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

11.1 • Nitrogenous Bases
11.2 • The Pentoses of Nucleotides and Nucleic Acids
11.3 • Nucleosides Are Formed by Joining a Nitrogenous Base to a Sugar
11.4 • Nucleotides Are Nucleoside Phosphates
11.5 • Nucleic Acids Are Polynucleotides
11.6 • Classes of Nucleic Acids
11.7 • Hydrolysis of Nucleic Acids

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,

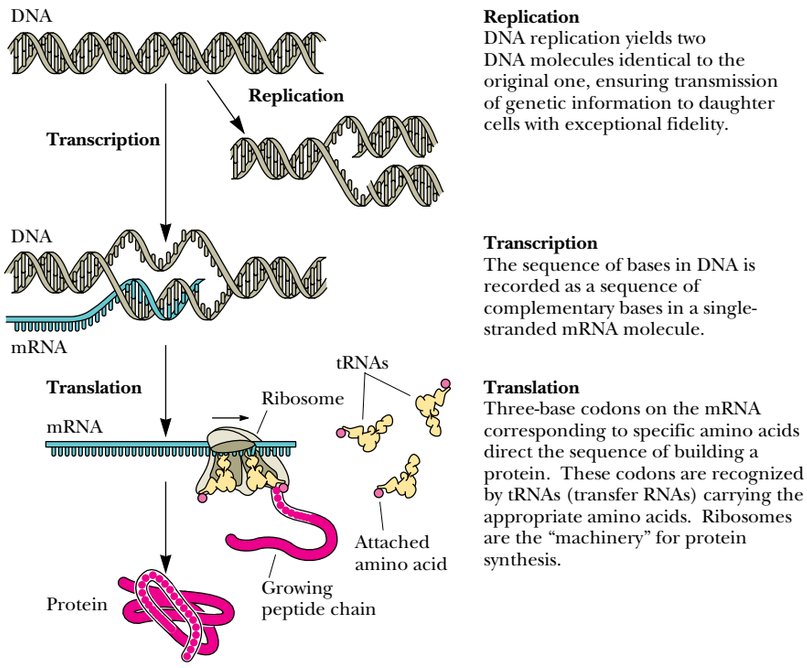
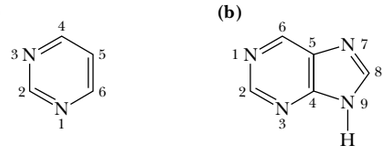


FIGURE 11.1 • The fundamental process of information transfer in cells. Information encoded in the nucleotide sequence of DNA is transcribed through synthesis of an RNA molecule whose sequence is dictated by the DNA sequence. As the sequence of this RNA is read (as groups of three consecutive nucleotides) 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

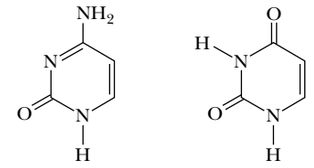
The bases of nucleotides and nucleic acids are derivatives of either **pyrimidine** or **purine.** Pyrimidines are six-membered heterocyclic aromatic rings containing two nitrogen atoms (Figure 11.2a). The atoms are numbered in a clockwise fashion, as shown in the figure. The purine ring structure is 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

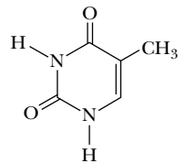
The common naturally occurring pyrimidines are **cytosine, uracil,** and **thymine** (5-methyluracil) (Figure 11.3). Cytosine and thymine are the pyrimidines typically found in DNA, whereas cytosine and uracil are common in RNA. To view this generality another way, the uracil component of DNA occurs 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.



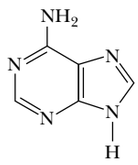
Cytosine (2-oxy-4-amino pyrimidine)

Uracil (2-oxy-4-oxy pyrimidine)

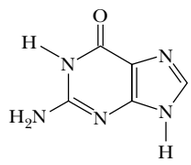


Thymine (2-oxy-4-oxy 5-methyl pyrimidine)

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

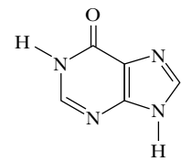


Adenine (6-amino purine)

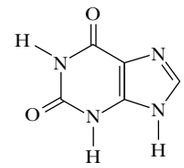


Guanine (2-amino-6-oxy purine)

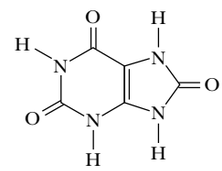
FIGURE 11.4 • The common purine bases—adenine and guanine—in the tautomeric forms predominant at pH 7.



Hypoxanthine



Xanthine



Uric acid

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

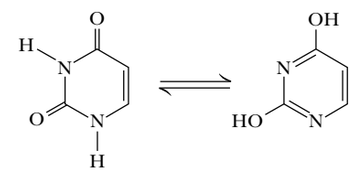
Properties of Pyrimidines and Purines

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.

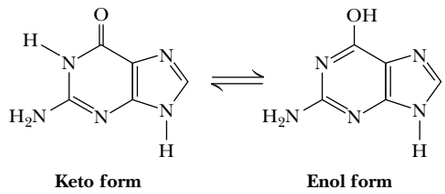


FIGURE 11.7 • The tautomerism of the purine, guanine.

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.

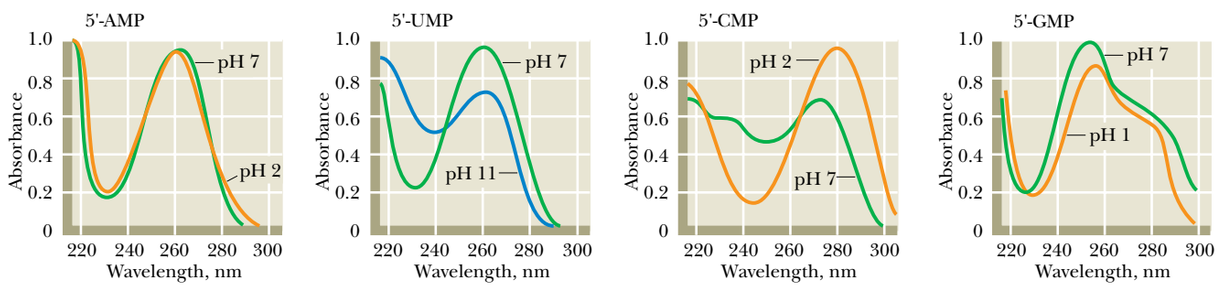


FIGURE 11.8 • The UV absorption spectra of the common ribonucleotides.

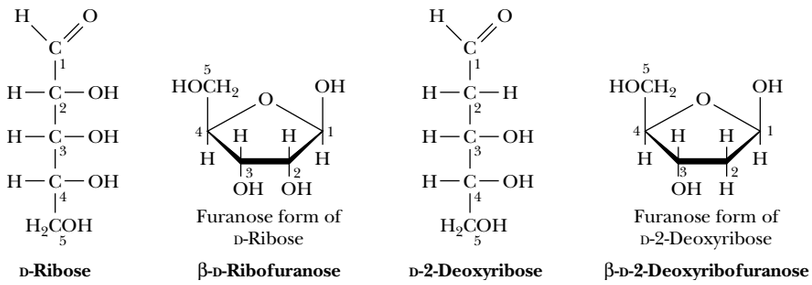
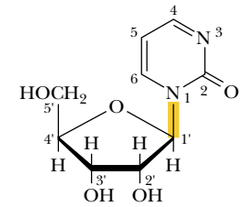


FIGURE 11.9 • Furanose structures—ribose and deoxyribose.

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.**



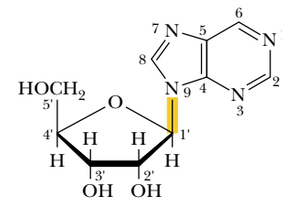


FIGURE 11.10 • β-Glycosidic bonds link nitrogenous bases and sugars to form nucleosides.

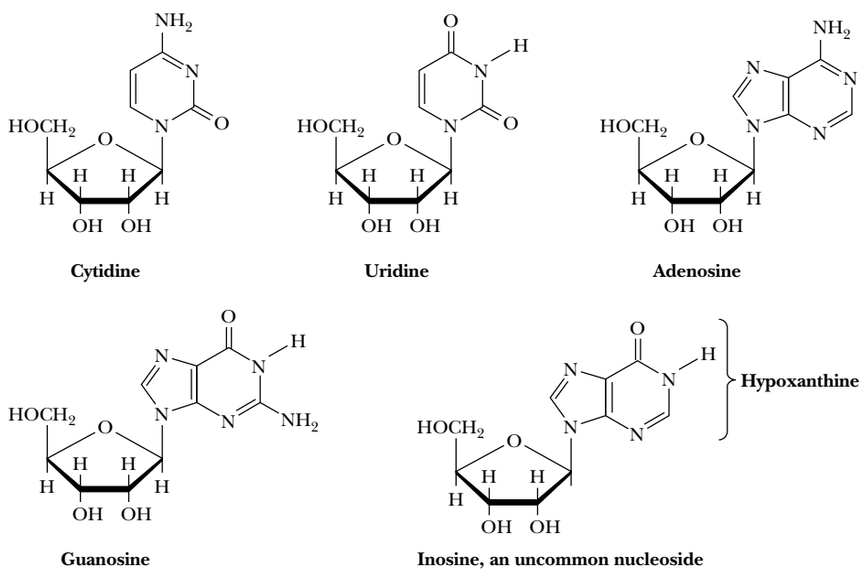
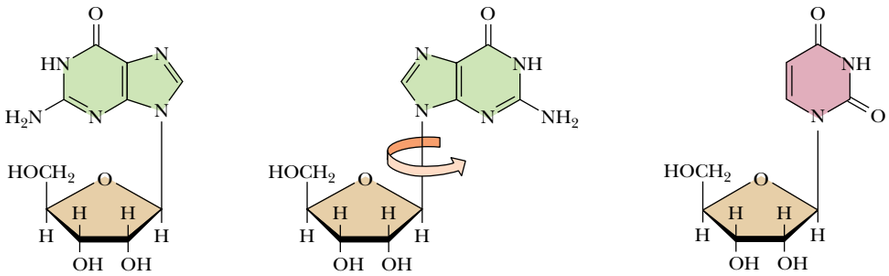


FIGURE 11.11 • The common ribonucleosides—cytidine, uridine, adenosine, and guanosine. Also, inosine drawn in anti conformation.



***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

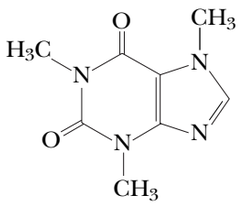
In nucleosides, rotation of the base about the glycosidic bond is sterically hindered, principally by the hydrogen atom on the C-2′ carbon of the furanose. (This hindrance is most easily seen and appreciated by manipulating accurate molecular models of these structures.) Consequently, nucleosides and 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,**

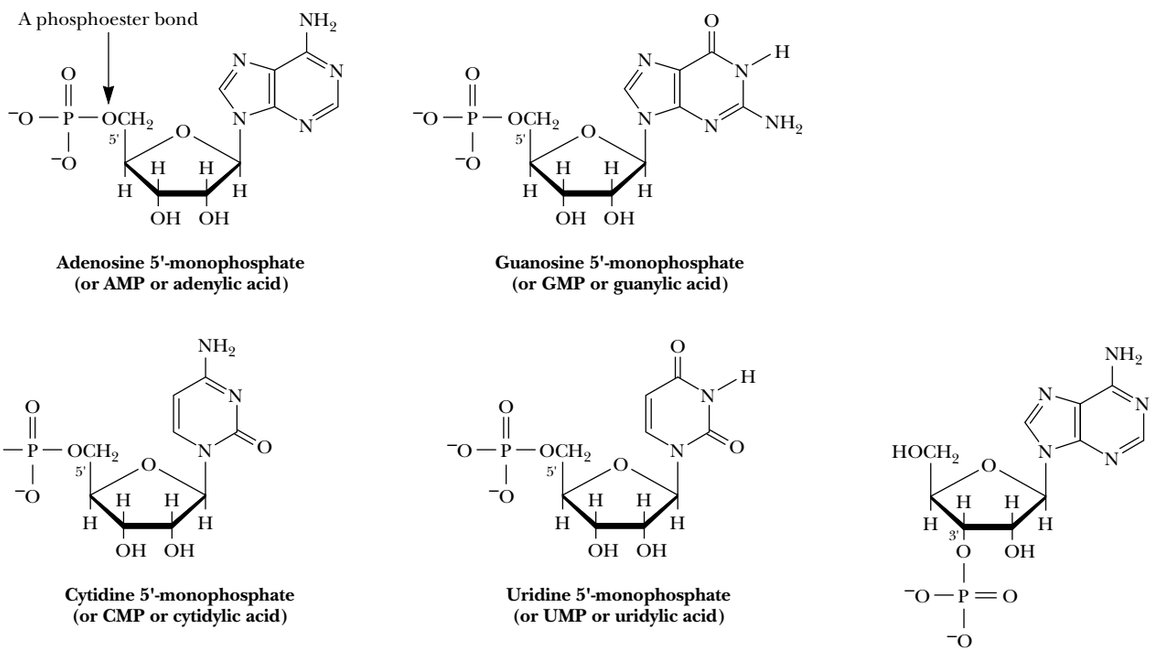
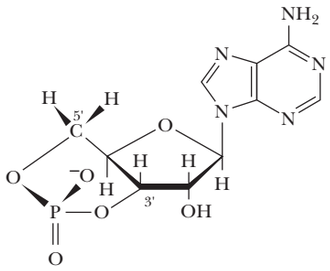
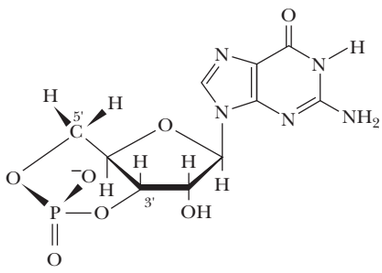


FIGURE 11.13 • Structures of the four common ribonucleotides—AMP, GMP, CMP, and UMP—together with their two sets of full names, for example, adenosine 5′-monophosphate and adenylic acid. Also shown is the nucleoside 3′-AMP.



3',5'-Cyclic AMP



3',5'-Cyclic GMP

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

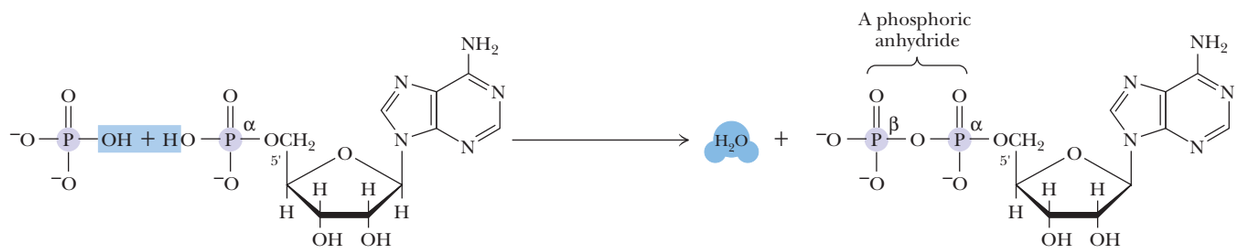
cytidylic acid, and **uridylic acid.** The pK_a value for the second dissociation, pK_2, is about 6.0, so at neutral pH or above, the net charge on a nucleoside monophosphate is −2. Nucleic acids, which are polymers of nucleoside monophosphates, derive their name from the acidity of these phosphate groups.

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.



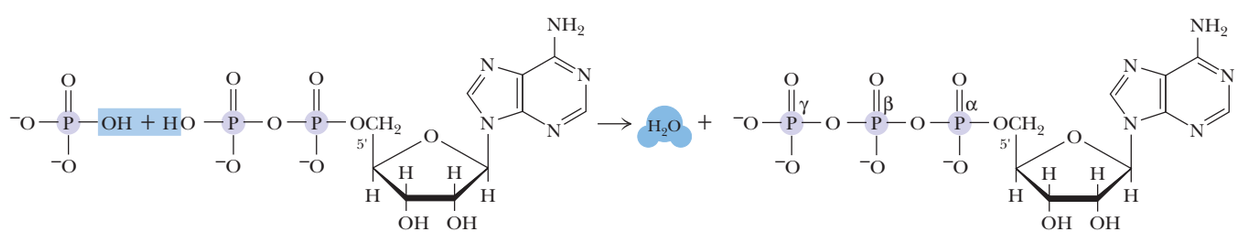


FIGURE 11.15 • Formation of ADP and ATP by the successive addition of phosphate groups via phosphoric anhydride linkages. Note the removal of equivalents of H_2O in these dehydration synthesis reactions.

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*

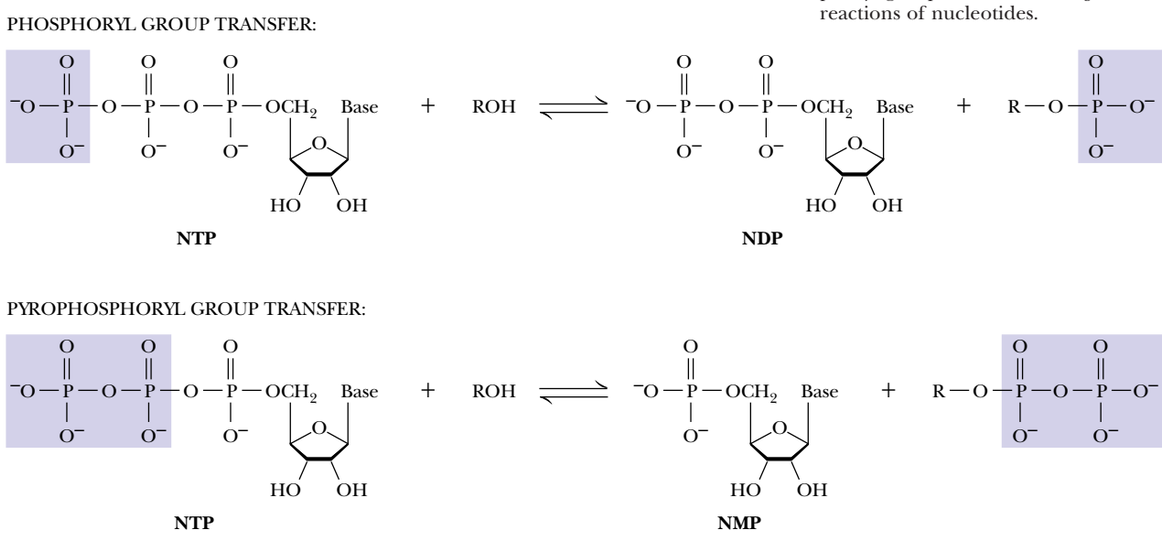


FIGURE 11.16 • Phosphoryl and pyrophosphoryl group transfer, the major biochemical reactions of nucleotides.

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

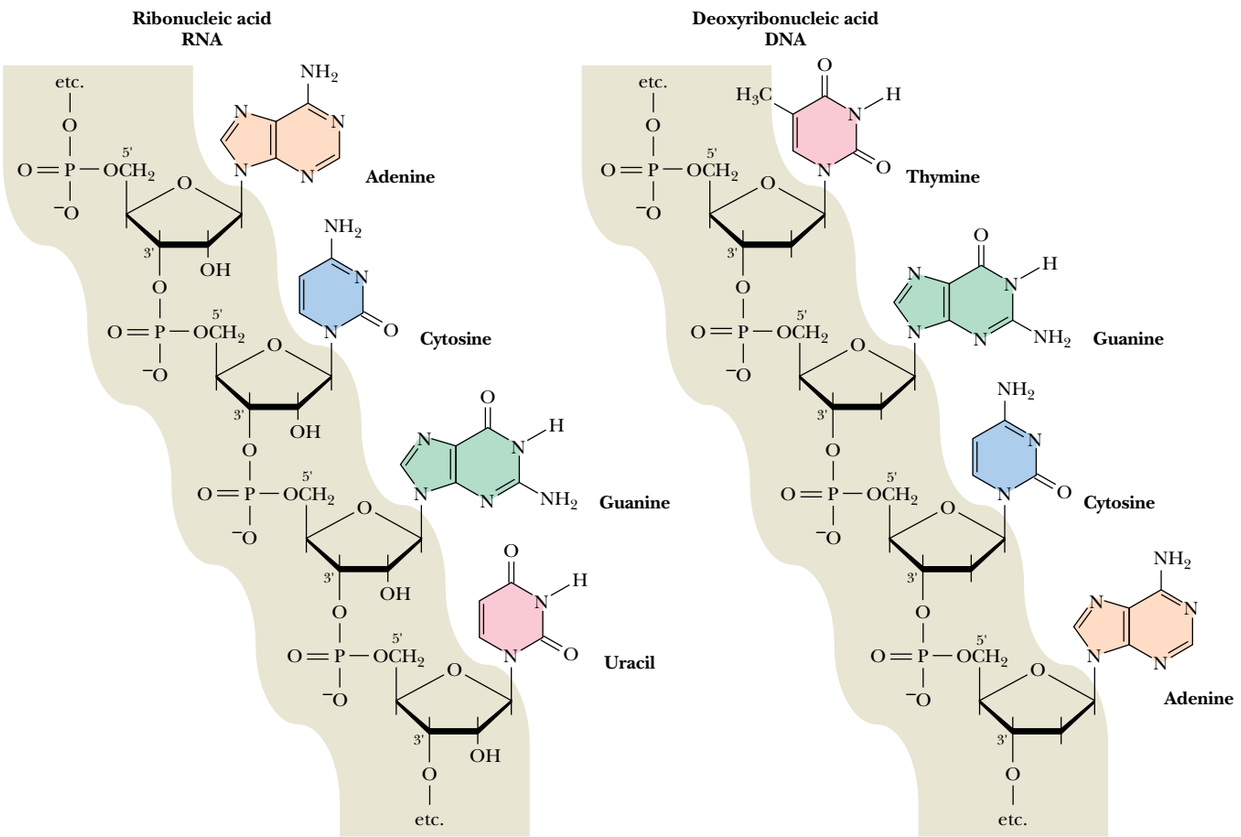


FIGURE 11.17 • 3′-5′ phosphodiester bridges link nucleotides together to form polynucleotide chains.

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.

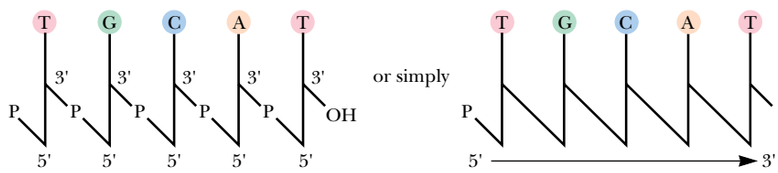


FIGURE 11.18 • Furanoses are represented by lines; phosphodiesters are represented by diagonal slashes in this shorthand notation for nucleic acid structures.

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.

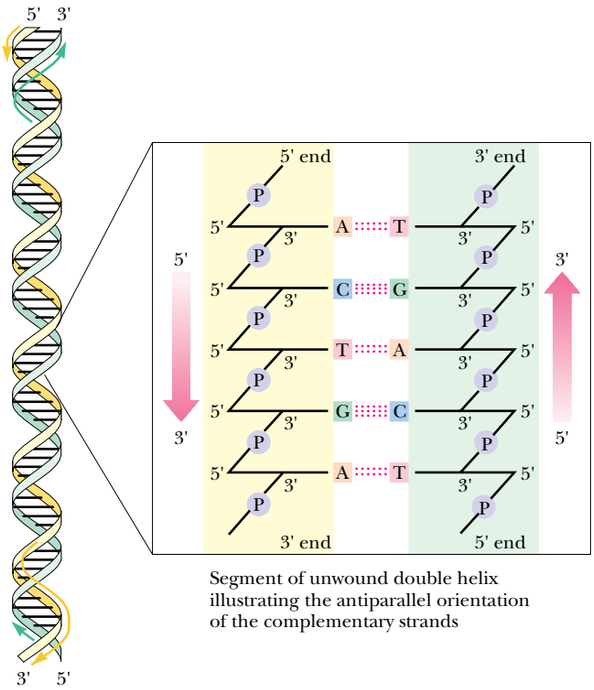
DNA

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.

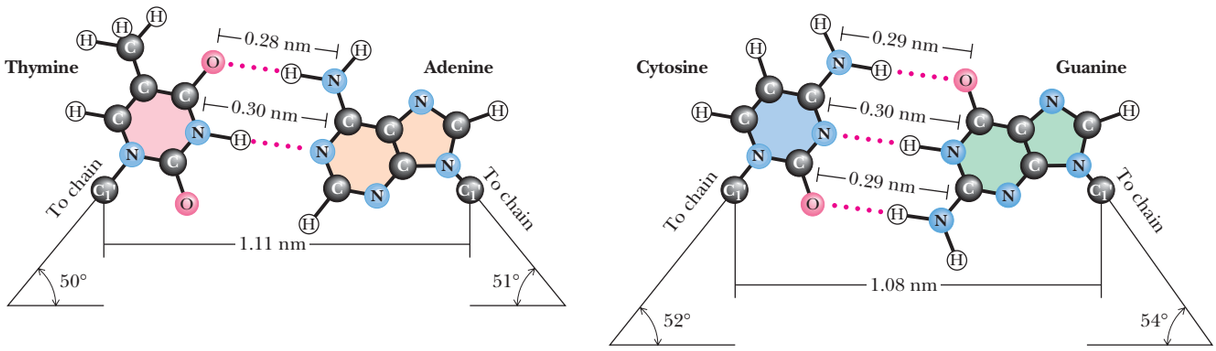


FIGURE 11.20 • The Watson–Crick base pairs A : T and G : C.

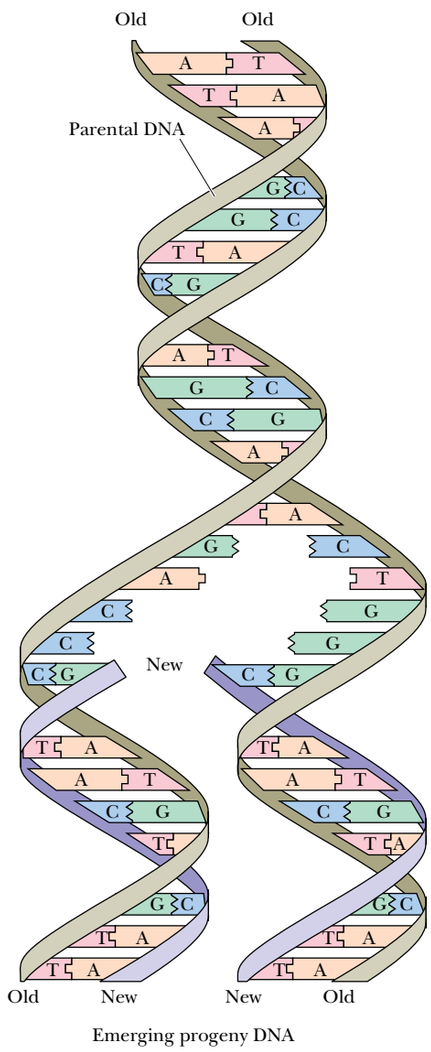


FIGURE 11.21 • Replication of DNA gives identical progeny molecules because base pairing is the mechanism determining the nucleotide sequence synthesized within each of the new strands during replication.

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!"