Chapter 11

Nucleotides and cleic Acids

We have discovered the secret of life!

Proclamation by Francis H. C. Crick to patrons of The Eagle, a pub in Cambridge, England (1953)

Francis Crick and James Watson point out features of their model for the structure of DNA. (© A. Barrington Brown/Science Source/Photo Researchers, Inc.)

Outline

Nucleotides and **nucleic acids** are biological molecules that possess heterocyclic nitrogenous bases as principal components of their structure. The biochemical roles of nucleotides are numerous; they participate as essential intermediates in virtually all aspects of cellular metabolism. Serving an even more central biological purpose are the nucleic acids, the elements of heredity and the agents of genetic information transfer. Just as proteins are linear polymers of amino acids, nucleic acids are linear polymers of nucleotides. Like the letters in this sentence, the orderly sequence of nucleotide residues in a nucleic acid can encode information. The two basic kinds of nucleic acids are **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA).** Complete hydrolysis of nucleic acids liberates nitrogenous bases, a five-carbon sugar, and phosphoric acid in equal amounts. The five-carbon sugar in DNA is 2-deoxyribose; in RNA,

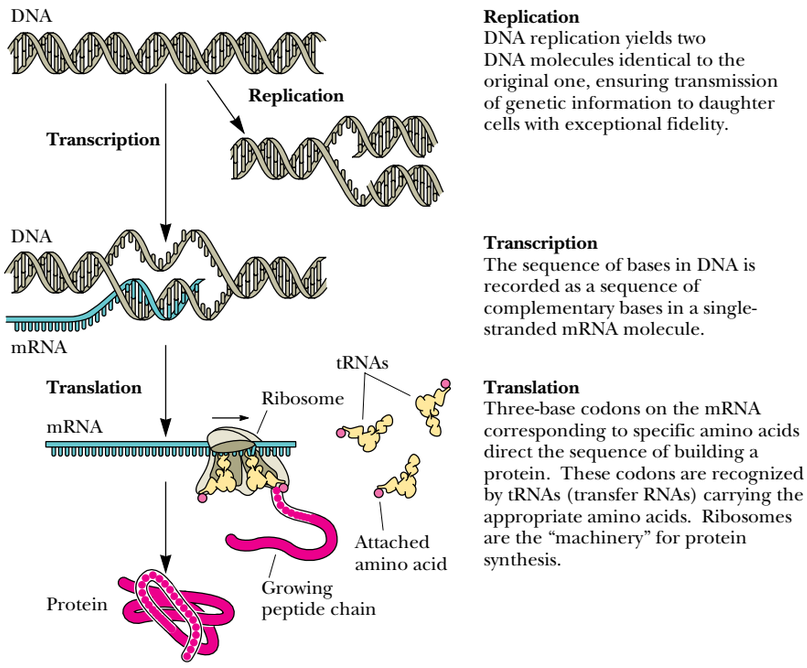
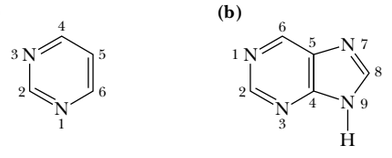


FIGURE 11.1 • The fundamental process of information transfer in cells. Information encoded in the nucleotide sequence of DNA is transcribed through synthesis of an RNA molecule whose sequence is dictated by the DNA sequence. As the sequence of this RNA is read (as groups of three consecutive nucleotides) by the protein synthesis machinery, it is translated into the sequence of amino acids in a protein. This information transfer system is encapsulated in the dogma: DNA → RNA → protein.

it is ribose. (See Chapter 7 for a detailed discussion of sugars and other carbohydrates.) DNA is the repository of genetic information in cells, while RNA serves in the transcription and translation of this information (Figure 11.1). An interesting exception to this rule is that some viruses have their genetic information stored as RNA.

This chapter describes the chemistry of nucleotides and the major classes of nucleic acids. Chapter 12 presents methods for determination of nucleic acid primary structure (nucleic acid sequencing) and describes the higher orders of nucleic acid structure. Chapter 13 introduces the *molecular biology of recombinant DNA:* the construction and uses of novel DNA molecules assembled by combining segments from other DNA molecules.

(a) **(b)**



The pyrimidine ring **The purine ring system**

FIGURE 11.2 • (a) The pyrimidine ring system; by convention, atoms are numbered as indicated. (b) The purine ring system, atoms numbered as shown.

11.1 • Nitrogenous Bases

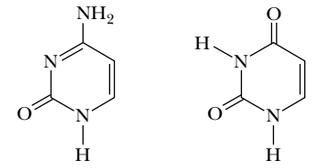
The bases of nucleotides and nucleic acids are derivatives of either **pyrimidine** or **purine.** Pyrimidines are six-membered heterocyclic aromatic rings containing two nitrogen atoms (Figure 11.2a). The atoms are numbered in a clockwise fashion, as shown in the figure. The purine ring structure is represented by the combination of a pyrimidine ring with a five-membered imidazole ring to yield a fused ring system (Figure 11.2b). The nine atoms in this system are numbered according to the convention shown.

The pyrimidine ring system is planar, while the purine system deviates somewhat from planarity in having a slight pucker between its imidazole and pyrimidine portions. Both are relatively insoluble in water, as might be expected from their pronounced aromatic character.

Common Pyrimidines and Purines

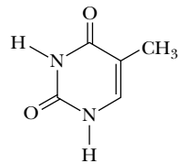
The common naturally occurring pyrimidines are **cytosine, uracil,** and **thymine** (5-methyluracil) (Figure 11.3). Cytosine and thymine are the pyrimidines typically found in DNA, whereas cytosine and uracil are common in RNA. To view this generality another way, the uracil component of DNA occurs as the 5-methyl variety, thymine. Various pyrimidine derivatives, such as dihydrouracil, are present as minor constituents in certain RNA molecules.

Adenine (6-amino purine) and **guanine** (2-amino-6-oxy purine), the two common purines, are found in both DNA and RNA (Figure 11.4). Other naturally occurring purine derivatives include **hypoxanthine, xanthine,** and **uric acid** (Figure 11.5). Hypoxanthine and xanthine are found only rarely as constituents of nucleic acids. Uric acid, the most oxidized state for a purine derivative, is never found in nucleic acids.



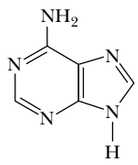
Cytosine (2-oxy-4-amino pyrimidine)

Uracil (2-oxy-4-oxy pyrimidine)

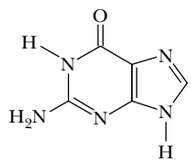


Thymine (2-oxy-4-oxy 5-methyl pyrimidine)

FIGURE 11.3 • The common pyrimidine bases—cytosine, uracil, and thymine—in the tautomeric forms predominant at pH 7.

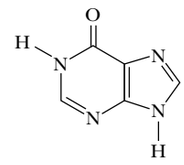


Adenine (6-amino purine)

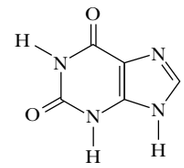


Guanine (2-amino-6-oxy purine)

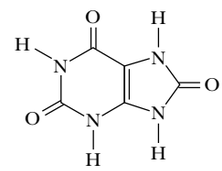
FIGURE 11.4 • The common purine bases—adenine and guanine—in the tautomeric forms predominant at pH 7.



Hypoxanthine



Xanthine



Uric acid

FIGURE 11.5 • Other naturally occurring purine derivatives—hypoxanthine, xanthine, and uric acid.

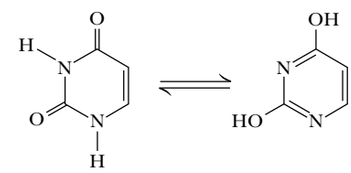
Properties of Pyrimidines and Purines

The aromaticity of the pyrimidine and purine ring systems and the electron-rich nature of their —OH and —NH_2 substituents endow them with the capacity to undergo **keto–enol tautomeric shifts.** That is, pyrimidines and purines exist as tautomeric pairs, as shown in Figure 11.6 for uracil. The keto tautomer is called a **lactam,** whereas the enol form is a **lactim.** The lactam form vastly predominates at neutral pH. In other words, pK_a values for ring nitrogen atoms 1 and 3 in uracil are greater than 8 (the pK_a value for N-3 is 9.5) (Table 11.1).

Table 11.1

Proton Dissociation Constants (pK_a Values) for Nucleotides

Nucleotide	pK_a Base-N	pK_1 Phosphate	pK_2 Phosphate
5′-AMP	3.8 (N-1)	0.9	6.1
5′-GMP	9.4 (N-1)	0.7	6.1
	2.4 (N-7)		
5′-CMP	4.5 (N-3)	0.8	6.3
5′-UMP	9.5 (N-3)	1.0	6.4



Lactam **Lactim**

FIGURE 11.6 • The keto/enol tautomerism of uracil.

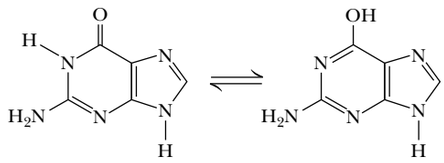


FIGURE 11.7 • The tautomerism of the purine, guanine.

In contrast, as might be expected from the form of cytosine that predominates at pH 7, the pK_a value for N-3 in this pyrimidine is 4.5. Similarly, tautomeric forms can be represented for purines, as given for guanine in Figure 11.7. Here, the pK_a value is 9.4 for N-1 and less than 5 for N-3. These pK_a values specify whether hydrogen atoms are associated with the various ring nitrogens at neutral pH. As such, they are important in determining whether these nitrogens serve as H-bond donors or acceptors. Hydrogen bonding between purine and pyrimidine bases is fundamental to the biological functions of nucleic acids, as in the formation of the double helix structure of DNA (see Section 11.6). The important functional groups participating in H-bond formation are the amino groups of cytosine, adenine, and guanine; the ring nitrogens at position 3 of pyrimidines and position 1 of purines; and the strongly electronegative oxygen atoms attached at position 4 of uracil and thymine, position 2 of cytosine, and position 6 of guanine (see Figure 11.21).

Another property of pyrimidines and purines is their strong absorbance of ultraviolet (UV) light, which is also a consequence of the aromaticity of their heterocyclic ring structures. Figure 11.8 shows characteristic absorption spectra of several of the common bases of nucleic acids—adenine, uracil, cytosine, and guanine—in their nucleotide forms: AMP, UMP, CMP, and GMP (see Section 11.4). This property is particularly useful in quantitative and qualitative analysis of nucleotides and nucleic acids.

11.2 • The Pentoses of Nucleotides and Nucleic Acids

Five-carbon sugars are called **pentoses** (see Chapter 7). RNA contains the pentose D-ribose, while 2-deoxy-D-ribose is found in DNA. In both instances, the pentose is in the five-membered ring form known as *furanose:* D-ribofuranose for RNA and 2-deoxy-D-ribofuranose for DNA (Figure 11.9). When these ribofuranoses are found in nucleotides, their atoms are numbered as 1′, 2′, 3′, and so on to distinguish them from the ring atoms of the nitrogenous bases. As we shall see, the seemingly minor difference of a hydroxyl group at the 2′-position has far-reaching effects on the secondary structures available to RNA and DNA, as well as their relative susceptibilities to chemical and enzymatic hydrolysis.

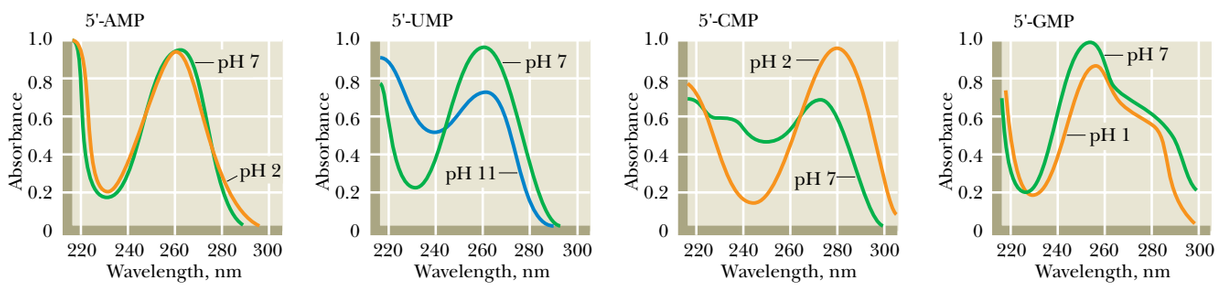


FIGURE 11.8 • The UV absorption spectra of the common ribonucleotides.

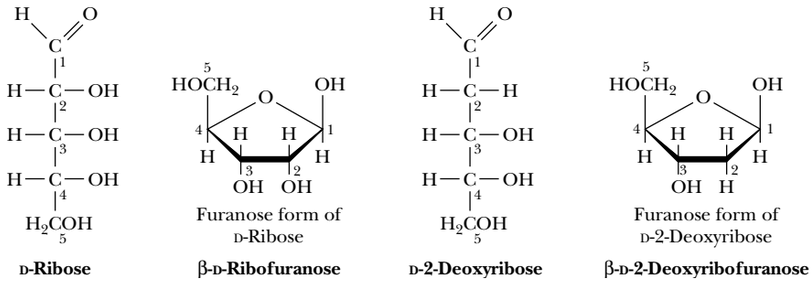
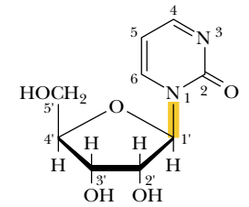


FIGURE 11.9 • Furanose structures—ribose and deoxyribose.

11.3 • Nucleosides Are Formed by Joining a Nitrogenous Base to a Sugar

Nucleosides are compounds formed when a base is linked to a sugar via a **glycosidic bond** (Figure 11.10). Glycosidic bonds by definition involve the carbonyl carbon atom of the sugar, which in cyclic structures is joined to the ring O atom (see Chapter 7). Such carbon atoms are called **anomeric.** In nucleosides, the bond is an *N*-glycoside because it connects the anomeric C-1′ to N-1 of a pyrimidine or to N-9 of a purine. Glycosidic bonds can be either α or β, depending on their orientation relative to the anomeric C atom. Glycosidic bonds in nucleosides and nucleotides are always of the β-configuration, as represented in Figure 11.10. Nucleosides are named by adding the ending *-idine* to the root name of a pyrimidine or *-osine* to the root name of a purine. The common nucleosides are thus **cytidine, uridine, thymidine, adenosine,** and **guanosine.** The structures shown in Figure 11.11 are *ribonucleosides. Deoxyribonucleosides,* in contrast, lack a 2′-OH group on the pentose. The nucleoside formed by hypoxanthine and ribose is **inosine.**



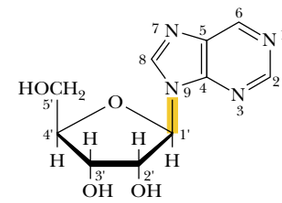


FIGURE 11.10 • β-Glycosidic bonds link nitrogenous bases and sugars to form nucleosides.

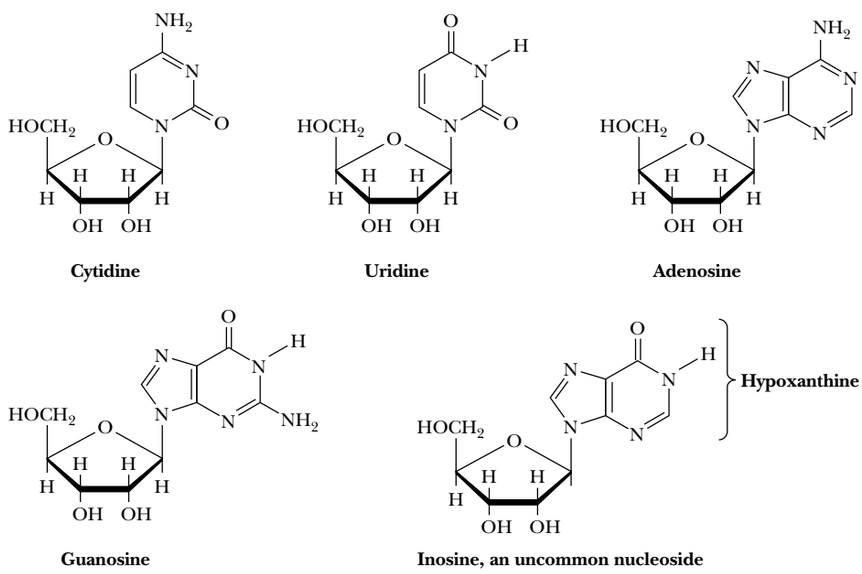
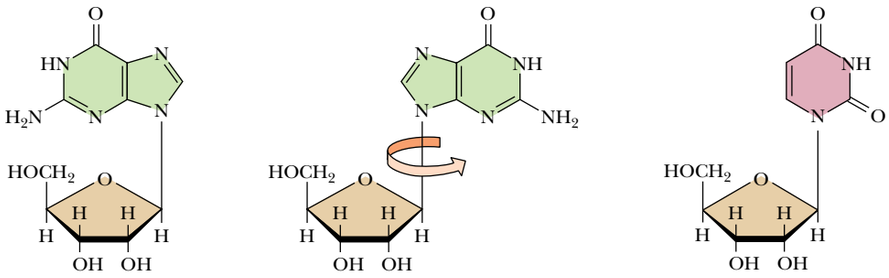


FIGURE 11.11 • The common ribonucleosides—cytidine, uridine, adenosine, and guanosine. Also, inosine drawn in anti conformation.



***syn* Guanosine** ***anti* Guanosine** ***anti* Uridine**

FIGURE 11.12 • Rotation around the glycosidic bond is sterically hindered; syn versus anti conformations in nucleosides are shown.

Nucleoside Conformation

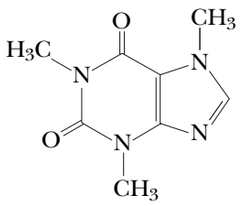
In nucleosides, rotation of the base about the glycosidic bond is sterically hindered, principally by the hydrogen atom on the C-2′ carbon of the furanose. (This hindrance is most easily seen and appreciated by manipulating accurate molecular models of these structures.) Consequently, nucleosides and nucleotides (see next section) exist in either of two conformations, designated *syn* and *anti* (Figure 11.12). For pyrimidines in the syn conformation, the oxygen substituent at position C-2 lies immediately above the furanose ring; in the anti conformation, this steric interference is avoided. Consequently, pyrimidine nucleosides favor the anti conformation. Purine nucleosides can adopt either the syn or anti conformation. In either conformation, the roughly planar furanose and base rings are not coplanar but lie at approximately right angles to one another.

HUMAN BIOCHEMISTRY

Adenosine: A Nucleoside with Physiological Activity

For the most part, nucleosides have no biological role other than to serve as component parts of nucleotides. Adenosine is an exception. In mammals, adenosine functions as an **autocoid,** or "local hormone." This nucleoside circulates in the bloodstream, acting locally on specific cells to influence such diverse physiological phenomena as blood vessel dilation, smooth muscle contraction, neuronal discharge, neurotransmitter release, and metabolism of fat. For example, when muscles work hard, they release adenosine, causing the surrounding blood vessels to dilate, which in turn increases the flow of blood and its delivery of O_2 and nutrients to the muscles. In a different autocoid role, adenosine acts in regulating heartbeat. The natural rhythm of the heart is controlled by a pacemaker, the sinoatrial node, that cyclically sends a wave of electrical excitation to the heart muscles. By blocking the flow of electrical current, adenosine slows the heart rate. *Supraventricular tachycardia* is a heart condition characterized by a rapid heartbeat. Intravenous injection of adenosine causes a momentary interruption of the rapid cycle of contraction and restores a normal heart rate. Adenosine is licensed and marketed as *Adenocard*™ to treat supraventricular tachycardia.

In addition, adenosine is implicated in sleep regulation. During periods of extended wakefulness, extracellular adenosine levels rise as a result of metabolic activity in the brain, and this increase promotes sleepiness. During sleep, adenosine levels fall. Caffeine promotes wakefulness by blocking the interaction of extracellular adenosine with its neuronal receptors.*



Caffeine

*Porrka-Heiskanen, T., et al., 1997. Adenosine: A mediator of the sleep-inducing effects of prolonged wakefulness. *Science* **276:**1265–1268.

Nucleosides Are More Water-Soluble Than Free Bases

Nucleosides are much more water-soluble than the free bases because of the hydrophilicity of the sugar moiety. Like glycosides (see Chapter 7), nucleosides are relatively stable in alkali. Pyrimidine nucleosides are also resistant to acid hydrolysis, but purine nucleosides are easily hydrolyzed in acid to yield the free base and pentose.

11.4 • Nucleotides Are Nucleoside Phosphates

A **nucleotide** results when phosphoric acid is esterified to a sugar —OH group of a nucleoside. The nucleoside ribose ring has three —OH groups available for esterification, at C-2′, C-3′, and C-5′ (although 2′-deoxyribose has only two). The vast majority of monomeric nucleotides in the cell are **ribonucleotides** having 5′-phosphate groups. Figure 11.13 shows the structures of the common four *ribonucleotides,* whose formal names are **adenosine 5′-monophosphate, guanosine 5′-monophosphate, cytidine 5′-monophosphate,** and **uridine 5′-monophosphate.** These compounds are more often referred to by their abbreviations: **5′-AMP, 5′-GMP, 5′-CMP,** and **5′-UMP,** or even more simply as **AMP, GMP, CMP,** and **UMP.** Nucleoside 3′-phosphates and nucleoside 2′-phosphates (3′-NMP and 2′-NMP, where N is a generic designation for "nucleoside") do not occur naturally, but are biochemically important as products of polynucleotide or nucleic acid hydrolysis. Because the pK_a value for the first dissociation of a proton from the phosphoric acid moiety is 1.0 or less (Table 11.1), the nucleotides have acidic properties. This acidity is implicit in the other names by which these substances are known—**adenylic acid, guanylic acid,**

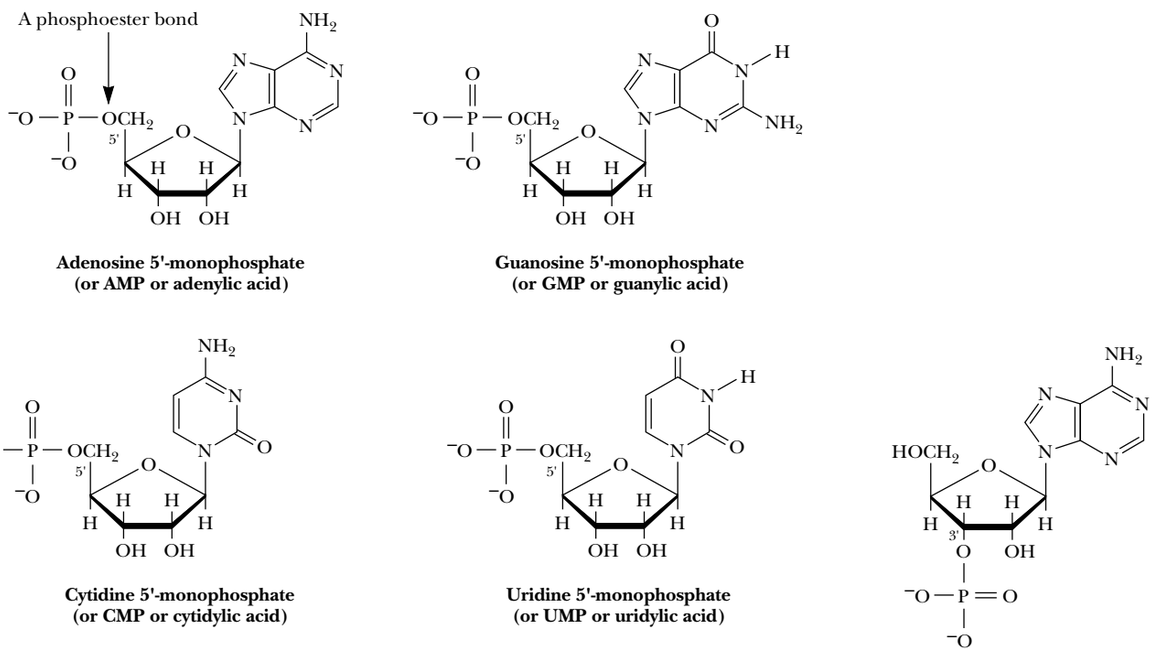
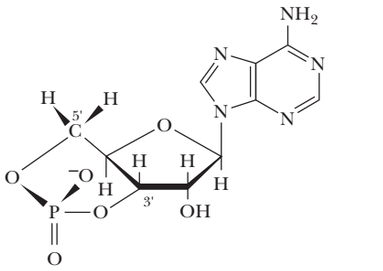
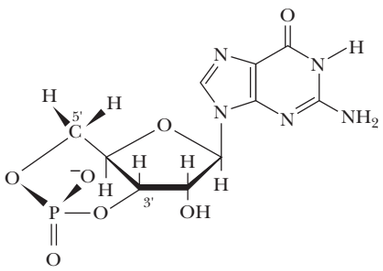


FIGURE 11.13 • Structures of the four common ribonucleotides—AMP, GMP, CMP, and UMP—together with their two sets of full names, for example, adenosine 5′-monophosphate and adenylic acid. Also shown is the nucleoside 3′-AMP.



3',5'-Cyclic AMP



3',5'-Cyclic GMP

FIGURE 11.14 • Structures of the cyclic nucleotides cAMP and cGMP.

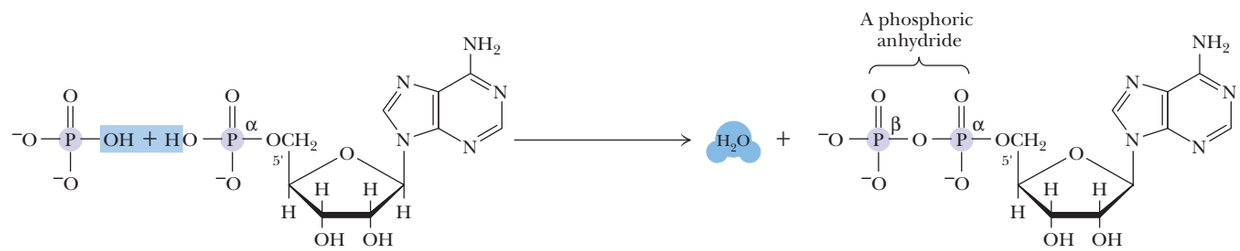
cytidylic acid, and **uridylic acid.** The pK_a value for the second dissociation, pK_2, is about 6.0, so at neutral pH or above, the net charge on a nucleoside monophosphate is −2. Nucleic acids, which are polymers of nucleoside monophosphates, derive their name from the acidity of these phosphate groups.

Cyclic Nucleotides

Nucleoside monophosphates in which the phosphoric acid is esterified to *two* of the available ribose hydroxyl groups (Figure 11.14) are found in all cells. Forming two such ester linkages with one phosphate results in a cyclic structure. **3′,5′-cyclic AMP,** often abbreviated **cAMP,** and its guanine analog **3′,5′-cyclic GMP,** or **cGMP,** are important regulators of cellular metabolism (see Part III: Metabolism and Its Regulation).

Nucleoside Diphosphates and Triphosphates

Additional phosphate groups can be linked to the phosphoryl group of a nucleotide through the formation of phosphoric anhydride linkages, as shown in Figure 11.15. Addition of a second phosphate to AMP creates **adenosine 5′-diphosphate,** or **ADP,** and adding a third yields **adenosine 5′-triphosphate,** or **ATP.** The respective phosphate groups are designated by the Greek letters α, β, and γ, starting with the α-phosphate as the one linked directly to the pentose. The abbreviations **GTP, CTP,** and **UTP** represent the other corresponding nucleoside 5′-triphosphates. Like the nucleoside 5′-monophosphates, the nucleoside 5′-diphosphates and 5′-triphosphates all occur in the free state in the cell, as do their deoxyribonucleoside phosphate counterparts, represented as dAMP, dADP, and dATP; dGMP, dGDP, and dGTP; dCMP, dCDP, and dCTP; dUMP, dUDP, and dUTP; and dTMP, dTDP, and dTTP.



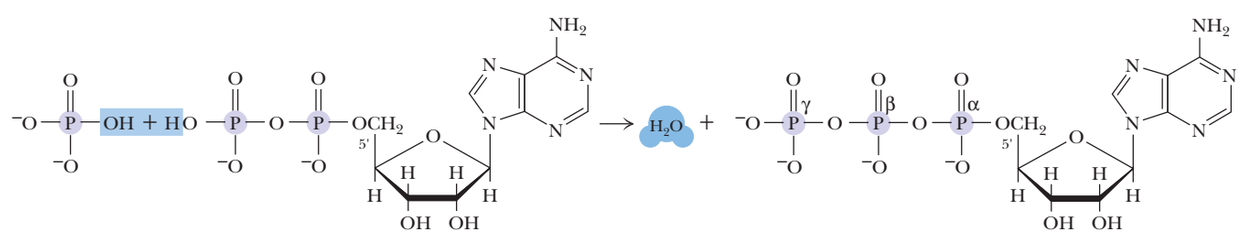


FIGURE 11.15 • Formation of ADP and ATP by the successive addition of phosphate groups via phosphoric anhydride linkages. Note the removal of equivalents of H_2O in these dehydration synthesis reactions.

NDPs and NTPs Are Polyprotic Acids

Nucleoside 5′-diphosphates (NDPs) and **nucleoside 5′-triphosphates (NTPs)** are relatively strong *polyprotic acids,* in that they dissociate three and four protons, respectively, from their phosphoric acid groups. The resulting phosphate anions on NDPs and NTPs form stable complexes with divalent cations such as Mg^{2+} and Ca^{2+}. Because Mg^{2+} is present at high concentrations (5 to 10 m*M*) intracellularly, NDPs and NTPs occur primarily as Mg^{2+} complexes in the cell. The phosphoric anhydride linkages in NDPs and NTPs are readily hydrolyzed by acid, liberating inorganic phosphate (often symbolized as P_i) and the corresponding NMP. A diagnostic test for NDPs and NTPs is quantitative liberation of P_i upon treatment with 1 *N* HCl at 100°C for 7 min.

Nucleoside 5′-Triphosphates Are Carriers of Chemical Energy

Nucleoside 5′-triphosphates are indispensable agents in metabolism because the phosphoric anhydride bonds they possess are a prime source of chemical energy to do biological work. ATP has been termed the energy currency of the cell (Chapter 3). GTP is the major energy source for protein synthesis (see Chapter 33), CTP is an essential metabolite in phospholipid synthesis (see Chapter 25), and UTP forms activated intermediates with sugars that go on to serve as substrates in the biosynthesis of complex carbohydrates and polysaccharides (see Chapter 23). The evolution of metabolism has led to the dedication of one of these four NTPs to each of the major branches of metabolism. To complete the picture, the four NTPs and their dNTP counterparts are the substrates for the synthesis of the remaining great class of biomolecules—the nucleic acids.

The Bases of Nucleotides Serve as "Information Symbols"

Virtually all of the biochemical reactions of nucleotides involve either *phosphate* or *pyrophosphate group transfer:* the release of a phosphoryl group from an NTP to give an NDP, the release of a pyrophosphoryl group to give an NMP unit, or the acceptance of a phosphoryl group by an NMP or an NDP to give an NDP or an NTP (Figure 11.16). Interestingly, the pentose and the base are *not*

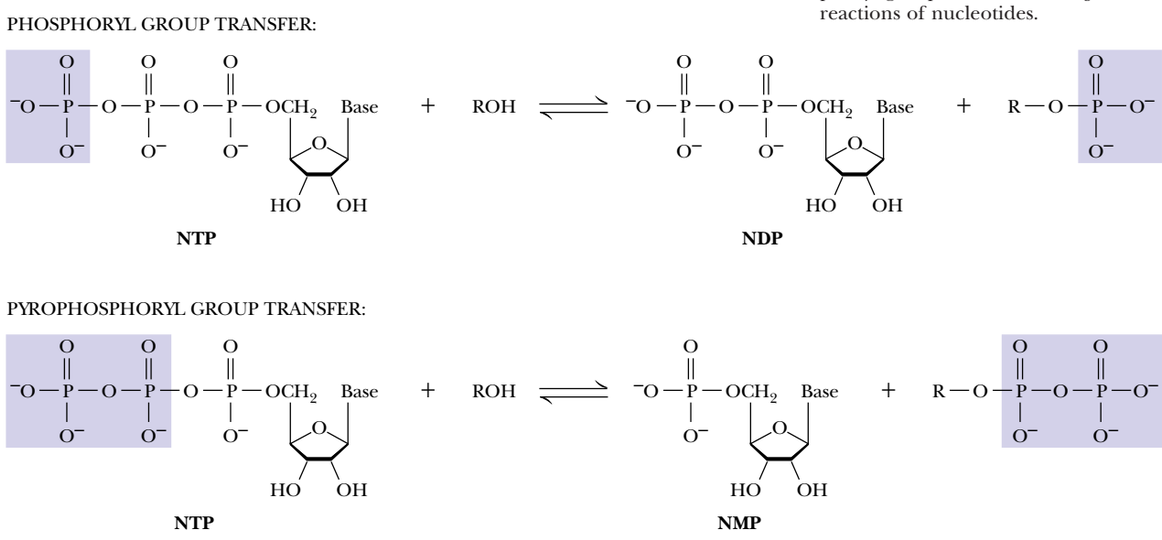


FIGURE 11.16 • Phosphoryl and pyrophosphoryl group transfer, the major biochemical reactions of nucleotides.

directly involved in this chemistry. However, a "division of labor" directs ATP to serve as the primary nucleotide in central pathways of energy metabolism, while GTP, for example, is used to drive protein synthesis. Thus, the various nucleotides are channeled in appropriate metabolic directions through specific recognition of the base of the nucleotide. That is, the bases of nucleotides serve solely as *information symbols* aloof from the covalent bond chemistry that goes on. This role as information symbols extends to nucleotide polymers, the nucleic acids, where the bases serve as the information symbols for the code of genetic information.

11.5 • Nucleic Acids Are Polynucleotides

Nucleic acids are linear polymers of nucleotides linked 3′ to 5′ by **phosphodiester bridges** (Figure 11.17). They are formed as 5′-nucleoside monophosphates are successively added to the 3′-OH group of the preceding nucleotide, a process that gives the polymer a directional sense. Polymers of ribonucleotides are named **ribonucleic acid,** or **RNA.** Deoxyribonucleotide polymers are called **deoxyribonucleic acid,** or **DNA.** Because C-1′ and C-4′ in deoxyribonucleotides are involved in furanose ring formation and because there is no 2′-OH, only

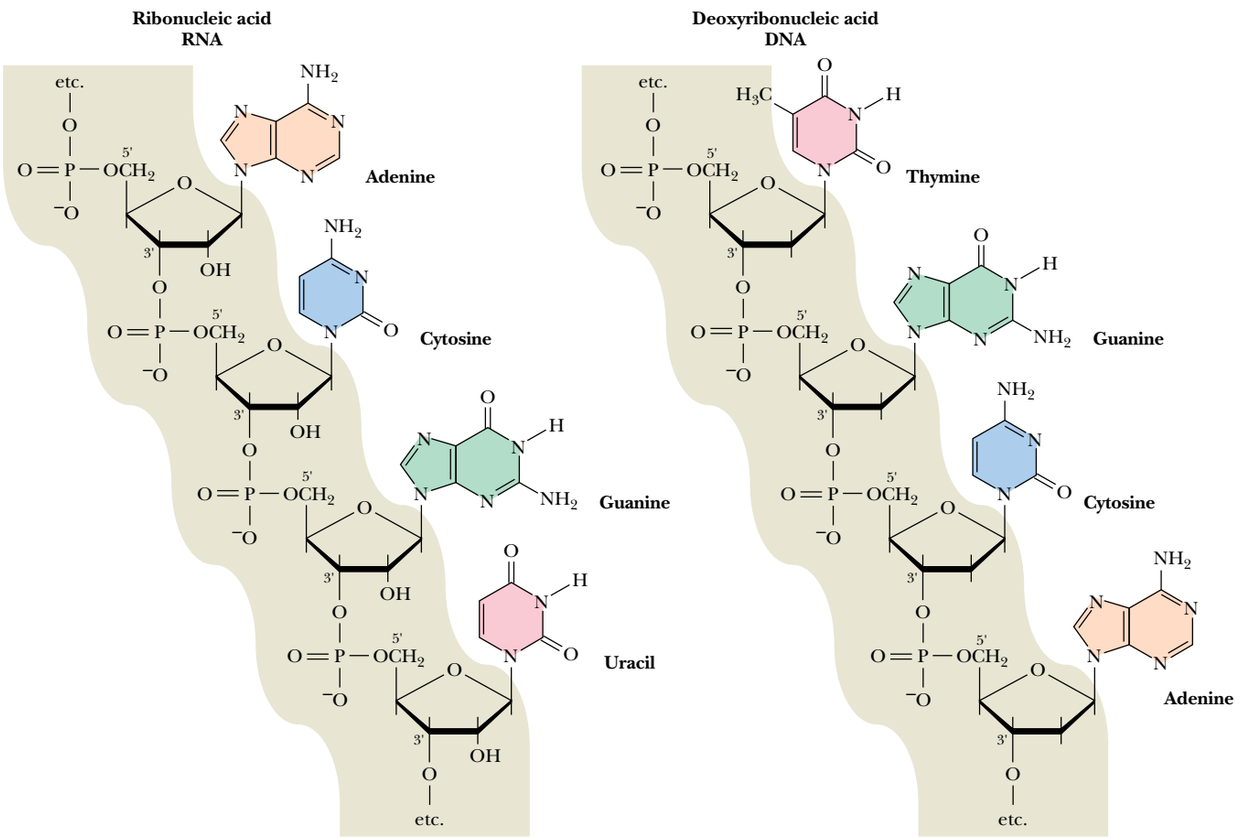


FIGURE 11.17 • 3′-5′ phosphodiester bridges link nucleotides together to form polynucleotide chains.

the 3′- and 5′-hydroxyl groups are available for internucleotide phosphodiester bonds. In the case of DNA, a polynucleotide chain may contain hundreds of millions of nucleotide units. Any structural representation of such molecules would be cumbersome at best, even for a short oligonucleotide stretch.

Shorthand Notations for Polynucleotide Structures

Several conventions have been adopted to convey the sense of polynucleotide structures. A repetitious uniformity exists in the covalent backbone of polynucleotides, in which the chain can be visualized as running from 5′ to 3′ along the atoms of one furanose and thence across the phosphodiester bridge to the furanose of the next nucleotide in line. Thus, this backbone can be portrayed by the symbol of a vertical line representing the furanose and a slash representing the phosphodiester link, as shown in Figure 11.18. The diagonal slash runs from the middle of a furanose line to the bottom of an adjacent one to indicate the 3′- (middle) to 5′- (bottom) carbons of neighboring furanoses joined by the phosphodiester bridge. The base attached to each furanose is indicated above it by a one-letter designation: A, C, G, or U (or T). *The convention in all notations of nucleic acid structure is to read the polynucleotide chain from the 5′-end of the polymer to the 3′-end.* Note that this reading direction actually passes through each phosphodiester from 3′ to 5′.

Base Sequence

The only significant variation that commonly occurs in the chemical structure of nucleic acids is the nature of the base at each nucleotide position. These bases are not part of the sugar–phosphate backbone but instead serve as distinctive side chains, much like the R groups of amino acids along a polypeptide backbone. They give the polymer its unique identity. A simple notation of these structures is merely to list the order of bases in the polynucleotide using single capital letters—A, G, C, and U (or T). Occasionally, a lowercase "p" is written between each successive base to indicate the phosphodiester bridge, as in GpApCpGpUpA. A "p" preceding the sequence indicates that the nucleic acid carries a PO_4 on its 5′-end, as in pGpApCpGpUpA; a "p" terminating the sequence connotes the presence of a phosphate on the 3′-OH end, as in GpApCpGpUpAp.

A more common method of representing nucleotide sequences is to omit the "p" and write only the order of bases, such as GACGUA. This notation assumes the presence of the phosphodiesters joining adjacent nucleotides. The presence of 3′- or 5′-phosphate termini, however, must still be specified, as in GACGUAp for a 3′-PO_4 terminus. To distinguish between RNA and DNA sequences, DNA sequences are typically preceded by a lowercase "d" to denote deoxy, as in d-GACGTA. From a simple string of letters such as this, any biochemistry student should be able to draw the unique chemical structure for a pentanucleotide, even though it may contain over 200 atoms.

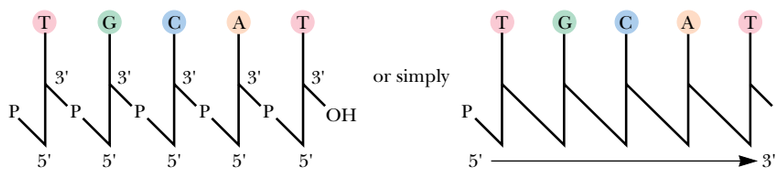


FIGURE 11.18 • Furanoses are represented by lines; phosphodiesters are represented by diagonal slashes in this shorthand notation for nucleic acid structures.

11.6 • Classes of Nucleic Acids

The two major classes of nucleic acids are DNA and RNA. DNA has only one biological role, but it is the more central one. The information to make all the functional macromolecules of the cell (even DNA itself) is preserved in DNA and accessed through transcription of the information into RNA copies. Coincident with its singular purpose, there is only a single DNA molecule (or "chromosome") in simple life forms such as viruses or bacteria. Such DNA molecules must be quite large in order to embrace enough information for making the macromolecules necessary to maintain a living cell. The *Escherichia coli* chromosome has a molecular mass of 2.9×10^9 D and contains over 9 million nucleotides. Eukaryotic cells have many chromosomes, and DNA is found principally in two copies in the diploid chromosomes of the nucleus, but it also occurs in mitochondria and in chloroplasts, where it encodes some of the proteins and RNAs unique to these organelles.

In contrast, RNA occurs in multiple copies and various forms (Table 11.2). Cells contain up to eight times as much RNA as DNA. RNA has a number of important biological functions, and on this basis, RNA molecules are categorized into several major types: **messenger RNA, ribosomal RNA,** and **transfer RNA.** Eukaryotic cells contain an additional type, **small nuclear RNA (snRNA).** With these basic definitions in mind, let's now briefly consider the chemical and structural nature of DNA and the various RNAs. Chapter 12 elaborates on methods to determine the primary structure of nucleic acids by sequencing methods and discusses the secondary and tertiary structures of DNA and RNA. Part IV, Information Transfer, includes a detailed treatment of the dynamic role of nucleic acids in the molecular biology of the cell.

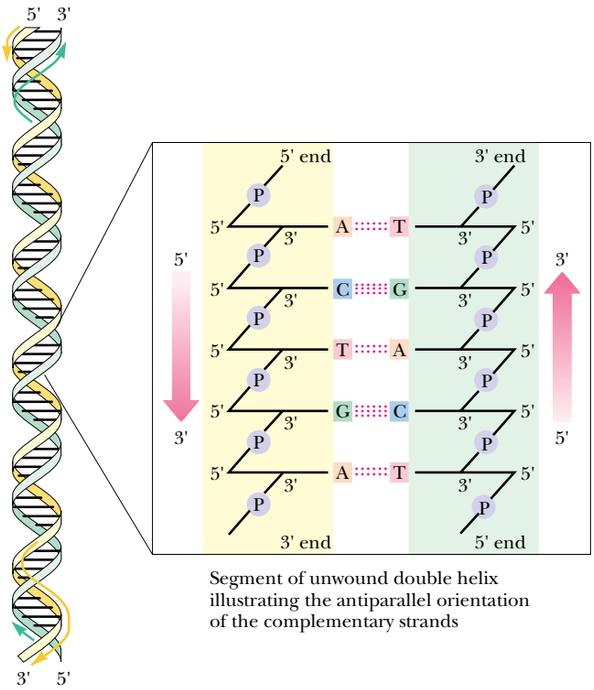
DNA

The DNA isolated from different cells and viruses characteristically consists of two polynucleotide strands wound together to form a long, slender, helical molecule, the **DNA double helix.** The strands run in opposite directions; that is, they are *antiparallel* and are held together in the double helical structure through *interchain hydrogen bonds* (Figure 11.19). These H bonds pair the bases of nucleotides in one chain to complementary bases in the other, a phenomenon called **base pairing.**

Table 11.2

Various Kinds of RNA Found in an *E. coli* Cell

Type	Sedimentation Coefficient	Molecular Weight	Number of Nucleotide Residues	Percentage of Total Cell RNA
mRNA	6–25	25,000–1,000,000	75–3,000	~2
tRNA	~4	23,000–30,000	73–94	16
rRNA	5	35,000	120	
	16	550,000	1542	82
	23	1,100,000	2904	



Segment of unwound double helix illustrating the antiparallel orientation of the complementary strands

FIGURE 11.19 • The antiparallel nature of the DNA double helix.

Chargaff's Rules

A clue to the chemical basis of base pairing in DNA came from the analysis of the base composition of various DNAs by Erwin Chargaff in the late 1940s. His data showed that the four bases commonly found in DNA (A, C, G, and T) do not occur in equimolar amounts and that the relative amounts of each vary from species to species (Table 11.3). Nevertheless, Chargaff noted that certain pairs of bases, namely, adenine and thymine, and guanine and cytosine, are

Table 11.3

Molar Ratios Leading to the Formulation of Chargaff's Rules

Source	Adenine to Guanine	Thymine to Cytosine	Adenine to Thymine	Guanine to Cytosine	Purines to Pyrimidines
Ox	1.29	1.43	1.04	1.00	1.1
Human	1.56	1.75	1.00	1.00	1.0
Hen	1.45	1.29	1.06	0.91	0.99
Salmon	1.43	1.43	1.02	1.02	1.02
Wheat	1.22	1.18	1.00	0.97	0.99
Yeast	1.67	1.92	1.03	1.20	1.0
Hemophilus influenzae	1.74	1.54	1.07	0.91	1.0
E. coli K-12	1.05	0.95	1.09	0.99	1.0
Avian tubercle bacillus	0.4	0.4	1.09	1.08	1.1
Serratia marcescens	0.7	0.7	0.95	0.86	0.9
Bacillus schatz	0.7	0.6	1.12	0.89	1.0

Source: After Chargaff, E., 1951. *Federation Proceedings* **10:**654–659.

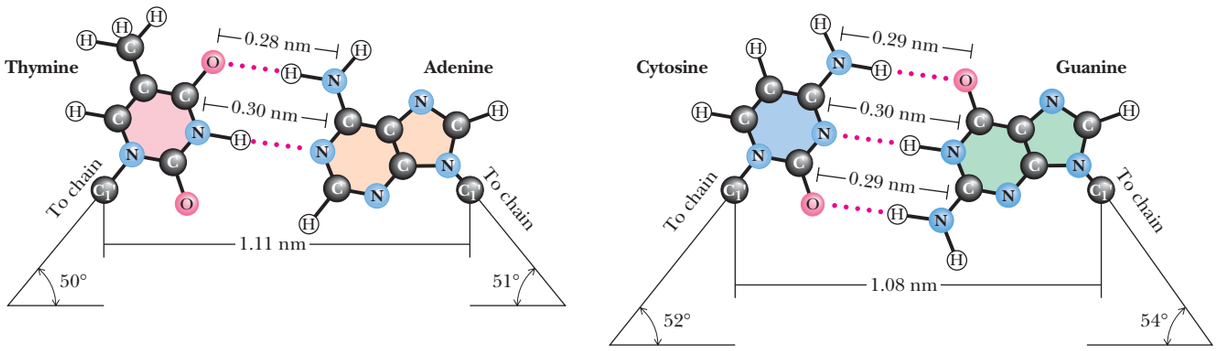


FIGURE 11.20 • The Watson–Crick base pairs A : T and G : C.

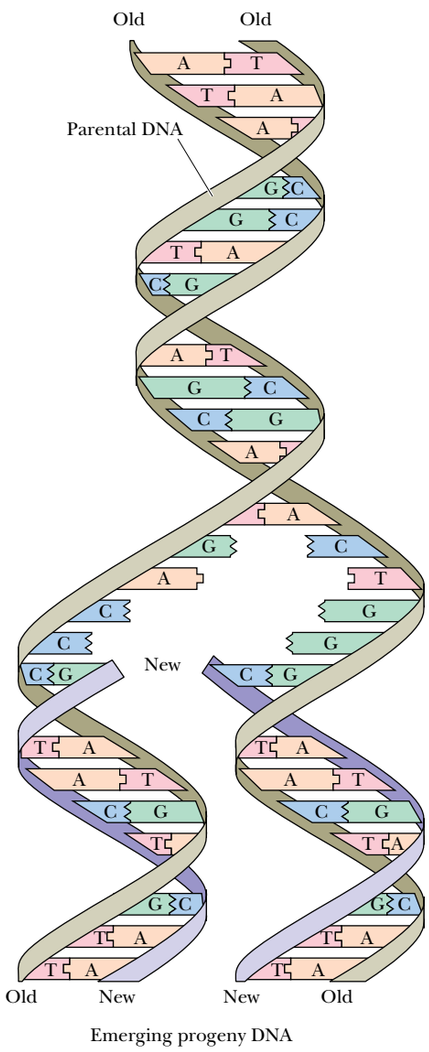


FIGURE 11.21 • Replication of DNA gives identical progeny molecules because base pairing is the mechanism determining the nucleotide sequence synthesized within each of the new strands during replication.

always found in a 1 : 1 ratio and that the number of pyrimidine residues always equals the number of purine residues. These findings are known as *Chargaff's rules:* **[A] = [T]; [C] = [G]; [pyrimidines] = [purines].**

Watson and Crick's Double Helix

James Watson and Francis Crick, working in the Cavendish Laboratory at Cambridge University in 1953, took advantage of Chargaff's results and the data obtained by Rosalind Franklin and Maurice Wilkins in X-ray diffraction studies on the structure of DNA to conclude that DNA was a *complementary double helix.* Two strands of deoxyribonucleic acid (sometimes referred to as the *Watson strand* and the *Crick strand*) are held together by hydrogen bonds formed between unique base pairs, always consisting of a purine in one strand and a pyrimidine in the other. Base pairing is very specific: if the purine is adenine, the pyrimidine must be thymine. Similarly, guanine pairs only with cytosine (Figure 11.20). Thus, if an A occurs in one strand of the helix, T must occupy the complementary position in the opposing strand. Likewise, a G in one dictates a C in the other. Because exceptions to this exclusive pairing of A only with T and G only with C are rare, these pairs are taken as the standard or accepted law, and the A:T and G:C base pairs are often referred to as canonical. As Watson recognized from testing various combinations of bases using structurally accurate models, the A:T pair and the G:C pair form spatially equivalent units (Figure 11.20). The backbone-to-backbone distance of an A:T pair is 1.11 nm, virtually identical to the 1.08 nm chain separation in G:C base pairs.

The DNA molecule not only conforms to Chargaff's rules but also has a profound property relating to heredity: *The sequence of bases in one strand has a complementary relationship to the sequence of bases in the other strand.* That is, the information contained in the sequence of one strand is conserved in the sequence of the other. Therefore, separation of the two strands and faithful replication of each, through a process in which base pairing specifies the nucleotide sequence in the newly synthesized strand, leads to two progeny molecules identical in every respect to the parental double helix (Figure 11.21). Elucidation of the double helical structure of DNA represented one of the most significant events in the history of science. This discovery more than any other marked the beginning of molecular biology. Indeed, upon solving the structure of DNA, Crick proclaimed in The Eagle, a pub just across from the Cavendish lab, "We have discovered the secret of life!"

FIGURE 11.22 • If the cell walls of bacteria such as *Escherichia coli* are partially digested and the cells are then osmotically shocked by dilution with water, the contents of the cells are extruded to the exterior. In electron micrographs, the most obvious extruded component is the bacterial chromosome, shown here surrounding the cell. *(Dr. Gopal Murti/CNRI/Phototake NYC)*

Size of DNA Molecules

Because of the double helical nature of DNA molecules, their size can be represented in terms of the numbers of nucleotide base pairs they contain. For example, the *E. coli* chromosome consists of 4.64×10^6 base pairs (abbreviated bp) or 4.64×10^3 kilobase pairs (kbp). DNA is a threadlike molecule. The diameter of the DNA double helix is only 2 nm, but the length of the DNA molecule forming the *E. coli* chromosome is over 1.6×10^6 nm (1.6 mm). Because the long dimension of an *E. coli* cell is only 2000 nm (0.002 mm), its chromosome must be highly folded. Because of their long, threadlike nature, DNA molecules are easily sheared into shorter fragments during isolation procedures, and it is difficult to obtain intact chromosomes even from the simple cells of prokaryotes.

DNA in the Form of Chromosomes

DNA occurs in various forms in different cells. The single chromosome of prokaryotic cells (Figure 11.22) is typically a circular DNA molecule. Relatively little protein is associated with prokaryotic chromosomes. In contrast, the DNA molecules of eukaryotic cells, each of which defines a chromosome, are linear and richly adorned with proteins. A class of arginine- and lysine-rich basic proteins called **histones** interact ionically with the anionic phosphate groups in the DNA backbone to form **nucleosomes,** structures in which the DNA double helix is wound around a protein "core" composed of pairs of four different histone polypeptides (Figure 11.23; see also Section 12.5 in Chapter 12). Chromosomes also contain a varying mixture of other proteins, so-called **nonhistone chromosomal proteins,** many of which are involved in regulating which genes in DNA are transcribed at any given moment. The amount of DNA in a diploid mammalian cell is typically more than 1000 times that found in an *E. coli* cell. Some higher plant cells contain more than 50,000 times as much.

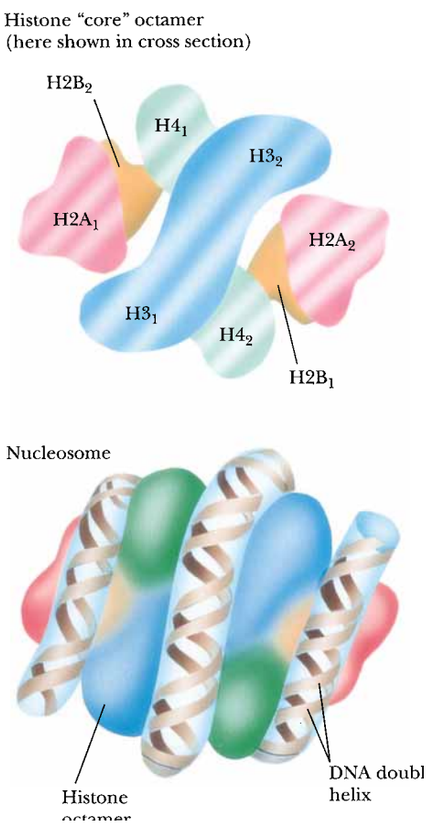


FIGURE 11.23 • A diagram of the histone octamer. Nucleosomes consist of two turns of DNA supercoiled about a histone "core" octamer.

RNA

Messenger RNA

Messenger RNA **(mRNA)** serves to carry the information or "message" that is encoded in genes to the sites of protein synthesis in the cell, where this information is translated into a polypeptide sequence. Because mRNA molecules are transcribed copies of the protein-coding genetic units that comprise most of DNA, mRNA is said to be "the DNA-like RNA."

Messenger RNA is synthesized during **transcription,** an enzymatic process in which an RNA copy is made of the sequence of bases along one strand of DNA. This mRNA then directs the synthesis of a polypeptide chain as the information that is contained within its nucleotide sequence is translated into an amino acid sequence by the protein-synthesizing machinery of the ribosomes. Ribosomal RNA and tRNA molecules are also synthesized by transcription of DNA sequences, but unlike mRNA molecules, these RNAs are not subsequently translated to form proteins. Only the genetic units of DNA sequence that encode proteins are transcribed into mRNA molecules. In prokaryotes, a single mRNA may contain the information for the synthesis of several polypeptide chains within its nucleotide sequence (Figure 11.24). In contrast, eukaryotic mRNAs encode only one polypeptide, but are more complex in that they are synthesized in the nucleus in the form of much larger precursor molecules called **heterogeneous nuclear RNA,** or **hnRNA.** hnRNA molecules contain stretches of nucleotide sequence that have no protein-coding capacity. These

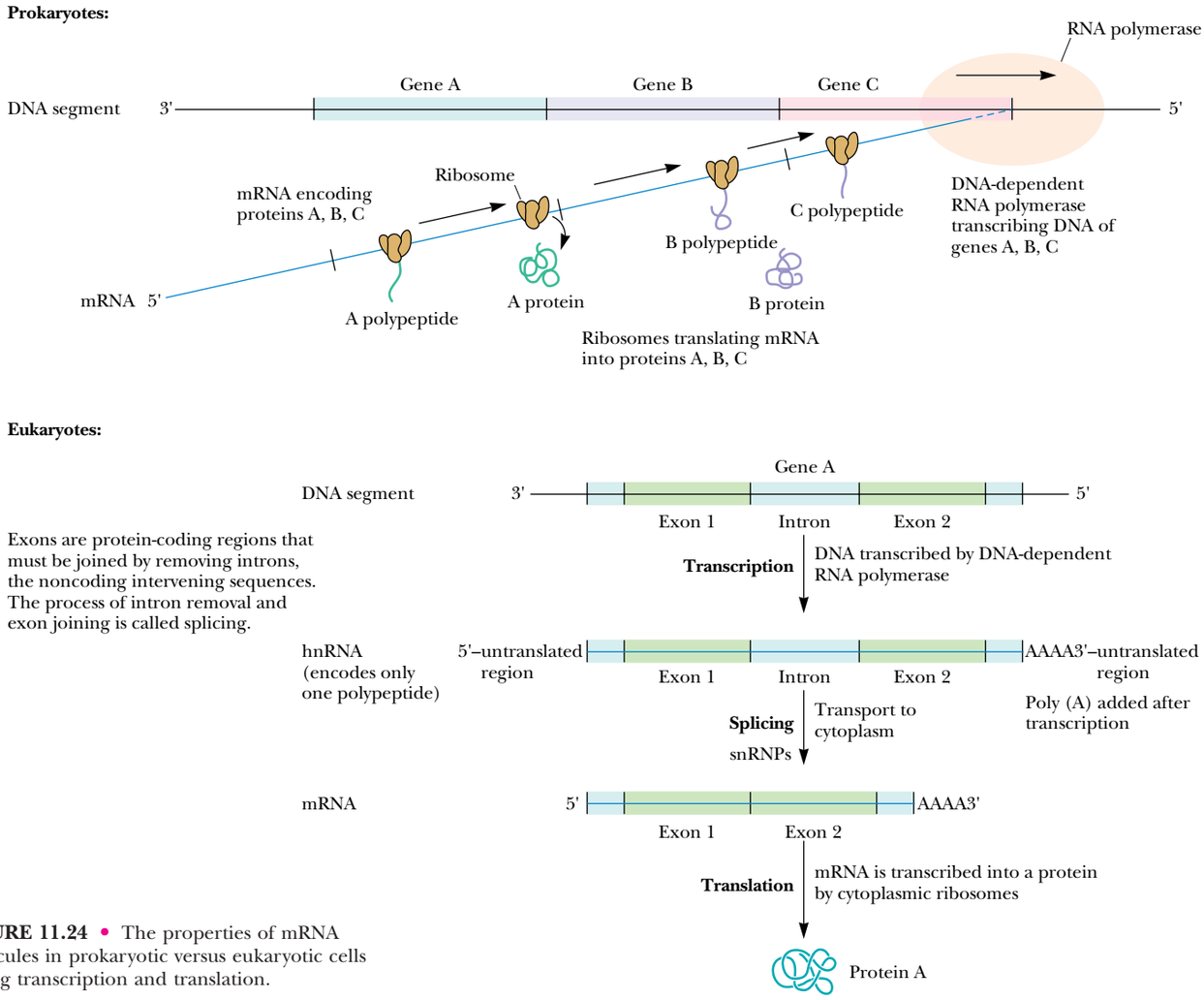


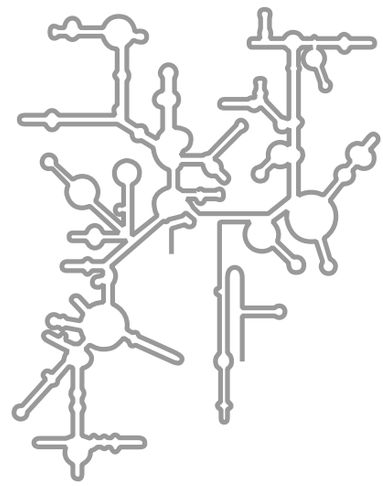
FIGURE 11.24 • The properties of mRNA molecules in prokaryotic versus eukaryotic cells during transcription and translation.

noncoding regions are called **intervening sequences** or **introns** because they intervene between coding regions, which are called **exons.** Introns interrupt the continuity of the information specifying the amino acid sequence of a protein and must be spliced out before the message can be translated. In addition, eukaryotic hnRNA and mRNA molecules have a run of 100 to 200 adenylic acid residues attached at their 3′-ends, so-called **poly(A) tails.** This polyadenylylation occurs after transcription has been completed and is believed to contribute to mRNA stability. The properties of messenger RNA molecules as they move through transcription and translation in prokaryotic versus eukaryotic cells are summarized in Figure 11.24.

Ribosomal RNA

Ribosomes, the supramolecular assemblies where protein synthesis occurs, are about 65% RNA of the ribosomal RNA type. Ribosomal RNA **(rRNA)** molecules fold into characteristic secondary structures as a consequence of intramolecular hydrogen bond interactions (marginal figure). The different species of rRNA are generally referred to according to their **sedimentation coefficients**[1] (see the Appendix to Chapter 5), which are a rough measure of their relative size (Table 11.2 and Figure 11.25).

Ribosomes are composed of two subunits of different sizes that dissociate from each other if the Mg^{2+} concentration is below 10^{-3} *M*. Each subunit is



Ribosomal RNA has a complex secondary structure due to many intrastrand hydrogen bonds.

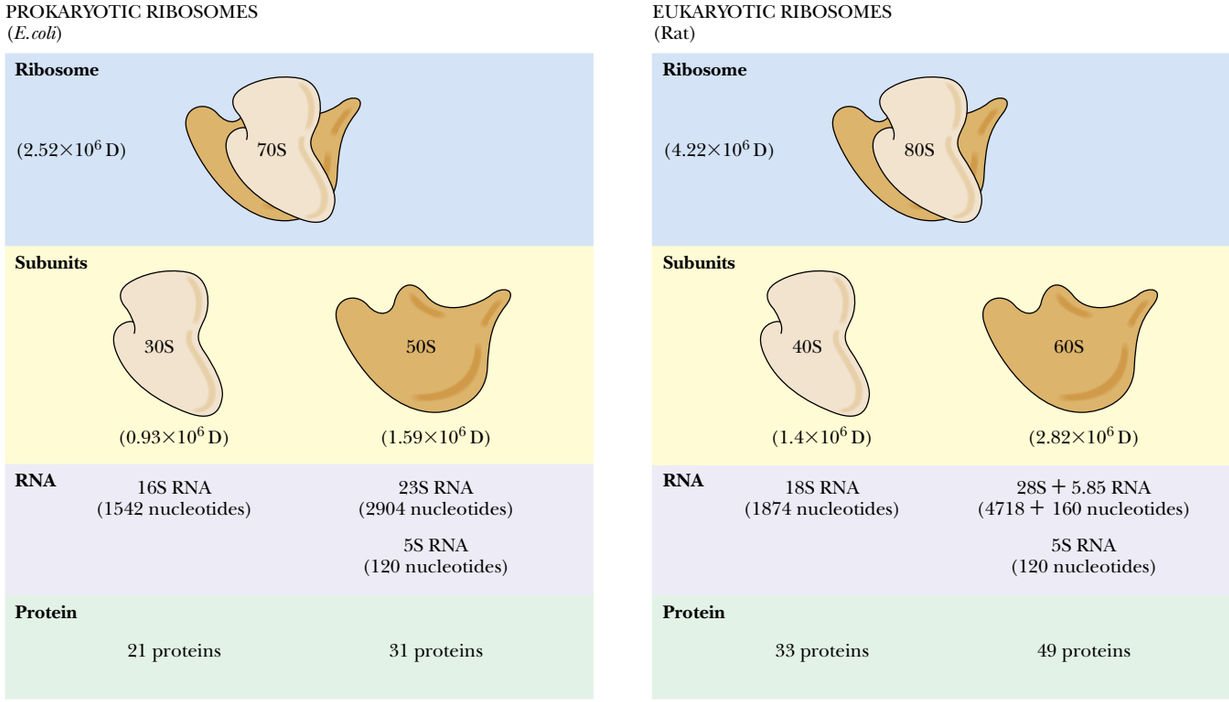


FIGURE 11.25 • The organization and composition of prokaryotic and eukaryotic ribosomes.

[1]Sedimentation coefficients are a measure of the velocity with which a particle sediments in a centrifugal force field. Sedimentation coefficients are typically expressed in **Svedbergs** (symbolized S), named to honor The Svedberg, developer of the ultracentrifuge. One S equals 10^{-13} sec.

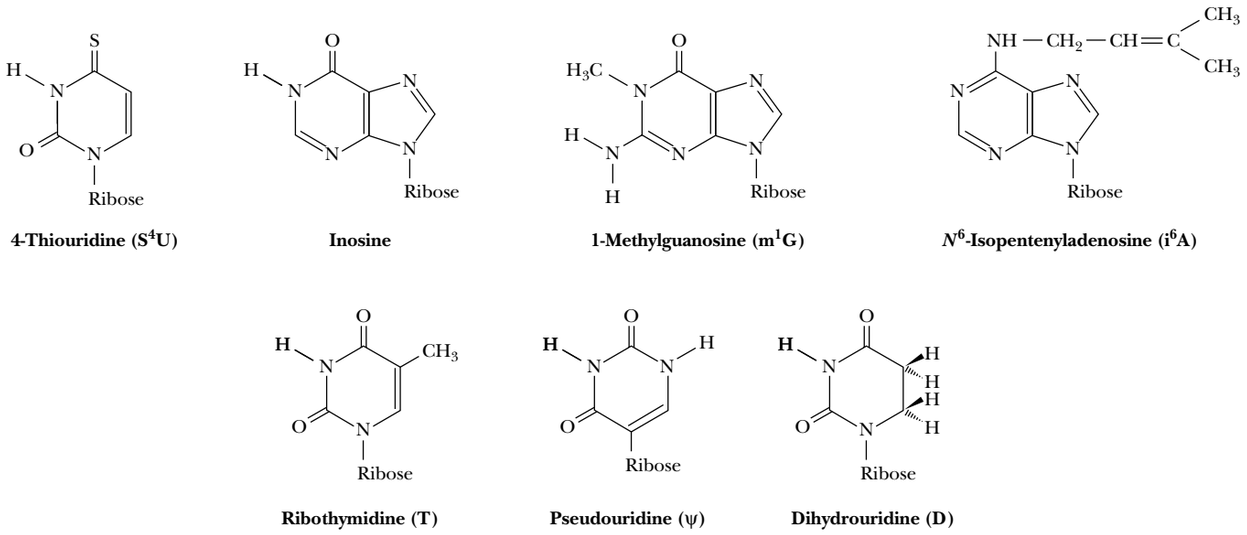


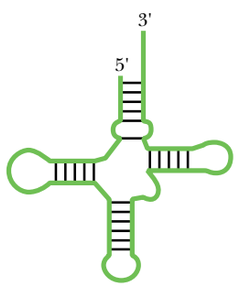
FIGURE 11.26 • Unusual bases of RNA—pseudouridine, ribothymidylic acid, and various methylated bases.

a supramolecular assembly of proteins and RNA and has a total mass of 10^6 daltons or more. *E. coli* ribosomal subunits have sedimentation coefficients of 30S (the small subunit) and 50S (the large subunit). Eukaryotic ribosomes are somewhat larger than prokaryotic ribosomes, consisting of 40S and 60S subunits. The properties of ribosomes and their rRNAs are summarized in Figure 11.25. The 30S subunit of *E. coli* contains a single RNA chain of 1542 nucleotides. This small subunit rRNA itself has a sedimentation coefficient of 16S. The large *E. coli* subunit has two rRNA molecules, a 23S (2904 nucleotides) and a 5S (120 nucleotides). The ribosomes of a typical eukaryote, the rat, have rRNA molecules of 18S (1874 nucleotides) and 28S (4718 bases), 5.8S (160 bases), and 5S (120 bases). The 18S rRNA is in the 40S subunit and the latter three are all part of the 60S subunit.

Ribosomal RNAs characteristically contain a number of specially modified nucleotides, including **pseudouridine** residues, **ribothymidylic acid,** and **methylated bases** (Figure 11.26). The central role of ribosomes in the biosynthesis of proteins is treated in detail in Chapter 33. Here we briefly note the significant point that genetic information in the nucleotide sequence of an mRNA is translated into the amino acid sequence of a polypeptide chain by ribosomes.

Transfer RNA

Transfer RNA **(tRNA)** serves as a carrier of amino acid residues for protein synthesis. Transfer RNA molecules also fold into a characteristic secondary structure (marginal figure). The amino acid is attached as an aminoacyl ester to the 3′-terminus of the tRNA. Aminoacyl-tRNAs are the substrates for protein biosynthesis. The tRNAs are the smallest RNAs (size range—23 to 30 kD) and contain 73 to 94 residues, a substantial number of which are methylated or otherwise unusually modified. Transfer RNA derives its name from its role as the carrier of amino acids during the process of protein synthesis (see Chapters 32 and 33). Each of the 20 amino acids of proteins has at least one unique tRNA species dedicated to chauffeuring its delivery to ribosomes for insertion into growing polypeptide chains, and some amino acids are served by several tRNAs. For example, five different tRNAs act in the transfer of leucine into



Transfer RNA also has a complex secondary structure due to many intrastrand hydrogen bonds.

proteins. In eukaryotes, there are even discrete sets of tRNA molecules for each site of protein synthesis—the cytoplasm, the mitochondrion, and, in plant cells, the chloroplast. All tRNA molecules possess a 3′-terminal nucleotide sequence that reads **-CCA,** and the amino acid is carried to the ribosome attached as an acyl ester to the free 3′-OH of the terminal A residue. These **aminoacyl-tRNAs** are the substrates of protein synthesis, the amino acid being transferred to the carboxyl end of a growing polypeptide. The peptide bond-forming reaction is a catalytic process intrinsic to ribosomes.

Small Nuclear RNAs

Small nuclear RNAs, or **snRNAs,** are a class of RNA molecules found in eukaryotic cells, principally in the nucleus. They are neither tRNA nor small rRNA molecules, although they are similar in size to these species. They contain from 100 to about 200 nucleotides, some of which, like tRNA and rRNA, are methylated or otherwise modified. No snRNA exists as naked RNA. Instead, snRNA is found in stable complexes with specific proteins forming **small nuclear ribonucleoprotein particles,** or **snRNPs,** which are about 10S in size. Their occurrence in eukaryotes, their location in the nucleus, and their relative abundance (1 to 10% of the number of ribosomes) are significant clues to their biological purpose: snRNPs are important in the processing of eukaryotic gene transcripts (hnRNA) into mature messenger RNA for export from the nucleus to the cytoplasm (Figure 11.24).

Significance of Chemical Differences Between DNA and RNA

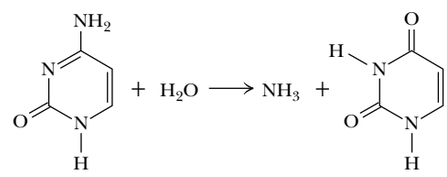
Two fundamental chemical differences distinguish DNA from RNA:

1. DNA contains 2-deoxyribose instead of ribose.
2. DNA contains thymine instead of uracil.

What are the consequences of these differences and do they hold any significance in common? An argument can be made that, because of these differences, DNA is a more stable polymeric form than RNA. The greater stability of DNA over RNA is consistent with the respective roles these macromolecules have assumed in heredity and information transfer.

Consider first why DNA contains thymine instead of uracil. The key observation is that *cytosine deaminates to form uracil* at a finite rate *in vivo* (Figure 11.27). Because C in one DNA strand pairs with G in the other strand, whereas U would pair with A, conversion of a C to a U could potentially result in a heritable change of a CG pair to a UA pair. Such a change in nucleotide sequence would constitute a *mutation* in the DNA. To prevent this reaction from leading to changes in nucleotide sequence, a cellular repair mechanism "proofreads" DNA, and when a U arising from C deamination is encountered, it is treated as inappropriate and is replaced by a C. If DNA normally contained U rather than T, this repair system could not readily distinguish U formed by C deamination from U correctly paired with A. However, the U in DNA is "5-methyl-U" or, as it is conventionally known, thymine (Figure 11.28). That is, the 5-methyl group on T labels it as if to say "this U belongs; do not replace it."

The ribose 2′-OH group of RNA is absent in DNA. Consequently, the ubiquitous 3′-O of polynucleotide backbones lacks a vicinal hydroxyl neighbor in DNA. This difference leads to a greater resistance of DNA to alkaline hydrolysis, examined in detail in the following section. To view it another way, RNA is less stable than DNA because its vicinal 2′-OH group makes the 3′-phosphodiester bond susceptible to nucleophilic cleavage (Figure 11.29). For just this reason, it is selectively advantageous for the heritable form of genetic information to be DNA rather than RNA.



Cytosine **Uracil**

FIGURE 11.27 • Deamination of cytosine forms uracil.

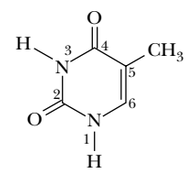
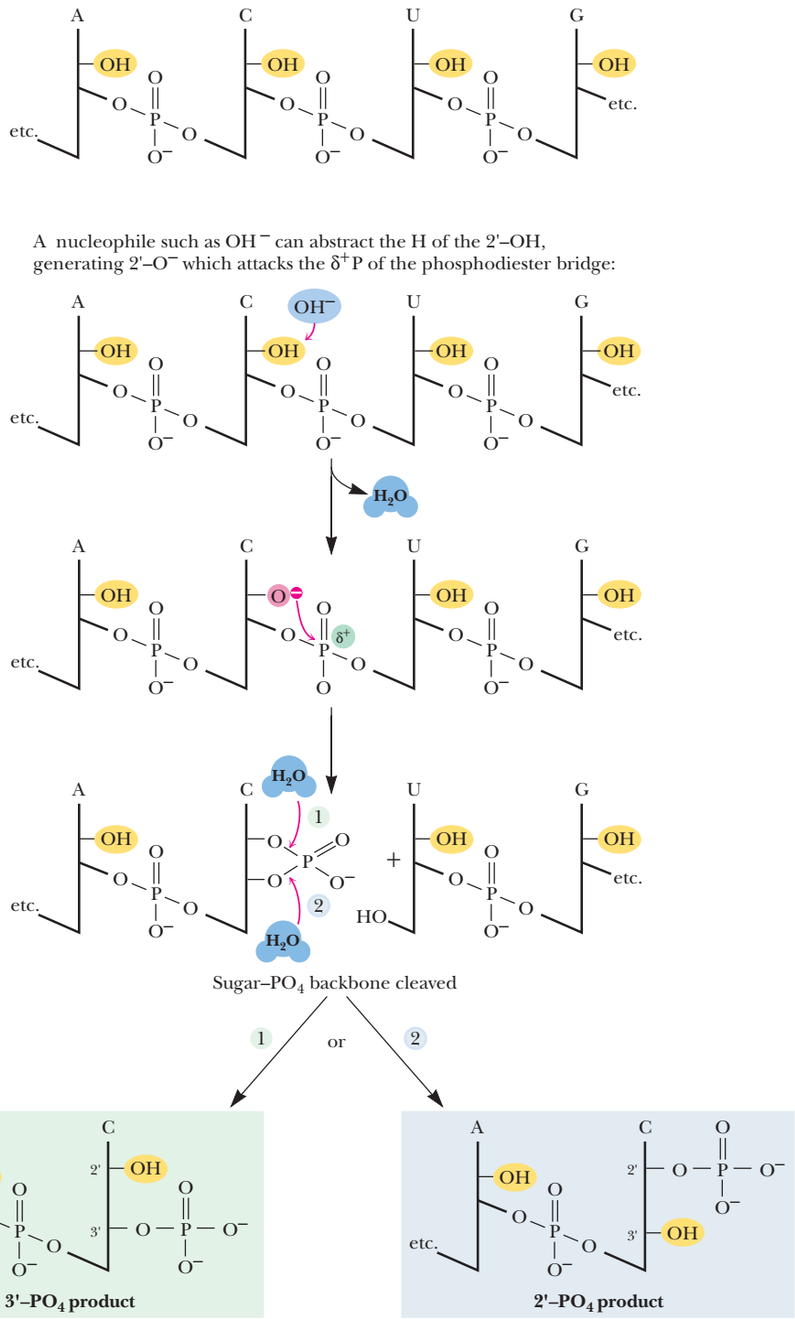


FIGURE 11.28 • The 5-methyl group on thymine labels it as a special kind of uracil.

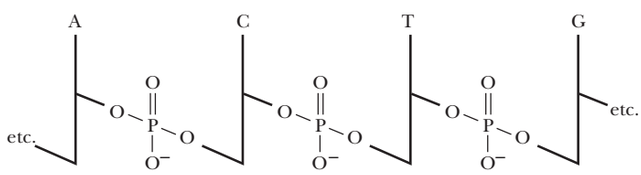
FIGURE 11.29 • The vicinal —OH groups of RNA are susceptible to nucleophilic attack leading to hydrolysis of the phosphodiester bond and fracture of the polynucleotide chain; DNA lacks a 2′-OH vicinal to its 3′-O-phosphodiester backbone. Alkaline hydrolysis of RNA results in the formation of a mixture of 2′- and 3′-nucleoside monophosphates.

RNA:



Complete hydrolysis of RNA by alkali yields a random mixture of 2'–NMPs and 3'–NMPs.

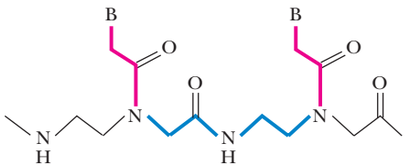
DNA: no 2'– OH; resistant to OH^-:



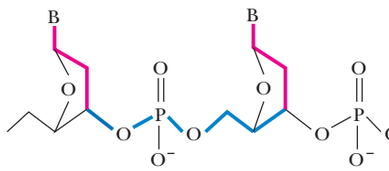
A DEEPER LOOK

Peptide Nucleic Acids (PNAs) Are Synthetic Mimics of DNA and RNA

Synthetic chemists have invented analogs of DNA (and RNA) in which the sugar–phosphate backbone is replaced by a peptide backbone, creating a polymer appropriately termed a *peptide nucleic acid,* or *PNA*. The PNA peptide backbone was designed so that the space between successive bases was the same as in natural DNA (see figure). PNA consists of repeating units of *N*-(2-aminoethyl)-glycine residues linked by peptide bonds; the bases are attached to this backbone through methylene carbonyl linkages. This chemistry provides six bonds along the backbone between bases and three bonds between the backbone and each base, just like natural DNA. PNA oligomers interact with DNA (and RNA) through specific base-pairing interactions, just as would be expected for a pair of complementary oligonucleotides. PNAs are resistant to nucleases and also are poor substrates for proteases. PNAs thus show great promise as specific diagnostic probes for unique DNA or RNA nucleotide sequences. PNAs also have potential application as antisense drugs (see problem 5 in the end-of-chapter problems).



PNA



Note the repeating six bonds (in bold) between base attachments and the three-bond linker between base (B) and backbone.

Buchardt, O., et al., 1993. Peptide nucleic acids and their potential applications in biotechnology. *Trends in Biotechnology* **11:***384–386.*

11.7 • Hydrolysis of Nucleic Acids

Most reactions of nucleic acid hydrolysis break bonds in the polynucleotide backbone. Such reactions are important because they can be used to manipulate these polymeric molecules. For example, hydrolysis of polynucleotides generates smaller fragments whose nucleotide sequence can be more easily determined.

Hydrolysis by Acid or Base

RNA is relatively resistant to the effects of dilute acid, but gentle treatment of DNA with 1 m*M* HCl leads to hydrolysis of purine glycosidic bonds and the loss of purine bases from the DNA. The glycosidic bonds between pyrimidine bases and 2′-deoxyribose are not affected, and, in this case, the polynucleotide's sugar–phosphate backbone remains intact. The purine-free polynucleotide product is called **apurinic acid.**

DNA is not susceptible to alkaline hydrolysis. On the other hand, RNA is alkali labile and is readily hydrolyzed by dilute sodium hydroxide. Cleavage is random in RNA, and the ultimate products are a mixture of nucleoside 2′- and 3′-monophosphates. These products provide a clue to the reaction mechanism (Figure 11.29). Abstraction of the 2′-OH hydrogen by hydroxyl anion leaves a 2′-O^- that carries out a nucleophilic attack on the δ^+ phosphorus atom of the phosphate moiety, resulting in cleavage of the 5′-phosphodiester bond and formation of a cyclic 2′,3′-phosphate. This cyclic 2′,3′-phosphodiester is unstable and decomposes randomly to either a 2′- or 3′-phosphate ester. DNA has no 2′-OH; therefore DNA is alkali stable.

Enzymatic Hydrolysis

Enzymes that hydrolyze nucleic acids are called **nucleases.** Virtually all cells contain various nucleases that serve important housekeeping roles in the normal course of nucleic acid metabolism. Organs that provide digestive fluids, such as the pancreas, are rich in nucleases and secrete substantial amounts to hydrolyze ingested nucleic acids. Fungi and snake venom are often good sources of nucleases. As a class, nucleases are **phosphodiesterases** because the reaction that they catalyze is the cleavage of phosphodiester bonds by H_2O. Because each internal phosphate in a polynucleotide backbone is involved in two phosphoester linkages, cleavage can potentially occur on either side of the phosphorus (Figure 11.30). Convention labels the 3′-side as *a* and the 5′-side as *b*. Cleavage on the *a* side leaves the phosphate attached to the 5′-position of the adjacent nucleotide, while *b*-side hydrolysis yields 3′-phosphate products. Enzymes or reactions that hydrolyze nucleic acids are characterized as acting at either *a* or *b*. A second convention denotes whether the nucleic acid chain was cleaved at some internal location, *endo,* or whether a terminal nucleotide residue was hydrolytically removed, *exo*. Note that exo *a* cleavage characteristically occurs at the 3′-end of the polymer, whereas exo *b* cleavage involves attack at the 5′-terminus (Figure 11.31).

Nuclease Specificity

Like most enzymes (see Chapter 14), nucleases exhibit selectivity or *specificity* for the nature of the substance on which they act. That is, some nucleases act only on DNA **(DNases),** while others are specific for RNA (the **RNases**). Still

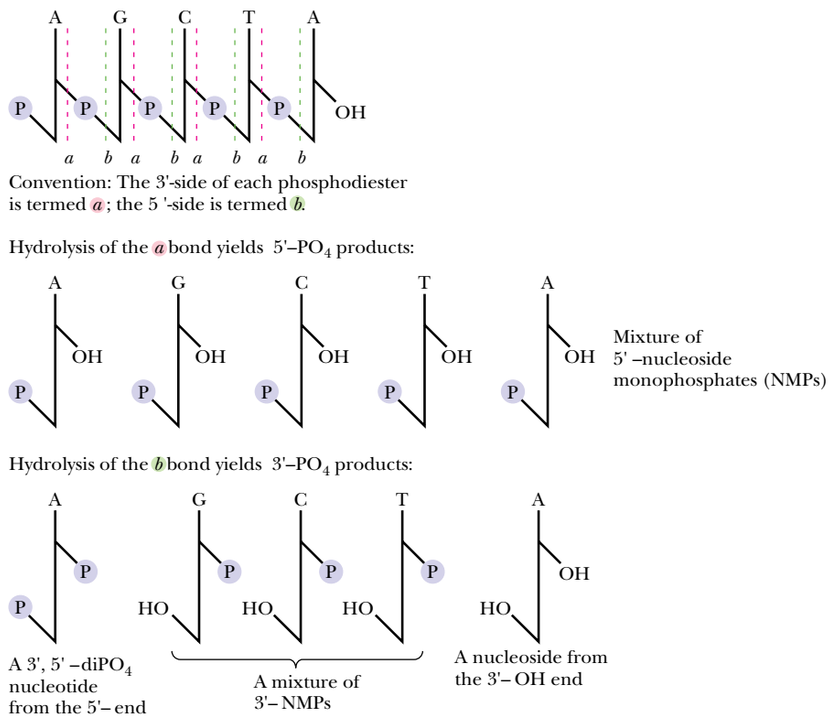


FIGURE 11.30 • Cleavage in polynucleotide chains: *a* cleavage yields 5′-phosphate products, whereas *b* cleavage gives 3′-phosphate products.

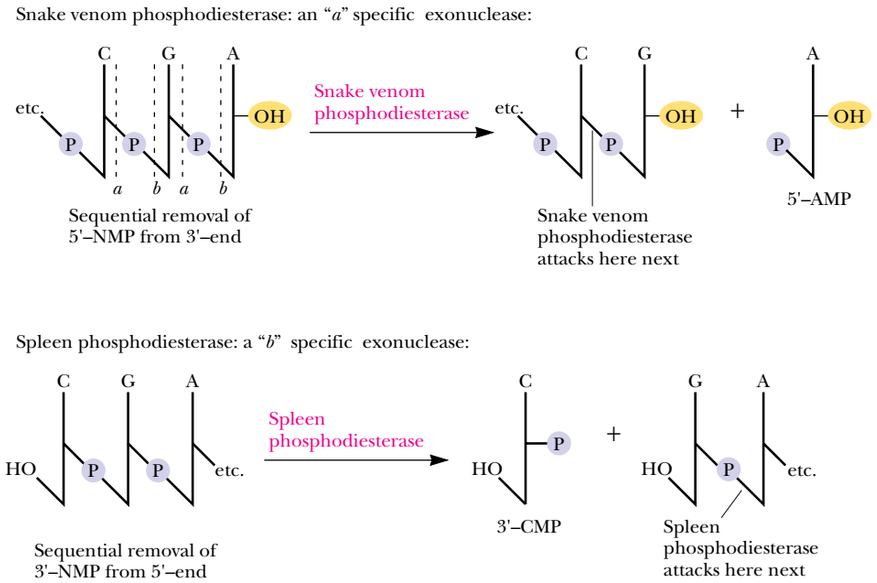


FIGURE 11.31 • Snake venom phosphodiesterase and spleen phosphodiesterase are exonucleases that degrade polynucleotides from opposite ends.

others are nonspecific and are referred to simply as **nucleases,** as in *nuclease S1* (see Table 11.4). Nucleases may also show specificity for only single-stranded nucleic acids or may only act on double helices. Single-stranded nucleic acids are abbreviated by an *ss* prefix, as in ssRNA; the prefix *ds* denotes double-stranded. Nucleases may also display a decided preference for acting only at

Table 11.4

Specificity of Various Nucleases

Enzyme	DNA, RNA, or Both	*a* or *b*	Specificity
Exonucleases			
Snake venom phosphodiesterase	Both	*a*	Starts at 3′-end, 5′-NMP products
Spleen phosphodiesterase	Both	*b*	Starts at 5′-end, 3′-NMP products
Endonucleases			
RNase A (pancreas)	RNA	*b*	Where 3′-PO_4 is to pyrimidine; oligos with pyrimidine 3′-PO_4 ends
Bacillus subtilis RNase	RNA	*b*	Where 3′-PO_4 is to purine; oligos with purine 3′-PO_4 ends
RNase T_1	RNA	*b*	Where 3′-PO_4 is to guanine
RNase T_2	RNA	*b*	Where 3′-PO_4 is to adenine
DNase I (pancreas)	DNA	*a*	Preferably between Py and Pu; nicks dsDNA, creating 3′-OH ends
DNase II (spleen, thymus, *Staphylococcus aureus*)	DNA	*b*	Oligo products
Nuclease S1	Both	*a*	Cleaves single-stranded but not double-stranded nucleic acids

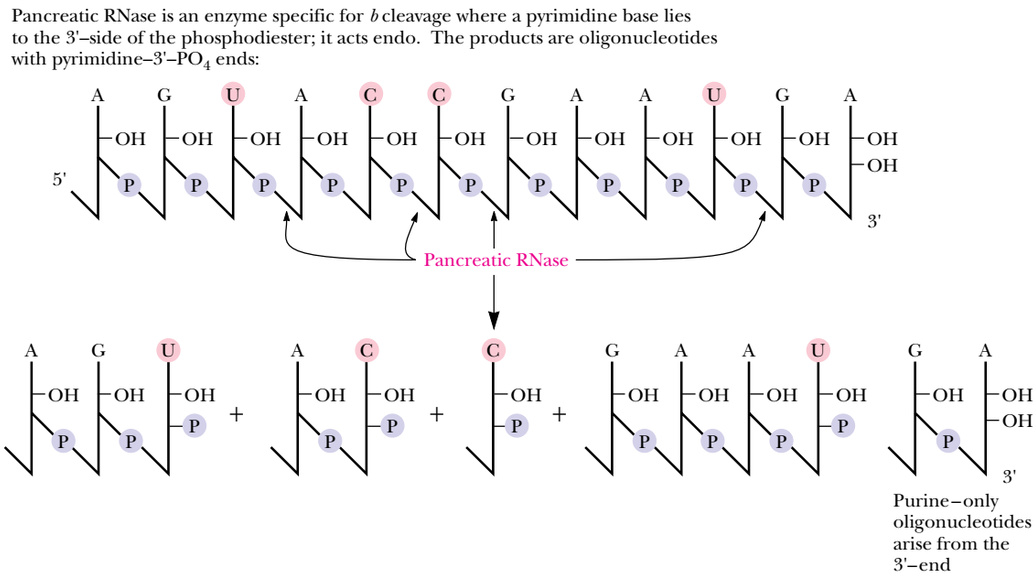


FIGURE 11.32 • An example of nuclease specificity: The specificity of RNA hydrolysis by bovine pancreatic RNase. This RNase cleaves *b* at 3′-pyrimidines, yielding oligonucleotides with pyrimidine 3′-PO_4 ends.

certain bases in a polynucleotide (Figure 11.32), or, as we shall see for *restriction endonucleases,* some nucleases will act only at a particular nucleotide sequence four to eight nucleotides in length. Table 11.4 lists the various permutations in specificity displayed by these nucleases and gives prominent examples of each. To the molecular biologist, nucleases are the surgical tools for the dissection and manipulation of nucleic acids in the laboratory.

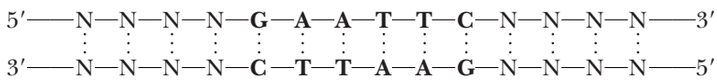
Exonucleases degrade nucleic acids by sequentially removing nucleotides from their ends. Two in common use are *snake venom phosphodiesterase* and *bovine spleen phosphodiesterase* (Figure 11.31). Because they act on either DNA or RNA, they are referred to by the generic name *phosphodiesterase.* These two enzymes have complementary specificities. Snake venom phosphodiesterase acts by *a* cleavage and starts at the free 3′-OH end of a polynucleotide chain, liberating nucleoside 5′-monophosphates. In contrast, the bovine spleen enzyme starts at the 5′-end of a nucleic acid, cleaving *b* and releasing 3′-NMPs.

Restriction Enzymes

Restriction endonucleases are enzymes, isolated chiefly from bacteria, that have the ability to cleave double-stranded DNA. The term *restriction* comes from the capacity of prokaryotes to defend against or "restrict" the possibility of takeover by foreign DNA that might gain entry into their cells. Prokaryotes degrade foreign DNA by using their unique restriction enzymes to chop it into relatively large but noninfective fragments. Restriction enzymes are classified into three types, I, II, or III. Types I and III require ATP to hydrolyze DNA and can also catalyze chemical modification of DNA through addition of methyl groups to specific bases. Type I restriction endonucleases cleave DNA randomly, while type III recognize specific nucleotide sequences within dsDNA and cut the DNA at or near these sites.

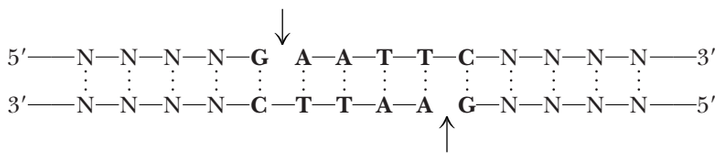
Type II Restriction Endonucleases

Type II restriction enzymes have received widespread application in the cloning and sequencing of DNA molecules. Their hydrolytic activity is not ATP-dependent, and they do not modify DNA by methylation or other means. Most importantly, they cut DNA within or near particular nucleotide sequences that they specifically recognize. These recognition sequences are typically four or six nucleotides in length and have a twofold axis of symmetry. For example, *E. coli* has a restriction enzyme, *Eco*RI, that recognizes the hexanucleotide sequence GAATTC:

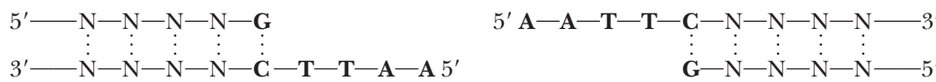


Note the twofold symmetry: the sequence read 5′→3′ is the same in both strands.

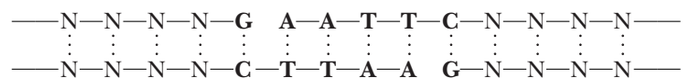
When *Eco*RI encounters this sequence in dsDNA, it causes a staggered, double-stranded break by hydrolyzing each chain between the G and A residues:



Staggered cleavage results in fragments with protruding single-stranded 5′-ends:



Because the protruding termini of *Eco*RI fragments have complementary base sequences, they can form base pairs with one another.



Therefore, DNA restriction fragments having such "sticky" ends can be joined together to create new combinations of DNA sequence. If the fragments are derived from DNA molecules of different origin, novel recombinant forms of DNA are created.

*Eco*RI leaves staggered 5′-termini. Other restriction enzymes, such as *Pst*I, which recognizes the sequence 5′-CTGCAG-3′ and cleaves between A and G, produce cohesive staggered 3′-ends. Still others, such as *Bal*I, act at the center of the twofold symmetry axis of their recognition site and generate blunt ends that are noncohesive. *Bal*I recognizes 5′-TGGCCA-3′ and cuts between G and C.

Table 11.5

Restriction Endonucleases

About 1000 restriction enzymes have been characterized. They are named by italicized three-letter codes, the first a capital letter denoting the genus of the organism of origin, while the next two letters are an abbreviation of the particular species. Because prokaryotes often contain more than one restriction enzyme, the various representatives are assigned letter and number codes as they are identified. Thus, *Eco*RI is the initial restriction endonuclease isolated from *Escherichia coli*, strain R. With one exception (*Nci*I), all known type II restriction endonucleases generate fragments with 5′-PO_4 and 3′-OH ends.

Enzyme	Common Isoschizomers	Recognition Sequence	Compatible Cohesive Ends
*Alu*I		AG↓CT	Blunt
*Apy*I	*Atu*I, *Eco*RII	CC↓$\binom{A}{T}$GG	
*Asu*II		TT↓CGAA	*Cla*I, *Hpa*II, *Taq*I
*Ava*I		G↓PyCGPuG	*Sal*I, *Xho*I, *Xma*I
*Avr*II		C↓CTAGG	
*Bal*I		TGG↓CCA	Blunt
*Bam*HI		G↓GATCC	*Bcl*I, *Bgl*II, *Mbo*I, *Sau*3A, *Xho*II
*Bcl*I		T↓GATCA	*Bam*HI, *Bgl*II, *Mbo*I, *Sau*3A, *Xho*II
*Bgl*II		A↓GATCT	*Bam*HI, *Bcl*I, *Mbo*I, *Sau*3A, *Xho*II
*Bst*EII		G↓GTNACC	
*Bst*XI		CCANNNNN↓NTGG	
*Cla*I		AT↓CGAT	*Acc*I, *Acy*I, *Asy*II, *Hpa*II, *Taq*I
*Dde*I		C↓TNAG	
*Eco*RI		G↓AATTC	
*Eco*RII	*Atu*I, *Apy*I	↓CC $\binom{A}{T}$GG	
*Fnu*DII	*Tha*I	CG↓CG	Blunt
*Hae*I		$\binom{A}{T}$GG↓CC$\binom{T}{A}$	Blunt
*Hae*II		PuGCGC↓Py	
*Hae*III		GG↓CC	Blunt
*Hin*cII		GTPy↓PuAC	Blunt
*Hin*dIII		A↓AGCTT	
*Hpa*I		GTT↓AAC	Blunt
*Hpa*II		C↓CGG	*Acc*I, *Acy*I, *Asu*II, *Cla*I, *Taq*I
*Kpn*I		GGTAC↓C	*Bam*HI, *Bcl*I, *Bgl*II, *Xho*II
*Mbo*I	*Sau*3A	↓GATC	
*Msp*I		C↓CGG	
*Mst*I		TGC↓GCA	Blunt
*Not*I		GC↓GGCCGC	
*Pst*I		CTGCA↓G	
*Sac*I	*Sst*I	GAGCT↓C	
*Sal*I		G↓TCGAC	*Ava*I, *Xho*I
*Sau*3A		↓GATC	*Bam*HI, *Bcl*I, *Bgl*II, *Mbo*I, *Xho*II
*Sfi*I		GGCCNNNN↓NGGCC	
*Sma*I	*Xma*I	CCC↓GGG	Blunt
*Sph*I		GCATG↓C	
*Sst*I	*Sac*I	GAGCT↓C	
*Taq*I		T↓CGA	*Acc*I, *Acy*I, *Asu*II, *Cla*I, *Hpa*II
*Xba*I		T↓CTAGA	
*Xho*I		C↓TCGAG	*Ava*I, *Sal*I
*Xho*II		$\binom{A}{G}$↓GATC$\binom{T}{C}$	*Bam*HI, *Bcl*I, *Bgl*II, *Mbo*I, *Sau*3A
*Xma*I	*Sma*I	C↓CCGGG	*Ava*I

Table 11.5 lists many of the commonly used restriction endonucleases and their recognition sites. Because these sites all have twofold symmetry, only the sequence on one strand needs to be designated.

ISOSCHIZOMERS. Different restriction enzymes sometimes recognize and cleave within identical target sequences. For example, *Mbo*I and *Sau*3A recognize the same tetranucleotide run: 5′-GATC-3′. Both cleave the DNA strands at the same position, namely, on the 5′-side of the G. Such enzymes are called **isoschizomers,** meaning that they cut at the same site. The enzyme *Bam*HI is an isoschizomer of *Mbo*I and *Sau*3A except that it has greater specificity because it acts only at hexanucleotide sequences reading GGATCC. *Bam*HI cuts between the two G's, leaving cohesive 5′-ends that can match up with *Mbo*I or *Sau*3A fragments.

schizo • from the Greek *schizein,* to split

RESTRICTION FRAGMENT SIZE. Assuming random distribution and equimolar proportions for the four nucleotides in DNA, a particular tetranucleotide sequence should occur every 4^4 nucleotides, or every 256 bases. Therefore, the fragments generated by a restriction enzyme that acts at a four-nucleotide sequence should average about 250 bp in length. "Six-cutters," enzymes such as *Eco*RI or *Bam*HI, will find their unique hexanucleotide sequences on the average once in every 4096 (4^6) bp of length. Because the genetic code is a triplet code with three bases of DNA specifying one amino acid in a polypeptide sequence, and because polypeptides typically contain at most 1000 amino acid residues, the fragments generated by six-cutters are approximately the size of prokaryotic genes. This property makes these enzymes useful in the construction and cloning of genetically useful recombinant DNA molecules. For the isolation of even larger nucleotide sequences, such as those of genes encoding large polypeptides (or those of eukaryotic genes that are disrupted by large introns), partial or limited digestion of DNA by restriction enzymes can be employed. However, restriction endonucleases that cut only at specific nucleotide sequences 8 or even 13 nucleotides in length are also available, such as *Not*I and *Sfi*I.

Restriction Mapping

The application of these sequence-specific nucleases to problems in molecular biology is considered in detail in Chapter 13, but one prominent application is described here. Because restriction endonucleases cut dsDNA at unique sites to generate large fragments, they provide a means for mapping DNA molecules that are many kilobase pairs in length. Restriction digestion of a DNA molecule is in many ways analogous to proteolytic digestion of a protein by an enzyme such as trypsin (see Chapter 5): the restriction endonuclease acts only at its specific sites so that a discrete set of nucleic acid fragments is generated. This action is analogous to trypsin cleavage only at Arg and Lys residues to yield a particular set of tryptic peptides from a given protein. The restriction fragments represent a unique collection of different-sized DNA pieces. Fortunately, this complex mixture can be resolved by *electrophoresis* (see the Appendix to Chapter 5). Electrophoresis of DNA molecules on gels of restricted pore size (as formed in agarose or polyacrylamide media) separates them according to size, the largest being retarded in their migration through the gel pores while the smallest move relatively unhindered. Figure 11.33 shows a hypothetical electrophoretogram obtained for a DNA molecule treated with two different restriction nucleases, alone and in combination. Just as cleavage of a protein with different proteases to generate overlapping fragments allows an ordering of the peptides, restriction fragments can be ordered or "mapped" according to their sizes, as deduced from the patterns depicted in Figure 11.33.

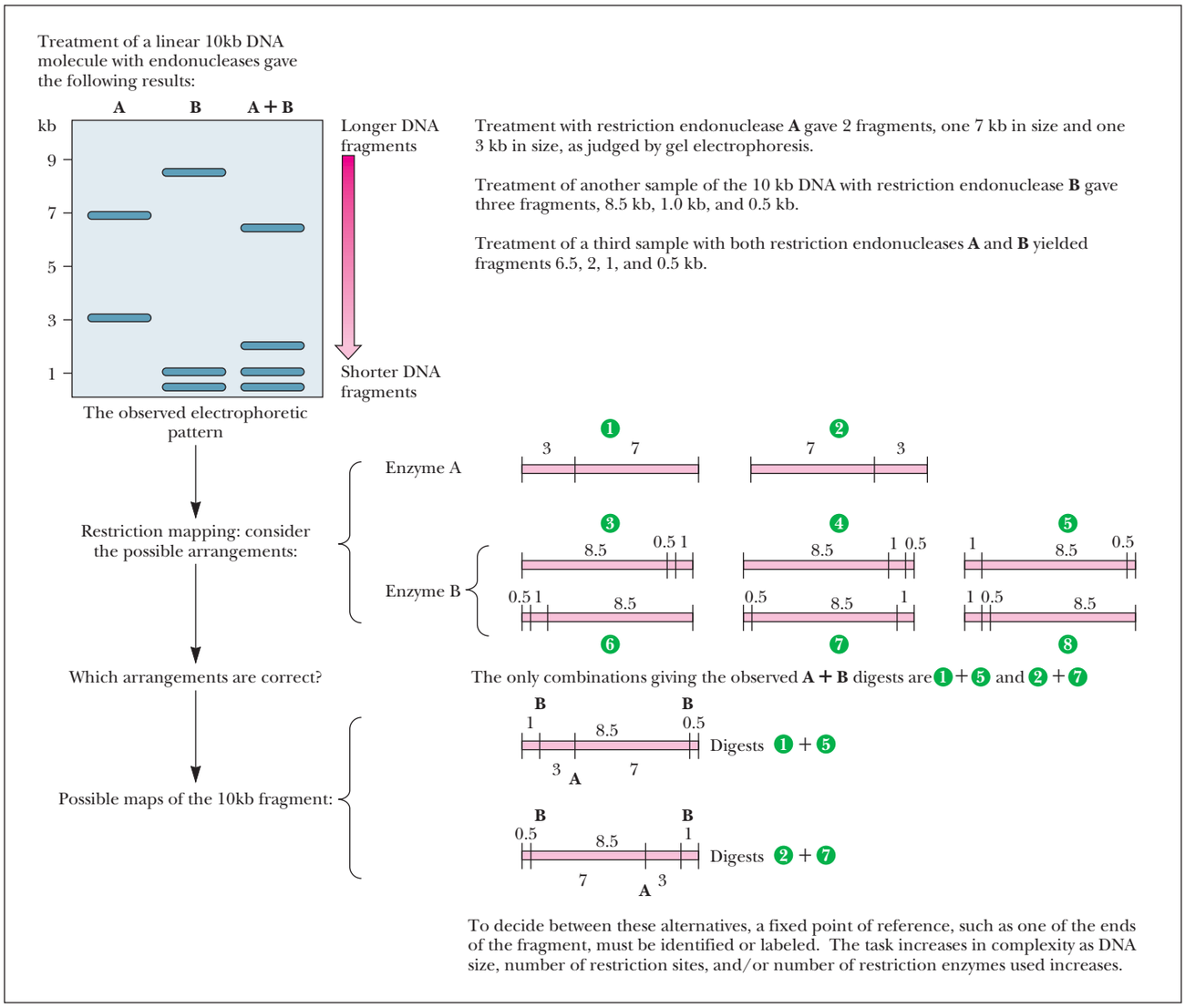


FIGURE 11.33 • Restriction mapping of a DNA molecule as determined by an analysis of the electrophoretic pattern obtained for different restriction endonuclease digests. (Keep in mind that a dsDNA molecule has a unique nucleotide sequence and therefore a definite polarity; thus, fragments from one end are distinctly different from fragments derived from the other end.)

PROBLEMS

1. Draw the chemical structure of pACG.

2. Chargaff's results (Table 11.3) yielded a molar ratio of 1.56 for A to G in human DNA, 1.75 for T to C, 1.00 for A to T, and 1.00 for G to C. Given these values, what are the mole fractions of A, C, G, and T in human DNA?

3. Adhering to the convention of writing nucleotide sequences in the 5′ → 3′ direction, what is the nucleotide sequence of the DNA strand that is complementary to d-ATCGCAACTGTCACTA?

4. Messenger RNAs are synthesized by RNA polymerases that read along a DNA template strand in the 3′ → 5′ direction, polymerizing ribonucleotides in the 5′ → 3′ direction (see Figure 11.24). Give the nucleotide sequence (5′ → 3′) of the DNA template strand from which the following mRNA segment was transcribed: 5′-UAGUGACAGUUGCGAU-3′.

5. The DNA strand that is complementary to the template strand copied by RNA polymerase during transcription has a nucleotide

sequence identical to that of the RNA being synthesized (except T residues are found in the DNA strand at sites where U residues occur in the RNA). An RNA transcribed from this nontemplate DNA strand would be complementary to the mRNA synthesized by RNA polymerase. Such an RNA is called antisense RNA. A promising strategy to thwart the deleterious effects of genes activated in disease states (such as cancer) is to generate antisense RNAs in affected cells. These antisense RNAs would form double-stranded hybrids with mRNAs transcribed from the activated genes and prevent their translation into protein. Suppose transcription of a cancer-activated gene yielded an mRNA whose sequence included the segment 5′-UACGGUCUAAGCUGA. What is the corresponding nucleotide sequence ($5' \rightarrow 3'$) of the template strand in a DNA duplex that might be introduced into these cells so that an antisense RNA could be transcribed from it?

6. A 10-kb DNA fragment digested with restriction endonuclease *Eco*RI yielded fragments 4 kb and 6 kb in size. When digested with *Bam*HI, fragments 1, 3.5, and 5.5 kb were generated. Concomitant digestion with both *Eco*RI and *Bam*HI yielded fragments 0.5, 1, 3, and 5.5 kb in size. Give a possible restriction map for the original fragment.

Further Reading

Adams, R. L. P., Knowler, J. T., and Leader, D. P., 1992. *The Biochemistry of the Nucleic Acids,* 11th ed. New York: Chapman and Hall (Methuen and Co., distrib.).

Gray, M. W., and Cedergren, R., eds., 1993. The New Age of RNA. *The FASEB Journal* **7:**4–239. A collection of articles emphasizing the new appreciation for RNA in protein synthesis, in evolution, and as a catalyst.

Judson, H. F., 1979. *The Eighth Day of Creation*. New York: Simon and Schuster.

Maniatis, T., Frisch, E. F., and Sambrook, J., 1989. *Molecular Cloning: A Laboratory Manual,* 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., and Weiner, A. M., 1987. *The Molecular Biology of the Gene,* Vol. I, *General Principles,* 4th ed. Menlo Park, CA: Benjamin/Cummings.

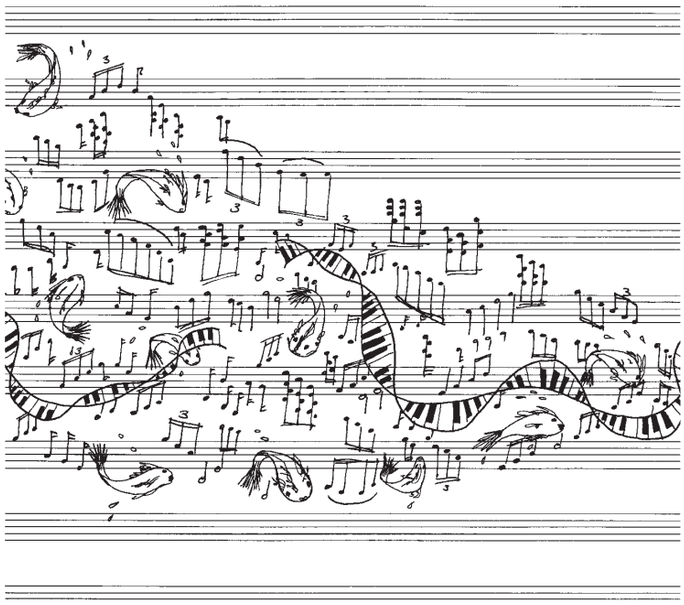
The Structure of DNA: "A melody for the eye of the intellect, with not a note wasted."

Horace Freeland Judson, *The Eighth Day of Creation*

Chapter 12

Structure of Nucleic Acids

Outline



"Scherzo in D&A" (detail) by David E. Rodale (1955–1985)

Chapter 11 presented the structure and chemistry of nucleotides and how these units are joined via phosphodiester bonds to form nucleic acids, the biological polymers for information storage and transmission. In this chapter, we investigate biochemical methods that reveal this information by determining the sequential order of nucleotides in a polynucleotide, the so-called **primary structure** of nucleic acids. Then, we consider the higher orders of structure in the nucleic acids, the secondary and tertiary levels. Although the focus here is primarily on the structural and chemical properties of these macromolecules, it is fruitful to keep in mind the biological roles of these remarkable substances. The sequence of nucleotides in nucleic acids is the embodiment of genetic information (see Part IV). We can anticipate that the cellular mechanisms for accessing this information, as well as reproducing it with high fidelity, will be illuminated by knowledge of the chemical and structural qualities of these polymers.

12.1 • The Primary Structure of Nucleic Acids

As recently as 1975, determining the primary structure of nucleic acids (the nucleotide sequence) was a more formidable problem than amino acid sequencing of proteins, simply because nucleic acids contain only four unique monomeric units whereas proteins have twenty. With only four, there are *apparently* fewer specific sites for selective cleavage, distinctive sequences are more difficult to recognize, and the likelihood of ambiguity is greater. The much greater number of monomeric units in most polynucleotides as compared to polypeptides is a further difficulty. Two important breakthroughs reversed this situation so that now sequencing nucleic acids is substantially easier than sequencing polypeptides. One was the discovery of *restriction endonucleases* that cleave DNA at specific oligonucleotide sites, generating unique fragments of manageable size (see Chapter 11). The second is the power of *polyacrylamide gel electrophoresis* separation methods to resolve nucleic acid fragments that differ from one another in length by just one nucleotide.

Sequencing Nucleic Acids

Two basic protocols for nucleic acid sequencing are in widespread use: the **chain termination** or **dideoxy method** of F. Sanger and the **base-specific chemical cleavage method** developed by A. M. Maxam and W. Gilbert. Because both methods are carried out on nanogram amounts of DNA, very sensitive analytical techniques are used to detect the DNA chains following electrophoretic separation on polyacrylamide gels. Typically, the DNA molecules are labeled with radioactive ^{32}P,[1] and following electrophoresis, the pattern of their separation is visualized by **autoradiography.** A piece of X-ray film is placed over the gel and the radioactive disintegrations emanating from ^{32}P decay create a pattern on the film that is an accurate image of the resolved oligonucleotides. Recently, sensitive biochemical and chemiluminescent methods have begun to supersede the use of radioisotopes as tracers in these experiments.

Chain Termination or Dideoxy Method

To appreciate the rationale of the chain termination or dideoxy method, we first must briefly examine the biochemistry of DNA replication. DNA is a double-helical molecule. In the course of its replication, the sequence of nucleotides in one strand is copied in a complementary fashion to form a new second strand by the enzyme **DNA polymerase.** Each original strand of the double helix serves as **template** for the biosynthesis that yields two daughter DNA duplexes from the parental double helix (Figure 12.1). DNA polymerase carries out this reaction *in vitro* in the presence of the four deoxynucleotide monomers and copies single-stranded DNA, provided a double-stranded region of DNA is artificially generated by adding a **primer.** This primer is merely an oligonucleotide capable of forming a short stretch of dsDNA by base pairing with the ssDNA (Figure 12.2). The primer must have a free 3′-OH end from which the new polynucleotide chain can grow as the first residue is added in the initial step of the polymerization process. DNA polymerases synthesize new strands by adding successive nucleotides in the 5′ → 3′ direction.

[1]Because its longer half-life and lower energy make it more convenient to handle, ^{35}S is replacing ^{32}P as the radioactive tracer of choice in sequencing by the Sanger method. ^{35}S-α-labeled deoxynucleotide analogs provide the source for incorporating radioactivity into DNA.

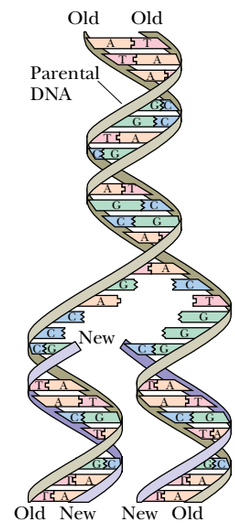


FIGURE 12.1 • DNA replication yields two daughter DNA duplexes identical to the parental DNA molecule. Each original strand of the double helix serves as a template, and the sequence of nucleotides in each of these strands is copied to form a new complementary strand by the enzyme DNA polymerase. By this process, biosynthesis yields two daughter DNA duplexes from the parental double helix.

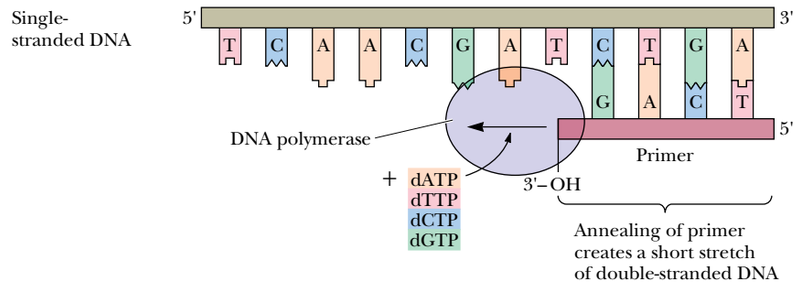


FIGURE 12.2 • DNA polymerase copies ssDNA *in vitro* in the presence of the four deoxynucleotide monomers, provided a double-stranded region of DNA is artificially generated by adding a primer, an oligonucleotide capable of forming a short stretch of dsDNA by base pairing with the ssDNA. The primer must have a free 3′-OH end from which the new polynucleotide chain can grow as the first residue is added in the initial step of the polymerization process.

Chain Termination Protocol

In the chain termination method of DNA sequencing, a DNA fragment of unknown sequence serves as template in a polymerization reaction using some type of DNA polymerase, usually *Sequenase 2*®, a genetically engineered version of bacteriophage T7 DNA polymerase that lacks all traces of exonuclease activity that might otherwise degrade the DNA. The primer requirement is met by an appropriate oligonucleotide (this method is also known as the **primed synthesis method** for this reason). Four parallel reactions are run; all four contain the four deoxynucleoside triphosphates dATP, dGTP, dCTP, and dTTP, which are the substrates for DNA polymerase (Figure 12.3 on the facing page). In each of the four reactions, a different 2′,3′-**di**deoxynucleotide is included, and it is these dideoxynucleotides that give the method its name.

Because dideoxynucleotides lack 3′-OH groups, these nucleotides cannot serve as acceptors for 5′-nucleotide addition in the polymerization reaction, and thus the chain is terminated where they become incorporated. The concentrations of the four deoxynucleotides and the single dideoxynucleotide in each reaction mixture are adjusted so that the dideoxynucleotide is incorporated infrequently. Therefore, base-specific premature chain termination is only a random, occasional event, and a population of new strands of varying length is synthesized. Four reactions are run, one for each dideoxynucleotide, so that termination, although random, can occur everywhere in the sequence. In each mixture, each newly synthesized strand has a dideoxynucleotide at its 3′-end, and its presence at that position demonstrates that a base of that particular kind was specified by the template. A radioactively labeled dNTP is included in each reaction mixture to provide a tracer for the products of the polymerization process.

Reading Dideoxy Sequencing Gels

The sequencing products are visualized by autoradiography (or similar means) following their separation according to size by polyacrylamide gel electrophoresis (Figure 12.3). Because the smallest fragments migrate fastest upon electrophoresis and because fragments differing by only single nucleotides in length are readily resolved, the autoradiogram of the gel can be read from bottom to top, noting which lane has the next largest band at each step. Thus, the gel in Figure 12.3 is read AGCGTAGC (5′ → 3′). Because of the way DNA polymerase acts, this observed sequence is complementary to the corresponding unknown template sequence. Knowing this, the template sequence now can be written GCTACGCT (5′ → 3′).

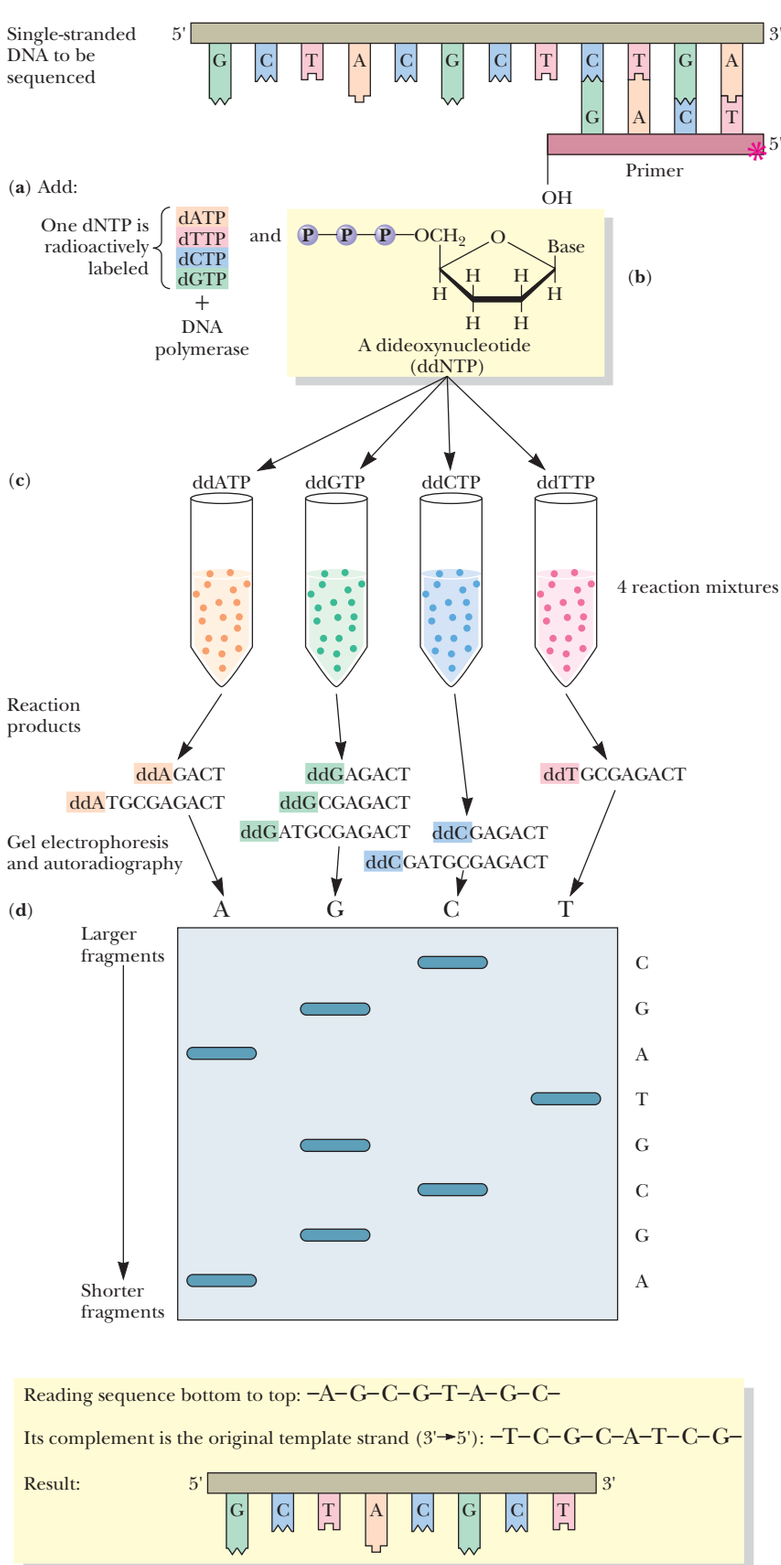


FIGURE 12.3 • The chain termination or dideoxy method of DNA sequencing. (a) DNA polymerase reaction. (b) Structure of dideoxynucleotide. (c) Four reaction mixtures with nucleoside triphosphates plus one dideoxynucleoside triphosphate. (d) Electrophoretogram. Note that the nucleotide sequence as read from the bottom to the top of the gel is the order of nucleotide addition carried out by DNA polymerase.

Reading sequence bottom to top: –A–G–C–G–T–A–G–C–

Its complement is the original template strand (3'→5'): –T–C–G–C–A–T–C–G–

Result:

Base-Specific Chemical Cleavage Method

The base-specific chemical cleavage (or Maxam–Gilbert) method starts with a single-stranded DNA that is labeled at one end with radioactive ^{32}P. (Double-stranded DNA can be used if only one strand is labeled at only one of its ends.) The DNA strand is then randomly cleaved by reactions that specifically fragment its sugar–phosphate backbone only where certain bases have been chemically removed. There is no unique reaction for each of the four bases. However,

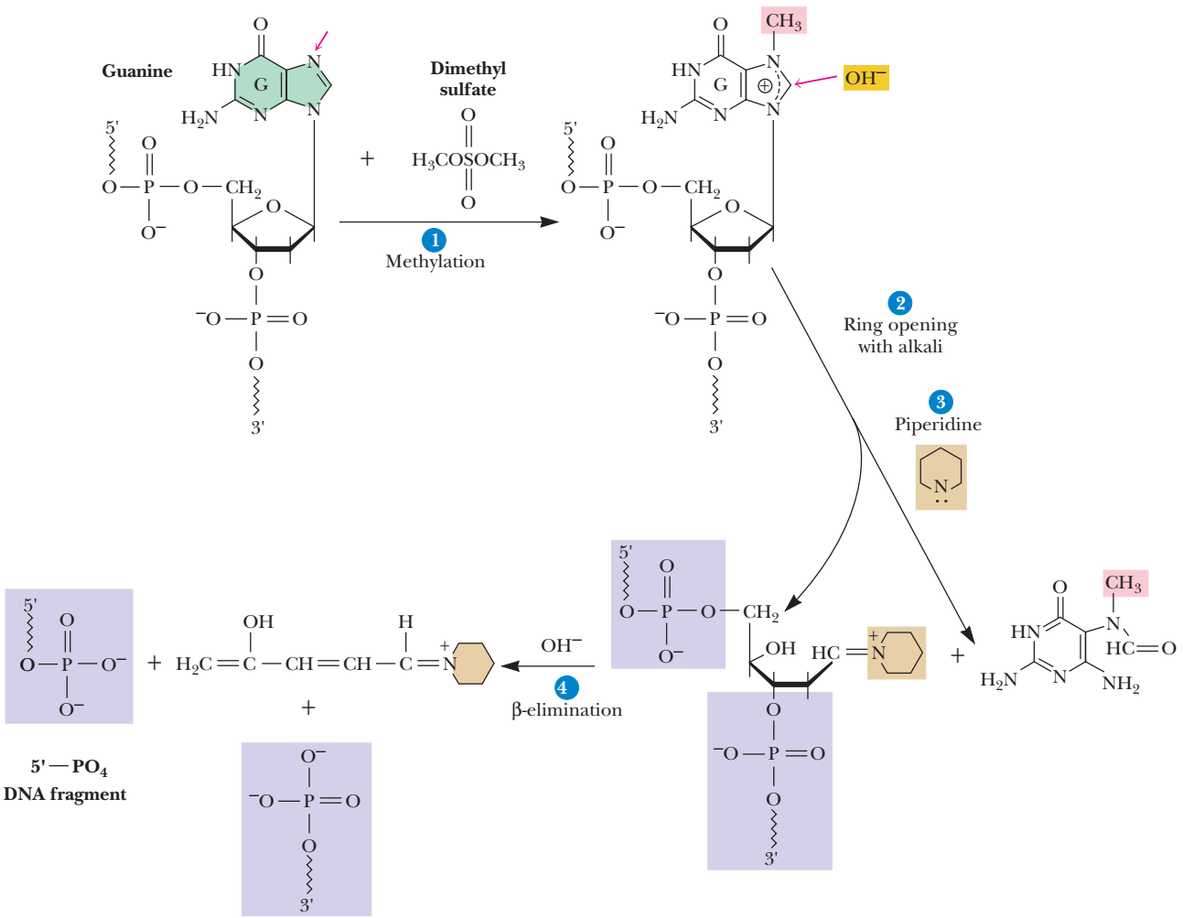


FIGURE 12.4 • Maxam–Gilbert sequencing of DNA: cleavage at purines uses dimethyl sulfate, followed by strand scission with piperidine.

Cleavage at G using dimethyl sulfate, followed by strand scission with piperidine: Under alkaline conditions, **dimethyl sulfate** reacts with guanine to methylate it at the 7-position **(1).** This substitution leads to instability of the N-9 glycosidic bond, so that in the presence of OH^- and the secondary amine **piperidine (2),** the purine ring is degraded and released. A β-elimination reaction facilitated by piperidine **(3)** then causes the excision of the naked deoxyribose moiety from the sugar–phosphate backbone, with consequent scission of the DNA strand to yield 5′- and 3′-fragments.

Cleavage at A or G: If the DNA is first treated with acid, dimethyl sulfate methylates adenine at the 3-position as well as guanine at the 7-position (not shown). Subsequent reaction with OH^- and piperidine triggers degradation and displacement of the methylated A or G purine base and strand scission, essentially as indicated here for reaction of dimethyl sulfate with guanine.

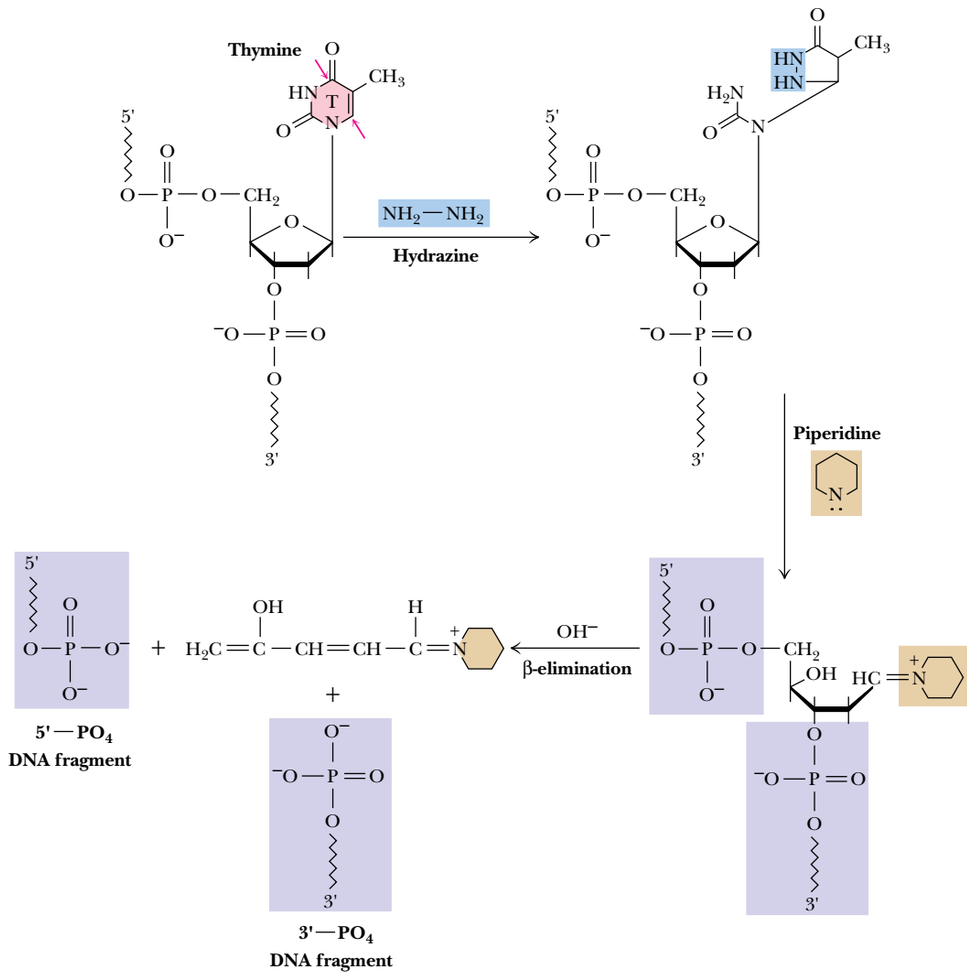
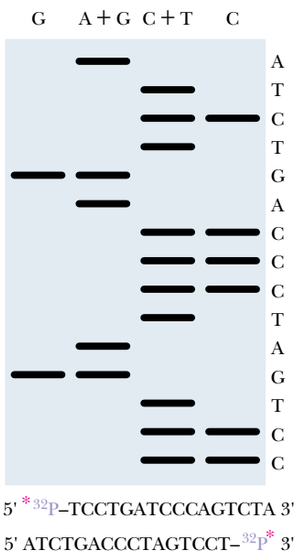


FIGURE 12.5 • Maxam–Gilbert sequencing of DNA: hydrolysis of pyrimidine rings by hydrazine. Hydrazine (H_2N-NH_2) attacks across the C-4 and C-6 atoms of pyrimidines to open the ring. This degradation subsequently leads to modification of the deoxyribose, rendering it susceptible to β-elimination by piperidine in the presence of hydroxide ion. Shown here is the excision of a T residue. As in Figure 12.4, 5′- and 3′-fragments are produced. The presence of high salt concentrations protects T (but not C) from reaction with hydrazine. In the presence of 2 *M* NaCl, the reaction shown here occurs only at C residues.

there is a reaction specific to G only and a purine-specific reaction that removes A or G (Figure 12.4). Thus, the difference in these two reactions is a specific indication of where A occurs. Similarly, there is a cleavage reaction specific for the pyrimidines (C+T) (Figure 12.5), which, if run in the presence of 1 or 2 *M* NaCl, works only with C. Differences in these two are thus attributable to the presence of T in the nucleotide sequence.

Note that the key to Maxam–Gilbert sequencing is to modify a base chemically so that it is removed from its sugar. Then piperidine excises the sugar from its 5′- and 3′-links in a β-elimination reaction. The conditions of chemical cleavage described in Figures 12.4 and 12.5 are generally adjusted so that,



◀ **FIGURE 12.6** • Autoradiogram of a hypothetical electrophoretic pattern obtained for four reaction mixtures, performed as described in Figures 12.4 and 12.5 and run in the four lanes G, A+G, C+T, and C, respectively. Reading this pattern from the bottom up yields the sequence CCTGATCCCAGTCTA. The correct 5′ → 3′ order is determined by knowing which end of the ssDNA was ^{32}P-labeled. If the 5′-end was ^{32}P-labeled, only the 5′-fragments will be evident on the autoradiogram; the 3′-ends will be invisible. Similarly, if the 3′-end was originally labeled, only the 3′-fragments light up the autoradiogram. Assuming that the 5′-end was labeled, the sequence would be CCTGATCCCAGTCTA. If it were the 3′-end, the sequence read in the 5′ → 3′ convention would be ATCTGACCC-TAGTCC. An interesting feature of the Maxam–Gilbert sequencing procedure is that the base that is "read" in the ladder is actually not present in the oligonucleotide that identifies it. Thus, an unidentified base bears the label at the end of the smallest fragment; this unidentified base is the one that preceded the first identified base. For example, an oligonucleotide of either

^{32}P-5′-**(A,C,G,T)CCTGATCCCAGTCTA**-3′

or

5′-**ATCTGACCCTAGTCC(A,C,G,T)**-3′-^{32}P

would yield the same pattern in the autoradiogram. (Indication here of T as the end-labeled nucleotide is arbitrary.)

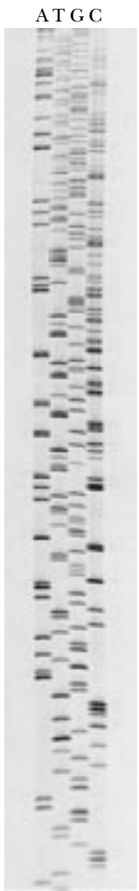
on average, only a single scission occurs per DNA molecule. However, because a very large number of DNA molecules exist in each reaction mixture, the products are a random collection of different-sized fragments wherein the occurrence of any base is represented by its unique pair of 5′- and 3′-cleavage products. These products form a complete set, the members of which differ in length by only one nucleotide, and they can be resolved by gel electrophoresis into a "ladder," which can be visualized by autoradiography of the gel if the DNA fragments are radioactively labeled (Figure 12.6).

In principle, the Maxam–Gilbert method can provide the total sequence of a dsDNA molecule just by determining the purine positions on one strand and then the purines on the complementary strand. Complementary base-pairing rules then reveal the pyrimidines along each strand, T complementary to where A is, C complementary to where G occurs. (The analogous approach of locating the pyrimidines on each strand would also provide sufficient information to write the total sequence.)

With current technology, it is possible to read the order of as many as 400 bases from the autoradiogram of a sequencing gel (Figure 12.7). The actual chemical or enzymatic reactions, electrophoresis, and autoradiography are now routine, and a skilled technician can sequence about 1 kbp per week using these manual techniques. The major effort in DNA sequencing is in the isolation and preparation of fragments of interest, such as cloned genes.

Automated DNA Sequencing

In recent years, automated DNA sequencing machines capable of identifying about 10^4 bases per day have become commercially available. One clever innovation has been the use of fluorescent dyes of different colors to uniquely label the primer DNA introduced into the four sequencing reactions; for example, red for the A reaction, blue for T, green for G, and yellow for C. Then, all four reaction mixtures can be combined and run together on one electrophoretic



◀ **FIGURE 12.7** • A photograph of the autoradiogram from an actual sequencing gel. A portion of the DNA sequence of *nit-6,* the *Neurospora* gene encoding the enzyme nitrite reductase. *(James D. Colandene, University of Virginia)*

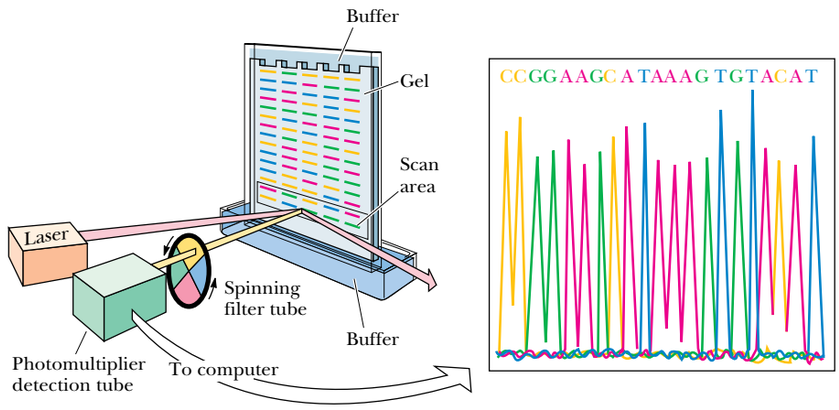


FIGURE 12.8 • Schematic diagram of the methodology used in fluorescent labeling and automated sequencing of DNA. Four reactions are set up, one for each base, and the primer in each is end-labeled with one of four different fluorescent dyes; the dyes serve to color-code the base-specific sequencing protocol (a unique dye is used in each dideoxynucleotide reaction). The four reaction mixtures are then combined and run in one lane. Thus, each lane in the gel represents a different sequencing experiment. As the differently sized fragments pass down the gel, a laser beam excites the dye in the scan area. The emitted energy passes through a rotating color filter and is detected by a fluorometer. The color of the emitted light identifies the final base in the fragment. *(Applied Biosystems, Inc., Foster City, CA)*

gel slab. As the oligonucleotides are separated and pass to the bottom of the gel, each is illuminated by a low-power argon laser beam that causes the dye attached to the primer to fluoresce. The color of the fluorescence is detected automatically, revealing the identity of the primer, and hence the base, immediately (Figure 12.8). The development of such automation has opened the possibility for sequencing the entire human genome, some 2.9 billion bp. Even so, if 100 automated machines operating at peak efficiency were dedicated to the task, it would still take at least 8 years to complete!

12.2 • The ABZs of DNA Secondary Structure

Double-stranded DNA molecules assume one of three secondary structures, termed A, B, and Z. Fundamentally, double-stranded DNA is a regular two-chain structure with hydrogen bonds formed between opposing bases on the two chains (see Chapter 11). Such H-bonding is possible only when the two chains are antiparallel. The polar sugar–phosphate backbones of the two chains are on the outside. The bases are stacked on the inside of the structure; these heterocyclic bases, as a consequence of their π-electron clouds, are hydrophobic on their flat sides. One purely hypothetical conformational possibility for a two-stranded arrangement would be a ladderlike structure (Figure 12.9) in which the base pairs are fixed at 0.6 nm apart because this is the distance between adjacent sugars in the DNA backbone. Because H_2O molecules would be accessible to the spaces between the hydrophobic surfaces of the bases, this conformation is energetically unfavorable. This ladderlike structure converts to a helix when given a simple right-handed twist. Helical twisting brings the base-pair rungs of the ladder closer together, stacking them 0.34 nm apart, without affecting the sugar–sugar distance of 0.6 nm. Because this helix repeats itself approximately every 10 bp, its **pitch** is 3.4 nm. This is the major conformation of DNA in solution and it is called **B-DNA.**

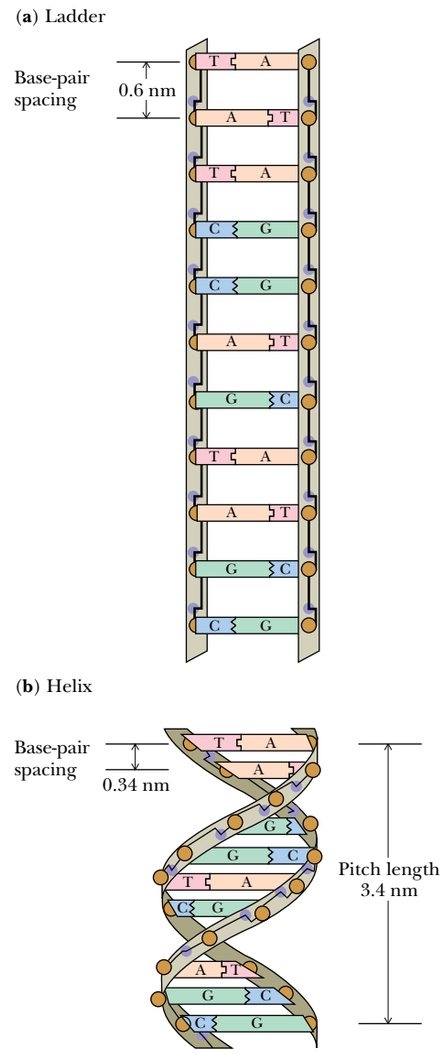


FIGURE 12.9 • (a) Double-stranded DNA as an imaginary ladderlike structure. (b) A simple right-handed twist converts the ladder to a helix.

Structural Equivalence of Watson–Crick Base Pairs

As indicated in Chapter 11, the base pairing in DNA is very specific: the purine adenine pairs with the pyrimidine thymine; the purine guanine pairs with the pyrimidine cytosine. Further, the A:T pair and G:C pair have virtually identical dimensions (Figure 12.10). Watson and Crick realized that units of such similarity could serve as spatially invariant substructures to build a polymer whose exterior dimensions would be uniform along its length, regardless of the sequence of bases.

The DNA Double Helix Is a Stable Structure

Several factors account for the stability of the double-helical structure of DNA. First, both internal and external hydrogen bonds stabilize the double helix. The two strands of DNA are held together by H-bonds that form between the complementary purines and pyrimidines, two in an A:T pair and three in a G:C pair (Figure 12.10), while polar atoms in the sugar–phosphate backbone form external H bonds with surrounding water molecules. Second, the negatively charged phosphate groups are all situated on the exterior surface of the helix in such a way that they have minimal effect on one another and are free to interact electrostatically with cations in solution such as Mg^{2+}. Third, the core of the helix consists of the base pairs, which, in addition to being H-bonded, stack together through hydrophobic interactions and van der Waals forces that contribute significantly to the overall stabilizing energy.

A stereochemical consequence of the way A:T and G:C base pairs form is that the sugars of the respective nucleotides have opposite orientations, and thus the sugar–phosphate backbones of the two chains run in opposite or

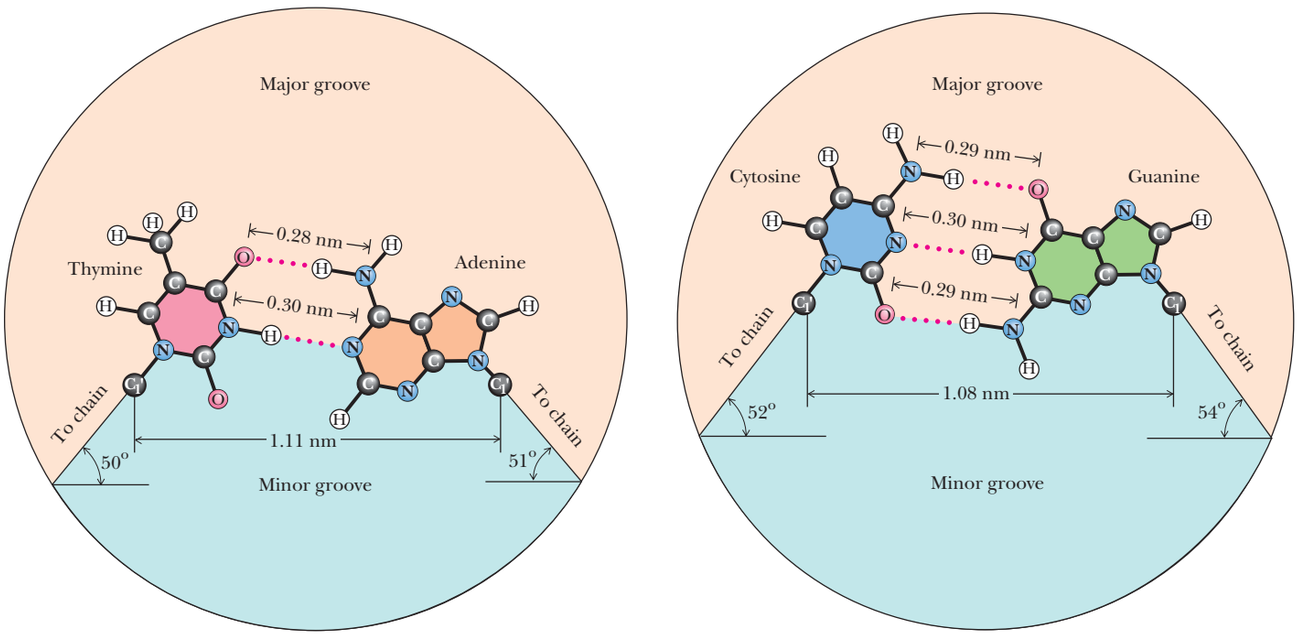


FIGURE 12.10 • Watson–Crick A:T and G:C base pairs. All H bonds in both base pairs are straight, with each H atom pointing directly at its acceptor N or O atom. Linear H bonds are the strongest. The mandatory binding of larger purines with smaller pyrimidines leads to base pairs that have virtually identical dimensions, allowing the two sugar–phosphate backbones to adopt identical helical conformations.

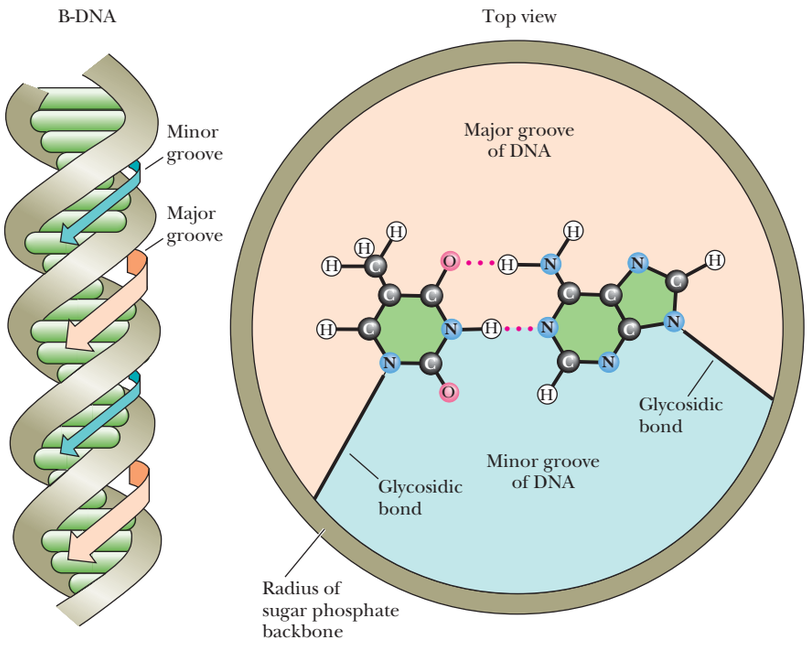
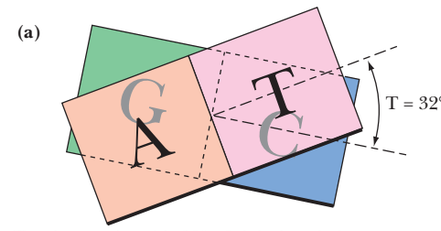


FIGURE 12.11 • The bases in a base pair are not directly across the helix axis from one another along some diameter but rather are slightly displaced. This displacement, and the relative orientation of the glycosidic bonds linking the bases to the sugar–phosphate backbone, leads to differently sized grooves in the cylindrical column created by the double helix, the major groove and the minor groove, each coursing along its length.

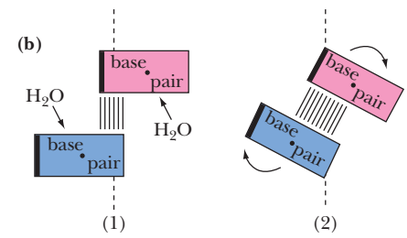
"antiparallel" directions. Furthermore, the two glycosidic bonds holding the bases in each base pair are not directly across the helix from each other, defining a common diameter (Figure 12.11). Consequently, the sugar–phosphate backbones of the helix are not equally spaced along the helix axis, and the grooves between them are not the same size. Instead, the intertwined chains create a **major groove** and a **minor groove** (Figure 12.11). The edges of the base pairs have a specific relationship to these grooves. The "top" edges of the base pairs ("top" as defined by placing the glycosidic bond at the bottom, as in Figure 12.10) are exposed along the interior surface or "floor" of the major groove; the base-pair edges nearest to the glycosidic bond form the interior surface of the minor groove. Some proteins that bind to DNA can actually recognize specific nucleotide sequences by "reading" the pattern of H-bonding possibilities presented by the edges of the bases in these grooves. Such DNA–protein interactions provide one step toward understanding how cells regulate the expression of genetic information encoded in DNA (see Chapter 32).

Conformational Variation in Double-Helical Structures

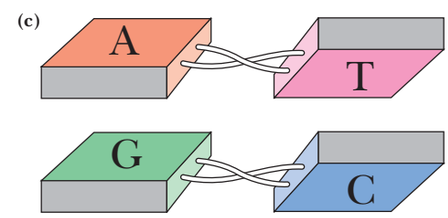
In solution, DNA ordinarily assumes the structure we have been discussing: B-DNA. However, nucleic acids also occur naturally in other double-helical forms. The base-pairing arrangement remains the same, but the sugar–phosphate groupings that constitute the backbone are inherently flexible and can adopt different conformations. One conformational variation is **propeller twist** (Figure 12.12). Propeller twist allows greater overlap between successive bases along a strand of DNA and diminishes the area of contact between bases and solvent water.



Two base pairs with 32° of right-handed helical twist: the *minor-groove edges are drawn with heavy shading.*

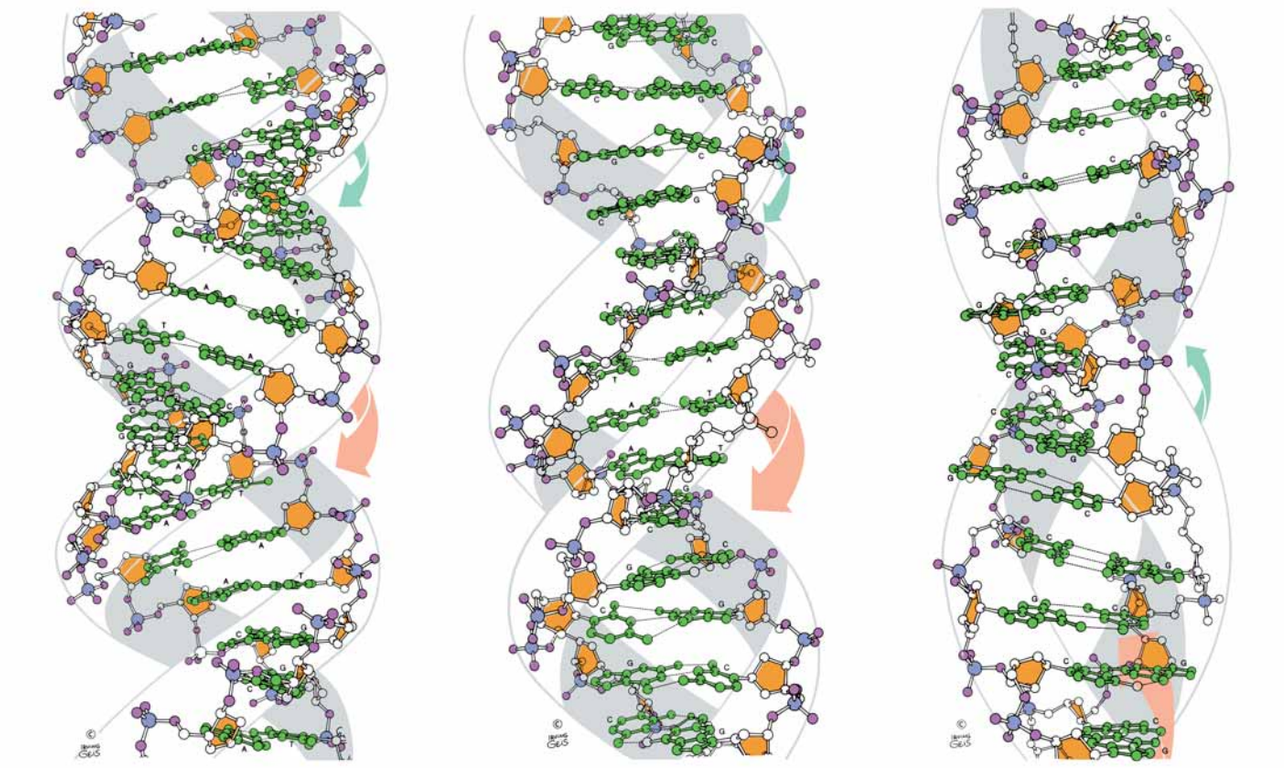


Propeller twist, as in (2), allows greater overlap of bases within the same strand and reduces the area of contact between the bases and water.



Propeller-twisted base pairs. Note how the hydrogen bonds between bases are distorted by this motion, yet remain intact. The minor-groove edges of the bases are shaded.

FIGURE 12.12 • Helical twist and propeller twist in DNA. (a) Successive base pairs in B-DNA show a rotation with respect to each other (so-called helical twist) of 36° or so, as viewed down the cylindrical axis of the DNA. (b) Rotation in a different dimension—**propellor twist**—allows the hydrophobic surfaces of bases to overlap better. The view here is edge-on to two successive bases in one DNA strand (as if the two bases on the right-hand strand of DNA in (a) were viewed from the right-hand margin of the page; dots represent end-on views down the glycosidic bonds). Clockwise rotation (as shown here) has a positive sign. (c) The two bases on the left-hand strand of DNA in (a) also show positive propellor twist (a clockwise rotation of the two bases in (a) as viewed from the left-hand margin of the paper). *(Adapted from Figure 3.4 in Callandine, C. R., and Drew, H. R., 1992.* Understanding DNA: The Molecule and How It Works. *London: Academic Press.)*



A
DNA

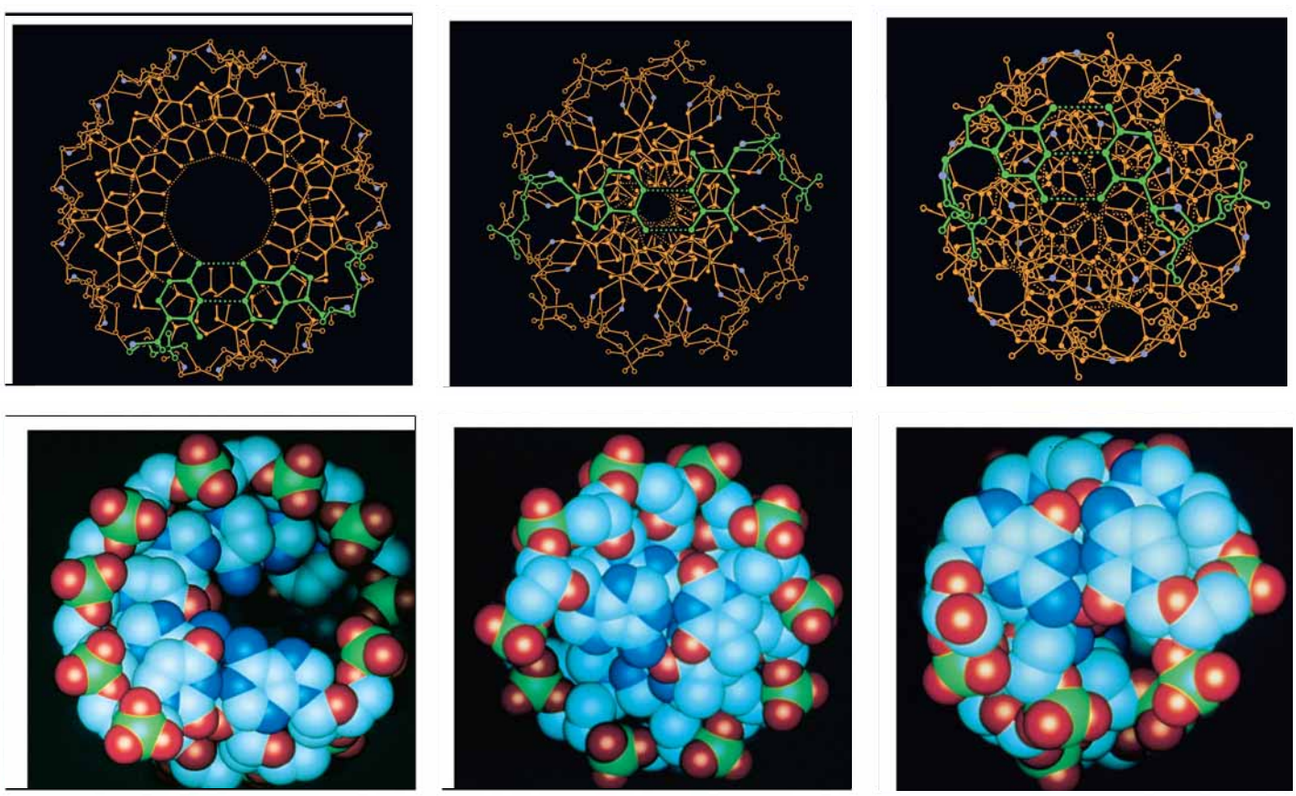
B
DNA

Z
DNA

Alternative Form of Right-Handed DNA

An alternative form of the right-handed double helix is **A-DNA.** A-DNA molecules differ in a number of ways from B-DNA. The pitch, or distance required to complete one helical turn, is different. In B-DNA, it is 3.4 nm, whereas in A-DNA it is 2.46 nm. One turn in A-DNA requires 11 bp to complete. Depending on local sequence, 10 to 10.6 bp define one helical turn in B-form DNA. In A-DNA, the base pairs are no longer nearly perpendicular to the helix axis but instead are tilted 19° with respect to this axis. Successive base pairs occur every 0.23 nm along the axis, as opposed to 0.332 nm in B-DNA. The B-form of DNA is thus longer and thinner than the short, squat A-form, which has its base pairs displaced around, rather than centered on, the helix axis. Figure 12.13 shows the relevant structural characteristics of the A- and B-forms of DNA. (Z-DNA, another form of DNA to be discussed shortly, is also depicted in Figure 12.13.) A comparison of the structural properties of A-, B-, and Z-DNA is summarized in Table 12.1.

Although relatively dehydrated DNA fibers can be shown to adopt the A-conformation under physiological conditions, it is unclear whether DNA ever assumes this form *in vivo*. However, double-helical DNA:RNA hybrids probably have an A-like conformation. The 2′-OH in RNA sterically prevents dou-



A DNA | B DNA | Z DNA

FIGURE 12.13 • *(here and on the facing page)* Comparison of the A-, B-, and Z-forms of the DNA double helix. The distance required to complete one helical turn is shorter in A-DNA than it is in B-DNA. The alternating pyrimidine–purine sequence of Z-DNA is the key to the "left-handedness" of this helix. *(Robert Stodola, Fox Chase Cancer Research Center, and Irving Geis)*

Table 12.1

Comparison of the Structural Properties of A-, B-, and Z-DNA

	Double Helix Type		
	A	B	Z
Overall proportions	Short and broad	Longer and thinner	Elongated and slim
Rise per base pair	2.3 Å	3.32 Å ± 0.19 Å	3.8 Å
Helix packing diameter	25.5 Å	23.7 Å	18.4 Å
Helix rotation sense	Right-handed	Right-handed	Left-handed
Base pairs per helix repeat	1	1	2
Base pairs per turn of helix	~11	~10	12
Mean rotation per base pair	33.6°	35.9° ± 4.2°	−60°/2
Pitch per turn of helix	24.6 Å	33.2 Å	45.6 Å
Base-pair tilt from the perpendicular	+19°	−1.2° ± 4.1°	−9°
Base-pair mean propeller twist	+18°	+16° ± 7°	~0°
Helix axis location	Major groove	Through base pairs	Minor groove
Major groove proportions	Extremely narrow but very deep	Wide and with intermediate depth	Flattened out on helix surface
Minor groove proportions	Very broad but shallow	Narrow and with intermediate depth	Extremely narrow but very deep
Glycosyl bond conformation	anti	anti	anti at C, syn at G

Adapted from Dickerson, R. L., et al., 1982. *Cold Spring Harbor Symposium on Quantitative Biology* **47:**14.

ble-helical regions of RNA chains from adopting the B-form helical arrangement. Importantly, double-stranded regions in RNA chains assume an A-like conformation, with their bases strongly tilted with respect to the helix axis.

Z-DNA: A Left-Handed Double Helix

Z-DNA was first recognized by Alexander Rich and his colleagues at MIT in X-ray analysis of the synthetic deoxynucleotide dCpGpCpGpCpG, which crystallized into an antiparallel double helix of unexpected conformation. The alternating pyrimidine–purine (Py–Pu) sequence of this oligonucleotide is the key to its unusual properties. The *N*-glycosyl bonds of G residues in this alternating copolymer are rotated 180° with respect to their conformation in B-DNA, so that now the purine ring is in the syn rather than the anti conformation (Figure 12.14). The C residues remain in the anti form. Because the G ring is "flipped," the C ring must also flip to maintain normal Watson–Crick base pairing. However, pyrimidine nucleosides do not readily adopt the syn conformation because it creates steric interference between the pyrimidine C-2 oxy substituent and atoms of the pentose. Because the cytosine ring does not rotate relative to the pentose, the whole C nucleoside (base and sugar) must flip 180° (Figure 12.15). It is topologically possible for the G to go syn and the C nucleoside to undergo rotation by 180° without breaking and re-forming the G:C hydrogen bonds. In other words, the B to Z structural transition can take place without disruption of the bonding relationships among the atoms involved.

Because alternate nucleotides assume different conformations, the repeating unit on a given strand in the Z-helix is the dinucleotide. That is, for any number of bases, n, along one strand, $n - 1$ dinucleotides must be considered. For example, a GpCpGpC subset of sequence along one strand is comprised

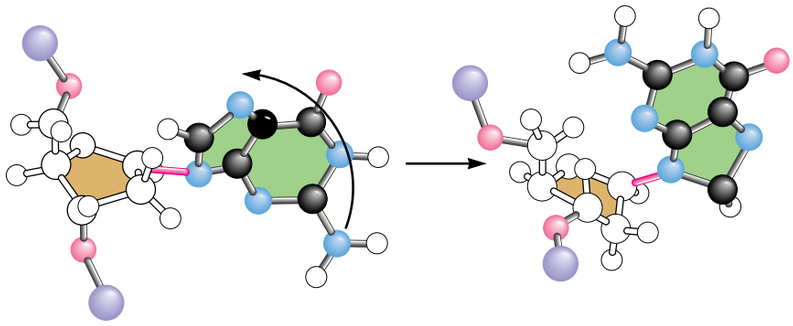


FIGURE 12.14 • Comparison of the deoxyguanosine conformation in B- and Z-DNA. In B-DNA, the Cl′–N-9 glycosyl bond is always in the anti position *(left)*. In contrast, in the left-handed Z-DNA structure, this bond rotates (as shown) to adopt the syn conformation.

of *three* successive dinucleotide units: GpC, CpG, and GpC. (In B-DNA, the nucleotide conformations are essentially uniform and the repeating unit is the mononucleotide.) It follows that the CpG sequence is distinct conformationally from the GpC sequence along the alternating copolymer chains in the Z-double helix. The conformational alterations going from B to Z realign the sugar–phosphate backbone along a zigzag course that has a left-handed orientation (Figure 12.13), thus the designation *Z-DNA*. Note that in any GpCpGp subset, the sugar–phosphates of GpC form the horizontal "zig" while the CpG backbone segment forms the vertical "zag." The mean rotation angle circumscribed around the helix axis is −15° for a CpG step and −45° for a GpC step (giving −60° for the dinucleotide repeat). The minus sign denotes a left-handed or counterclockwise rotation about the helix axis. Z-DNA is more elongated and slimmer than B-DNA.

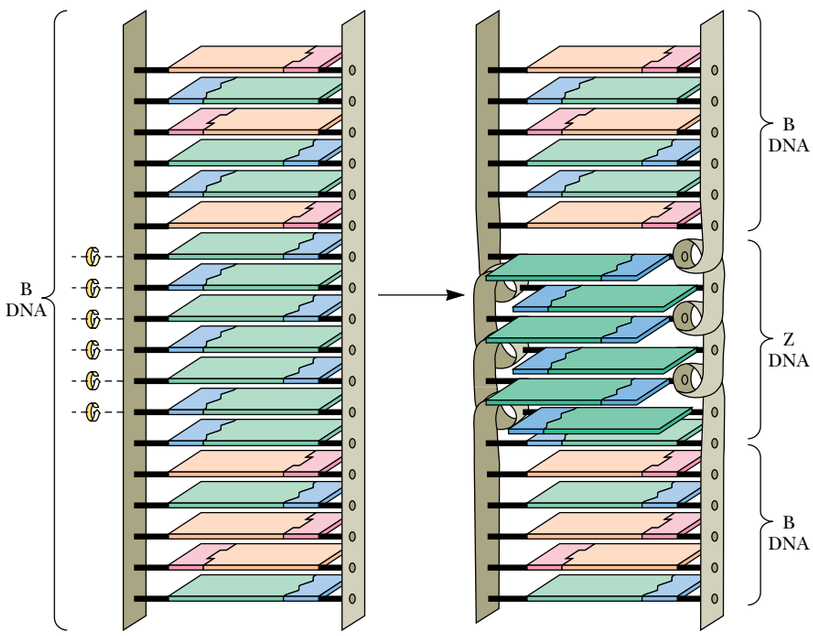


FIGURE 12.15 • The change in topological relationships of base pairs from B- to Z-DNA. A six-base-pair segment of B-DNA is converted to Z-DNA through rotation of the base pairs, as indicated by the curved arrows. The purine rings (green) of the deoxyguanosine nucleosides rotate via an anti to syn change in the conformation of the guanine–deoxyribose glycosidic bond; the pyrimidine rings (blue) are rotated by flipping the entire deoxycytidine nucleoside (base *and* deoxyribose). As a consequence of these conformational changes, the base pairs in the Z-DNA region no longer share π,π stacking interactions with adjacent B-DNA regions.

Cytosine Methylation and Z-DNA

The Z-form can arise in sequences that are not strictly alternating Py–Pu. For example, the hexanucleotide ^{m5}CGAT^{m5}CG, a Py-Pu-Pu-Py-Py-Pu sequence containing two 5-methylcytosines (^{m5}C), crystallizes as Z-DNA. Indeed, the *in vivo* methylation of C at the 5-position is believed to favor a B to Z switch because, in B-DNA, these hydrophobic methyl groups would protrude into the aqueous environment of the major groove and destabilize its structure. In Z-DNA, the same methyl groups can form a stabilizing hydrophobic patch. It is likely that the Z-conformation naturally occurs in specific regions of cellular DNA, which otherwise is predominantly in the B-form. Furthermore, because methylation is implicated in gene regulation, the occurrence of Z-DNA may affect the expression of genetic information (see Part IV, Information Transfer).

The Double Helix in Solution

The long-range structure of B-DNA in solution is not a rigid, linear rod. Instead, DNA behaves as a dynamic, flexible molecule. Localized thermal fluctuations temporarily distort and deform DNA structure over short regions. Base and backbone ensembles of atoms undergo elastic motions on a time scale of nanoseconds. To some extent, these effects represent changes in rotational angles of the bonds comprising the polynucleotide backbone. These changes are also influenced by sequence-dependent variations in base-pair stacking. The consequence is that the helix bends gently. When these variations are summed over the great length of a DNA molecule, the net result of these bending motions is that, at any given time, the double helix assumes a roughly spherical shape, as might be expected for a long, semi-rigid rod undergoing apparently random coiling. It is also worth noting that, on close scrutiny, the surface of the double helix is *not* that of a totally featureless, smooth, regular "barber pole" structure. Different base sequences impart their own special signatures to the molecule by subtle influences on such factors as the groove width, the angle between the helix axis and base planes, and the mechanical rigidity. Certain regulatory proteins bind to specific DNA sequences and participate in activating or suppressing expression of the information encoded therein. These proteins bind at unique sites by virtue of their ability to recognize novel structural characteristics imposed on the DNA by the local nucleotide sequence.

Intercalating Agents Distort the Double Helix

intercalate • to insert between others

Aromatic macrocycles, flat hydrophobic molecules composed of fused, heterocyclic rings, such as **ethidium bromide, acridine orange,** and **actinomycin D** (Figure 12.16), can insert between the stacked base pairs of DNA. The bases are forced apart to accommodate these so-called **intercalating agents,** causing an unwinding of the helix to a more ladderlike structure. The deoxyribose–phosphate backbone is almost fully extended as successive base pairs are displaced 0.7 nm from one another, and the rotational angle about the helix axis between adjacent base pairs is reduced from 36° to 10°.

Dynamic Nature of the DNA Double Helix in Solution

Intercalating substances insert with ease into the double helix, indicating that the van der Waals bonds they form with the bases sandwiching them are more favorable than similar bonds between the bases themselves. Furthermore, the fact that these agents slip in suggests that the double helix must temporarily unwind and present gaps for these agents to occupy. That is, the DNA double helix in solution must be represented by a set of metastable alternatives to the standard B-conformation. These alternatives constitute a flickering repertoire of dynamic structures.

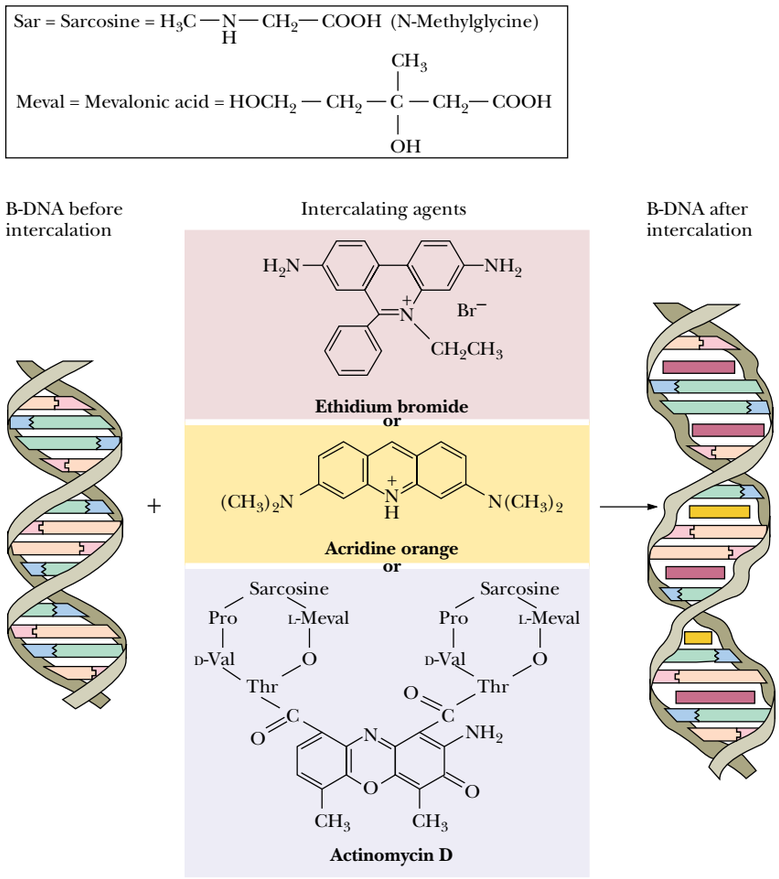


FIGURE 12.16 • The structures of ethidium bromide, acridine orange, and actinomycin D, three intercalating agents, and their effects on DNA structure.

12.3 • Denaturation and Renaturation of DNA

Thermal Denaturation and Hyperchromic Shift

When duplex DNA molecules are subjected to conditions of pH, temperature, or ionic strength that disrupt hydrogen bonds, the strands are no longer held together. That is, the double helix is **denatured** and the strands separate as individual random coils. If temperature is the denaturing agent, the double helix is said to *melt*. The course of this dissociation can be followed spectrophotometrically because the relative absorbance of the DNA solution at 260 nm increases as much as 40% as the bases unstack. This absorbance increase, or **hyperchromic shift,** is due to the fact that the aromatic bases in DNA interact via their π electron clouds when stacked together in the double helix. Because the UV absorbance of the bases is a consequence of π electron transitions, and because the potential for these transitions is diminished when the bases stack, the bases in duplex DNA absorb less 260-nm radiation than expected for their numbers. Unstacking alleviates this suppression of UV absorbance. The rise in absorbance coincides with strand separation, and the midpoint of the absorbance increase is termed the **melting temperature,** T_m

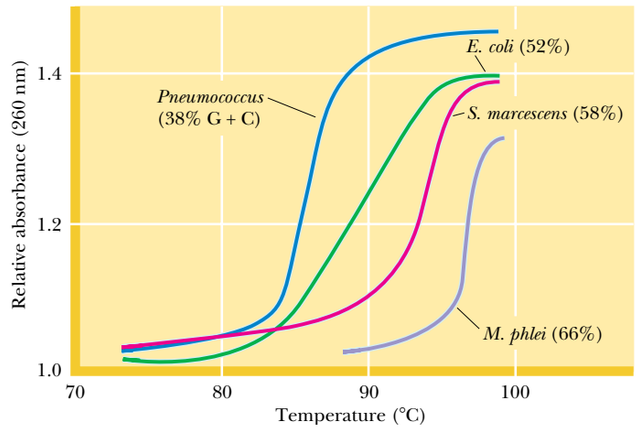


FIGURE 12.17 • Heat denaturation of DNA from various sources, so-called melting curves. The midpoint of the melting curve is defined as the melting temperature, T_m. *(From Marmur, J., 1959.* Nature ***183:****1427–1429.)*

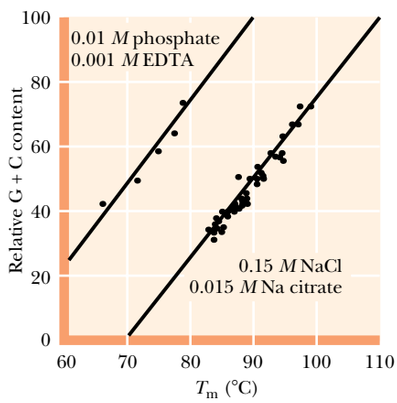


FIGURE 12.18 • The dependence of melting temperature on relative (G + C) content in DNA. Note that T_m increases if ionic strength is raised at constant pH (pH 7); 0.01 *M* phosphate+0.001 *M* EDTA versus 0.15 *M* NaCl/0.015 *M* Na citrate. In 0.15 *M* NaCl/0.015 *M* Na citrate, duplex DNA consisting of 100% A:T pairs melts at less than 70°C, whereas DNA of 100% G:C has a T_m greater than 110°C. *(From Marmur, J., and Doty, P., 1962.* Journal of Molecular Biology ***5:****120.)*

(Figure 12.17). DNAs differ in their T_m values because they differ in relative G + C content. The higher the G + C content of a DNA, the higher its melting temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The dependence of T_m on the G + C content is depicted in Figure 12.18. Also note that T_m is dependent on the ionic strength of the solution; the lower the ionic strength, the lower the melting temperature. At 0.2 *M* Na^+, $T_m = 69.3 + 0.41(\%\ G + C)$. Ions suppress the electrostatic repulsion between the negatively charged phosphate groups in the complementary strands of the helix, thereby stabilizing it. (DNA in pure water melts even at room temperature.) At high concentrations of ions, T_m is raised and the transition between helix and coil is sharp.

pH Extremes or Strong H-Bonding Solutes Also Denature DNA Duplexes

At pH values greater than 10, extensive deprotonation of the bases occurs, destroying their hydrogen bonding potential and denaturing the DNA duplex. Similarly, extensive protonation of the bases below pH 2.3 disrupts base pairing. Alkali is the preferred denaturant because, unlike acid, it does not hydrolyze the glycosidic linkages in the sugar–phosphate backbone. Small solutes that readily form H bonds are also DNA denaturants at temperatures below T_m if present in sufficiently high concentrations to compete effectively with the H-bonding between the base pairs. Examples include formamide and urea.

DNA Renaturation

Denatured DNA will **renature** to re-form the duplex structure if the denaturing conditions are removed (that is, if the solution is cooled, the pH is returned to neutrality, or the denaturants are diluted out). Renaturation requires reassociation of the DNA strands into a double helix, a process termed **reannealing.** For this to occur, the strands must realign themselves so that their complementary bases are once again in register and the helix can be zippered up (Figure 12.19). Renaturation is dependent both on DNA concentration and time. Many of the realignments are imperfect, and thus the strands must dissociate again to allow for proper pairings to be formed. The process occurs more quickly if the temperature is warm enough to promote diffusion of the large DNA molecules but not so warm as to cause melting.

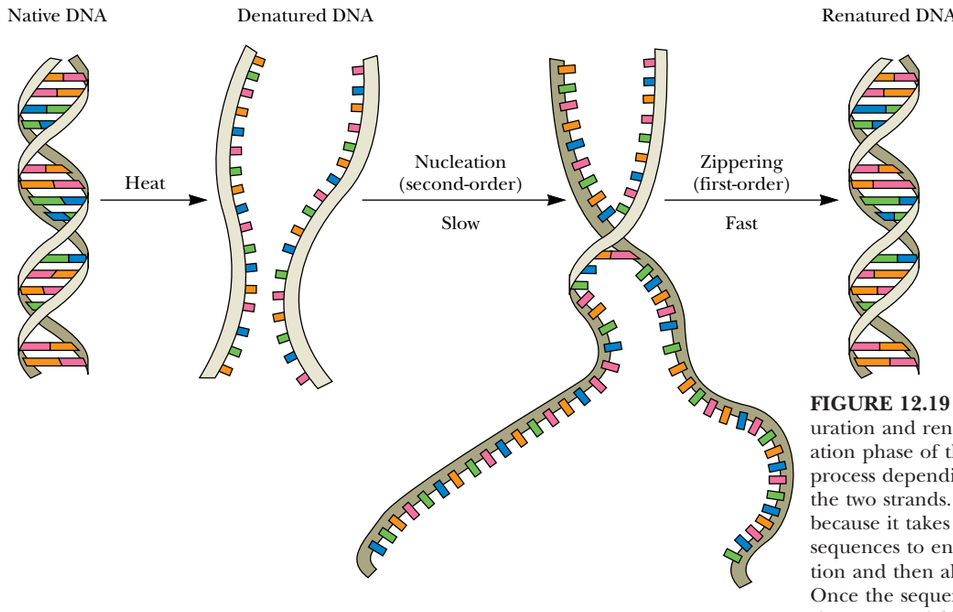


FIGURE 12.19 • Steps in the thermal denaturation and renaturation of DNA. The nucleation phase of the reaction is a second-order process depending on sequence alignment of the two strands. This process takes place slowly because it takes time for complementary sequences to encounter one another in solution and then align themselves in register. Once the sequences are aligned, the strands zipper up quickly.

Renaturation Rate and DNA Sequence Complexity—c_0t Curves

The renaturation rate of DNA is an excellent indicator of the sequence complexity of DNA. For example, bacteriophage T_4 DNA contains about 2×10^5 nucleotide pairs, whereas *Escherichia coli* DNA possesses 4.64×10^6. *E. coli* DNA is considerably more complex in that it encodes more information. Expressed another way, for any given amount of DNA (in grams), the sequences represented in an *E. coli* sample are more heterogeneous, that is, more dissimilar from one another, than those in an equal weight of phage T_4 DNA. Therefore, it will take the *E. coli* DNA strands longer to find their complementary partners and reanneal. This situation can be analyzed quantitatively.

If c is the concentration of single-stranded DNA at time t, then the second-order rate equation for two complementary strands coming together is given by the rate of decrease in c:

$$-dc/dt = k_2c^2$$

where k_2 is the second-order rate constant. Starting with a concentration, c_0, of completely denatured DNA at $t = 0$, the amount of single-stranded DNA remaining at some time t is

$$c/c_0 = 1/(1 + k_2c_0t)$$

where the units of c are mol of nucleotide per L and t is in seconds. The time for half of the DNA to renature (when $c/c_0 = 0.5$) is defined as $t = t_{1/2}$. Then,

$$0.5 = 1/(1 + k_2c_0t_{1/2}) \quad \text{and thus} \quad 1 + k_2c_0t_{1/2} = 2$$

yielding

$$c_0t_{1/2} = 1/k_2$$

A graph of the fraction of single-stranded DNA reannealed (c/c_0) as a function of c_0t on a semilogarithmic plot is referred to as a **c_0t** (pronounced "cot") **curve** (Figure 12.20). The rate of reassociation can be followed spectrophotometrically by the UV absorbance decrease as duplex DNA is formed. Note that

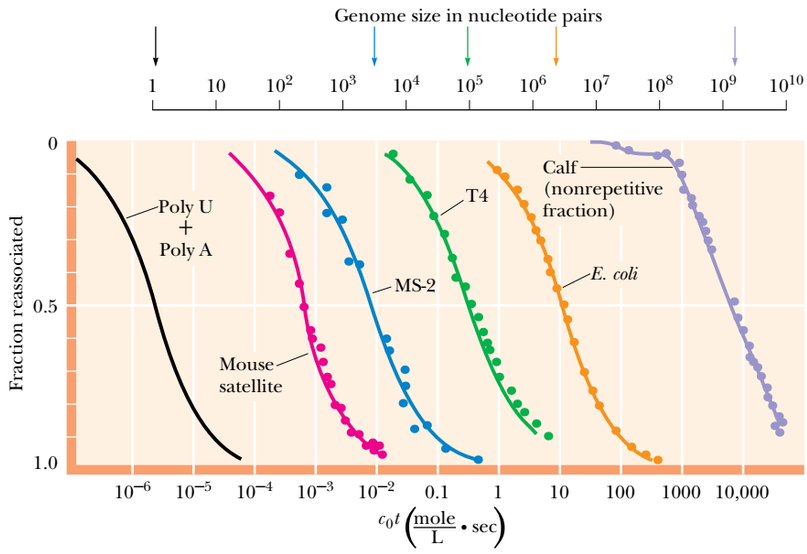


FIGURE 12.20 • These c_0t curves show the rates of reassociation of denatured DNA from various sources and illustrate how the rate of reassociation is inversely proportional to genome complexity. The DNA sources are as follows: poly A+poly U, a synthetic DNA duplex of poly A and poly U polynucleotide chains; mouse satellite DNA, a fraction of mouse DNA in which the same sequence is repeated many thousands of times; MS-2 dsRNA, the double-stranded form of RNA found during replication of MS-2, a simple bacteriophage; T4 DNA, the DNA of a more complex bacteriophage; *E. coli* DNA, bacterial DNA; calf DNA (nonrepetitive fraction), mammalian DNA (calf) from which the highly repetitive DNA fraction (satellite DNA) has been removed. Arrows indicate the genome size (in bp) of the various DNAs. *(From Britten, R. J., and Kohne, D. E., 1968.* Science ***161****:529–540.)*

relatively more complex DNAs take longer to renature, as reflected by their greater $c_0t_{1/2}$ values. Poly A and poly U (Figure 12.20) are minimally complex in sequence and anneal rapidly to form a double-stranded A:U polynucleotide. *Mouse satellite DNA* is a highly repetitive subfraction of mouse DNA. Its lack of sequence heterogeneity is seen in its low $c_0t_{1/2}$ value. MS-2 is a small bacteriophage whose genetic material is RNA. Calf thymus DNA is the mammalian representative in Figure 12.20.

Nucleic Acid Hybridization

If DNA from two different species are mixed, denatured, and allowed to cool slowly so that reannealing can occur, artificial **hybrid duplexes** may form, provided the DNA from one species is similar in nucleotide sequence to the DNA of the other. The degree of hybridization is a measure of the sequence similarity or *relatedness* between the two species. Depending on the conditions of the experiment, about 25% of the DNA from a human forms hybrids with mouse DNA, implying that some of the nucleotide sequences (genes) in humans are very similar to those in mice. Mixed RNA:DNA hybrids can be created *in vitro* if single-stranded DNA is allowed to anneal with RNA copies of itself, such as those formed when genes are transcribed into mRNA molecules.

Nucleic acid hybridization is a commonly employed procedure in molecular biology. First, it can reveal evolutionary relationships. Second, it gives researchers the power to identify specific genes selectively against a vast background of irrelevant genetic material. An appropriately labeled oligo- or polynucleotide, referred to as a **probe,** is constructed so that its sequence is complementary to a target gene. The probe specifically base pairs with the target gene, allowing identification and subsequent isolation of the gene. Also, the quantitative expression of genes (in terms of the amount of mRNA synthesized) can be assayed by hybridization experiments.

Buoyant Density of DNA

Not only the melting temperature of DNA but also its density in solution is dependent on relative G:C content. G:C-rich DNA has a significantly higher density than A:T-rich DNA. Furthermore, a linear relationship exists between the buoyant densities of DNA from different sources and their G:C content

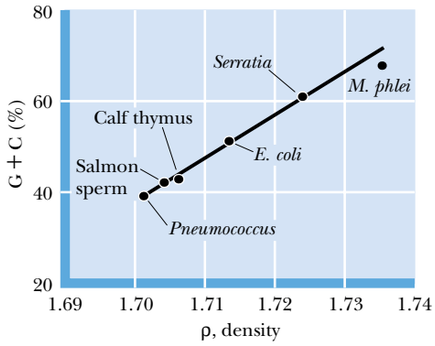


FIGURE 12.21 • The relationship of the densities (in g/mL) of DNAs from various sources and their G:C content. *(From Doty, P., 1961.* Harvey Lectures ***55:****103.)*

(Figure 12.21). The density of DNA, ρ (in g/mL), as a function of its G:C content is given by the equation $\rho = 1.660 + 0.098(GC)$, where (GC) is the mole fraction of (G + C) in the DNA. Because of its relatively high density, DNA can be purified from cellular material by a form of density gradient centrifugation known as *isopycnic centrifugation* (see Appendix to this chapter).

12.4 • Supercoils and Cruciforms: Tertiary Structure in DNA

The conformations of DNA discussed thus far are variations sharing a common secondary structural theme, the double helix, in which the DNA is assumed to be in a regular, linear form. DNA can also adopt regular structures of higher complexity in several ways. For example, many DNA molecules are circular. Most, if not all, bacterial chromosomes are covalently closed, circular DNA duplexes, as are almost all plasmid DNAs. **Plasmids** are naturally occurring, self-replicating, circular, extrachromosomal DNA molecules found in bacteria; plasmids carry genes specifying novel metabolic capacities advantageous to the host bacterium. Various animal virus DNAs are circular as well.

Supercoils

In duplex DNA, the two strands are wound about each other once every 10 bp, that is, once every turn of the helix. Double-stranded circular DNA (or linear DNA duplexes whose ends are not free to rotate), form **supercoils** if the strands are underwound (*negatively supercoiled*) or overwound (*positively supercoiled*) (Figure 12.22). Underwound duplex DNA has fewer than the natural number of turns, whereas overwound DNA has more. DNA supercoiling is analogous to twisting or untwisting a two-stranded rope so that it is torsionally stressed. Negative supercoiling introduces a torsional stress that favors unwinding of the right-handed B-DNA double helix, while positive supercoiling overwinds such a helix. Both forms of supercoiling compact the DNA so that it sediments faster upon ultracentrifugation or migrates more rapidly in an electrophoretic gel in comparison to **relaxed DNA** (DNA that is not supercoiled).

Linking Number

The basic parameter characterizing supercoiled DNA is the **linking number** (L). This is the number of times the two strands are intertwined, and, provided both strands remain covalently intact, L cannot change. In a relaxed circular

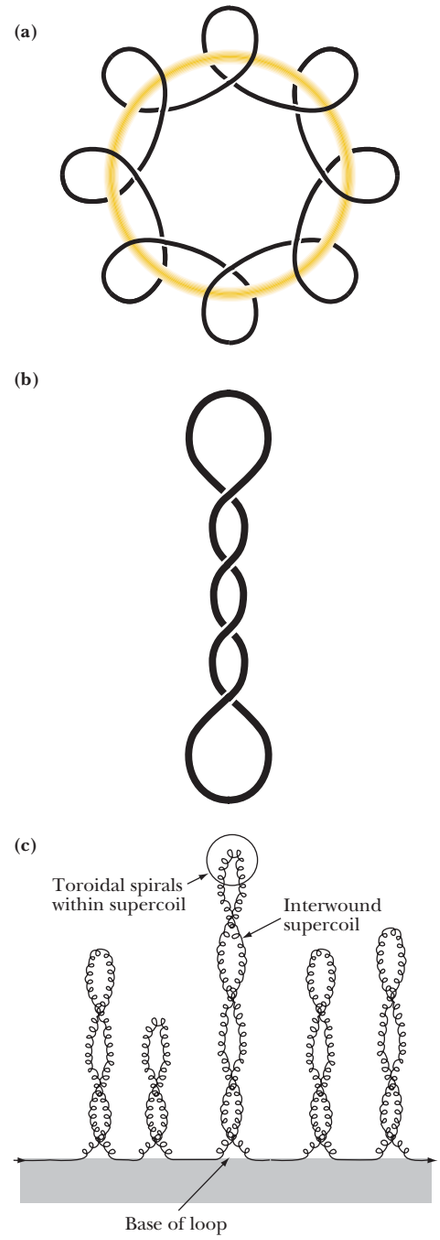


FIGURE 12.22 • Toroidal and interwound varieties of DNA supercoiling. (a) The DNA is coiled in a spiral fashion about an imaginary toroid. (b) The DNA interwinds and wraps about itself. (c) Supercoils in long, linear DNA arranged into loops whose ends are restrained —a model for chromosomal DNA. *(Adapted from Figures 6.1 and 6.2 in Callandine, C. R., and Drew, H. R., 1992.* Understanding DNA: The Molecule and How It Works. *London: Academic Press.)*

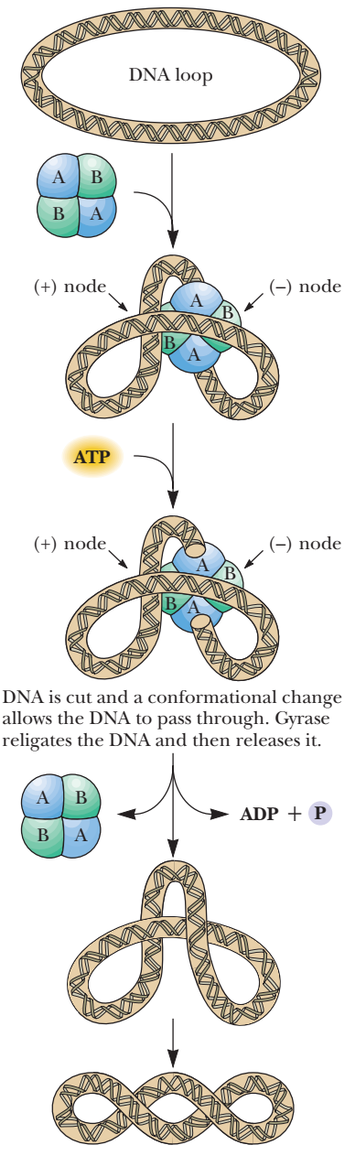


FIGURE 12.24 • A simple model for the action of bacterial DNA gyrase (topoisomerase II). The A-subunits cut the DNA duplex and then hold onto the cut ends. Conformational changes occur in the enzyme that allow a continuous region of the DNA duplex to pass between the cut ends and into an internal cavity of the protein. The cut ends are then re-ligated, and the intact DNA duplex is released from the enzyme. The released intact circular DNA now contains two negative supercoils as a consequence of DNA gyrase action.

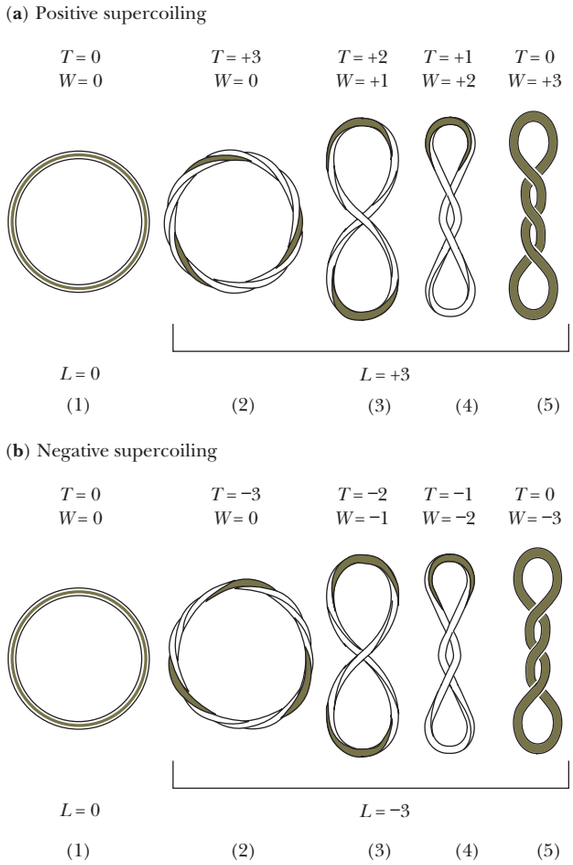


FIGURE 12.23 • Supercoiled DNA topology. *(Adapted from Figures 6.5 and 6.6 in Callandine, C. R., and Drew, H. R., 1992.* Understanding DNA: The Molecule and How It Works. *London: Academic Press.)*

DNA duplex of 400 bp, L is 40 (assuming 10 bp per turn in B-DNA). The linking number for relaxed DNA is usually taken as the reference parameter and is written as L_0. L can be equated to the **twist** (T) and **writhe** (W) of the duplex, where twist is the number of helical turns and writhe is the number of supercoils:

$$L = T + W$$

Figure 12.23a shows the values of T and W for various positively and negatively supercoiled circular DNAs. In any closed, circular DNA duplex that is relaxed, $W = 0$. A relaxed circular DNA of 400 bp has 40 helical turns, $T = L = 40$. This linking number can only be changed by breaking one or both strands of the DNA, winding them tighter or looser, and rejoining the ends. Enzymes capable of carrying out such reactions are called **topoisomerases** because they change the topological state of DNA. Topoisomerase falls into two basic classes, I and II. Topoisomerases of the I type cut one strand of a DNA double helix, pass the other strand through, and then rejoin the cut ends. Topoisomerase II enzymes cut both strands of a dsDNA, pass a region of the DNA duplex between the cut ends, and then rejoin the ends (Figure 12.24). Topoisomerases are important players in DNA replication (see Chapter 30).

DNA Gyrase

The bacterial enzyme **DNA gyrase** is a topoisomerase that introduces negative supercoils into DNA in the manner shown in Figure 12.24. Suppose DNA gyrase puts four negative supercoils into the 400-bp circular duplex, then $W = -4$, T remains the same, and $L = 36$ (Figure 12.25). In actuality, the negative supercoils cause a torsional stress on the molecule so that T tends to decrease; that is, the helix becomes a bit unwound so that base pairs are separated. The extreme would be that T would decrease by 4 and the supercoiling would be removed ($T = 36$, $L = 36$, and $W = 0$). Usually the real situation is a compromise in which the negative value of W is reduced, T decreases slightly, and these changes are distributed over the length of the circular duplex so that no localized unwinding of the helix ensues. Although the parameters T and W are conceptually useful, neither can be measured experimentally at the present time.

Superhelix Density

The difference between the linking number of a DNA and the linking number of its relaxed form is ΔL: $\Delta L = (L - L_0)$. In our example with four negative supercoils, $\Delta L = -4$. The **superhelix density** or **specific linking difference** is defined as $\Delta L/L_0$ and is sometimes termed *sigma*, σ. For our example, $\sigma = -4/40$, or -0.1. As a ratio, σ is a measure of supercoiling that is independent of length. Its sign reflects whether the supercoiling tends to unwind (*negative* σ) or overwind (*positive* σ) the helix. In other words, the superhelix density states the number of supercoils per 10 bp, which also is the same as the number of supercoils per B-DNA repeat. Circular DNA isolated from natural sources is always found in the underwound, negatively supercoiled state.

Toroidal Supercoiled DNA

Negatively supercoiled DNA can arrange into a toroidal state (Figure 12.26). The toroidal state of negatively supercoiled DNA is stabilized by wrapping around proteins which serve as spools for the DNA "ribbon." This toroidal con-

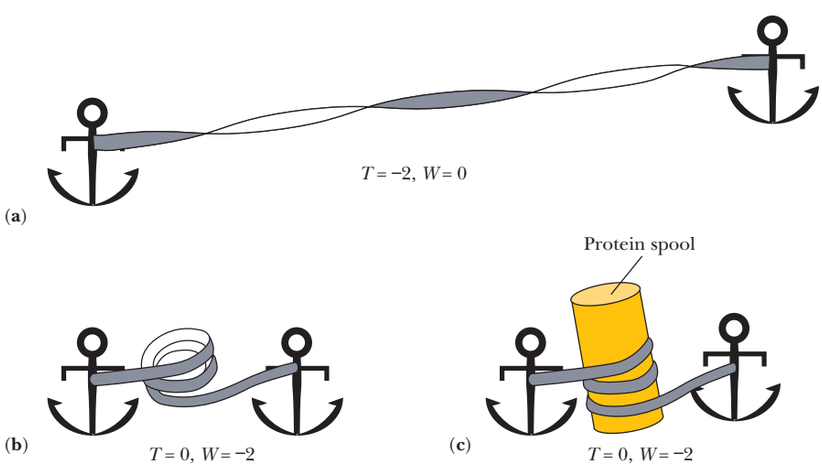


FIGURE 12.26 • Supercoiled DNA in a toroidal form wraps readily around protein "spools." A twisted segment of linear DNA with two negative supercoils (a) can collapse into a toroidal conformation if its ends are brought closer together (b). Wrapping the DNA toroid around a protein "spool" stabilizes this conformation of supercoiled DNA (c). *(Adapted from Figure 6.6 in Callandine, C. R., and Drew, H. R., 1992.* Understanding DNA: The Molecule and How It Works. *London: Academic Press.)*

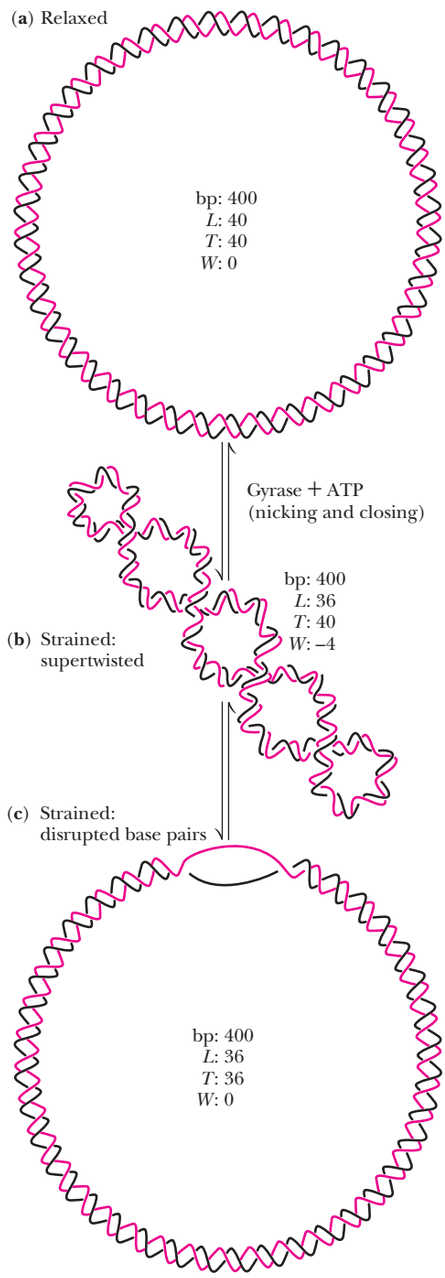


FIGURE 12.25 • A 400-bp circular DNA molecule in different topological states: (a) relaxed, (b) negative supercoils distributed over the entire length, and (c) negative supercoils creating a localized single-stranded region. Negative supercoiling has the potential to cause localized unwinding of the DNA double helix so that single-stranded regions (or bubbles) are created.

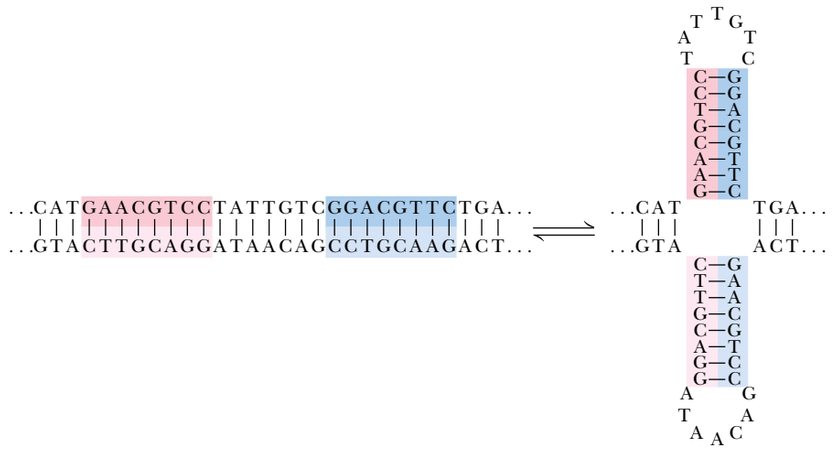


FIGURE 12.27 • The formation of a cruciform structure from a palindromic sequence within DNA. The self-complementary inverted repeats can rearrange to form hydrogen-bonded cruciform loops.

formation of DNA is found in protein : DNA interactions that are the basis of phenomena as diverse as chromosome structure (see Figure 12.31) and gene expression.

Cruciforms

Palindromes are words, phrases, or sentences that are the same when read backward or forward, such as "radar," "sex at noon taxes," "Madam, I'm Adam," and "a man, a plan, a canal, Panama." DNA sequences that are **inverted repeats,** or palindromes, have the potential to form a tertiary structure known as a **cruciform** (literally meaning "cross-shaped") if the normal interstrand base pairing is replaced by intrastrand pairing (Figure 12.27). In effect, each DNA strand folds back on itself in a hairpin structure to align the palindrome in base-pairing register. Such cruciforms are never as stable as normal DNA duplexes because an unpaired segment must exist in the loop region. However, negative supercoiling causes a localized disruption of hydrogen bonding between base pairs in DNA and may promote formation of cruciform loops. Cruciform structures have a twofold rotational symmetry about their centers and potentially create distinctive recognition sites for specific DNA-binding proteins.

12.5 • Chromosome Structure

A typical human cell is 20 μm in diameter. Its genetic material consists of 23 pairs of dsDNA molecules in the form of **chromosomes,** the average length of which is 3×10^9 bp/23 or 1.3×10^8 nucleotide pairs. At 0.34 nm/bp in B-DNA, this represents a DNA molecule 5 cm long. Together, these 46 dsDNA molecules amount to more than 2 m of DNA that must be packaged into a nucleus perhaps 5μm in diameter! Clearly, the DNA must be condensed by a factor of more than 10^5. This remarkable task is accomplished by neatly wrapping the DNA around protein spools called **nucleosomes** and then packing the nucleosomes to form a helical filament that is arranged in loops associated with the **nuclear matrix,** a skeleton or scaffold of proteins providing a structural framework within the nucleus.

Table 12.2

Properties of Histones

Histone	Ratio of Lysine to Arginine	M_r	Copies per Nucleosome
H1	59/3	21,200	1 (not in bead)
H2A	13/13	14,100	2 (in bead)
H2B	20/8	13,900	2 (in bead)
H3	13/17	15,100	2 (in bead)
H4	11/14	11,400	2 (in bead)

Nucleosomes

The DNA in a eukaryotic cell nucleus during the interphase between cell divisions exists as a nucleoprotein complex called **chromatin.** The proteins of chromatin fall into two classes: **histones** and **nonhistone chromosomal proteins.** Histones are abundant structural proteins, whereas the nonhistone class is represented only by a few copies each of many diverse proteins involved in genetic regulation. The histones are relatively small, positively charged arginine- or lysine-rich proteins that interact via ionic bonds with the negatively charged phosphate groups on the polynucleotide backbone. Five distinct histones are known: **H1, H2A, H2B, H3,** and **H4** (Table 12.2). Pairs of histones H2A, H2B, H3, and H4 aggregate to form an octameric core structure, which is the core of the **nucleosome,** around which the DNA helix is wound (see Figure 11.23).

If chromatin is swelled suddenly in water and prepared for viewing in the electron microscope, the nucleosomes are evident as "beads on a string," dsDNA being the string (Figure 12.28). The structure of the histone octamer core has been determined by X-ray crystallography without DNA by E. N. Moudrianakis's laboratory (Figure 12.29) and wrapped with DNA by T. J.

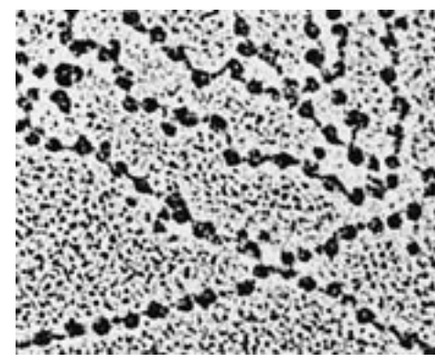
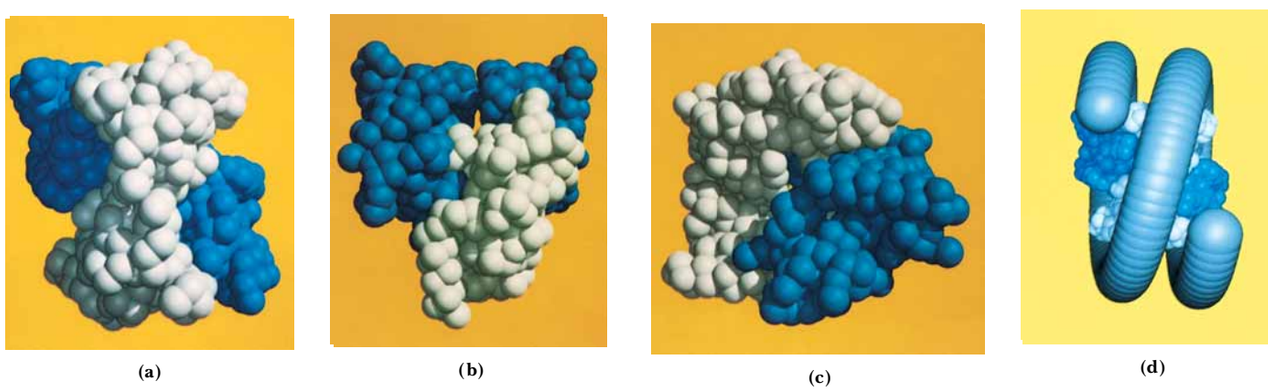


FIGURE 12.28 • Electron micrograph of *Drosophila melanogaster* chromatin after swelling reveals the presence of nucleosomes as "beads on a string." *(Electron micrograph courtesy of Oscar L. Miller, Jr., of the University of Virginia)*



(a) **(b)** **(c)** **(d)**

FIGURE 12.29 • Four orthogonal views of the histone octamer as determined by X-ray crystallography: (a) front view; (b) top view; and (c) disk view, that is, as viewed down the long axis of the chromatin fiber. In the (c) perspective, the DNA duplex would wrap around the octamer, with the axis of the DNA supercoil perpendicular to the plane of the picture. (d) Suggested appearance of the nucleosome when wrapped with DNA. *(Photographs courtesy of Evangelos N. Moudrianakis of Johns Hopkins University)*

(**a**) (**b**)

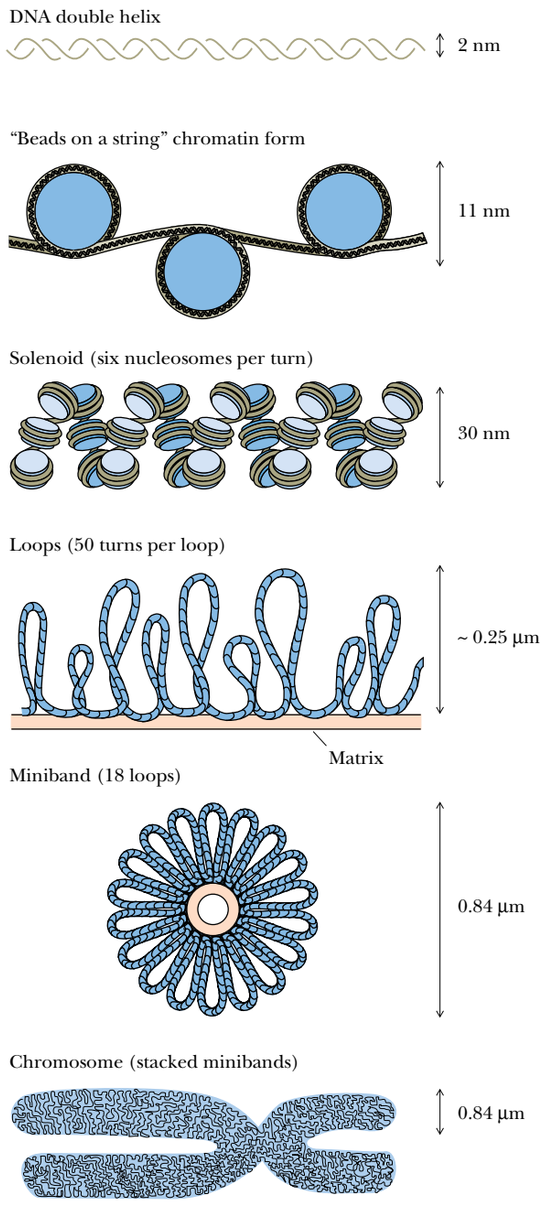
FIGURE 12.30 • (a) Deduced structure of the nucleosome core particle wrapped with 1.65 turns of DNA (146 bp). The DNA is shown as a ribbon. (*left*) View down the axis of the nucleosome; *(right)* view perpendicular to the axis. (b) One-half of the nucleosome core particle with 73 bp of DNA, as viewed down the nucleosome axis. Note that the DNA does not wrap in a uniform circle about the histone core, but instead follows a course consisting of a series of somewhat straight segments separated by bends. *(Adapted from Luger, C., et al., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution.* Nature ***389:****251–260. Photos courtesy of T. J. Richmond, ETH-Hönggerberg, Zurich, Switzerland.)*

Richmond and collaborators (Figure 12.30). The octamer (Figure 12.29) has surface landmarks that guide the course of the DNA around the octamer; 146 bp of B-DNA in a flat, left-handed superhelical conformation make 1.65 turns around the histone core (Figure 12.30), which itself is a protein superhelix consisting of a spiral array of the four histone dimers. Histone 1, a three-domain protein, serves to seal the ends of the DNA turns to the nucleosome core and to organize the additional 40 to 60 bp of DNA that link consecutive nucleosomes.

Organization of Chromatin and Chromosomes

solenoid • a coil wound in the form of a helix

A higher order of chromatin structure is created when the nucleosomes, in their characteristic beads-on-a-string motif, are wound in the fashion of a *solenoid* having six nucleosomes per turn (Figure 12.31). The resulting 30-nm filament contains about 1200 bp in each of its solenoid turns. Interactions between the respective H1 components of successive nucleosomes stabilize the 30-nm filament. This 30-nm filament then forms long DNA loops of variable length, each containing on average between 60,000 and 150,000 bp. Electron microscopic analysis of human chromosome 4 suggests that 18 such loops are then arranged radially about the circumference of a single turn to form a **miniband unit** of the chromosome. According to this model, approximately 10^6 of these minibands are arranged along a central axis in each of the chromatids of human chromosome 4 that form at mitosis (Figure 12.31). Despite intensive study, much of the higher-order structure of chromosomes remains a mystery.



Base pairs per turn	Packing ratio
10	1
80	6–7
1200	~40
60,000	680
~1.1×10^6	1.2×10^4
18 loops/ miniband	1.2×10^4

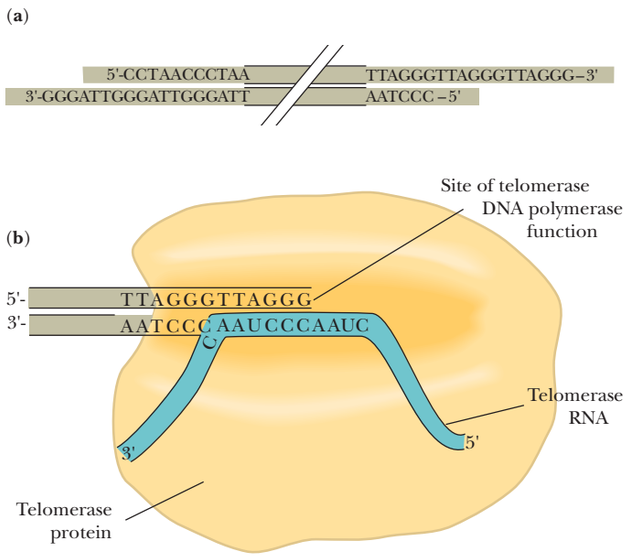
FIGURE 12.31 • A model for chromosome structure, human chromosome 4. The 2-nm DNA helix is wound twice around histone octamers to form 10-nm nucleosomes, each of which contains 160 bp (80 per turn). These nucleosomes are then wound in solenoid fashion with six nucleosomes per turn to form a 30-nm filament. In this model, the 30-nm filament forms long DNA loops, each containing about 60,000 bp, which are attached at their base to the nuclear matrix. Eighteen of these loops are then wound radially around the circumference of a single turn to form a miniband unit of a chromosome. Approximately 10^6 of these minibands occur in each chromatid of human chromosome 4 at mitosis.

HUMAN BIOCHEMISTRY

Telomeres and Tumors

Eukaryotic chromosomes are linear. The ends of chromosomes have specialized structures known as **telomeres.** The telomeres of virtually all eukaryotic chromosomes consist of short, tandemly repeated nucleotide sequences at the ends of the chromosomal DNA. For example, the telomeres of human germline (sperm and egg) cells contain between 1000 and 1700 copies of the hexameric repeat TTAGGG (see figure). Telomeres are believed to be responsible for maintaining chromosomal integrity by protecting against DNA degradation or rearrangement. Telomeres are added to the ends of chromosomal DNA by an RNA-containing enzyme known as **telomerase** (Chapter 30); telomerase is an unusual DNA polymerase that was discovered in 1985 by Elizabeth Blackburn and Carol Greider of the University of California, San Francisco.

However, most normal somatic cells lack telomerase. Consequently, upon every cycle of cell division when the cell replicates its DNA, about 50-nucleotide portions are lost from the end of each telomere. Thus, over time, the telomeres of somatic cells in animals become shorter and shorter, eventually leading to chromosome instability and cell death. This phenomenon has led some scientists to espouse a "telomere theory of aging" that implicates telomere shortening as the principal factor in cell, tissue, and even organism aging. Interestingly, cancer cells appear "immortal" because they continue to reproduce indefinitely. A survey of 20 different tumor types by Geron Corporation of Menlo Park, California, revealed that all contained telomerase activity.



(a) Telomeres on human chromosomes consist of the hexanucleotide sequence TTAGGG repeated between 1000 and 1700 times. These TTAGGG tandem repeats are attached to the 3′-ends of the DNA strands and are paired with the complementary sequence 3′-AATCCC-5′ on the other DNA strand. Thus, a G-rich region is created at the 3′-end of each DNA strand and a C-rich region is created at the 5′-end of each DNA strand. Typically, at each end of the chromosome, the G-rich strand protrudes 12 to 16 nucleotides beyond its complementary C-rich strand. (b) Like other telomerases, human telomerase is a ribonucleoprotein. The ribonucleic acid of human telomerase is an RNA molecule 962 nucleotides long. This RNA serves as the template for the DNA polymerase activity of telomerase. Nucleotides 46 to 56 of this RNA are CUAA**CCCUAA**C and provide the template function for the telomerase-catalyzed addition of TTAGGG units to the 3′-end of a DNA strand.

12.6 • Chemical Synthesis of Nucleic Acids

Laboratory synthesis of oligonucleotide chains of defined sequence presents some of the same problems encountered in chemical synthesis of polypeptides (see Chapter 5). First, functional groups on the monomeric units (in this case, bases) are reactive under conditions of polymerization and therefore must be protected by blocking agents. Second, to generate the desired sequence, a phosphodiester bridge must be formed between the 3′-O of one nucleotide (B) and the 5′-O of the preceding one (A) in a way that precludes the unwanted bridging of the 3′-O of A with the 5′-O of B. Finally, recoveries at each step must be high so that overall yields in the multistep process are acceptable. As in peptide synthesis (see Chapter 5), *solid phase methods* are used to overcome some

of these problems. Commercially available automated instruments, called **DNA synthesizers** or "gene machines," are capable of carrying out the synthesis of oligonucleotides of 150 bases or more.

Phosphoramidite Chemistry

Phosphoramidite chemistry is currently the accepted method of oligonucleotide synthesis. The general strategy involves the sequential addition of nucleotide units as *nucleoside phosphoramidite* derivatives to a nucleoside covalently attached to the insoluble resin. Excess reagents, starting materials, and side products are removed after each step by filtration. After the desired oligonucleotide has been formed, it is freed of all blocking groups, hydrolyzed from the resin, and purified by gel electrophoresis. The four-step cycle is shown in Figure 12.32. Chemical synthesis takes place in the 3′ → 5′ direction (the reverse of the biological polymerization direction).

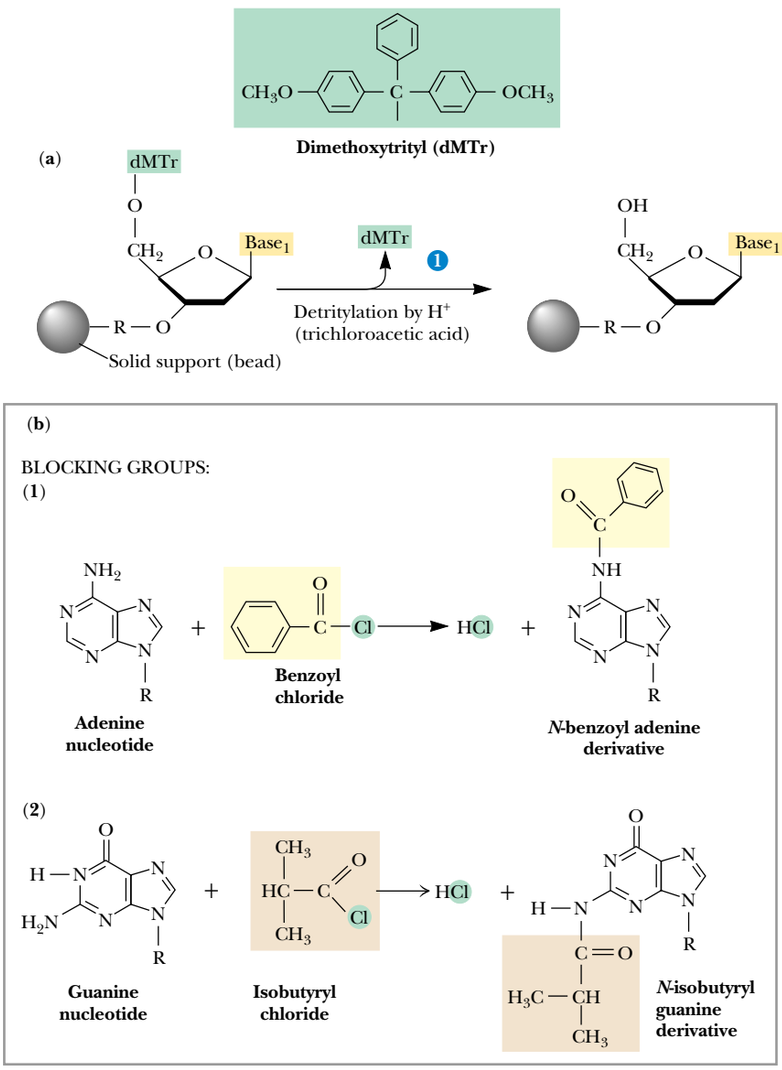


FIGURE 12.32 • Solid phase oligonucleotide synthesis. The four-step cycle starts with the first base in nucleoside form (N-1) attached by its 3′-OH group to an insoluble, inert resin or matrix, typically either controlled pore glass (CPG) or silica beads. Its 5′-OH is blocked with a dimethoxytrityl (DMTr) group (a). If the base has reactive —NH_2 functions, as in A, G, or C, then *N*-benzoyl or *N*-isobutyryl derivatives are used to prevent their reaction (b). In step 1, the DMTr protecting group is removed by trichloroacetic acid treatment. Step 2 is the coupling step: the second base (N-2) is added in the form of a nucleoside phosphoramidite derivative whose 5′-OH bears a DMTr blocking group so it cannot polymerize with itself (c). *(Figure continued on following page.)*

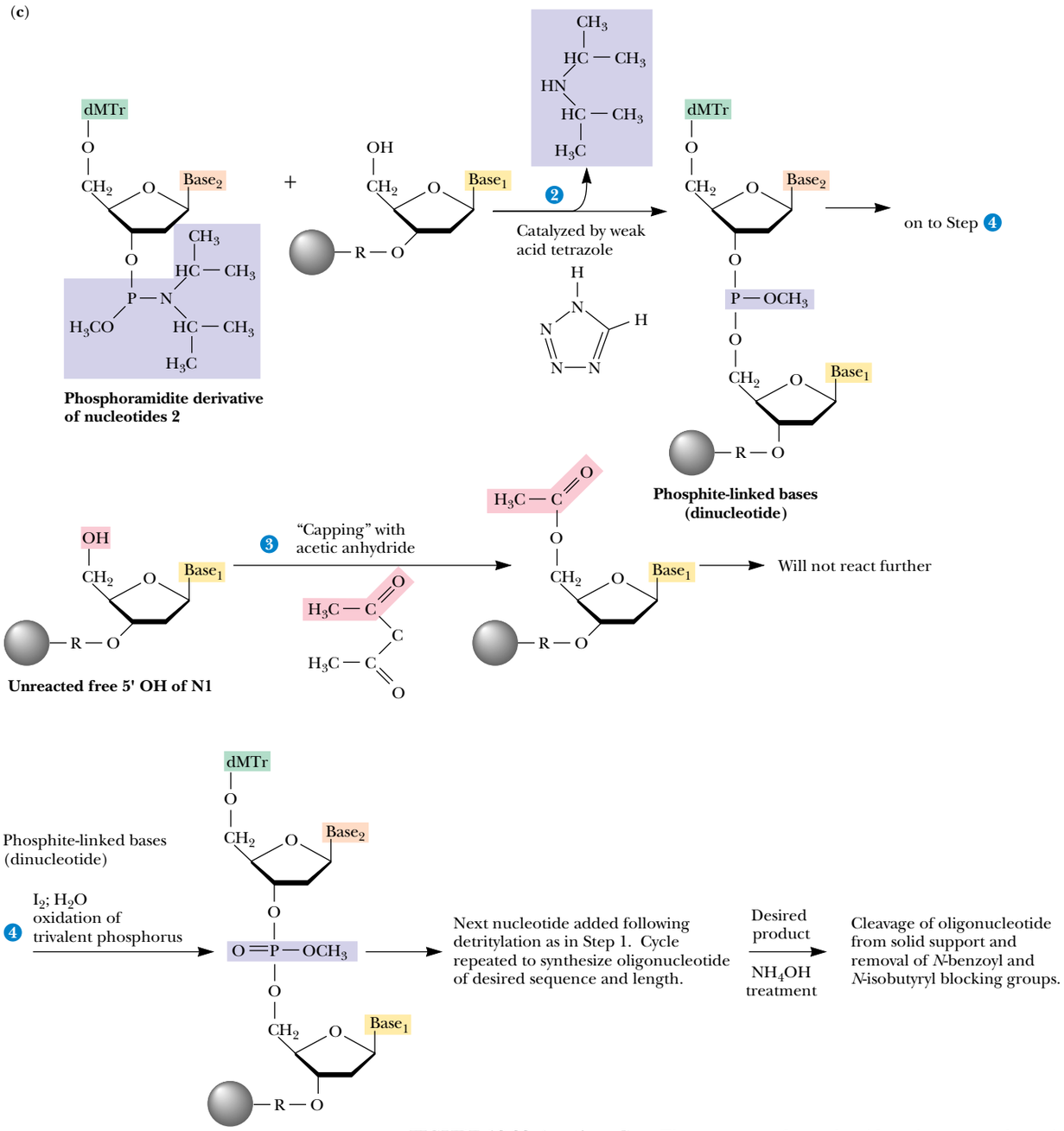


FIGURE 12.32 ***(continued)*** • The presence of a weak acid, such as tetrazole, activates the phosphoramidite, and it rapidly reacts with the free 5′-OH of N-1, forming a dinucleotide linked by a phosphite group. Chemical synthesis thus takes place in the 3′ → 5′ direction. Unreacted free 5′-OHs of N-1 (usually only 2–6% of the total) are blocked from further participation in the polymerization process by acetylation with acetic anhydride in step 3, referred to as *capping*. The phosphite linkage between N-1 and N-2 is highly reactive and, in step 4, it is oxidized by aqueous iodine (I_2) to form the desired more stable phosphate group. This completes the cycle. Subsequent cycles add successive residues to the resin-immobilized chain. When the chain is complete, it is cleaved from the support with NH_4OH, which also removes the *N*-benzoyl and *N*-isobutyryl protecting groups from the amino functions on the A, G, and C residues.

Chemically Synthesized Genes

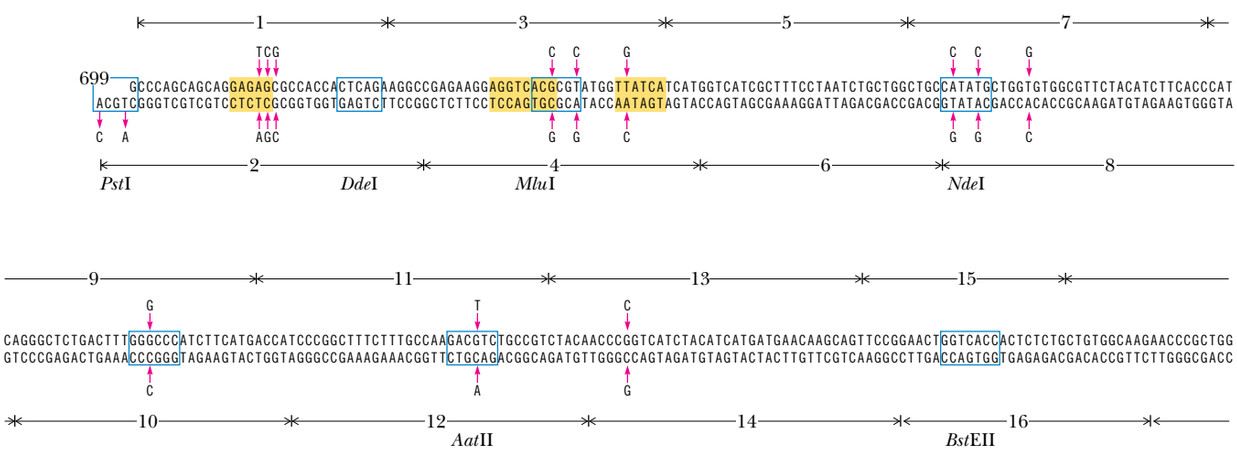
Table 12.3 lists some of the genes that have been chemically synthesized. Because protein-coding genes are characteristically much larger than the 150-bp practical limit on oligonucleotide synthesis, their synthesis involves joining a series of oligonucleotides to assemble the overall sequence. A prime example of such synthesis is the gene for rhodopsin.

Figure 12.33 illustrates the strategy used in the total synthesis of the gene for bovine rhodopsin. This gene, which is 1057 base pairs long, encodes the 348 amino-acid photoreceptor protein of the vertebrate retina. Theoretically, no gene is beyond the scope of these methods, a fact that opens the door to an incredibly exciting range of possibilities for investigating structure–function relationships in the organization and expression of hereditary material.

Table 12.3

Some Chemically Synthesized Genes

Gene	Size (bp)
tRNA	126
α-Interferon	542
Secretin	81
γ-Interferon	453
Rhodopsin	1057
Proenkephalin	77
Connective tissue activating peptide III	280
Lysozyme	385
Tissue plasminogen activator	1610
c-Ha-ras	576
RNase T1	324
Cytochrome b_5	330
Bovine intestinal Ca-binding protein	298
Hirudin	226
RNase A	375



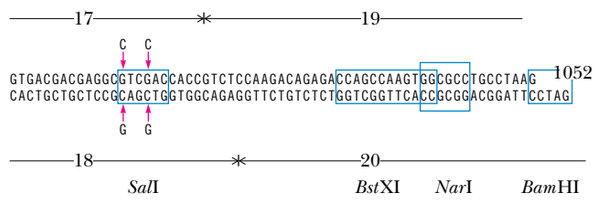


FIGURE 12.33 • Total synthesis of the bovine rhodopsin gene was achieved by joining 72 synthetic oligonucleotides, 36 representing one strand and 36 the complementary strand. These oligonucleotides are overlapping. Once synthesized, the various oligonucleotides, each 15 to 40 nucleotides long, were assembled by annealing and enzymatic ligation into three large fragments, representing nucleotides −5 to 338 (−5 meaning 5 nucleotides before the start of the coding region), 335 to 702, and 699 to 1052. The total gene was then created by joining these fragments. This figure shows only one fragment (fragment PB, comprising nucleotides 699 through 1052), assembled from 20 complementary oligonucleotides whose ends overlap. Odd-numbered oligonucleotides (1, 3, 5, . . .) compose the 5′ → 3′ strand; even-numbered ones (2, 4, 6, . . .) represent the 3′ → 5′ strand. (Vertical arrows indicate nucleotides that were changed from the native gene sequence. Restriction sites are shown boxed in blue lines; those removed from the gene through nucleotide substitutions are shown as yellow shaded boxes.) Note the single-stranded overhangs at either end of the 3′ → 5′ strand. The sequences at these overhangs correspond to restriction endonuclease sites (*PstI* and *BamH1*), which facilitate subsequent manipulation of the fragment in gene assembly and cloning.

12.7 • Secondary and Tertiary Structure of RNA

RNA molecules (see Chapter 11) are typically single-stranded. Nevertheless, they are often rich in double-stranded regions that form when complementary sequences within the chain come together and join via **intrastrand hydrogen bonding.** RNA strands cannot fold to form B-DNA type double helices because their 2′-OH groups are a steric hindrance to this conformation. Instead, RNA double helices adopt a conformation similar to the A-form of DNA, having about 11 bp per turn, and the bases strongly tilted from the plane perpendicular to the helix axis (see Figure 12.13). Both tRNA and rRNA have characteristic secondary structures formed in this manner. Secondary structures are presumed to exist in mRNA species as well, although their nature is as yet little understood. (The functions of tRNA, rRNA, and mRNA are discussed in detail in Part IV: Information Transfer.)

Transfer RNA

In tRNA molecules, which contain from 73 to 94 nucleotides in a single chain, a majority of the bases are hydrogen-bonded to one another. Figure 12.34 shows the structure that typifies tRNAs. *Hairpin turns* bring complementary stretches of bases in the chain into contact so that double-helical regions form. Because of the arrangement of the complementary stretches along the chain, the overall pattern of H-bonding can be represented as a *cloverleaf*. Each cloverleaf consists of four H-bonded segments—three loops and the stem where the 3′- and 5′-ends of the molecule meet. These four segments are designated the ***acceptor stem***, the ***D loop***, the ***anticodon loop***, and the ***TψC loop***.

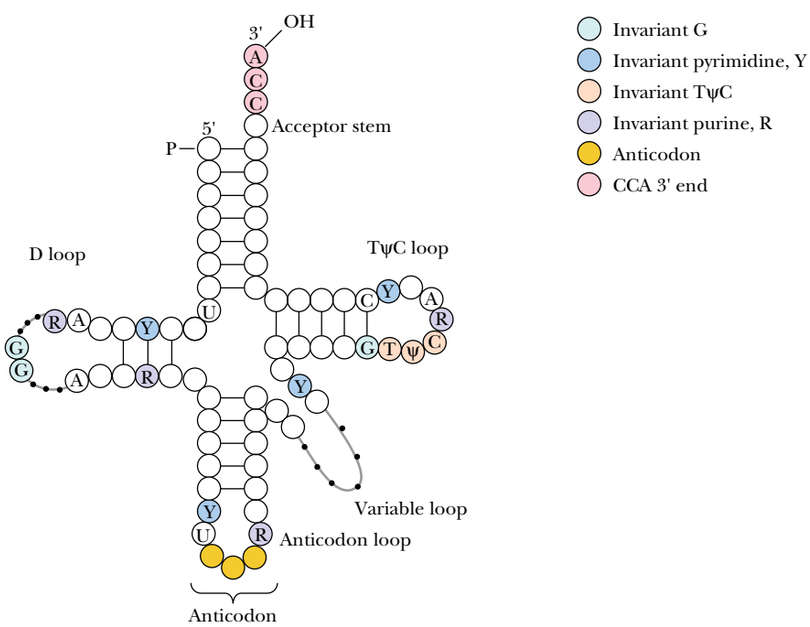


FIGURE 12.34 • A general diagram for the structure of tRNA. The positions of invariant bases as well as bases that seldom vary are shown in color. The numbering system is based on yeast $tRNA^{Phe}$. R = purine; Y = pyrimidine. Dotted lines denote sites in the D loop and variable loop regions where varying numbers of nucleotides are found in different tRNAs.

tRNA Secondary Structure

The *acceptor stem* is where the amino acid is linked to form the aminoacyl-tRNA derivative, which serves as the amino acid–donating species in protein synthesis; this is the physiological role of tRNA. The amino acid adds to the 3′-OH of the 3′-terminal A nucleotide (Figure 12.35). The 3′-end of tRNA is invariantly CCA-3′-OH. This CCA sequence plus a fourth nucleotide extends beyond the double-helical portion of the acceptor stem. The *D loop* is so named because this tRNA loop often contains dihydrouridine, or D, residues. In addition to dihydrouridine, tRNAs characteristically contain a number of unusual bases, including inosine, thiouridine, pseudouridine, and hypermethylated purines (see Figure 11.26). The *anticodon loop* consists of a double-helical segment and seven unpaired bases, three of which are the **anticodon.** (The anticodon is the three-nucleotide unit that recognizes and base pairs with a particular mRNA **codon,** a complementary three-base unit in mRNA which is the genetic information that specifies an amino acid.) Reading 3′ → 5′, the anticodon is invariably preceded by a purine (often an alkylated one) and followed by a U. Anticodon base pairing to the codon on mRNA allows a particular tRNA species to deliver its amino acid to the protein-synthesizing apparatus. It represents the key event in translating the information in the nucleic acid sequence so that the appropriate amino acid is inserted at the right place in the amino acid sequence of the protein being synthesized. Next along the tRNA sequence in the 5′ → 3′ direction comes a loop that varies from tRNA to tRNA in the number of residues that it has, the so-called **extra** or **variable loop.** The last loop in the tRNA, reading 5′ → 3′, is the **TψC loop,** which contains seven unpaired bases including the sequence TψC, where ψ is the symbol for **pseudouridine.** Ribosomes bind tRNAs through recognition of this TψC loop. Almost all of the invariant residues common to tRNAs lie within the non–hydrogen-bonded regions of the cloverleaf structure (Figure 12.34). Figure 12.36 depicts the complete nucleotide sequence and cloverleaf structure of yeast alanine tRNA.

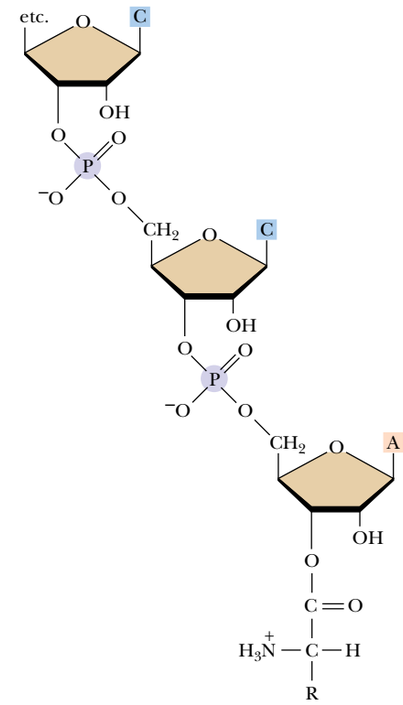


FIGURE 12.35 • Amino acids are linked to the 3′-OH end of tRNA molecules by an ester bond formed between the carboxyl group of the amino acid and the 3′-OH of the terminal ribose of the tRNA.

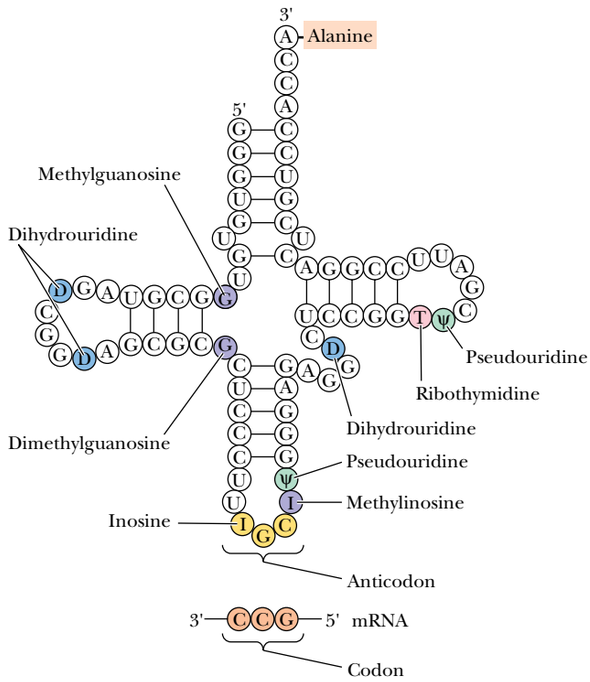


FIGURE 12.36 • The complete nucleotide sequence and cloverleaf structure of yeast alanine tRNA.

tRNA Tertiary Structure

Tertiary structure in tRNA arises from hydrogen-bonding interactions between bases in the D loop with bases in the variable and TψC loops, as shown for yeast phenylalanine tRNA in Figure 12.37. Note that these H bonds involve the invariant nucleotides of tRNAs, thus emphasizing the importance of the tertiary structure they create to the function of tRNAs in general. These H bonds fold the D and TψC arms together and bend the cloverleaf into the stable L-shaped tertiary form (Figure 12.38). Many of these H bonds involve base pairs that are not canonical A:T or G:C pairings (Figure 12.38). The amino acid acceptor stem is at one end of the L, separated by 7 nm or so from the anticodon at the opposite end of the L. The D and TψC loops form the corner of the L. In the L-conformation, the bases are oriented to maximize hydrophobic stacking interactions between their flat faces. Such stacking is a second major factor contributing to L-form stabilization.

Ribosomal RNA

rRNA Secondary Structure

Ribosomes, the protein-synthesizing machinery of cells, are composed of two **subunits,** called **small** and **large,** and ribosomal RNAs are integral components of these subunits (see Table 11.2). A large degree of *intrastrand sequence com-*

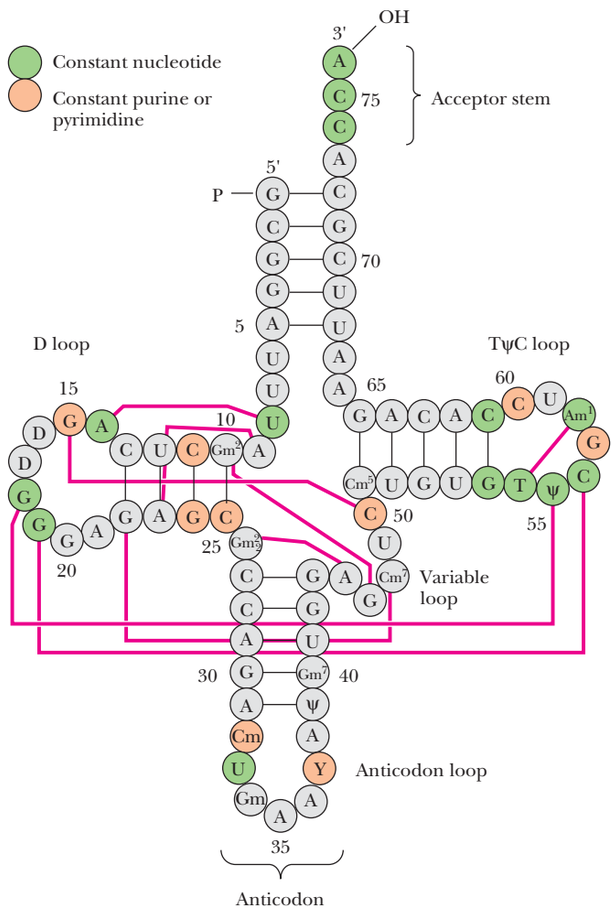
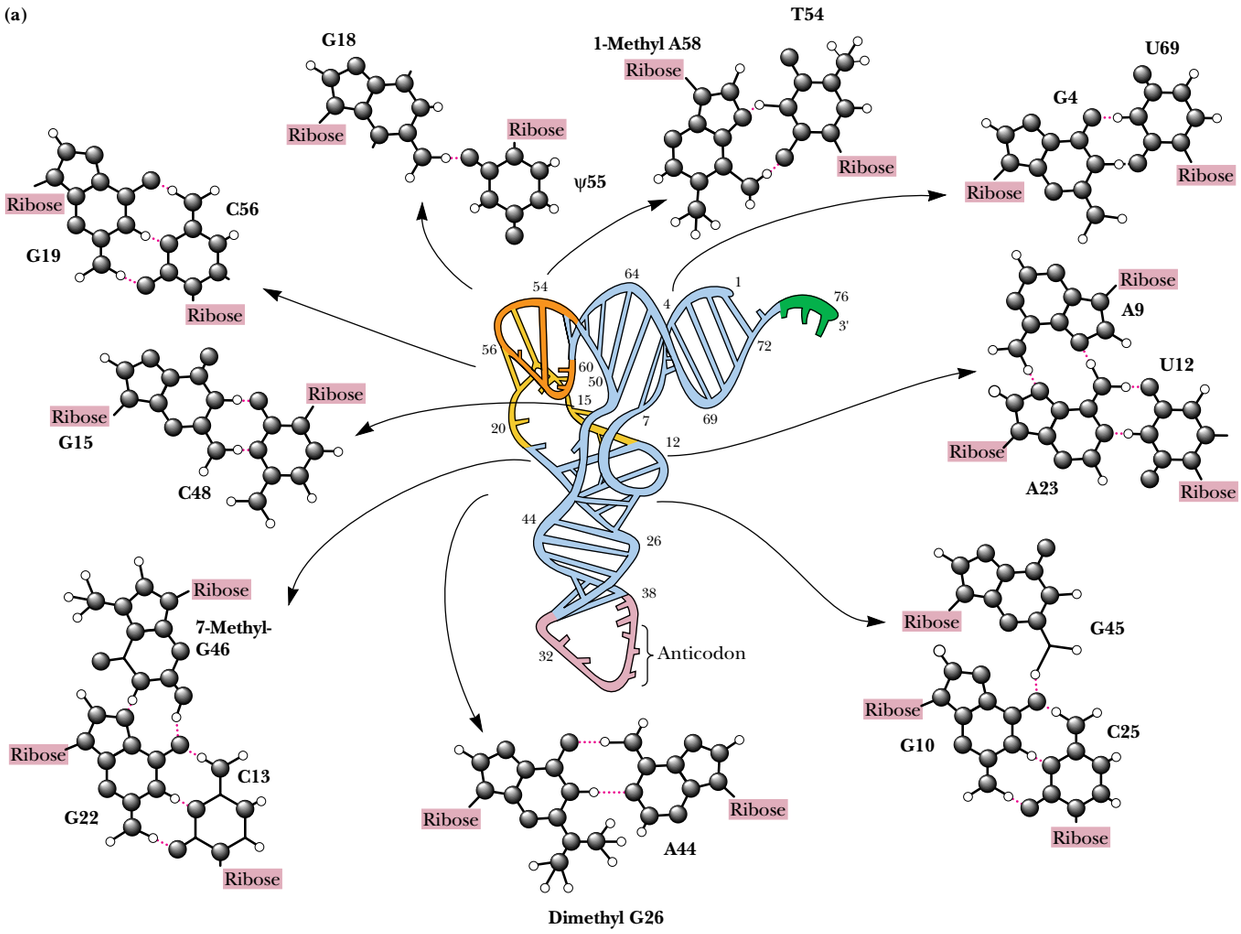
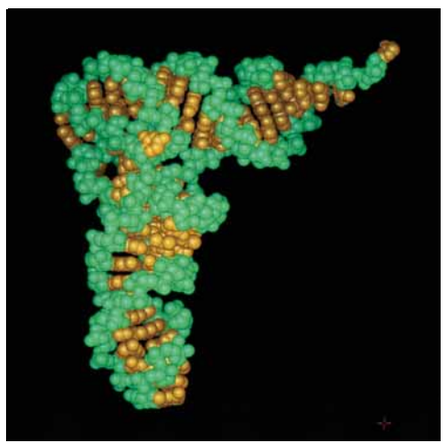


FIGURE 12.37 • Tertiary interactions in yeast phenylalanine tRNA. The molecule is presented in the conventional cloverleaf secondary structure generated by intrastrand hydrogen bonding. Solid lines connect bases that are hydrogen-bonded when this cloverleaf pattern is folded into the characteristic tRNA tertiary structure (see also Figure 12.36).





(b)

FIGURE 12.38 • (a) The three-dimensional structure of yeast phenylalanine tRNA as deduced from X-ray diffraction studies of its crystals. The tertiary folding is illustrated in the center of the diagram with the ribose–phosphate backbone presented as a continuous ribbon; H bonds are indicated by crossbars. Unpaired bases are shown as short, unconnected rods. The anticodon loop is at the bottom and the -CCA 3′-OH acceptor end is at the top right. The various types of noncanonical hydrogen-bonding interactions observed between bases surround the central molecule. Three of these structures show examples of unusual H-bonded interactions involving three bases; these interactions aid in establishing tRNA tertiary structure. (b) A space-filling model of the molecule. *(After Kim, S. H., in Schimmel, P., Söll, D., and Abelson, J. N., eds., 1979.* Transfer RNA: Structure, Properties, and Recognition. *New York: Cold Spring Harbor Laboratory.)*

plementarity is found in all rRNA strands, and all assume a highly folded pattern that allows base pairing between these complementary segments. Figure 12.39 shows the secondary structure assigned to the *E. coli* 16S rRNA. This structure is based on alignment of the nucleotide sequence into H-bonding segments. The reliability of these alignments is then tested through a comparative analysis of whether identical secondary structures can be predicted from

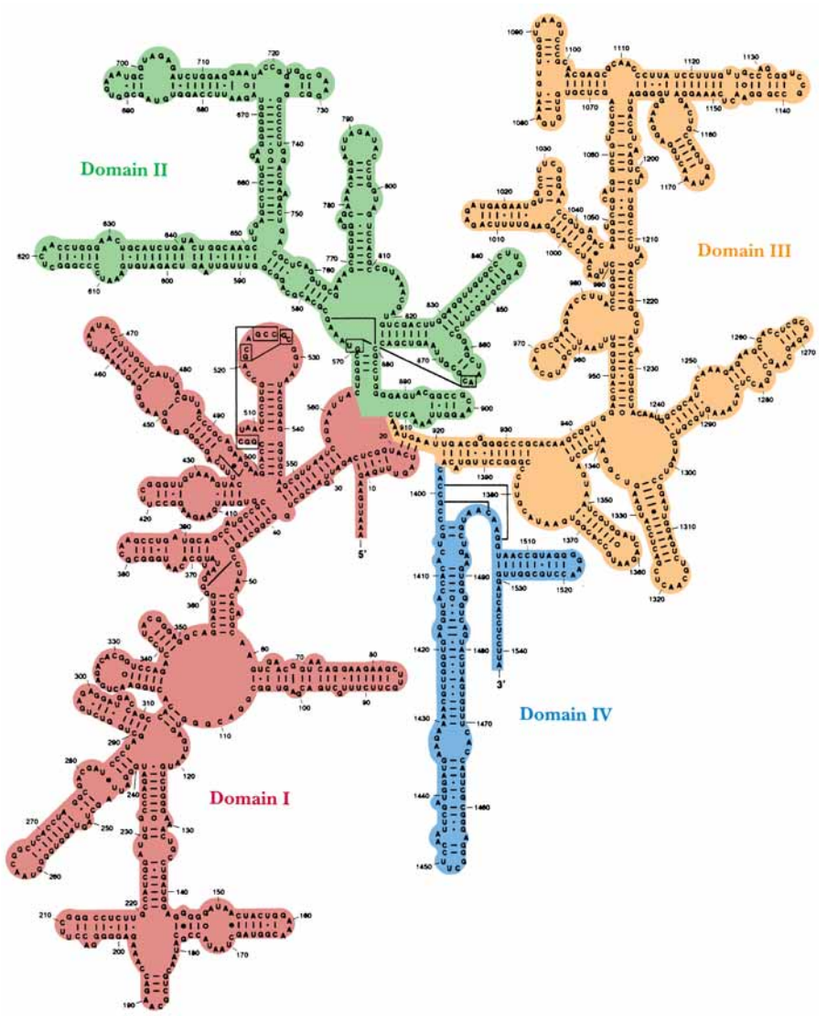


FIGURE 12.39 • The proposed secondary structure for *E. coli* 16S rRNA, based on comparative sequence analysis in which the folding pattern is assumed to be conserved across different species. The molecule can be subdivided into four domains—**I, II, III,** and **IV**—on the basis of contiguous stretches of the chain that are closed by long-range base-pairing interactions. **I,** the 5′-domain, includes nucleotides 27 through 556. **II,** the central domain, runs from nucleotide 564 to 912. Two domains comprise the 3′-end of the molecule. **III,** the major one, comprises nucleotides 923 to 1391. **IV,** the 3′-terminal domain, covers residues 1392 to 1541.

nucleotide sequences of 16S-like rRNAs from other species. If so, then such structures are apparently conserved. The approach is based on the thesis that, because ribosomal RNA species (regardless of source) serve common roles in protein synthesis, it may be anticipated that they share structural features. The structure is marvelously rich in short, helical segments separated and punctuated by single-stranded loops.

Comparison of rRNAs from Various Species

If a phylogenetic comparison is made of the 16S-like rRNAs from an archaebacterium (*Halobacterium volcanii*), a eubacterium (*E. coli*), and a eukaryote (the yeast *Saccharomyces cerevisiae*), a striking similarity in secondary structure emerges (Figure 12.40). Remarkably, these secondary structures are similar despite the fact that the nucleotide sequences of these rRNAs themselves exhibit a low degree of similarity. Apparently, evolution is acting at the level of rRNA secondary structure, not rRNA nucleotide sequence. Similar conserved folding patterns are seen for the 23S-like and 5S-like rRNAs that reside in the

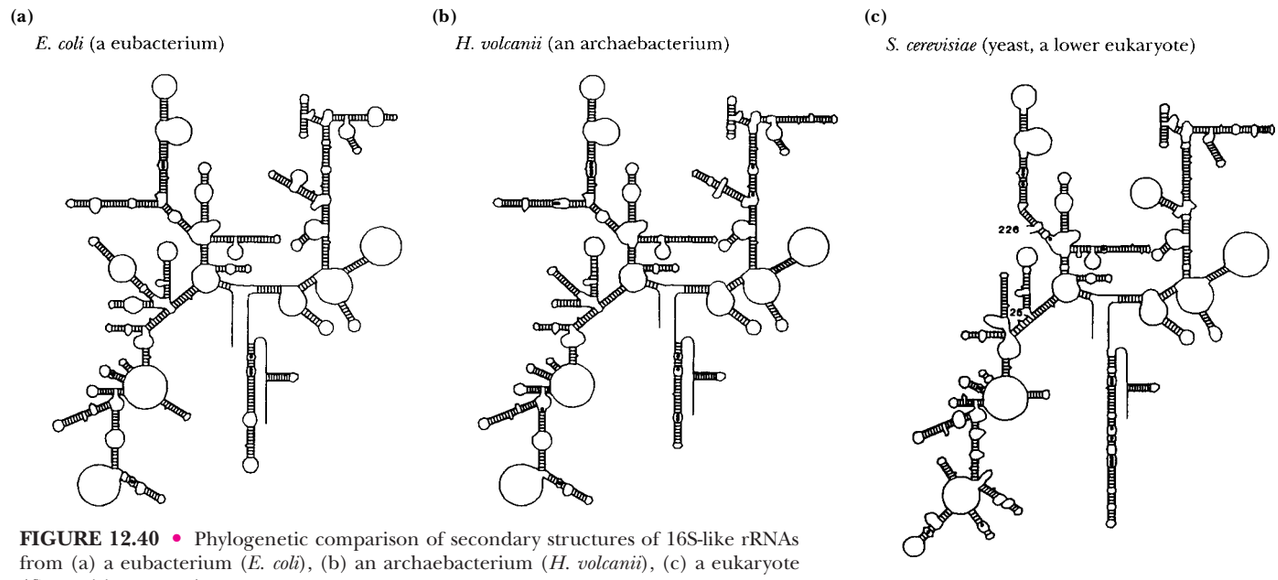


FIGURE 12.40 • Phylogenetic comparison of secondary structures of 16S-like rRNAs from (a) a eubacterium (*E. coli*), (b) an archaebacterium (*H. volcanii*), (c) a eukaryote (*S. cerevisiae,* a yeast).

large ribosomal subunits of various species. An insightful conclusion may be drawn regarding the persistence of such strong secondary structure conservation despite the millennia that have passed since these organisms diverged: *all ribosomes are constructed to a common design and all function in a similar manner.*

rRNA Tertiary Structure

Despite the unity in secondary structural patterns, little is known about the three-dimensional, or tertiary, structure of rRNAs. Even less is known about the quaternary interactions that occur when ribosomal proteins combine with rRNAs and when the ensuing ribonucleoprotein complexes, the small and large subunits, come together to form the complete ribosome. Furthermore, assignments of functional roles to rRNA molecules are still tentative and approximate. (We return to these topics in Chapter 33.)

Problems

1. The oligonucleotide d-ATGCCTGACT was subjected to sequencing by (a) Sanger's dideoxy method and (b) Maxam and Gilbert's chemical cleavage method, and the products were analyzed by electrophoresis on a polyacrylamide gel. Draw diagrams of the gel-banding patterns obtained for (a) and (b).

2. The result of sequence determination of an oligonucleotide as performed by the Sanger dideoxy chain termination method is displayed at right.

What is the sequence of the original oligonucleotide? A second sample of the oligonucleotide was 3′-end labeled with ^{32}P and then subjected to the Maxam–Gilbert chemical cleavage sequencing protocol. Draw a diagram depicting the pattern seen on the autoradiogram of the Maxam–Gilbert sequencing gel.

A	C	G	T

---			

3. X-ray diffraction studies indicate the existence of a novel double-stranded DNA helical conformation in which ΔZ (the rise per base pair) = 0.32 nm and P (the pitch) = 3.36 nm. What are the other parameters of this novel helix: (a) the number of base pairs per turn, (b) $\Delta\phi$ (the mean rotation per base pair), and (c) c (the true repeat)?

4. A 41.5-nm-long duplex DNA molecule in the B-conformation adopts the A-conformation upon dehydration. How long is it now? What is its approximate number of base pairs?

5. If 80% of the base pairs in a duplex DNA molecule (12.5 kbp) are in the B-conformation and 20% are in the Z-conformation, what is the length of the molecule?

6. A "relaxed," circular, double-stranded DNA molecule (1600 bp) is in a solution where conditions favor 10 bp per turn. What is the value of L_0 for this DNA molecule? Suppose DNA gyrase introduces 12 negative supercoils into this molecule. What are the values of L, W, and T now? What is the superhelical density, σ?

7. Suppose one double-helical turn of a superhelical DNA molecule changes conformation from B-form to Z-form. What are the changes in L, W, and T? Why do you suppose the transition of DNA from B-form to Z-form is favored by negative supercoiling?

8. There is one nucleosome for every 200 bp of eukaryotic DNA. How many nucleosomes are in a diploid human cell? Nucleosomes can be approximated as disks 11 nm in diameter and 6 nm long. If all the DNA molecules in a diploid human cell are in the B-conformation, what is the sum of their lengths? If this DNA is now arrayed on nucleosomes in the "beads-on-a-string" motif, what is its approximate total length?

9. The characteristic secondary structures of tRNA and rRNA molecules are achieved through intrastrand hydrogen bonding. Even for the small tRNAs, remote regions of the nucleotide sequence interact via H-bonding when the molecule adopts the cloverleaf pattern. Using Figure 12.34 as a guide, draw the primary structure of a tRNA and label the positions of its various self-complementary regions.

10. Using the data in Table 11.3, arrange the DNAs from the following sources in order of increasing T_m: human, salmon, wheat, yeast, *E. coli*.

11. The DNAs from mice and rats have (G+C) contents of 44% and 40%, respectively. Calculate the T_ms for these DNAs in 0.2 *M* NaCl. If samples of these DNAs were inadvertently mixed, how might they be separated from one another? Describe the procedure and the results (hint: see the Appendix to this chapter).

12. Calculate the density (ρ) of avian tubercle bacillus DNA from the data presented in Table 11.3 and the equation $\rho = 1.660 + 0.098(\text{GC})$, where (GC) is the mole fraction of (G+C) in DNA.

Further Reading

Adams, R. L. P., Knowler, J. T., and Leader, D. P., 1992. *The Biochemistry of the Nucleic Acids,* 11th ed. London: Chapman and Hall.

Arents, G., et al., 1991. The nucleosome core histone octamer at 3.1 Å resolution: A tripartite protein assembly and a left-hand superhelix. *Proceedings of the National Academy of Sciences U.S.A.* **88:**10148–10152.

Axelrod, N., 1996. Of telomeres and tumors. *Nature Medicine* **2:**158–159.

Callandine, C. R., and Drew, H. R., 1992. *Understanding DNA: The Molecule and How It Works.* London: Academic Press.

Ferretti, L., Karnik, S. S., Khorana, H. G., Nassal, M., and Oprian, D. D., 1986. Total synthesis of a gene for bovine rhodopsin. *Proceedings of the National Academy of Sciences U.S.A.* **83:**599–603.

Kornberg, A., and Baker, T. A., 1991. *DNA Replication,* 2nd ed. New York: W.H. Freeman and Co.

Luger, C., et al., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389:**251–260.

Noller, H. F., 1984. Structure of ribosomal RNA. *Annual Review of Biochemistry* **53:**119–162.

Pienta, K. J., and Coffey, D. S., 1984. A structural analysis of the role of the nuclear matrix and DNA loops in the organization of the nucleus and chromosomes. In Cook, P. R., and Laskey, R. A., eds., Higher Order Structure in the Nucleus. *Journal of Cell Science* Supplement **1:**123–135.

Rhodes, D., 1997. The nucleosome core all wrapped up. *Nature* **389:**231–233.

Rich, A., Nordheim, A., and Wang, A. H.-J., 1984. The chemistry and biology of left-handed Z-DNA. *Annual Review of Biochemistry* **53:**791–846.

Wand, B.-C., et al., 1994. The octameric histone core of the nucleosome. *Journal of Molecular Biology* **236:**179–188.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., and Weiner, A. M., 1987. *The Molecular Biology of the Gene,* Vol. I, *General Principles,* 4th ed. Menlo Park, CA: Benjamin/Cummings.

Watson, J. D., ed., 1983. Structures of DNA. *Cold Spring Harbor Symposia on Quantitative Biology,* Volume XLVII. New York: Cold Spring Harbor Laboratory.

Wu, R., 1993. Development of enzyme-based methods for DNA sequence analysis and their application in genome projects. *Methods in Enzymology* **67:**431–468.