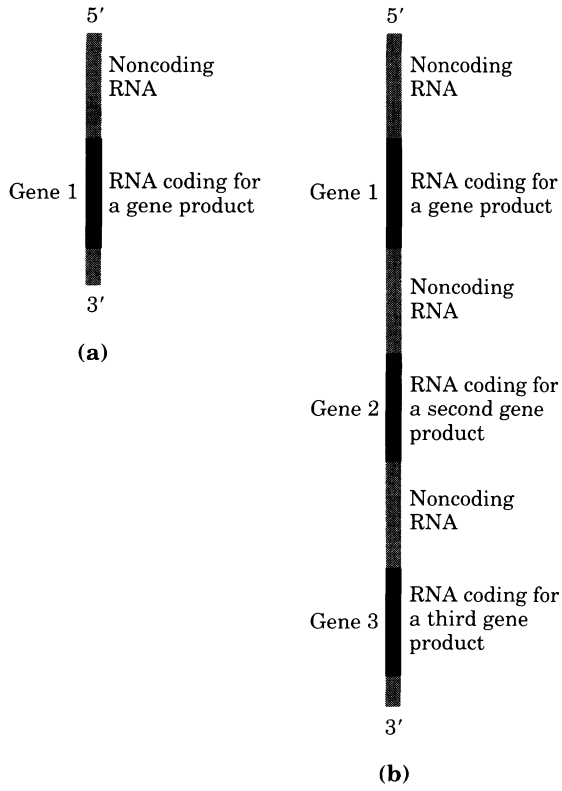


Figure 12–23 Schematic diagram of monocistronic **(a)** and polycistronic **(b)** mRNAs of prokaryotes. The polycistronic transcript shown here contains three genes. Noncoding RNA separates the genes.

In prokaryotes 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. (The term **cistron,** for purposes of this discussion, 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 (Chapter 26). However, mRNAs transcribed from DNA are always somewhat longer than needed simply to specify the code for the polypeptide sequence(s). The additional noncoding RNA includes sequences that regulate protein synthesis (Chapter 26). Figure 12–23 summarizes the general structure of prokaryotic mRNAs.

Many RNAs Have More Complex Structures

Messenger RNA is only one of several classes of cellular RNA. Transfer RNAs serve as adapter molecules in protein synthesis; covalently linked to an amino acid at one end, they pair with the mRNA in such a way that the amino acids are joined in the correct sequence. Ribosomal RNAs are structural components of ribosomes. There is also a wide variety of special-function RNAs. All of these are considered in detail in Chapter 25.

Regardless of the class of RNA being synthesized, the product of transcription is always a single strand of RNA. The single-stranded nature of these molecules does not mean their structure is random. The single strands tend to take up a right-handed helical conformation that is dominated by base-stacking interactions (Fig. 12–24). The stacking interactions are stronger between two purines than between a purine and a pyrimidine or between two pyrimidines. The purine–purine interaction is so strong that a pyrimidine separating two purines will often be displaced from the stacking pattern so that the purines can interact. Any self-complementary sequences in the molecule will lead to more complex and specific structures. RNA can base-pair with complementary strands of either RNA or DNA. The standard base-pairing rules are identical to those for DNA: guanine pairs with cytosine and adenine pairs with uracil (or thymine). One difference is that one unusual base pairing—between guanine and uracil—is fairly common between two strands of RNA; see Fig. 12–26. The paired strands in RNA or RNA–DNA are antiparallel, as in DNA.

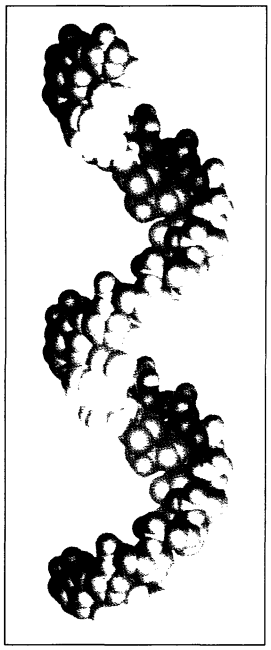
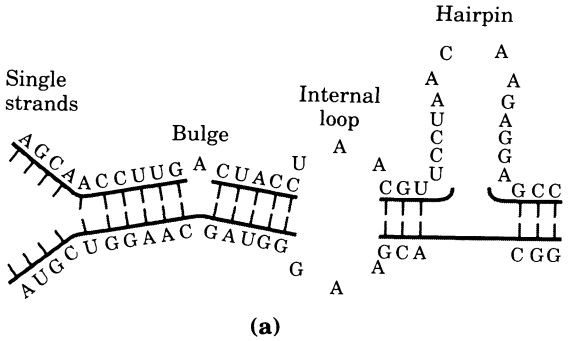


Figure 12–24 Typical right-handed stacking pattern found in a single strand of RNA. The bases are shown in white, the ribose rings in green, and the phosphate atoms in yellow. The right-handed twist of the backbone is evident.



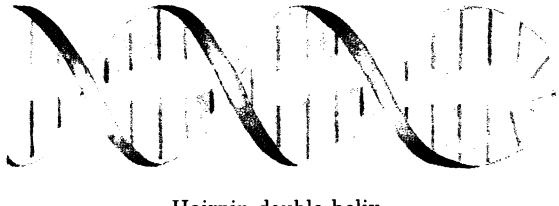
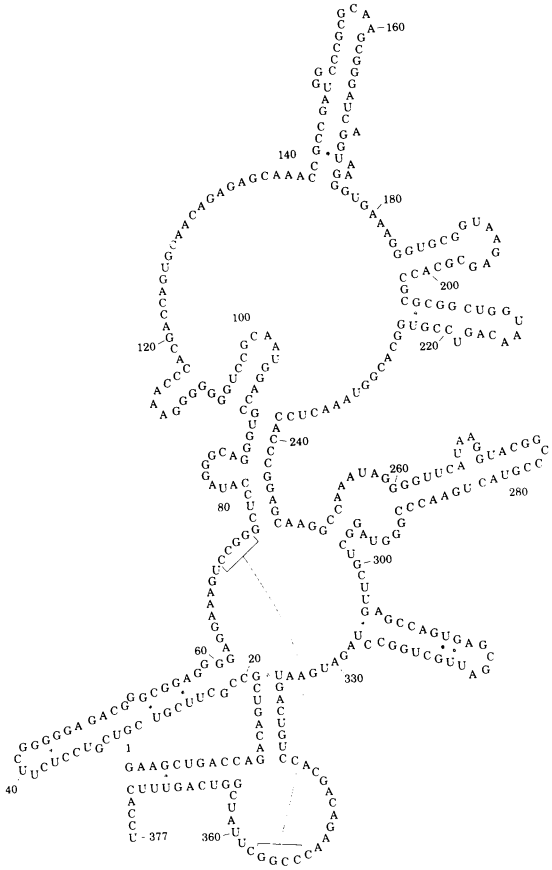


Figure 12–25 (a) Types of secondary structure found in some RNAs. The paired regions generally have an A-form right-handed helix, as shown for a hairpin **(b).**

Unlike the double helix of DNA, there is no simple, regular secondary structure that forms a reference point for RNA structure. The three-dimensional structures of many RNAs, like those of proteins, are complex and unique. Weak interactions, especially base-stacking (hydrophobic) interactions, again play a major role in stabilizing structures. Where complementary sequences are present, the predominant double-stranded structure is an A-form right-handed double helix. 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. 12–25). Hairpin loops form between nearby self-complementary sequences in the RNA strand (Fig. 12–25). The potential for base-paired helical structures in many RNAs is extensive (Fig. 12–26), and the resulting hairpins can be considered the most common type of secondary structure in RNA. Certain 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 play an important role in nucleating the folding of an RNA molecule into its precise three-dimensional structure. Important additional structural 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 form a hydrogen bond with other groups, and a variety of nonstandard base-pairing patterns are also observed. Some of these properties are evident in the structure of the phenylalanine transfer RNA of yeast (Fig. 12–27).



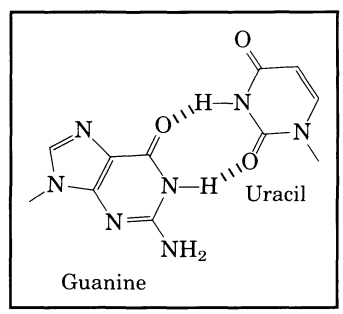
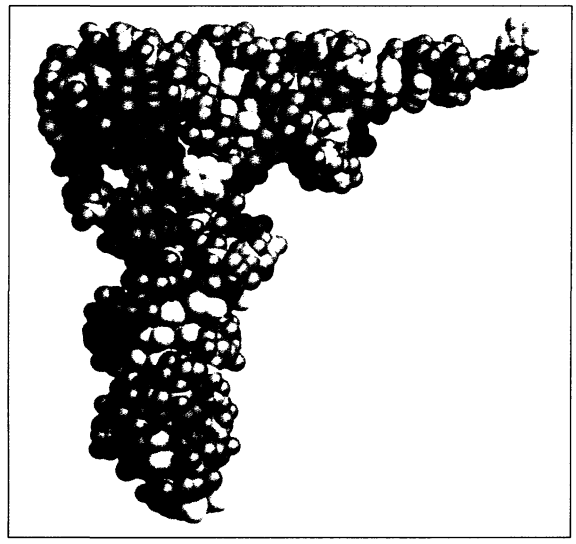
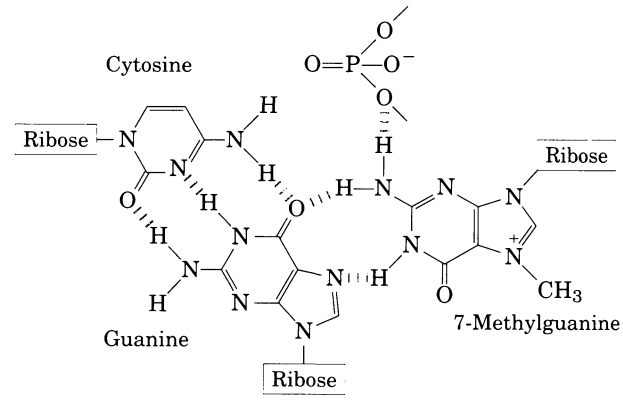


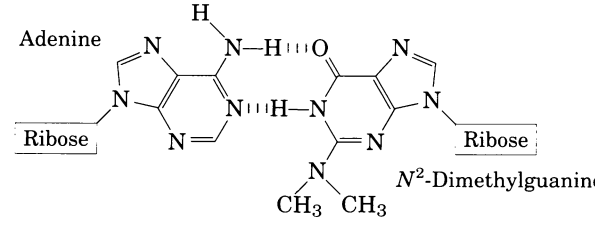
Figure 12–26 Possible secondary structure of the M1 RNA component of the enzyme RNase P of *E. coli,* showing many hairpins. RNase P also contains a protein component (not shown). This enzyme functions in the processing of transfer RNAs, as described in Chapter 25. Brackets indicate additional complementary sequences that may be paired in the three-dimensional structure. The dots (•) indicate non–Watson–Crick G=U base pairs, as shown above.

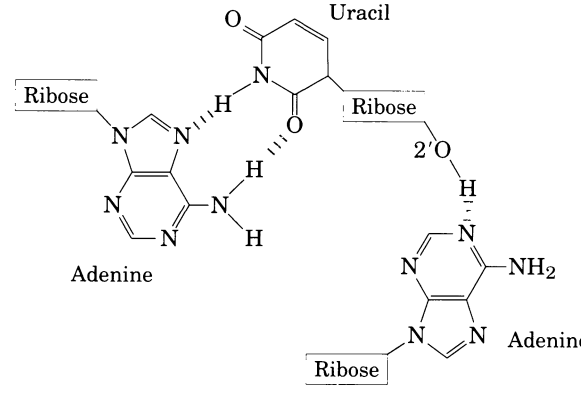


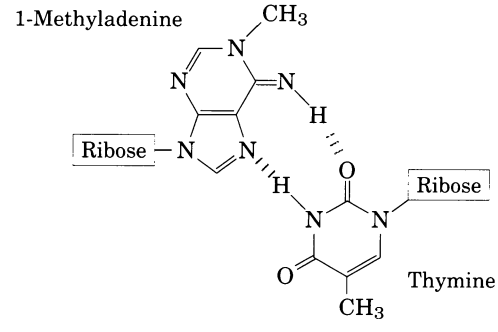
(a)

Figure 12–27 Phenylalanine tRNA of yeast. **(a)** Three-dimensional structure. **(b)** Some unusual base-pairing patterns. Note also the involvement of a phosphodiester bond oxygen in one hydrogen-bonding arrangement, and the involvement of a 2′-hydroxyl group in another (both in red).









(b)

The analysis of RNA structure and its relationship to 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 aware of an increasing number of functions of RNA molecules.

Nucleic Acid Chemistry

To understand how nucleic acids function, we must understand their chemical properties as well as their structures. DNA functions well as a repository of genetic information in part because of 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. Nondestructive alterations, such as the strand separation that must precede DNA replication or transcription, are also important. In addition to providing these 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 (20 to 25 °C). When such a solution is subjected to extremes of pH or to temperatures above 80 to 90 °C, its viscosity decreases sharply, indicating that the DNA has undergone a physical change. Just as heat and extremes of pH cause denaturation of globular proteins, so too will they cause denaturation or melting of double-helical DNA. This involves disruption of the hydrogen bonds between the paired bases and the hydrophobic interactions between the stacked bases. As a result, the double helix unwinds to form two single strands,

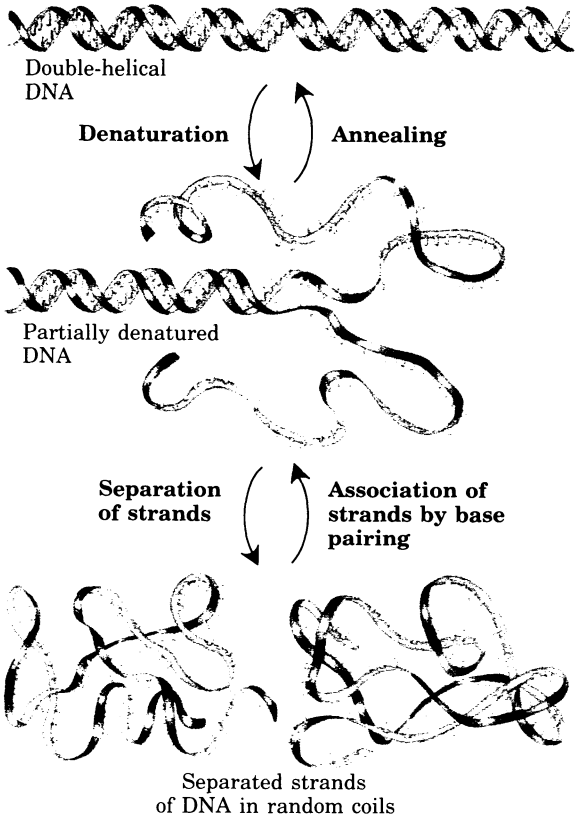


Figure 12–28 Stages in the reversible denaturation and annealing (renaturation) of DNA.

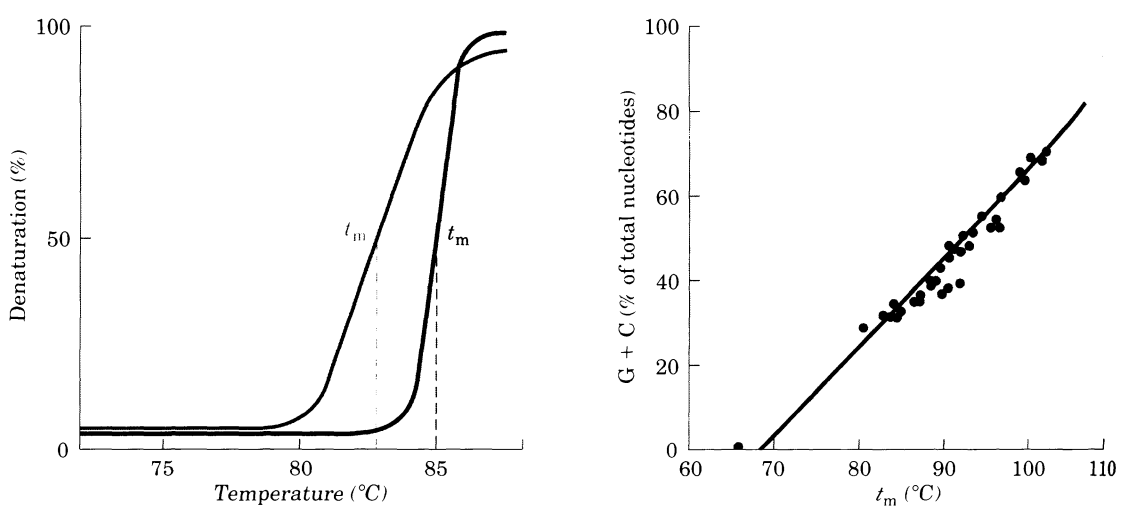
completely separate from each other along the entire length, or part o the length (partial denaturation), of the molecule. No covalent bonds i the DNA are broken (Fig. 12–28).

Renaturation of DNA is a rapid one-step process as long as a double-helical segment of a dozen or more residues still unites the tw strands. When the temperature or pH is returned to the biologica range, the unwound segments of the two strands spontaneously re wind or **anneal** to yield the intact duplex (Fig. 12–28). However, if th two strands are completely separated, renaturation occurs in tw steps. The first step is relatively slow, because the two strands mus first "find" each other by random collisions and form a short segment o complementary double helix. The second step is much faster: the re maining unpaired bases successively come into register as base pairs and the two strands "zipper" themselves together to form the doubl helix.

Viral or bacterial DNA molecules in solution denature at charac teristic temperatures when they are heated slowly (Fig. 12–29). The transition from double-stranded DNA to the single-stranded, dena tured form can be detected by an increase in the absorption of UV light (the hyperchromic effect) or a decrease in the viscosity of the DNA solution. Each species of DNA has a characteristic denaturation tem perature or melting point: 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, are more stable and require more heat energy to dissociate than A=T base pairs. Careful determination of the melting point of a DNA specimen, 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 denatured regions can be visualized with electron microscopy (Fig. 12–30). Strand separation of DNA *must* occur in vivo during processes such as DNA replication and transcription. As we will see, the DNA sites where these processes are initiated are often rich in A=T base pairs.

Double-stranded nucleic acids with two RNA strands or with one RNA strand and one DNA strand (RNA–DNA hybrids) can also be denatured. Notably, RNA duplexes are more stable than DNA du-

Figure 12–29 (a) The denaturation or melting curve of two DNA specimens. The temperature at the midpoint 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, in a solution containing 0.15 M NaCl and 0.015 M sodium citrate.



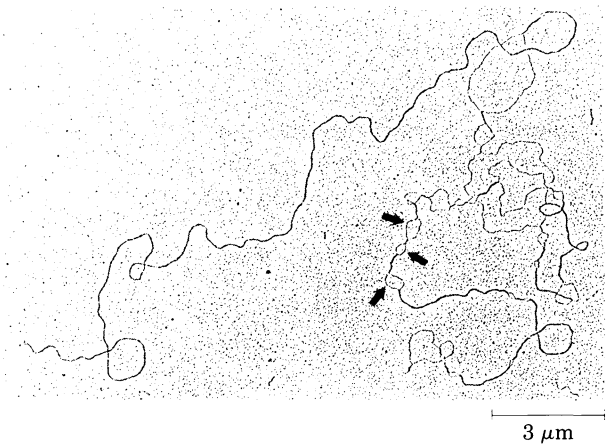


Figure 12–30 Electron micrograph of partially denatured DNA. Few structural details are evident. The shadowing method used to visualize the DNA increases its diameter approximately fivefold and obliterates the details of the helix. However, length measurements can be obtained, and single-stranded regions are readily distinguished from double-stranded regions. The arrows point to some single-stranded bubbles in the DNA, where denaturation has occurred. The regions that denature are highly reproducible and are rich in A=T base pairs.

plexes. At neutral pH, a double-helical RNA will often denature at temperatures 20 °C or more higher than a DNA molecule with a comparable sequence. The stability of an RNA–DNA hybrid is generally intermediate between that of RNA and that of DNA. The physical basis for these differences in stability is not known.

Nucleic Acids from Different Species Can Form Hybrids

The capability 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 65 °C for many hours, much of the DNA will anneal. Most of the mouse DNA strands anneal with complementary mouse DNA strands to form mouse duplex DNA; similarly, many of the 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 the mouse DNA strand form base-paired regions with segments of the human DNA strand (Fig. 12–31). This reflects the fact that different organisms have some common evolutionary heritage; they generally have some proteins and RNAs with similar functions and, often, similar structures. In many cases, the DNA encoding these proteins and RNAs will have similar (homologous) sequences. The closer the evolutionary relationship between the species, the more extensively will their DNAs 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 of a powerful set of techniques essential to the modern practice of molecular genetics. It is possible to detect a specific DNA sequence or gene 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 28). The complementary DNA can be from a different species or from the same species; in some cases it is synthesized 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 relies on these techniques, and new applications of this technology are making it possible to accurately identify an individual on the basis of a single hair left at the scene of a crime or predict the onset of some diseases in an individual decades before symptoms appear (see Box 28–1).

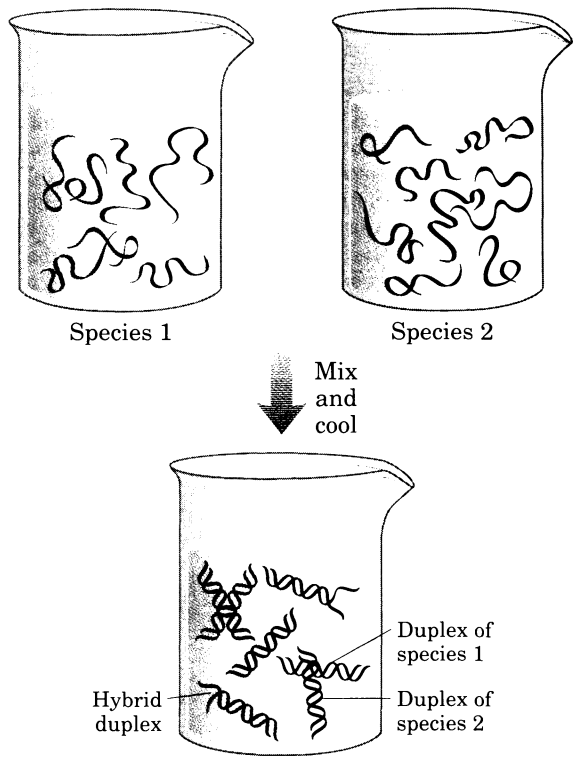
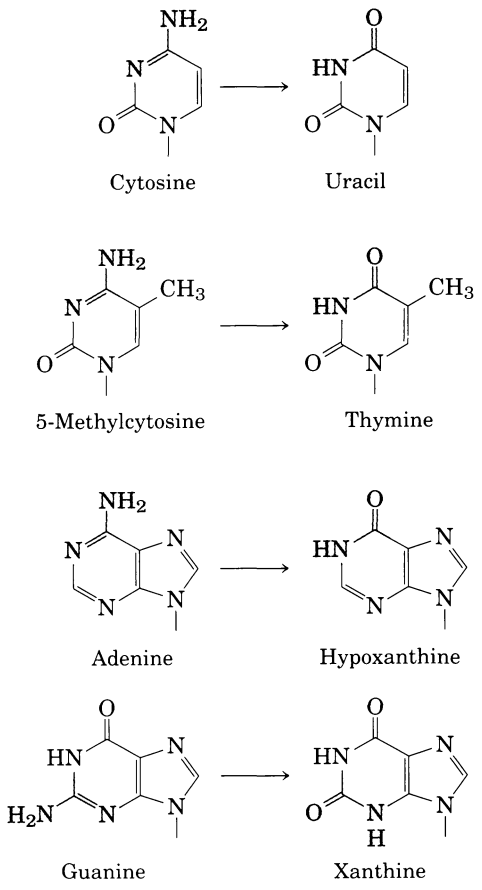


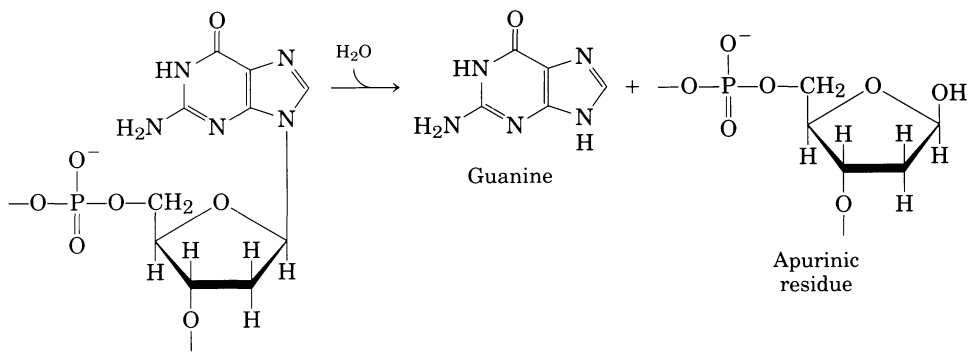
Figure 12–31 Principle of the hybridization test. Two DNAs from different species are completely denatured by heating. When mixed and slowly cooled, complementary DNA strands of each species will associate and anneal to form normal duplexes. If the different DNAs have significant sequence homology, they will tend to form partial duplexes or hybrids with each other: the greater the sequence homology between the two DNAs, the greater the number of hybrids formed. Hybrid formation can be measured by different procedures, such as chromatography or isopycnic centrifugation. Usually one of the DNAs is labeled with a radioactive isotope to simplify the measurements.

Deamination



(a)

Depurination



(b)

Figure 12–32 Some well-characterized reactions of nucleotides. **(a)** Deamination reactions. Only the base is shown. **(b)** Depurination, in which a purine is lost by hydrolysis of the *N*-glycosyl bond. The deoxyribose remaining after depurination is readily converted from the β-furanose to the aldehyde form (see Fig. 12–3).

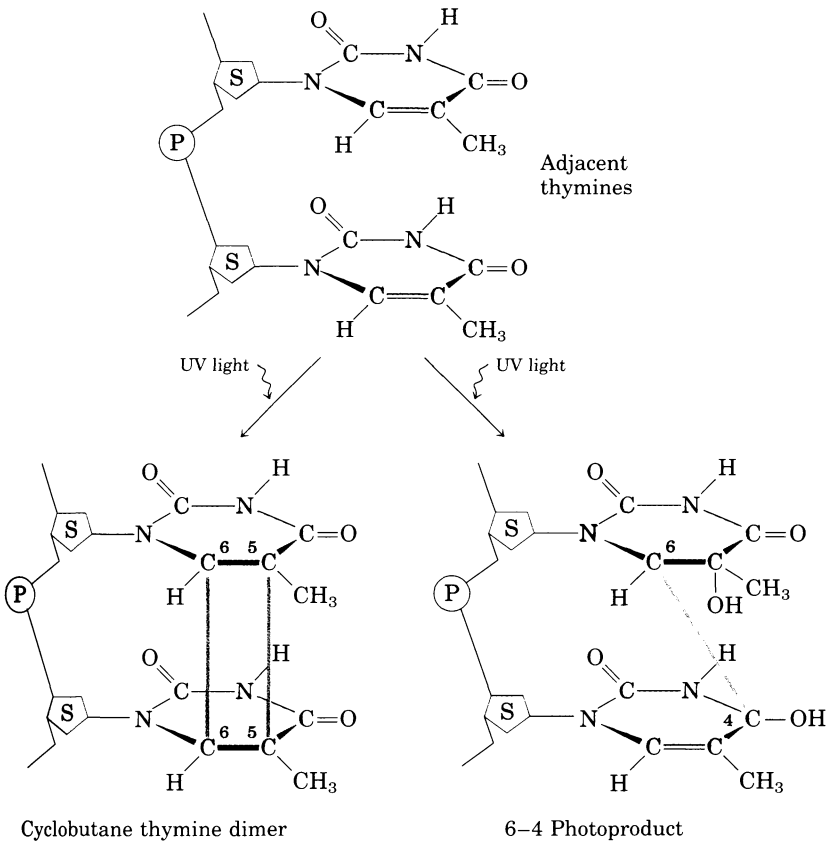
Nucleotides and Nucleic Acids Undergo Nonenzymatic Transformations

Purines and pyrimidines, along with the nucleotides of which they are a part, undergo a number of reactions involving spontaneous alteration of their covalent structure. These reactions are generally *very slow,* but they are physiologically significant because of the cell's very low tolerance for alterations in genetic information. Alterations in DNA structure that lead to permanent changes in the genetic information encoded therein are called **mutations,** and much evidence suggests an intimate link between the accumulation of mutations and the processes of aging and cancer.

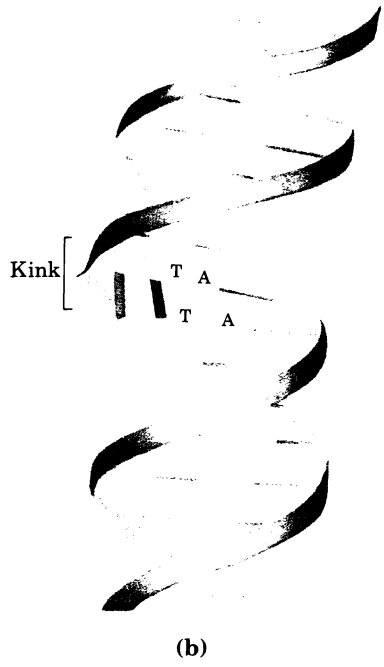
Several bases undergo spontaneous loss of their exocyclic amino groups (deamination) (Fig. 12–32a). For example, under conditions found in a typical cell, deamination of cytosine (in DNA) to uracil will occur in about one of every 10^7 cytosines in 24 h. This corresponds to about 100 spontaneous events per day in an average mammalian cell. Deamination of adenine and guanine is about 100 times slower.

The slow cytosine deamination reaction seems innocuous enough, but it 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 24). 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≡C base pairs and an increase in A=U base pairs in the DNA of all cells. Over the millennia, the cytosine deamination reaction could eliminate G≡C base pairs and the genetic code that depends on them. Establishing thymine as one of the four bases in DNA may well have been one of the key turning points in evolution, making the long-term storage of genetic information possible.

Another important reaction in deoxynucleotides is the hydrolysis of the glycosyl bond between the base and the pentose (Fig. 12–32b). This occurs much faster for purines than for pyrimidines. In DNA as many as one in 10^5 purines (10,000 per mammalian cell) are lost every 24 h 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.**



(a)



(b)

Figure 12–33 Formation of thymine dimers induced by UV light. **(a)** The reaction on the left results in the formation of a cyclobutyl ring involving C-5 and C-6 of each thymine residue, just as UV light will induce the formation of cyclobutane from two molecules of ethylene. An alternative light-induced reaction between adjacent thymines results in a linkage between C-6 of one residue and C-4 of its neighbor, as shown on the right. A bend or kink is introduced into the DNA on formation of a cyclobutane thymine dimer **(b)**.

Other reactions are promoted by certain types of radiation. In the laboratory, UV light will induce the condensation of two ethylene groups to form a cyclobutane ring. In the cell, the same reaction occurs between adjacent pyrimidine bases in nucleic acids to form cyclobutane pyrimidine dimers. This happens most frequently between adjacent thymine residues on the same DNA strand (Fig. 12–33). A second type of pyrimidine dimer formed during UV irradiation, called a 6–4 photoproduct, is also shown in Figure 12–33. 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. UV radiation (having wavelengths of 200 to 400 nm), which makes up a significant portion of the solar spectrum, can cause pyrimidine dimer formation and other chemical changes in the DNA of bacteria and of human skin cells. There is a constant field of ionizing radiation around us 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}C, and ^{3}H. X rays used in medical or 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 nonbiological agents.

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 three major classes of such reactive chemical agents

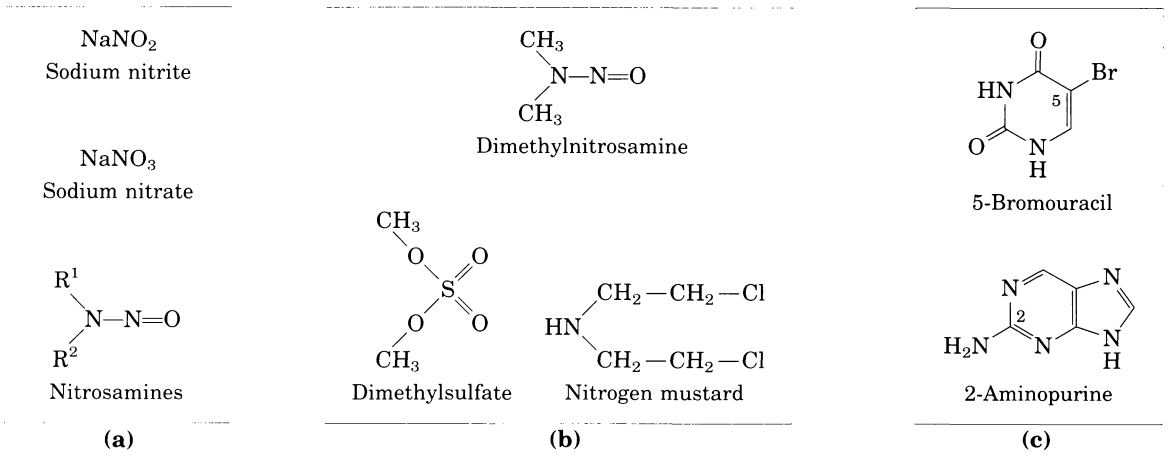
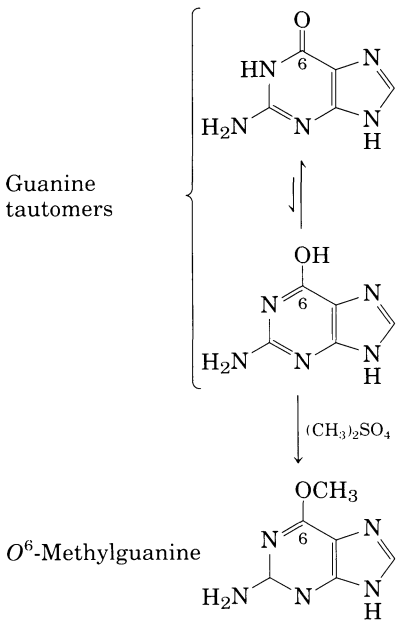


Figure 12–34 Chemical agents that cause some types of DNA damage. **(a)** Nitrous acid precursors. **(b)** Alkylating agents. **(c)** Base analogs.



(Fig. 12–34): (1) deaminating agents, particularly nitrous acid (HNO_2) or compounds that can be metabolized to nitrous acid or nitrites, (2) alkylating agents, and (3) compounds that can simulate or mimic the normal bases present in DNA.

Nitrous acid, formed from organic precursors such as nitrosamines and from nitrite and nitrate salts, is a potent reagent that accelerates the deamination of bases described above. Bisulfite has similar effects. Both agents are used as preservatives in processed foods to prevent the growth of toxic bacteria. They do not appear to significantly increase cancer risks when used in this way, perhaps because they are used in small amounts and their contribution to the overall levels of DNA damage is minor. (The potential health risk from food spoilage if these preservatives were not used is much greater.)

Alkylating agents can alter certain bases of DNA. For example, the highly reactive chemical dimethylsulfate (Fig. 12–34b) can methylate a guanine residue to yield O^6-methylguanine, which is unable to base-pair with cytosine. Many similar reactions are brought about by alkylating agents normally present in cells, such as *S*-adenosylmethionine (see Fig. 17–20) and other compounds.

Possibly 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. Cells possess an elaborate defense system to destroy these reactive species, including enzymes such as catalase and superoxide dismutase. A fraction of these oxidants inevitably escapes cellular defenses, however, and damage to DNA involves a large, complex group of reactions ranging from oxidation of sugar and base moieties to breaking strands. Accurate estimates for the extent of this damage are not yet available, but it is clear that each day the DNA in each human cell is subject to thousands of damaging oxidative reactions.

This is merely a sampling of the best-understood reactions. Many carcinogenic compounds present in food, water, or air exert their cancer-causing effects by modifying bases in DNA. In the cell, the integrity of DNA as a polymer is nevertheless maintained better than that of either RNA or protein, because DNA is the only macromolecule having biochemical repair systems. These repair processes (described in Chapter 24) greatly lessen the impact of damage to DNA.

DNA Is Often Methylated

Certain nucleotide bases in DNA molecules are often enzymatically methylated. Adenine and cytosine are methylated more often than guanine and thymine. Methylation of these bases is not random but 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 is still unclear. All known DNA methylases use *S*-adenosylmethionine as a methyl group donor. In *E. coli* there are two prominent methylation systems. One serves as part of a cellular defense mechanism that helps to distinguish the cell's own DNA from foreign DNA (restriction modification, described in Chapter 28). The other system methylates adenine to N^6-methyladenine (see Fig. 12–5a) within the sequence (5′)GATC(3′). This is mediated by an enzyme called the Dam methylase, which functions as part of a system that repairs mismatched base pairs formed occasionally during DNA replication (Chapter 24).

In eukaryotic cells, about 5% of cytosine residues are methylated to form 5-methylcytosine (see Fig. 12–5a). 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 in different regions of large eukaryotic DNA molecules, and is often inversely related to the degree of gene expression. These methylations have structural as well as regulatory significance. The presence of 5-methylcytosine in an alternating CpG sequence markedly increases the tendency for that sequence to take up the Z conformation.

Long DNA Sequences Can Be Determined

In its capacity as a repository of information, the most important property of a DNA molecule is its nucleotide sequence. Until the late 1970s, obtaining the sequence of a nucleic acid containing even five or ten nucleotides was difficult and very laborious. The development of two new techniques in 1977, one by Alan Maxam and Walter Gilbert and the other by Frederick Sanger, has made it possible to sequence ever larger DNA molecules with an ease unimagined just a few decades ago. The techniques depend upon an improved understanding of nucleotide chemistry and DNA metabolism, and on electrophoretic methods that allow the separation of DNA strands differing in size by only one nucleotide. Electrophoresis of DNA is similar to the electrophoresis of proteins (see Fig. 6–4). Polyacrylamide is often used as the gel matrix for short DNAs (up to a few hundred nucleotides). Agarose is generally used as the gel matrix for separating longer DNAs.

In both Sanger (dideoxy) and Maxam–Gilbert sequencing, the general principle is to reduce the DNA to be sequenced to four sets of labeled fragments. The reaction producing each set is base-specific, so that 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, a reaction that produces only fragments ending in C will generate fragments four and seven nucleotides long, whereas a reaction producing fragments ending in G will produce only a five-nucleotide fragment. 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 (Figs. 12–35, 12–

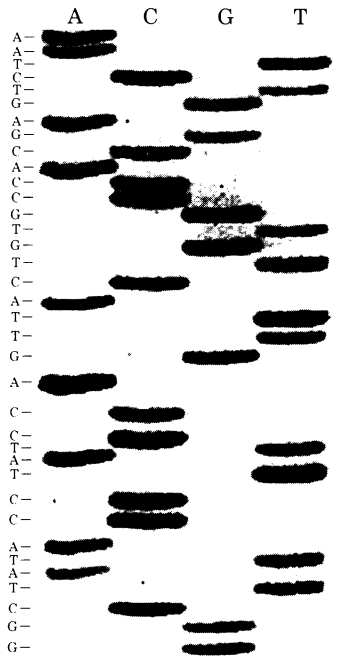


Figure 12–35 Section of an autoradiogram produced by the method developed by Sanger and colleagues. Side-by-side electrophoresis of the DNA fragments generated by each dideoxynucleotide generates a ladder of bands. Each band on the film corresponds to a population of DNA fragments of a specific length produced in the sequencing reactions (see Fig. 12–36). The identity of the base at each position in the sequence is determined from the lane in which a band is observed; the order of the bands read from the bottom of the gel corresponds to the DNA sequence.

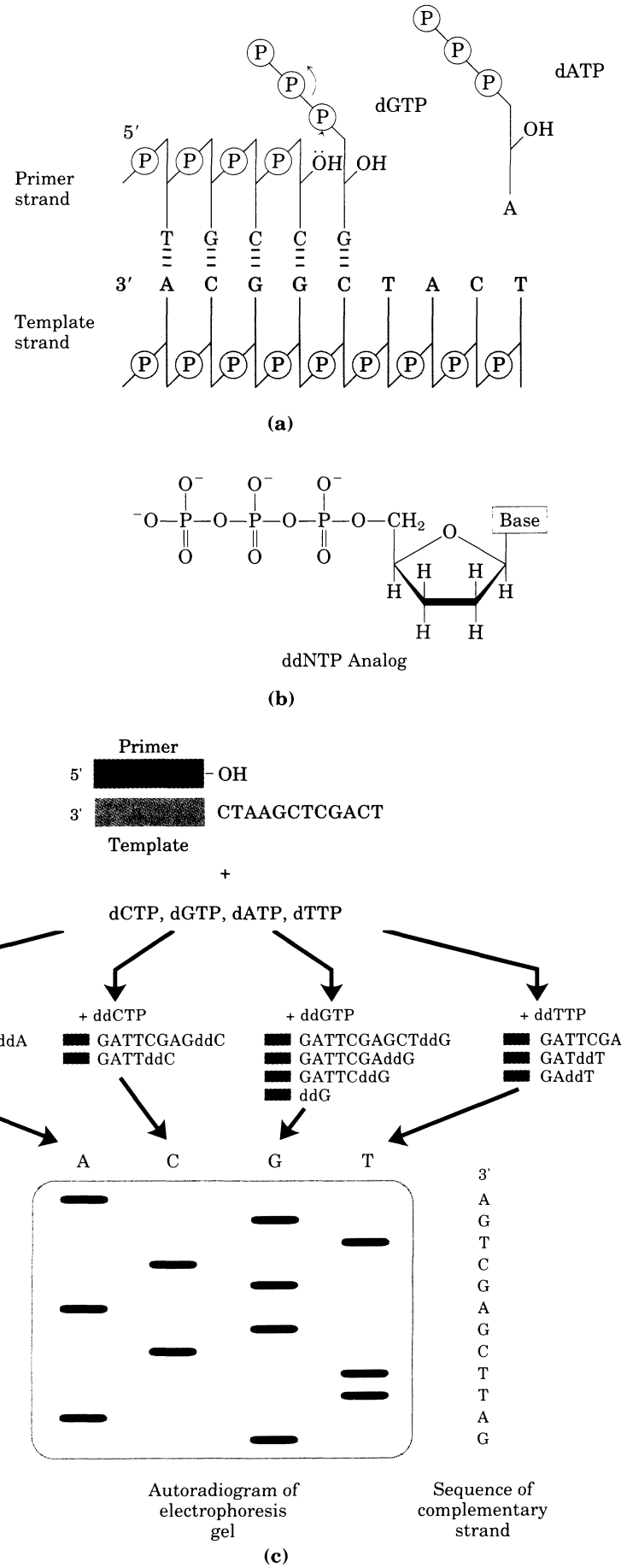


Figure 12–36 DNA sequencing by the Sanger (dideoxy) method. This method makes use of the mechanism of DNA synthesis by DNA polymerases (Chapter 24). DNA polymerases require both a primer, to which nucleotides are added, and a template strand to guide selection of each new nucleotide **(a).** The 3′-hydroxyl group of the primer reacts with the incoming deoxynucleoside triphosphate (dNTP), forming a new phosphodiester bond. The Sanger sequencing procedure uses dideoxynucleoside triphosphate (ddNTP) analogs **(b)** to interrupt DNA synthesis. When the dNTP is replaced by the ddNTP, strand elongation is halted after the analog is added because it lacks the 3′-hydroxyl group needed for the next step. The DNA to be sequenced is used as the template strand, and a short primer (usually radioactively labeled) is annealed to it **(c).** By adding small amounts of a single ddNTP, for example ddCTP, to an otherwise normal reaction system, the synthesized strands will be prematurely terminated at locations where dC normally occurs. Because there is much more dCTP than ddCTP, there is only a small chance that the analog will be incorporated whenever a dC is to be added, but there is generally enough ddCTP that each new strand has a high probability of acquiring one ddC at some point during synthesis. The result is a solution containing fragments representing each C residue in the sequence. The size of the fragments, separated by electrophoresis, reveals the location of C residues in the sequence. This procedure is repeated separately for each of the four ddNTPs, and the sequence can be read directly from an autoradiogram of the gel **(c).** Because shorter DNA fragments migrate faster, the fragments near the bottom represent the nucleotide positions closest to the primer (the 5′ end), and the sequence is read from bottom to top. Note that the sequence obtained is that of the strand *complementary* to the strand being analyzed. An actual sequencing gel is shown in Fig. 12–35.

36). The Sanger method (Fig. 12–36) is in more widespread use because it has proven to be technically easier. It involves the enzymatic synthesis of a DNA strand complementary to the strand to be analyzed.

DNA sequencing is now automated, using a variation of Sanger's sequencing method in which the primer used for each reaction is labeled with a differently colored fluorescent tag (Fig. 12–37). This technology allows sequences of thousands of nucleotides to be obtained in a few hours, and very large DNA-sequencing projects are being contemplated. The most ambitious of these, now underway, is the Human Genome Initiative, in which all of the 3 billion base pairs of DNA in a human cell will be sequenced.

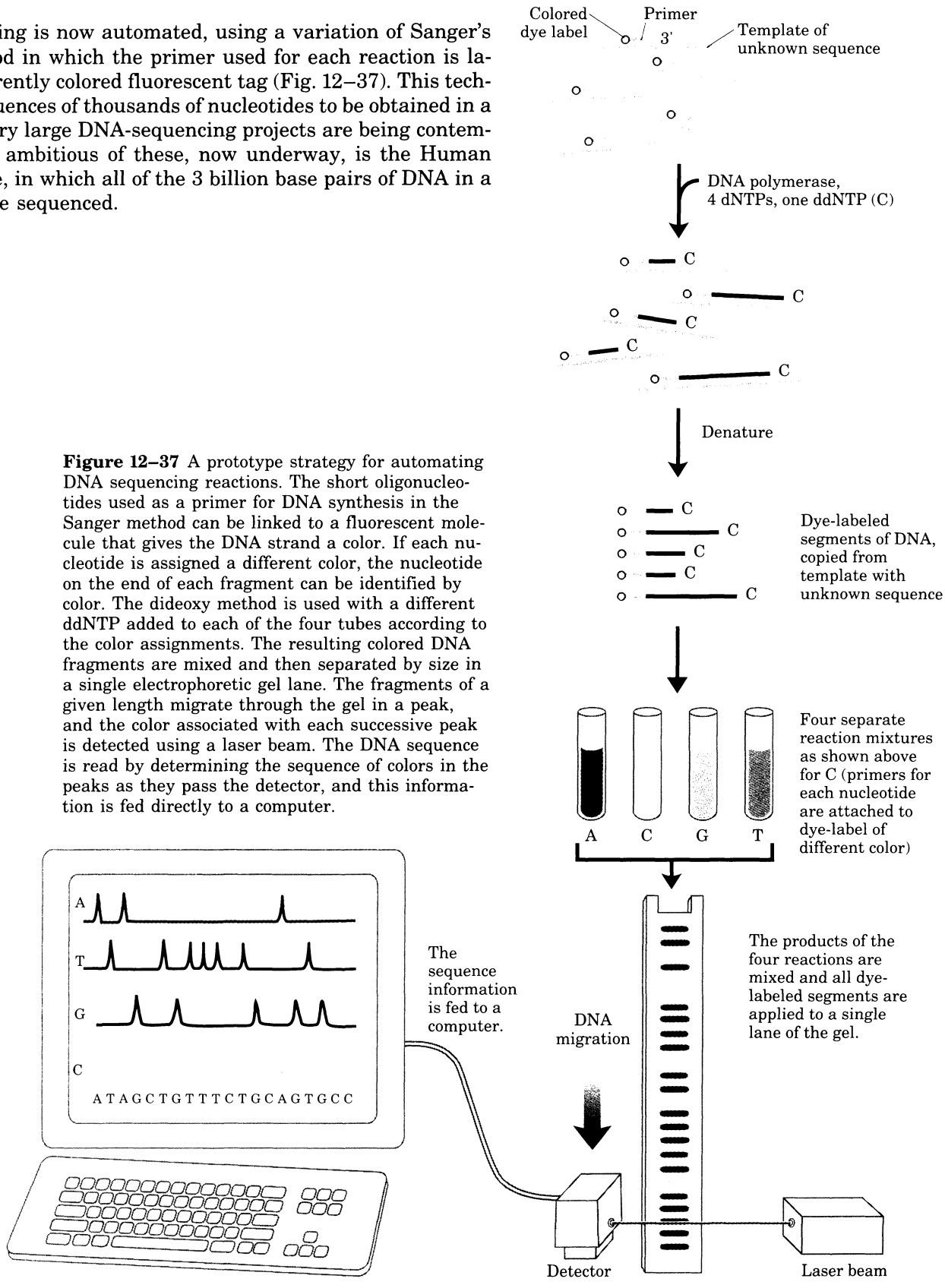


Figure 12–37 A prototype strategy for automating DNA sequencing reactions. The short oligonucleotides used as a primer for DNA synthesis in the Sanger method can be linked to a fluorescent molecule that gives the DNA strand a color. If each nucleotide is assigned a different color, the nucleotide on the end of each fragment can be identified by color. The dideoxy method is used with a different ddNTP added to each of the four tubes according to the color assignments. The resulting colored DNA fragments are mixed and then separated by size in a single electrophoretic gel lane. The fragments of a given length migrate through the gel in a peak, and the color associated with each successive 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, and this information is fed directly to a computer.

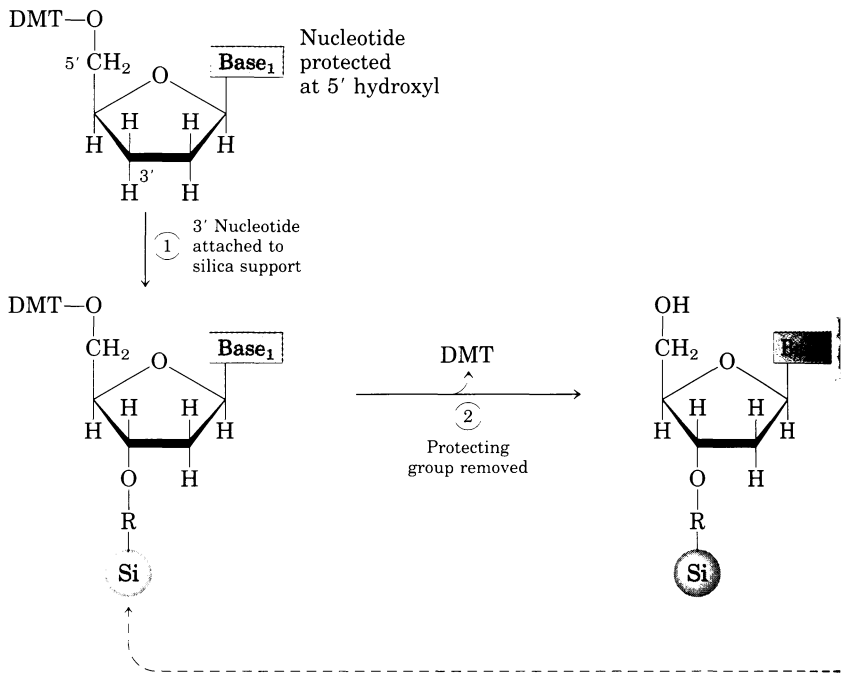


Figure 12–38 Automated synthesis of DNA is conceptually similar to the solid-state synthesis of polypeptides. The desired oligonucleotide is built up on a solid support (silica) one nucleotide at a time in a repeated series of chemical reactions with suitably protected nucleotide precursors. ① The first nucleotide (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 protecting group (dimethoxytrityl, DMT). The reactive groups on all bases are also chemically blocked. 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 is activated and reacted with the bound nucleotide to form a 5′-3′ linkage, which in ④ is oxidized with iodine to produce a phosphotriester linkage. (One of the phosphate oxygens is methylated.) 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 ⑤ and ⑥ the remaining blocking groups on the bases and the methyl groups on the phosphates are removed, and in ⑦ the oligonucleotide is separated from the solid support and purified. The chemistry of RNA synthesis has lagged far behind the procedures for DNA synthesis because of difficulties in protecting the 2′ hydroxyl of ribose without adverse effects on the reactivity of the 3′ hydroxyl.

The Chemical Synthesis of DNA Has Been Automated

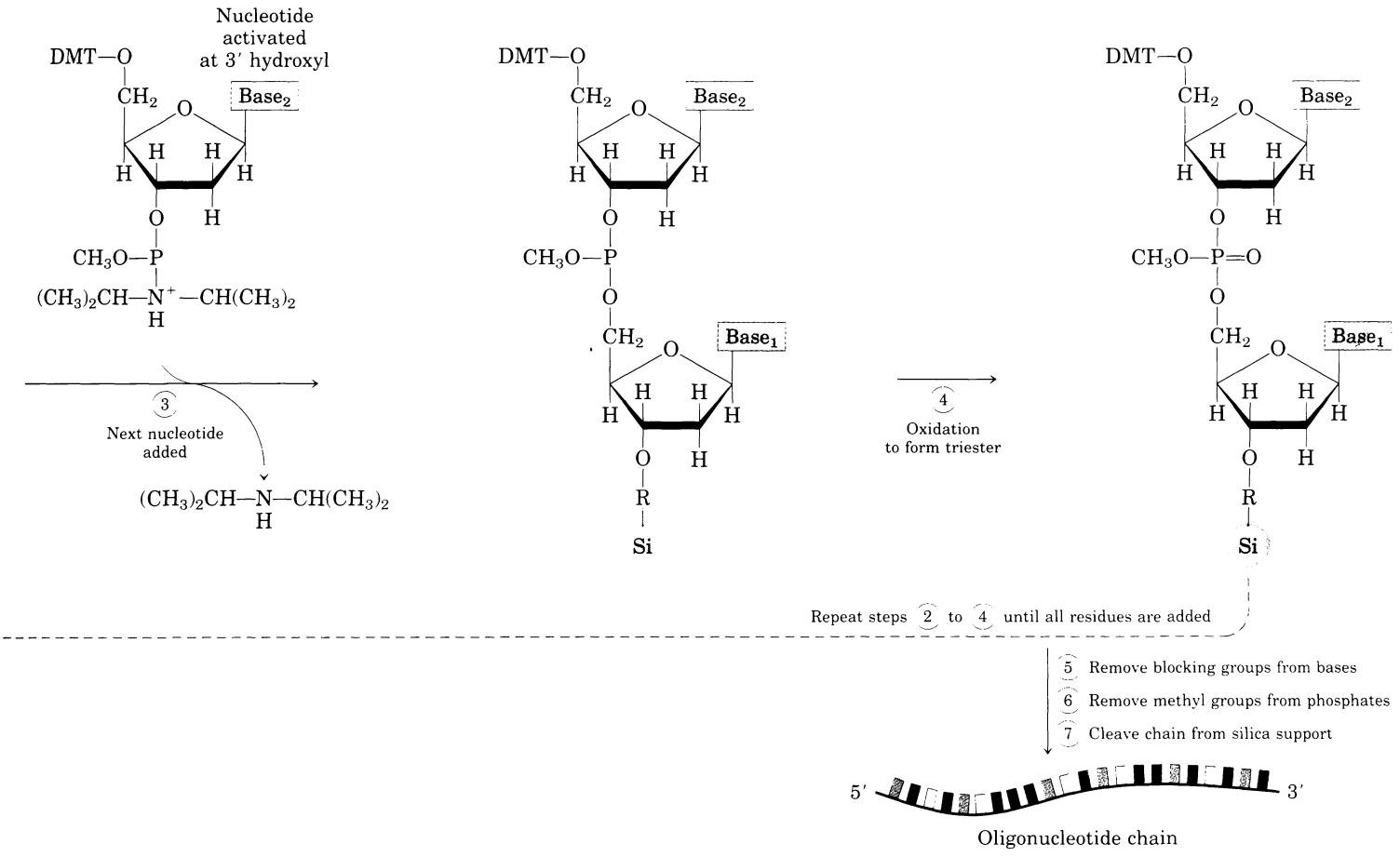
Another technology that has paved the way for many biochemical advances is the chemical synthesis of oligonucleotides with any chosen sequence. The chemical methods for synthesizing nucleic acids were developed primarily by H. Gobind Khorana in the 1970s. Refinement and automation of these methods has made it possible to synthesize DNA strands rapidly and accurately. The synthesis is carried out with the growing strand attached to a solid support (Fig. 12–38), using principles similar to those used by Merrifield in peptide synthesis (see Box 5–2). The efficiency of each addition step is very high, allowing the routine laboratory synthesis of polymers of 70 or 80 nucleotides. In some laboratories much longer strands are synthesized. The availability of relatively inexpensive DNA polymers with predesigned sequences is having a powerful impact on all areas of biochemistry (Chapter 28).

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

Nucleotides may have one, two, or three phosphate groups covalently linked at the 5′ hydroxyl of ribose. These are referred to as nucleoside mono-, di-, and triphosphates, respectively (Fig. 12–39). Starting from



the ribose, the three phosphates are generally labeled α, β, and γ. Nucleoside triphosphates are used as a source of chemical energy to drive a wide variety of biochemical reactions. ATP is by far the most widely used, but UTP, GTP, and CTP are used in specific reactions. Nucleoside triphosphates also serve as the activated precursors of DNA and RNA synthesis, as will be described in Chapters 24 and 25.

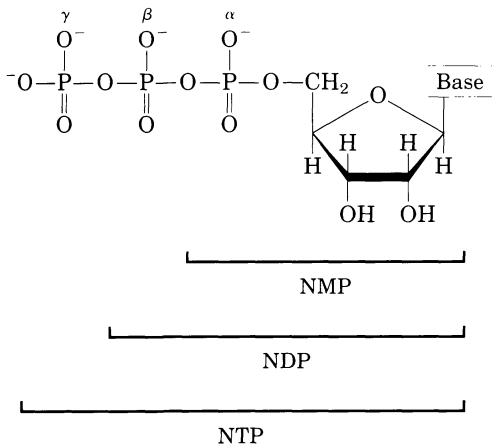


Figure 12–39 General structure of nucleoside 5′-mono-, 5′-di-, and 5′-triphosphates (NMPs, NDPs, and NTPs) and their standard abbreviations. In the deoxyribonucleoside phosphates (dNMPs, dNDPs, and dNTPs) the pentose is 2′-deoxy-D-ribose.

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

The hydrolysis of ATP and the other nucleoside triphosphates is an energy-yielding reaction because of the chemistry of the triphosphate structure. The bond between the ribose and the α phosphate is an ester linkage. The α–β and β–γ linkages are phosphoric acid anhydrides (Fig. 12–40). Hydrolysis of the ester linkage yields about 14 kJ/mol, whereas hydrolysis of each of the anhydride bonds yields about 30 kJ/mol. In biosynthesis, ATP hydrolysis often drives less favorable metabolic reactions (i.e., those with $\Delta G^{\circ\prime} > 0$). 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 and free-energy change described in Chapter 8).

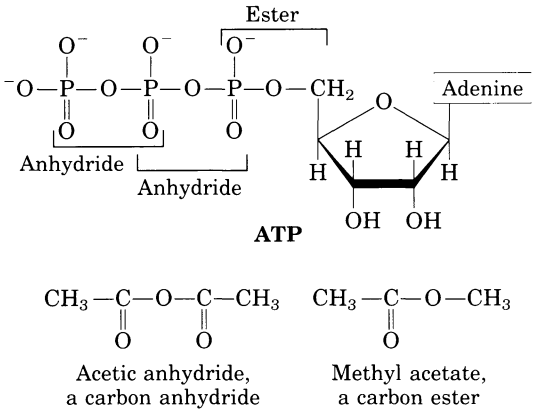


Figure 12–40 The phosphate ester and phosphoric acid anhydride bonds of ATP. Hydrolysis of an anhydride bond yields more energy than hydrolysis of the ester. A carbon anhydride and ester are shown for comparison.

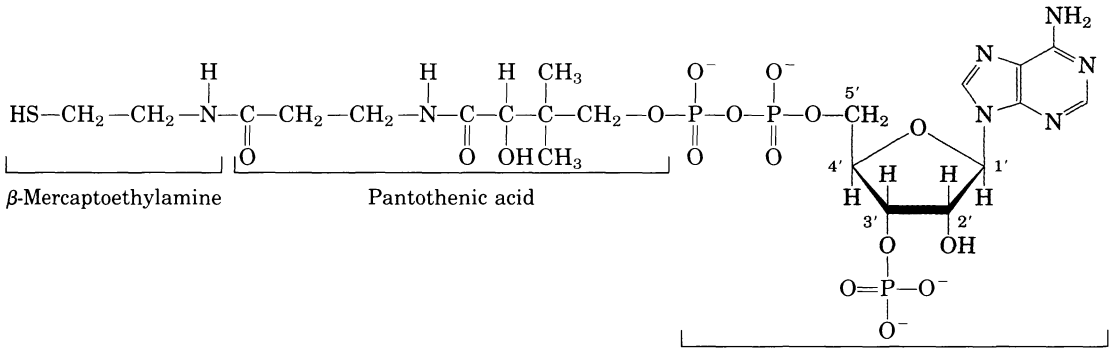
It is appropriate to ask why ATP serves as the primary carrier of energy in the cell. The chemical energy potential of pyrophosphate (~33 kJ/mol), a much simpler molecule, is almost identical to that of ATP because pyrophosphate also contains a phosphoric acid anhydride. Pyrophosphate would be so much easier to synthesize than ATP that the selection of ATP at first seems to contradict evolutionary logic.

The explanation can be found in the fundamental energetic principles governing every chemical reaction. In promoting chemically unfavorable reactions such as those in many biosynthetic processes, the cell must deal with both the *equilibrium* and the *rate* of the reaction. We have seen that an unfavorable equilibrium can be overcome by coupling such a reaction to one with a favorable equilibrium, such as the hydrolysis of an anhydride. Pyrophosphate would be just as effective as ATP in its potential effects on reaction equilibria. Therefore, the advantage to cells in using ATP rather than pyrophosphate must lie in reaction rates. In Chapter 8 we described how the energy used in catalysis is derived from binding energy, the multiple weak interactions that occur between substrate and enzyme. ATP, because of its larger structure, clearly can contribute many more of these weak interactions than pyrophosphate. In other words, the potential for reaction *rate* enhancement is much greater for ATP than pyrophosphate. A reaction with a favorable energetic equilibrium will not be of benefit to a cell if it takes several years to occur. This principle can be illustrated by the simple empirical observation that pyrophosphate will rarely function in an enzymatic reaction requiring ATP, even though it should fit into any enzyme active site that can accommodate ATP.

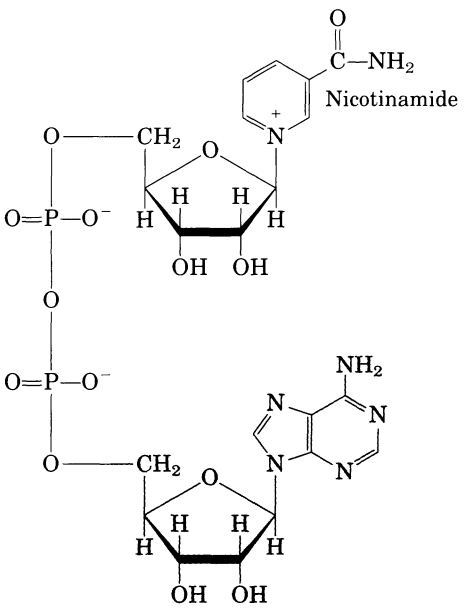
Nucleotides Are Components of Many Enzyme Cofactors

A variety of enzyme cofactors serving a wide range of chemical functions include adenosine as part of their structure (Fig. 12–41). 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 from these structures generally results in a drastic reduction of their activities. For example, removal of the adenosine nucleotide (3′-P-ADP; see Fig. 12–41) from acetoacetyl-CoA reduces its reactivity as a substrate for β-ketoacyl-CoA transferase (an enzyme of lipid metabolism) by a factor of 10^6. Although the reason for this requirement for adenosine has not been examined in detail, it must involve the binding energy between enzyme and substrate (or cofactor) that is used both in catalysis and to stabilize the initial ES complex (Chapter 8). In the case of CoA transferase, the nucleotide appears to be a binding "handle" that helps to

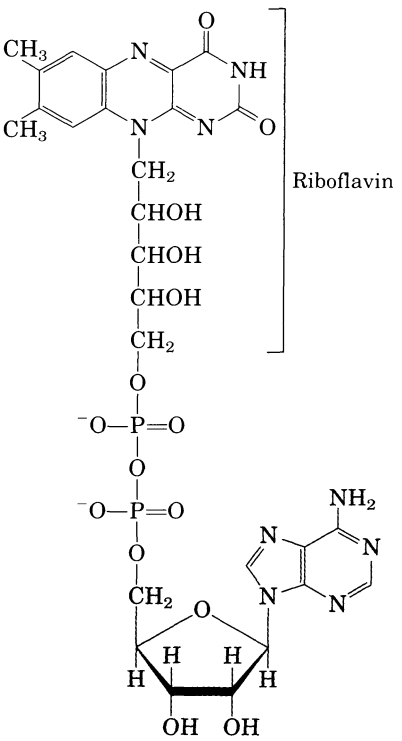
Figure 12–41 Enzyme cofactors and coenzymes incorporating adenosine in their structure. The adenosine portion is shaded in red. Coenzyme A functions in acyl group transfer reactions; NAD^+ participates in hydride transfers; FAD, the active form of vitamin B_2 (riboflavin), participates in electron transfers. Another coenzyme incorporating adenosine in its structure is 5′-deoxyadenosylcobalamin, the active form of vitamin B_{12} (see Box 16–2). This coenzyme is involved in intramolecular group transfers between adjacent carbons.



Coenzyme A



Nicotinamide adenine dinucleotide (NAD^+)



Flavin adenine dinucleotide (FAD)

pull the substrate into the active site. Similar roles may be found for the nucleoside portion of other nucleotide cofactors.

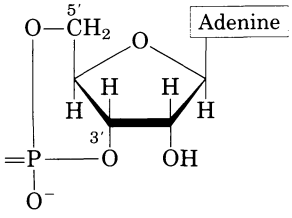
Now we may ask why adenosine, rather than some other large molecule, is used in these structures. The answer here may involve a kind of evolutionary economy. Adenosine is certainly not unique in the amount of potential binding energy it can contribute. The importance of adenosine probably lies not so much in some special chemical characteristic, but rather that an advantage existed in making one compound a standard. Once ATP became the standard source of chemical energy, systems developed to synthesize ATP more efficiently 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 protein domain that binds adenosine can be used in a wide variety of different enzymes. Such a structure, called a **nucleotide-binding fold,** is found in many enzymes that bind ATP and nucleotide cofactors.

Some Nucleotides Are Intermediates in Cellular Communication

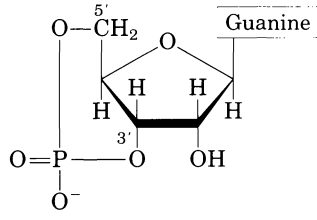
Cells respond to their environment by taking cues from hormones or other chemical signals in the surrounding medium. The interaction of these extracellular chemical signals (first messengers) with receptors on the cell surface often leads to the production of **second messengers** inside the cell, which in turn lead to adaptive changes in the cell interior (Chapter 22). Often, the second messenger is a nucleotide.

One of the most common second messengers is the nucleotide **adenosine 3′,5′-cyclic monophosphate (cyclic AMP,** or **cAMP),** formed from ATP in a reaction catalyzed by adenylate cyclase, associated with the inner face of the plasma membrane. Cyclic AMP serves regulatory functions in virtually every cell outside the plant kingdom, and these are described in detail in Chapter 22. Guanosine 3′,5′-cyclic monophosphate (cGMP) occurs in many cells and also has regulatory functions.

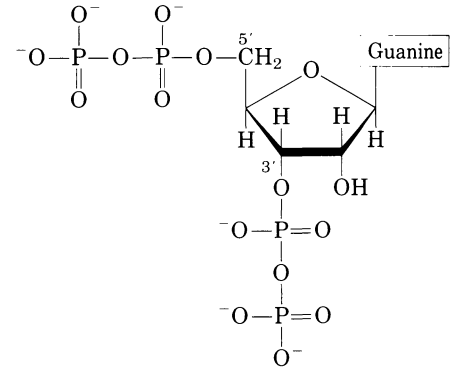
Another regulatory nucleotide, ppGpp, is produced in bacteria in response to the slowdown in protein synthesis that occurs during amino acid starvation. This nucleotide inhibits the synthesis of the rRNA and tRNA molecules (Chapter 27) needed for protein synthesis, preventing the unnecessary production of nucleic acids.



osine 3′,5′-cyclic monophosphate
(cyclic AMP; cAMP)



Guanosine 3′,5′-cyclic monophosphate
(cyclic GMP; cGMP)



Guanosine 3′-diphosphate,5′-diphosphate
(guanosine tetraphosphate)
(ppGpp)

Summary

Nucleotides serve a diverse set of important functions in cells. As subunits of nucleic acids they carry genetic information. They are also the primary carriers of chemical energy in cells, structural components of many enzyme cofactors, and cellular second messengers.

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, linked together by phosphodiester bridges between the 5′ hydroxyl of one pentose and the 3′ hydroxyl 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 pyrimidine bases are generally thymine and cytosine. The primary purines are adenosine and guanine in both RNA and DNA.

Many lines of evidence show that DNA bears genetic information. In particular, the Avery–MacLeod–McCarty experiment showed that DNA isolated from one strain of a bacterium 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 the host cell.

From x-ray diffraction studies of DNA fibers and the base equivalences in DNA discovered by Chargaff (A = T and G = C), 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, and the hydrophilic sugar–phosphate backbones are located on the outside. The base pairs are stacked perpendicular to the long axis, 0.34 nm apart; there are about 10 base pairs in each complete turn of the double helix.

DNA can exist in several structural forms. Two variations from the Watson–Crick B-form DNA, the A and Z forms, have been characterized in DNA crystal structures. The A-form helix is shorter and of greater diameter than a B-form helix with the same sequence. The Z form is a left-handed helix. Some sequence-dependent structural variations cause bends in the DNA. DNA strands with self-complementary inverted repeats can form hairpin or cruciform structures. Polypyrimidine tracts arranged in mirror repeats can take up a triple-helical structure called H-DNA.

Messenger RNA is the vehicle by which genetic information is transferred to ribosomes for protein synthesis. Transfer RNA and ribosomal RNA are also involved in protein synthesis. RNA can be structurally complex, with single RNA strands often folded into hairpins, double-stranded regions, and complex loops.

Native DNA undergoes reversible unwinding and separation (melting) of strands on heating or at extremes of pH. Because G≡C base pairs are more stable than A=T pairs, the melting point of DNAs rich in G≡C pairs is higher than that of 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 homology. Hybridization is the basis for important techniques used to study and isolate specific genes and RNAs.

DNA is a relatively stable polymer. Very slow, spontaneous reactions such as deamination of certain bases, hydrolysis of base–sugar *N*-glycosidic bonds, formation of pyrimidine dimers (radiation damage), and oxidative damage are important because of the very low tolerance of cells for changes in genetic material. DNA sequences can be determined and DNA polymers synthesized using simple protocols involving chemical and enzymatic methods.

ATP is the central carrier of chemical energy in cells, probably reflecting the requirement for binding energy in catalysis. The presence of adenosine in the structure of a variety of enzyme cofactors may also be related to binding energy requirements. Cyclic AMP is a common second messenger produced in response to hormones and other chemical signals. It is formed from ATP in a reaction catalyzed by adenylate cyclase.

Further Reading

General

Friedberg, E.C. (1985) *DNA Repair,* W.H. Freeman and Company, New York.
A good source for more information on the chemistry of nucleotides and nucleic acids.

Kornberg, A. & Baker, T.A. (1991) *DNA Replication,* 2nd edn, W.H. Freeman and Company, New York.
The best place to start for learning more about DNA structure.

Saenger, W. (1984) *Principles of Nucleic Acid Structure,* Springer-Verlag, New York.
A more detailed treatment.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A., & Weiner, A.M. (1987) *Molecular Biology of the Gene,* 4th edn, The Benjamin/Cummings Publishing Company, Menlo Park, CA.
Excellent general reference.

Variations in DNA Structure

Dickerson, R.E. (1983) The DNA helix and how it is read. *Sci. Am.* **249** (December), 94–111.

Htun, H. & Dahlberg, J.E. (1989) Topology and formation of triple-stranded H-DNA. *Science* **243,** 1571–1576.

Rich, A., Nordheim, A., & Wang, A.H.-J. (1984) The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.* **53,** 791–846.

Wells, R.D. & Harvey, S.C. (eds) (1988) *Unusual DNA Structures,* Springer-Verlag, New York.

Wells, R.D. (1988) Unusual DNA structures. *J. Biol. Chem.* **263,** 1095–1098.
Minireview; a concise summary.

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.

Historical

Olby, R. (1974) *The Path to the Double Helix,* University of Washington Press, Seattle.

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

Watson, J.D. (1968) *The Double Helix,* Atheneum Publishers, New York.
A personal account of the human aspects of the discovery.

Problems

1. *Determination of Protein Concentration by UV Absorption in a Solution Containing Nucleic Acids* The concentration of protein or nucleic acid in solutions containing both can be estimated by using their light absorption properties. Proteins have a strong absorption centered at a wavelength of 280 nm, whereas nucleic acids absorb most strongly at 260 nm. When both proteins and nucleic acids are present in a solution, their respective concentrations can be estimated by measuring the absorbance (A) of the solution at 280 nm and 260 nm and using the table at the top of page 357. $R_{280/260}$ is the ratio of the absorbance at 280 and 260 nm. The table indicates the percentage of total mass that is nucleic acid, and provides a factor, F, to correct the A_{280} reading and give a more accurate protein estimate. The protein concentration (in mg/ml) is equal to $F \times A_{280}$ (assuming the cuvette is 1 cm wide). What are the protein and nucleic acid concentration if $A_{280} = 0.69$ 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

2. *Nucleotide Structure* What positions in a purine ring have the potential to form hydrogen bonds, but are not involved in the hydrogen bonds of Watson–Crick base pairs?

3. *Base Sequence of Complementary DNA Strands* Write the base sequence of the complementary strand of double-helical DNA in which one strand has the sequence (5′)ATGCCCGTATGCATTC(3′).

4. *DNA of the Human Body* Calculate the weight in grams of a double-helical DNA molecule stretching from the earth to the moon (~320,000 km). The DNA double helix weighs about 1×10^{-18} g per 1,000 nucleotide pairs; each base pair extends 0.34 nm. For an interesting comparison, your body contains about 0.5 g of DNA!

5. *DNA Bending* Assume that a poly(A) tract five base pairs long produces a bend of about 20°. Calculate the total (net) bend produced in the DNA if the center base pairs (the third of five) of two successive $(dA)_5$ tracts are located (a) 10 or (b) 15 base pairs apart. Assume that there are 10 base pairs per turn in the DNA double helix.

6. *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 hairpin in RNA different from that of a hairpin in DNA?

7. *Nucleotide Chemistry* In the cells of many eukaryotic organisms, there are highly specialized systems that specifically repair G–T mismatches in DNA. The mismatch is repaired to form a G≡C base pair (not A=T). This G–T mismatch repair system occurs in addition to a more general system that repairs virtually all mismatches. Can you think of a reason why cells require a specialized system to repair G–T mismatches?

8. *Nucleic Acid Structure* Explain why there is an increase in the absorption of UV light (hyperchromic effect) when double-stranded DNA is denatured.

9. *Base Pairing in DNA* In samples of DNA isolated from two unidentified species of bacteria, 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 bacteria was isolated from a hot spring (64 °C). Which DNA came from this thermophilic bacterium? What is the basis for your answer?

PART

Information Pathways

The fourth and final part of this book considers biochemical questions raised by the genetic continuity and the evolution of living organisms. What is the molecular nature of the genetic material? How is genetic information transmitted with such fidelity? How is it ultimately translated into the amino acid sequence of protein molecules?

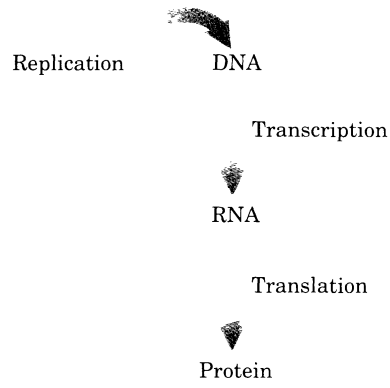
The fundamental unit of information in living systems is the **gene.** A gene is defined biochemically as that segment of DNA (or in a few cases RNA) that encodes the information required to produce a functional biological product. This product is most often a protein, and much of the material in the chapters to follow concerns genes that encode proteins. However, a gene product can also be one of several classes of RNA molecules. The storage and metabolism of these informational units now becomes the focal point of our discussion.

Modern biochemical research on gene structure and function has brought to biology a revolution comparable to that evoked over 100 years ago by Darwin's theory on the origin of species. An understanding of how information is stored and used in cells has brought penetrating new insights into some of the most fundamental problems concerning the structure and function of cells. Moreover, it has led to a more comprehensive conceptual framework for the science of biochemistry.

Today's knowledge of information pathways has arisen from the convergence of three different disciplines: genetics, physics, and biochemistry. The contributions of these three fields are epitomized by the discovery that opened the modern era of genetic biochemistry: the double-helical structure of DNA, as postulated by James Watson and Francis Crick in 1953 (see Fig. 12–15). Genetic theory contributed the concept of coding by genes. Physics made possible the determination of molecular structure by x-ray diffraction analysis. Biochemistry revealed the chemical composition of DNA. The great impact of the Watson–Crick hypothesis was largely due to its ability to account for a wide range of results derived from these varied sources.

A vastly improved understanding of DNA structure inevitably led to questions about its function. The structure itself suggested how DNA might be copied so that the information contained therein could be transmitted from one generation to the next. Understanding how the information in DNA was converted into functional proteins became possible through the discovery of messenger RNA and transfer RNA and the solution of the genetic code. These and other major advances led to the central dogma of molecular genetics, which defines three major processes in the cellular utilization of genetic information. The first is **replication,** the copying of parental DNA to form daughter

Facing page: The two β subunits of *E. coli* DNA polymerase III bound to DNA. The subunits, shown as gray ribbon structures, form a circle around the DNA, tethering the DNA polymerase III (which has at least 9 other subunits) to the DNA. This permits the enzyme to synthesize long stretches of DNA without dissociation. The complex set of operations by which macromolecules containing information are faithfully synthesized requires a great many enzymes, of which this is just part of one.



The central dogma of molecular genetics, showing the general pathways of information flow via the processes of replication, transcription, and translation. The term "dogma" is a misnomer here. It was introduced by Francis Crick at a time when little evidence supported these ideas. The "dogma" is now a well-established principle.

DNA molecules having identical nucleotide sequences. The second is **transcription,** the process by which parts of the coded genetic message in DNA are copied precisely in the form of RNA. The third is **translation,** in which the genetic message coded in messenger RNA is translated on the ribosomes into a protein with a specific sequence of amino acids.

Part IV is devoted to an explanation of these and related processes. First (Chapter 23) we will examine the structure, topology, and packaging of chromosomes and genes. The processes that make up the central dogma will be elaborated in Chapters 24 through 26. Then, as we have done for biosynthetic pathways, we will turn to regulation and examine how the expression of genetic information is controlled (Chapter 27).

A major theme running through these chapters is the added complexity encountered in the biosynthesis of a macromolecule when that macromolecule contains information. Assembling nucleic acids and proteins with the correct sequences of nucleotides and amino acids, respectively, represents nothing less than preserving the faithful expression of the template upon which life itself is based. The formation of phosphodiester bonds in DNA or peptide bonds in proteins might be expected to be a trivial feat for cells, given the arsenal of enzymatic and chemical tools described in Part III of this book. Nevertheless, the framework of patterns and rules established in the examination of metabolic pathways must be enlarged considerably when information is added to the equation. Forming *specific* bonds and preventing sequence errors in these polymers has an enormous impact on the thermodynamics, chemistry, and enzymology of the synthetic processes. For example, formation of a peptide bond should require an input of only about 21 kJ, and relatively simple enzymes that catalyze comparable reactions are known. To synthesize the correct peptide bond between two specific amino acids at a given point in a protein, however, the cell invests about 125 kJ in chemical energy and makes use of the combined activities of over 200 RNA molecules, enzymes, and specialized proteins. Information is expensive.

The dynamic interaction between nucleic acids and proteins is another central theme of Part IV. With the important exception of a few catalytic RNA molecules (discussed in Chapter 25), the processes that make up the pathways of cellular information flow are catalyzed and regulated by proteins. An understanding of these enzymes and proteins can have practical as well as intellectual rewards because they form the basis of the development of recombinant DNA technology. This technology is making possible the prenatal diagnosis of genetic disease; the production of a wide range of potent new pharmaceutical agents; the sequencing of the entire human genome; the introduction of new traits into bacteria, plants, and animals for industry and agriculture; human gene therapy; and many other advances. We finish our tour of the information pathways, and indeed the entire book, in Chapter 28 with a look at this technology and its implications for the future.