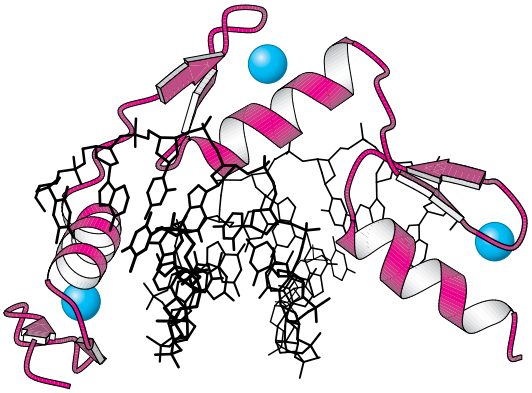
CHAPTER

4

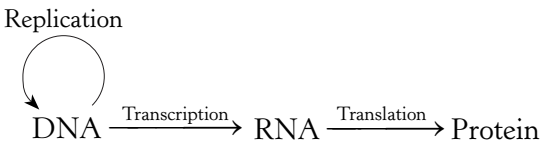
DNA, RNA, and the Flow of Genetic Information



Family resemblance, very evident in this photograph of four sisters, results from having genes in common. Genes must be expressed to exert an effect, and proteins regulate such expression. One such regulatory protein, a zinc-finger protein (zinc ion is blue, protein is red), is shown bound to a control region of DNA (black). [(Left) © Nicholas Nixon, courtesy Fraenkel Gallery, San Francisco. (Right) Drawn from 1AAY.pdb.]

DNA and RNA are long linear polymers, called nucleic acids, that carry information in a form that can be passed from one generation to the next. These macromolecules consist of a large number of linked nucleotides, each composed of a sugar, a phosphate, and a base. Sugars linked by phosphates form a common backbone that plays a structural role, whereas the sequence of bases along a nucleic acid strand carries genetic information. The DNA molecule has the form of a *double helix*, a helical structure consisting of two complementary nucleic acid strands. Each strand serves as the template for the other in DNA replication. The genes of all cells and many viruses are made of DNA.

Genes specify the kinds of proteins that are made by cells, but DNA is not the direct template for protein synthesis. Rather, a DNA strand is copied into a class of RNA molecules called messenger RNA (mRNA), the information-carrying intermediates in protein synthesis. This process of transcription is followed by translation, the synthesis of proteins according to instructions given by mRNA templates. Information processing in all cells is quite complex. The scheme that underlies information processing at the level of gene expression was first proposed by Francis Crick in 1958.



Crick called this scheme the *central dogma*. The basic tenets of this dogma are true, but, as we will see later, this scheme is not as simple as depicted.

OUTLINE

This flow of information depends on the genetic code, which defines the relation between the sequence of bases in DNA (or its mRNA transcript) and the sequence of amino acids in a protein. The code is nearly the same in all organisms: a sequence of three bases, called a *codon,* specifies an amino acid. There is another step in the expression of most eukaryotic genes, which are mosaics of nucleic acid sequences called *introns* and *exons*. Both are transcribed, but before translation takes place, introns are cut out of newly synthesized RNA molecules, leaving mature RNA molecules with continuous exons. The existence of introns and exons has crucial implications for the evolution of proteins.

4.1 A Nucleic Acid Consists of Four Kinds of Bases Linked to a Sugar–Phosphate Backbone

The nucleic acids DNA and RNA are well suited to function as the carriers of genetic information by virtue of their covalent structures. These macromolecules are *linear polymers* built up from similar units connected end to end (Figure 4.1). Each monomer unit within the polymer is a *nucleotide*. A single nucleotide unit consists of three components: a sugar, a phosphate, and one of four bases. *The sequence of bases in the polymer uniquely characterizes a nucleic acid and constitutes a form of linear information*—information analogous to the letters that spell a person's name.

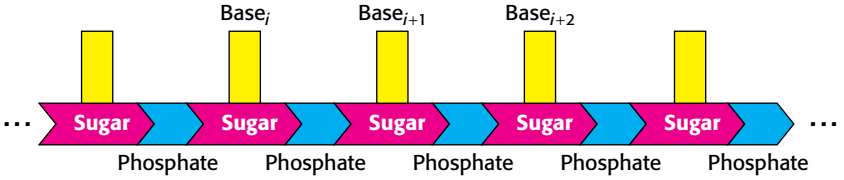


FIGURE 4.1 Polymeric structure of nucleic acids.

RNA and DNA differ in the sugar component and one of the bases

The sugar in *deoxyribonucleic acid* (DNA) is *deoxyribose*. The prefix deoxy indicates that the 2′-carbon atom of the sugar lacks the oxygen atom that is linked to the 2′-carbon atom of *ribose,* as shown in Figure 4.2. Note that sugar carbons are numbered with primes to differentiate them from atoms in the bases. The sugars in both nucleic acids are linked to one another by phosphodiester bridges. Specifically, the 3′-hydroxyl (3′-OH) group of the sugar moiety of one nucleotide is esterified to a phosphate group, which is, in turn, joined to the 5′-hydroxyl group of the adjacent sugar. The chain of sugars linked by phosphodiester bridges is referred to as the *backbone* of the nucleic acid (Figure 4.3). Whereas the backbone is constant in a nucleic

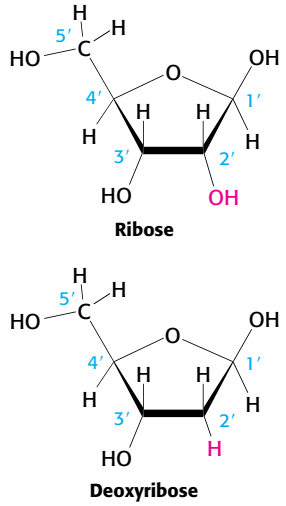


FIGURE 4.2 Ribose and deoxyribose. Atoms in sugar units are numbered with primes to distinguish them from atoms in bases (see Figure 4.4).

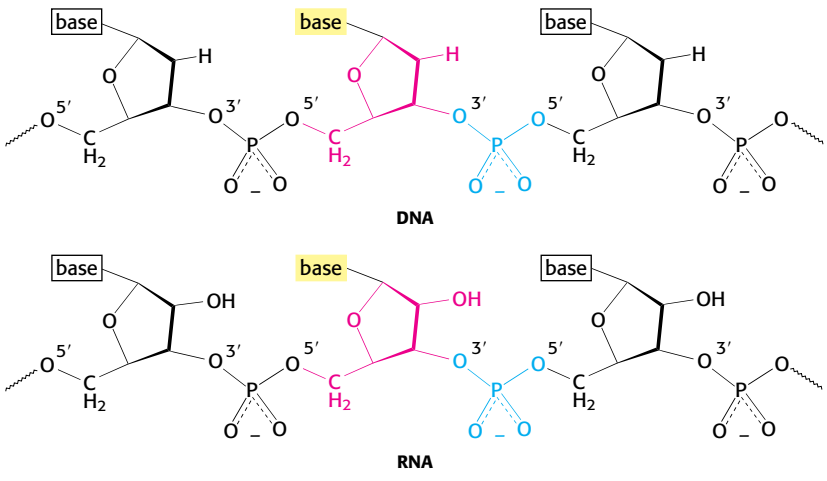


FIGURE 4.3 Backbones of DNA and RNA. The backbones of these nucleic acids are formed by 3′-to-5′ phosphodiester linkages. A sugar unit is highlighted in red and a phosphate group in blue.

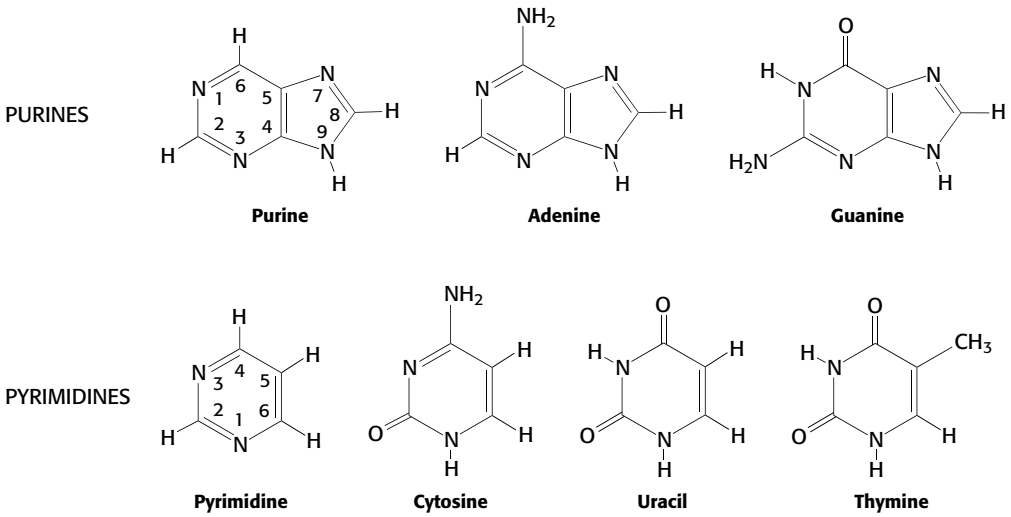


FIGURE 4.4 Purines and pyrimidines. Atoms within bases are numbered without primes. Uracil is present in RNA instead of thymine.

acid, the bases vary from one monomer to the next. Two of the bases of DNA are derivatives of *purine*—adenine (A) and guanine (G)—and two of *pyrimidine*—cytosine (C) and thymine (T), as shown in Figure 4.4.

Ribonucleic acid (RNA), like DNA, is a long unbranched polymer consisting of nucleotides joined by 3′-to-5′ phosphodiester linkages (Figure 4.3). The covalent structure of RNA differs from that of DNA in two respects. First, the sugar units in RNA are riboses rather than deoxyriboses. Ribose contains a 2′-hydroxyl group not present in deoxyribose. Second, one of the four major bases in RNA is uracil (U) instead of thymine (T).

Note that each phosphodiester bridge has a negative charge. This negative charge repels nucleophilic species such as hydroxide ions, which are capable of hydrolytic attack on the phosphate backbone. This resistance is crucial for maintaining the integrity of information stored in nucleic acids. The absence of the 2′-hydroxyl group in DNA further increases its resistance to hydrolysis. The greater stability of DNA probably accounts for its use rather than RNA as the hereditary material in all modern cells and in many viruses.

Nucleotides are the monomeric units of nucleic acids

The building blocks of nucleic acids and the precursors of these building blocks play many other roles throughout the cell—for instance, as energy currency and as molecular signals. Consequently, it is important to be familiar with the nomenclature of nucleotides and their precursors. A unit consisting of a base bonded to a sugar is referred to as a *nucleoside*. The four nucleoside units in RNA are called *adenosine, guanosine, cytidine*, and *uridine*, whereas those in DNA are called *deoxyadenosine, deoxyguanosine, deoxycytidine*, and *thymidine*. In each case, N-9 of a purine or N-1 of a pyrimidine is attached to C-1′ of the sugar by an *N*-glycosidic linkage (Figure 4.5). The base lies above the plane of the sugar when the structure is written in the standard orientation; that is, the configuration of the *N*-glycosidic linkage is β (Section 11.1). Note that thymidine contains deoxyribose; by convention, the prefix deoxy is not added because thymine-containing nucleosides are only rarely found in RNA.

A *nucleotide* is a nucleoside joined to one or more phosphoryl groups by an ester linkage. *Nucleotide triphosphates, nucleosides joined to three phosphoryl groups, are the monomers—the building blocks—that are linked to form RNA and DNA*. The four nucleotide units that link to form DNA are nucleotide monophosphates called *deoxyadenylate, deoxyguanylate, deoxycytidylate*, and *thymidylate*. Similarly, the most common nucleotides

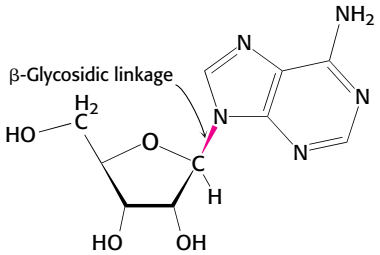


FIGURE 4.5 β-Glycosidic linkage in a nucleoside.

that link to form RNA are nucleotide monophosphates adenylate, guanylate, cytidylate and uridylate.

This nomenclature does not describe the number of phosphoryl groups or the site of attachment to carbon of the ribose. A more precise nomenclature is also commonly used. A compound formed by the attachment of a phosphoryl group to C-5′ of a nucleoside sugar (the most common site of phosphate esterification) is called a *nucleoside 5′-phosphate* or a *5′-nucleotide*. In this naming system for nucleotides, the number of phosphoryl groups and the attachment site are designated. Look, for example, at *adenosine 5′-triphosphate* (ATP; Figure 4.6). This nucleotide is tremendously important because, in addition to being a building block for RNA, it is the most commonly used energy currency. The energy released from cleavage of the triphosphate group is used to power many cellular processes (Chapter 15). Another nucleotide is deoxyguanosine 3′-monophosphate (3′-dGMP; Figure 4.6). This nucleotide differs from ATP in that it contains guanine rather than adenine, deoxyribose rather than ribose (indicated by the prefix "d"), and one rather than three phosphoryl groups. In addition, the phosphoryl group is esterified to the hydroxyl group in the 3′ rather than the 5′ position.

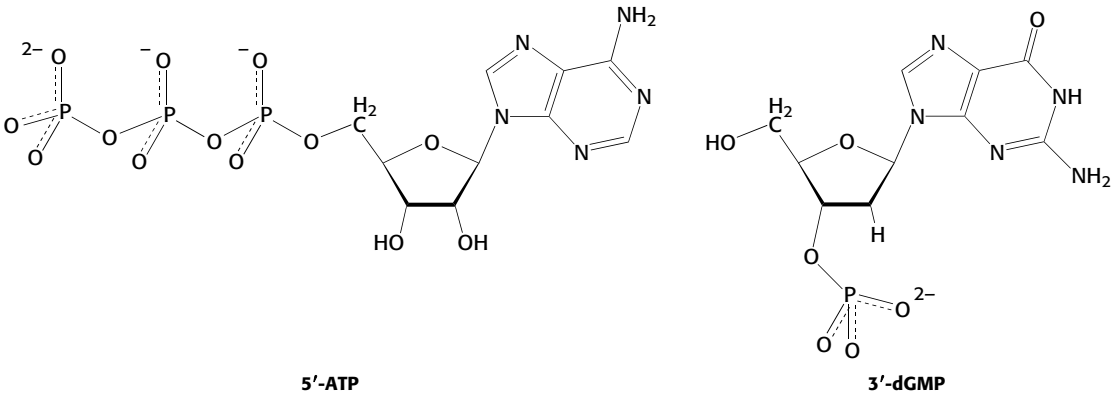


FIGURE 4.6 Nucleotides adenosine 5′-triphosphate (5′-ATP) and deoxyguanosine 3′-monophosphate (3′-dGMP).

DNA molecules are very long and have directionality

Scientific communication frequently requires the sequence of a nucleic acid—in some cases, a sequence thousands of nucleotides in length—to be written like that on page 17. Rather than writing the cumbersome chemical structures, scientists have adopted the use of abbreviations. The abbreviated notations pApCpG or ACG denote a trinucleotide of DNA consisting of the building blocks deoxyadenylate monophosphate, deoxycytidylate monophosphate, and deoxyguanylate monophosphate linked by a phosphodiester bridge, where "p" denotes a phosphoryl group (Figure 4.7). The 5′ end will often have a phosphoryl group attached to the 5′-OH group. Note that, like a polypeptide (Section 2.2), *a DNA chain has directionality,* commonly called polarity. One end of the chain has a free 5′-OH group (or a 5′-OH group attached to a phosphoryl group) and the other end has a free 3′-OH group, neither of which is linked to another nucleotide. By convention, *the base sequence is written in the 5′-to-3′ direction*. Thus, ACG indicates that the unlinked 5′-OH group is on deoxyadenylate, whereas the unlinked 3′-OH group is on deoxyguanylate. Because of this polarity, ACG and GCA correspond to different compounds.

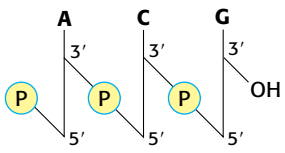


FIGURE 4.7 Structure of a DNA strand. The strand has a 5′ end, which is usually attached to a phosphoryl group, and a 3′ end, which is usually a free hydroxyl group.

A striking characteristic of naturally occurring DNA molecules is their length. A DNA molecule must comprise many nucleotides to carry the genetic information necessary for even the simplest organisms. For example, the DNA of a virus such as polyoma, which can cause cancer in certain organisms, consists of two paired strands of DNA, each 5100 nucleotides in length. The *E. coli* genome is a single DNA molecule consisting of two strands of 4.6 million nucleotides each (Figure 4.8).

The DNA molecules of higher organisms can be much larger. The human genome comprises approximately 3 billion nucleotides in each strand of DNA, divided among 24 distinct molecules of DNA called chromosomes (22 autosomal chromosomes plus the X and Y sex chromosomes) of different sizes. One of the largest known DNA molecules is found in the Indian muntjac, an Asiatic deer; its genome is nearly as large as the human genome but is distributed on only 3 chromosomes (Figure 4.9). The largest of these chromosomes has two strands of more than 1 billion nucleotides each. If such a DNA molecule could be fully extended, it would stretch more than 1 foot in length. Some plants contain even larger DNA molecules.

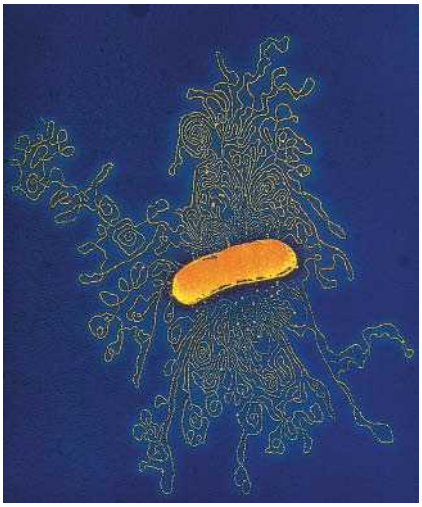


FIGURE 4.8 Electron micrograph of part of the *E. coli* genome. [Dr. Gopal Murti/ Science Photo Library/Photo Researchers.]

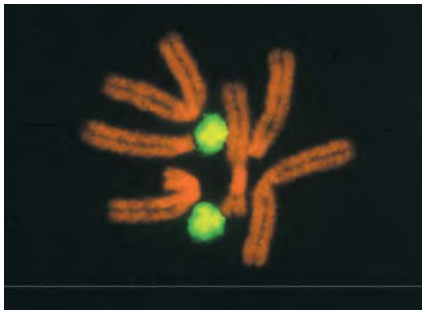


FIGURE 4.9 The Indian muntjac and its chromosomes. Cells from a female Indian muntjac (right) contain three pairs of very large chromosomes (stained orange). The cell shown is a hybrid containing a pair of human chromosomes (stained green) for comparison. [(Left) Hugh Lansdown/Shutterstock. (Right) J.–Y. Lee, M. Koi, E. J. Stanbridge, M. Oshimura, A. T. Kumamoto, and A. P. Feinberg. *Nat. Genet.* 7:30, 1994.]

4.2 A Pair of Nucleic Acid Strands with Complementary Sequences Can Form a Double-Helical Structure

As discussed in Chapter 1, the covalent structure of nucleic acids accounts for their ability to carry information in the form of a sequence of bases along a nucleic acid strand. The bases on the two separate nucleic acid strands form *specific base pairs* in such a way that a helical structure is formed. The double-helical structure of DNA facilitates the *replication* of the genetic material—that is, the generation of two copies of a nucleic acid from one.

The double helix is stabilized by hydrogen bonds and van der Waals interactions

The ability of nucleic acids to form specific base pairs was discovered in the course of studies directed at determining the three-dimensional structure of DNA. Maurice Wilkins and Rosalind Franklin obtained x-ray diffraction photographs of fibers of DNA (Figure 4.10). The characteristics of these diffraction patterns indicated that DNA is formed of two strands that wind in a regular helical structure. From these data and others, James Watson and Francis Crick deduced a structural model for DNA that accounted for the diffraction pattern

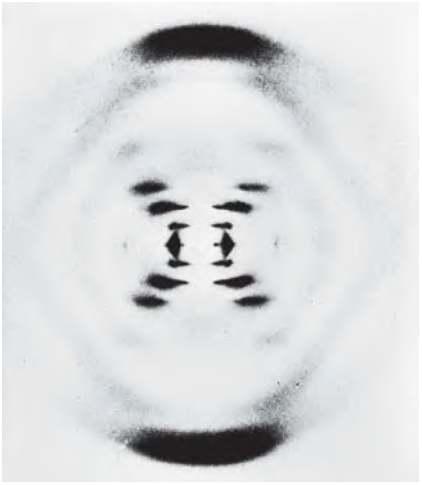


FIGURE 4.10 X-ray diffraction photograph of a hydrated DNA fiber. When crystals of a biomolecule are irradiated with x-rays, the x-rays are diffracted and these diffracted x-rays are seen as a series of spots, called reflections, on a screen behind the crystal. The structure of the molecule can be determined by the pattern of the reflections (Section 3.5). In regard to DNA crystals, the central cross is diagnostic of a helical structure. The strong arcs on the meridian arise from the stack of nucleotide bases, which are 3.4 Å apart. [Science Photo Library.]

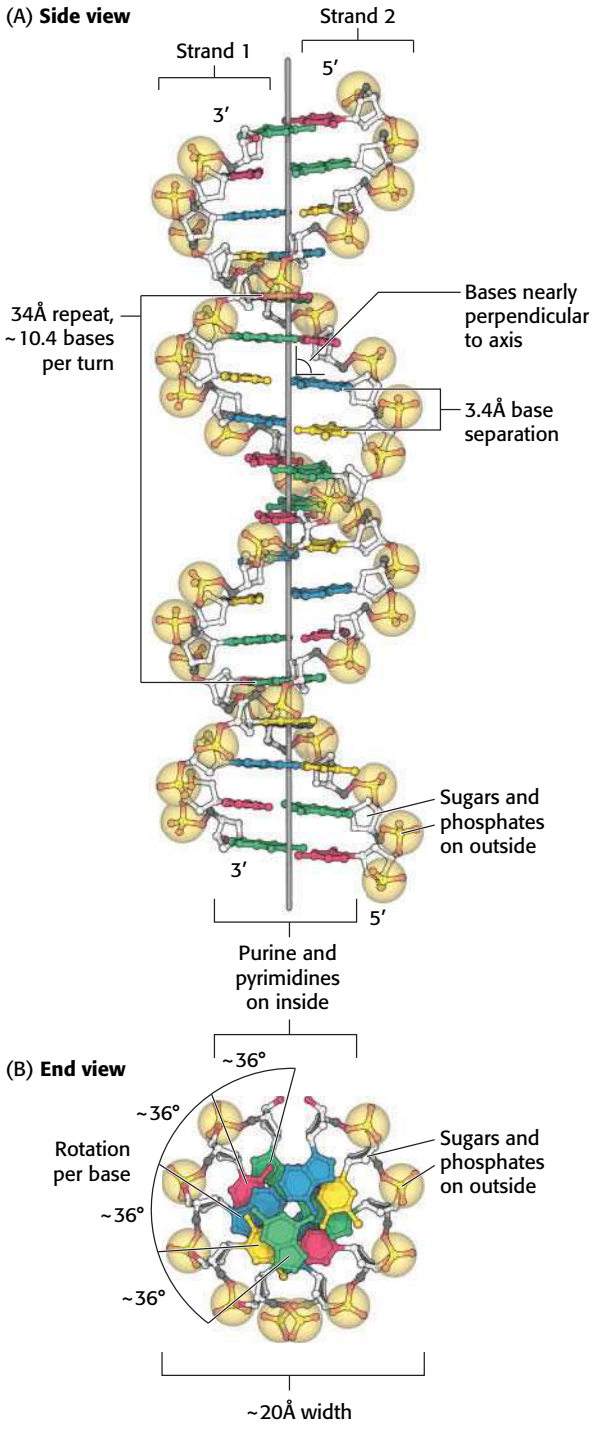


FIGURE 4.11 Watson–Crick model of double-helical DNA. (A) Side view. Adjacent bases are separated by 3.4 Å. The structure repeats along the helical axis (vertical) at intervals of 34 Å, which corresponds to approximately 10 nucleotides on each chain. (B) Axial view, looking down the helix axis, reveals a rotation of 36° per base and shows that the bases are stacked on top of one another [Source: J. L. Tymoczko, J. Berg, and L. Stryer, *Biochemistry*: *A Short Course*, 2nd ed. (W. H. Freeman and Company, 2013), Fig. 33.11.].

and was the source of some remarkable insights into the functional properties of nucleic acids (Figure 4.11).

The features of the Watson–Crick model of DNA deduced from the diffraction patterns are:

1. Two helical polynucleotide strands are coiled around a common axis with a right-handed screw sense (pp. 40-41). The strands are antiparallel, meaning that they have opposite directionality.
2. The sugar–phosphate backbones are on the outside and the purine and pyrimidine bases lie on the inside of the helix.
3. The bases are nearly perpendicular to the helix axis, and adjacent bases are separated by approximately 3.4 Å. The helical structure repeats on the order of every 34 Å, with about 10.4 bases per turn of helix. There is a rotation of nearly 36 degrees per base (360 degrees per full turn/10.4 bases per turn).
4. The diameter of the helix is about 20 Å.

How is such a regular structure able to accommodate an arbitrary sequence of bases, given the different sizes and shapes of the purines and pyrimidines? In attempting to answer this question, Watson and Crick discovered that guanine can be paired with cytosine and adenine with thymine to form base pairs that have essentially the same shape (Figure 4.12). These base pairs are held together by specific hydrogen bonds, which, although weak (4–21 kJ mol^{-1}, or 1–5 kcal mol^{-1}), stabilize the helix because of their large numbers in a DNA molecule. These *base-pairing rules* account for the observation, originally made by Erwin Chargaff in 1950, that the ratios of adenine to thymine and of guanine to cytosine are nearly the same in all species studied, whereas the adenine-to-guanine ratio varies considerably (Table 4.1).

Inside the helix, the bases are essentially stacked one on top of another (Figure 4.11B). The stacking of base pairs contributes to the stability of the double helix in two ways. First, the formation of the double helix is facilitated by the hydrophobic effect (p. 9). The hydrophobic bases cluster in the interior of the helix away

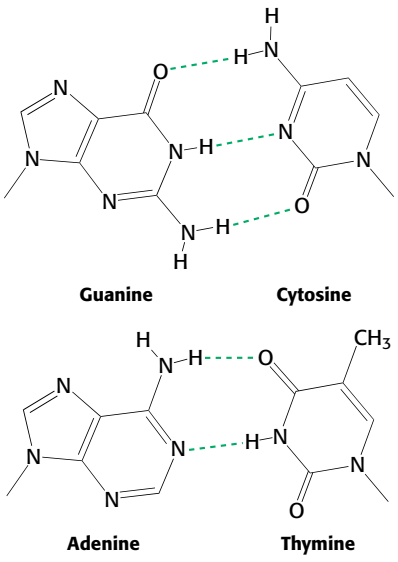


FIGURE 4.12 Structures of the base pairs proposed by Watson and Crick.

TABLE 4.1 Base compositions experimentally determined for a variety of organisms

Organism	A : T	G : C	A : G
Human being	1.00	1.00	1.56
Salmon	1.02	1.02	1.43
Wheat	1.00	0.97	1.22
Yeast	1.03	1.02	1.67
Escherichia coli	1.09	0.99	1.05
Serratia marcescens	0.95	0.86	0.70

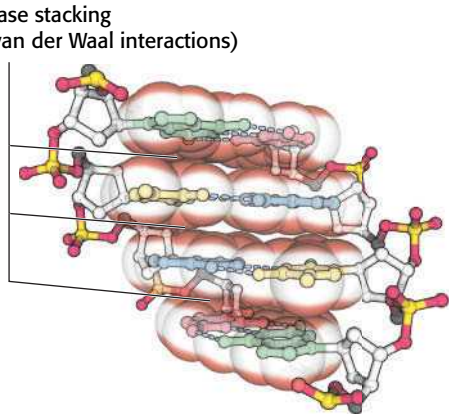


FIGURE 4.13 A side view of DNA. Base pairs are stacked nearly one on top of another in the double helix. The stacked bases interact with van der Waals forces. Such stacking forces help stabilize the double helix. [Source: J. L. Tymoczko, J. Berg, and L. Stryer, *Biochemistry: A Short Course*, 2nd ed. (W. H. Freeman and Company, 2013), Fig. 33.13.].

from the surrounding water, whereas the more polar surfaces are exposed to water. This arrangement is reminiscent of protein folding, where hydrophobic amino acids are in the protein's interior and the hydrophilic amino acids are on the exterior (Section 2.4). Second, the stacked base pairs attract one another through van der Waals forces (p. 8), appropriately referred to as *stacking forces,* further contributing to stabilization of the helix (Figure 4.13). The energy associated with a single van der Waals interaction is quite small, typically from 2 to 4 kJ mol^{-1} (0.5–1.0 kcal mol^{-1}). In the double helix, however, a large number of atoms are in van der Waals contact, and the net effect, summed over these atom pairs, is substantial. In addition, base stacking in DNA is favored by the conformations of the somewhat rigid five-membered rings of the backbone sugars.

DNA can assume a variety of structural forms

Watson and Crick based their model (known as the *B-DNA helix*) on x-ray diffraction patterns of highly hydrated DNA fibers, which provided information about properties of the double helix that are averaged over its constituent residues. Under physiological conditions, most DNA is in the B form. X-ray diffraction studies of less-hydrated DNA fibers revealed a different form called *A-DNA*. Like B-DNA, A-DNA is a right-handed double helix made up of anti-parallel strands held together by Watson–Crick base-pairing. The A-form helix is wider and shorter than the B-form helix, and its base pairs are tilted rather than perpendicular to the helix axis (Figure 4.14).

If the A-form helix were simply a property of dehydrated DNA, it would be of little significance. However, double-stranded regions of RNA and at least some RNA–DNA hybrids adopt a double-helical form very similar to that of A-DNA. What is the biochemical basis for differences between the two forms of DNA? Many of the structural differences between B-DNA and A-DNA arise from different puckerings of their ribose units (Figure 4.15). In A-DNA, C-3′ lies out of the plane (a conformation referred to as C-3′ endo) formed by the other four atoms of the ring; in B-DNA, C-2′ lies out of the plane (a conformation called C-2′ endo). The C-3′-endo puckering in A-DNA leads to an 11-degree tilting of

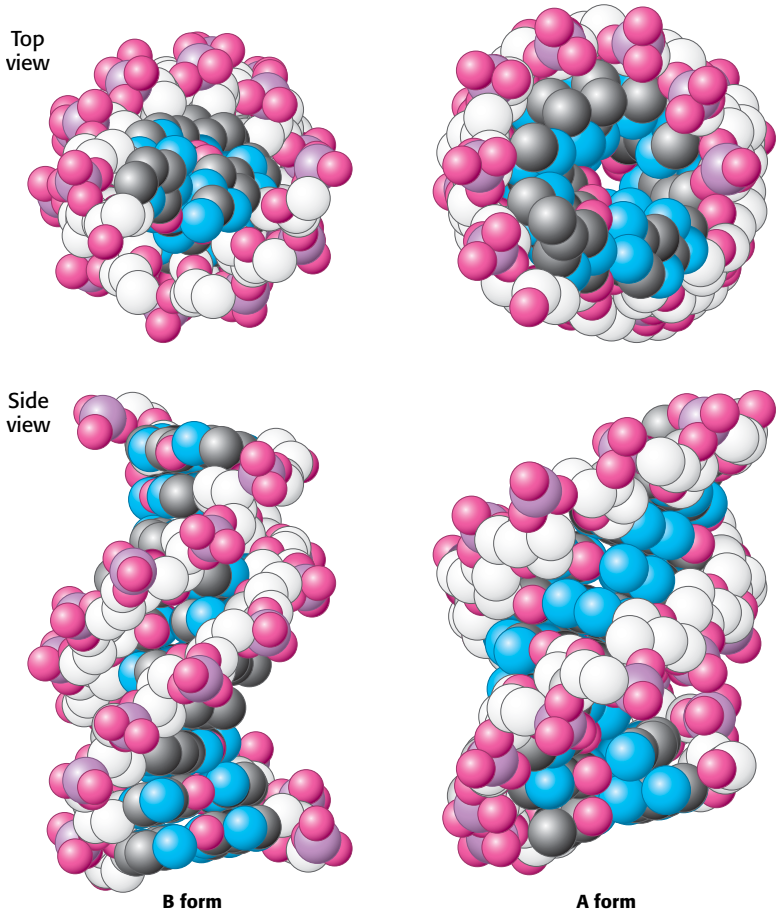


FIGURE 4.14 B-form and A-form DNA. Space-filling models of 10 base pairs of B-form and A-form DNA depict their right-handed helical structures. *Notice* that the B-form helix is longer and narrower than the A-form helix. The carbon atoms of the backbone are shown in white. [Drawn from 1BNA.pdb and 1DNZ.pdb.]

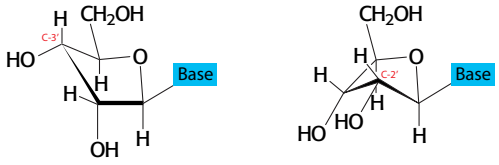


FIGURE 4.15 Sugar pucker. In A-form DNA, the C-3′ carbon atom lies above the approximate plane defined by the four other sugar nonhydrogen atoms (called C-3′ endo). In B-form DNA, each deoxyribose is in a C-2′-endo conformation, in which C-2′ lies out of the plane.

the base pairs away from perpendicular to the helix. RNA helices are further induced to take the A-DNA form because of steric hindrance from the 2′-hydroxyl group: the 2′-oxygen atom would be too close to three atoms of the adjoining phosphoryl group and to one atom in the next base. In an A-form helix, in contrast, the 2′-oxygen atom projects outward, away from other atoms. The phosphoryl and other groups in the A-form helix bind fewer H_2O molecules than do those in B-DNA. Hence, dehydration favors the A form.

Z-DNA is a left-handed double helix in which backbone phosphates zigzag

A third type of double helix is *left-handed,* in contrast with the *right-handed* screw sense of the A and B helices. Furthermore, the phosphoryl groups in the backbone are *zigzagged;* hence, this form of DNA is called *Z-DNA* (Figure 4.16).

Although the biological role of Z-DNA is still under investigation, Z-DNA-binding proteins have been isolated, one of which is required for viral pathogenesis of poxviruses, including variola, the agent of smallpox. The existence of Z-DNA shows that DNA is a flexible, dynamic molecule whose parameters are not as fixed as depictions suggest. The properties of A-, B-, and Z-DNA are compared in Table 4.2.

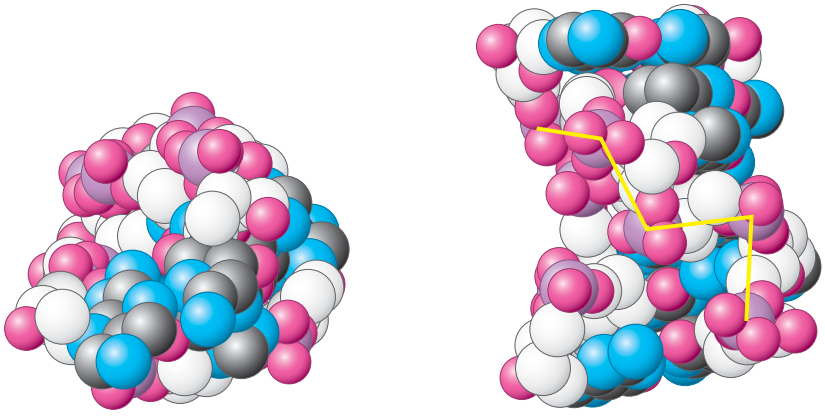


FIGURE 4.16 Z-DNA. DNA oligomers such as CGCGCG adopt an alternative conformation under some conditions. This conformation is called Z-DNA because the phosphoryl groups zigzag along the backbone. [Drawn from 131D.pdb.]

TABLE 4.2 Comparison of A-, B-, and Z-DNA

	A	B	Z
Shape	Broadest	Intermediate	Narrowest
Rise per base pair	2.3 Å	3.4 Å	3.8 Å
Helix diameter	~26 Å	~20 Å	~18 Å
Screw sense	Right-handed	Right-handed	Left-handed
Glycosidic bond*	*anti*	*anti*	Alternating *anti* and *syn*
Base pairs per turn of helix	11	10.4	12
Pitch per turn of helix	25.3 Å	35.4 Å	45.6 Å
Tilt of base pairs from perpendicular to helix axis	19 degrees	1 degree	9 degrees

**Syn* and *anti* refer to the orientation of the *N*-glycosidic bond between the base and deoxyribose. In the *anti* orientation, the base extends away from the deoxyribose. In the *syn* orientation, the base is above the deoxyribose. Pyrimidines can be in *anti* orientations only, whereas purines can be *anti* or *syn*.

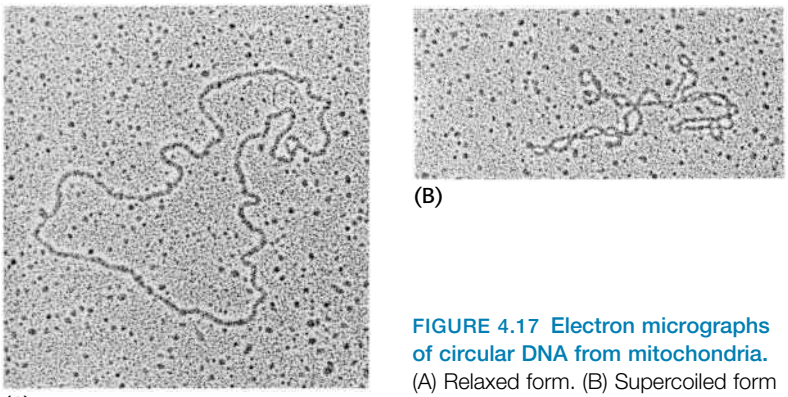


FIGURE 4.17 Electron micrographs of circular DNA from mitochondria. (A) Relaxed form. (B) Supercoiled form [Courtesy of Dr. David Clayton.]

Some DNA molecules are circular and supercoiled

The DNA molecules in human chromosomes are linear. However, electron microscopic and other studies have shown that intact DNA molecules from bacteria and archaea are circular (Figure 4.17A). The term *circular* refers to the continuity of the DNA strands, not to their geometric form. DNA molecules inside cells necessarily have a very compact shape. Note that the *E. coli* chromosome, fully extended, would be about 1000 times as long as the greatest diameter of the bacterium.

A closed DNA molecule has a property unique to circular DNA. The axis of the double helix can itself be twisted or supercoiled into a *superhelix* (Figure 4.17B). A circular DNA molecule without any superhelical turns is known as a *relaxed molecule*. Supercoiling is biologically important for two reasons. First, *a supercoiled DNA molecule is more compact than its relaxed counterpart*. Second, *supercoiling may hinder or favor the capacity of the double helix to unwind and thereby affect the interactions between DNA and other molecules.* These topological features of DNA will be considered further in Chapter 28.

Single-stranded nucleic acids can adopt elaborate structures

Single-stranded nucleic acids often fold back on themselves to form well-defined structures. Such structures are especially prominent in RNA and RNA-containing complexes such as the ribosome—a large complex of RNAs and proteins on which proteins are synthesized.

The simplest and most-common structural motif formed is a *stem-loop,* created when two complementary sequences within a single strand come together to form double-helical structures (Figure 4.18). In many cases, these double helices are made up entirely of Watson–Crick base pairs. In other cases, however, the structures include mismatched base pairs or unmatched bases that bulge out from the helix. Such mismatches destabilize

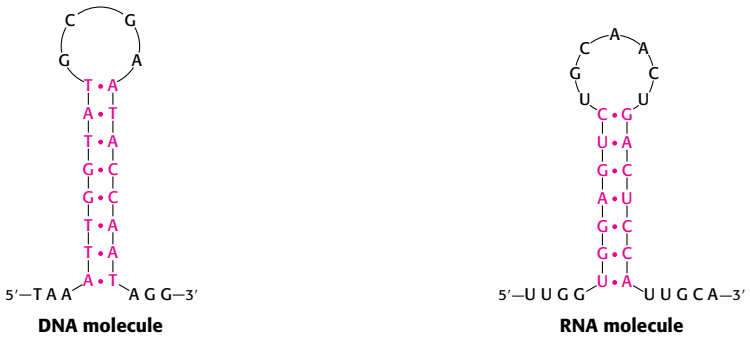


FIGURE 4.18 Stem-loop structures. Stem-loop structures can be formed from single-stranded DNA and RNA molecules.

the local structure but introduce deviations from the standard double-helical structure that can be important for higher-order folding and for function (Figure 4.19).

Single-stranded nucleic acids can adopt structures that are more complex than simple stem-loops through the interaction of more widely separated bases. Often, three or more bases interact to stabilize these structures. In such cases, hydrogen-bond donors and acceptors that do not participate in Watson–Crick base pairs participate in hydrogen bonds to form nonstandard pairings. Metal ions such as magnesium ion (Mg^{2+}) often assist in the stabilization of these more elaborate structures. These complex structures allow RNA to perform a host of functions that the double-stranded DNA molecule cannot. Indeed, the complexity of some RNA molecules rivals that of proteins, and these RNA molecules perform a number of functions that had formerly been thought the exclusive domain of proteins.

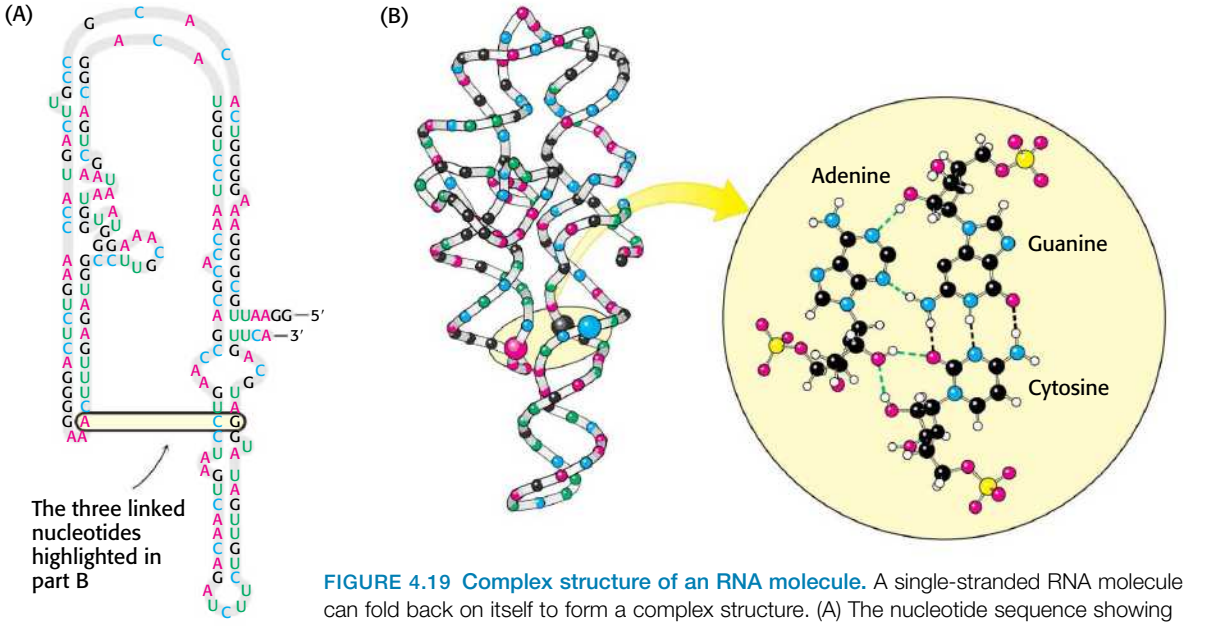


FIGURE 4.19 Complex structure of an RNA molecule. A single-stranded RNA molecule can fold back on itself to form a complex structure. (A) The nucleotide sequence showing Watson–Crick base pairs and other nonstandard base pairings in stem-loop structures. (B) The three-dimensional structure and one important long-range interaction between three bases. In the three-dimensional structure at the left, cytidine nucleotides are shown in blue, adenosine in red, guanosine in black, and uridine in green. In the detailed projection, hydrogen bonds within the Watson–Crick base pair are shown as dashed black lines; additional hydrogen bonds are shown as dashed green lines.

4.3 The Double Helix Facilitates the Accurate Transmission of Hereditary Information

The double-helical model of DNA and the presence of specific base pairs immediately suggested how the genetic material might replicate. *The sequence of bases of one strand of the double helix precisely determines the sequence of the other strand:* a guanine base on one strand is always paired with a cytosine base on the other strand, and so on. Thus, separation of a double helix into its two component strands would yield two single-stranded templates onto which new double helices could be constructed, each of which would have the same sequence of bases as the parent double helix. Consequently, as DNA is replicated, one of the strands of each daughter DNA molecule is newly synthesized, whereas the other is passed unchanged from the parent DNA molecule. This distribution of parental atoms is achieved by *semiconservative replication*.

Differences in DNA density established the validity of the semiconservative replication hypothesis

Matthew Meselson and Franklin Stahl carried out a critical test of this hypothesis in 1958. They labeled the parent DNA with ^{15}N, a heavy isotope of nitrogen, to make it denser than ordinary DNA. The labeled DNA was generated by growing *E. coli* for many generations in a medium that contained $^{15}NH_4Cl$ as the sole nitrogen source. After the incorporation of heavy nitrogen was complete, the bacteria were abruptly transferred to a medium that contained ^{14}N, the ordinary isotope of nitrogen. The question asked was: What is the distribution of ^{14}N and ^{15}N in the DNA molecules after successive rounds of replication?

The distribution of ^{14}N and ^{15}N was revealed by the technique of *density-gradient equilibrium sedimentation*. A small amount of DNA was dissolved in a concentrated solution of cesium chloride having a density close to that of the DNA (1.7 g cm^{-3}). This solution was centrifuged until it was nearly at equilibrium. At that point, the opposing processes of sedimentation and diffusion created a gradient in the concentration of cesium chloride across the centrifuge cell. The result was a stable density gradient ranging from 1.66 to 1.76 g cm^{-3}. The DNA molecules in this density gradient were driven by centrifugal force into the region where the solution's density was equal to their own. The DNA yielded a narrow band that was detected by its absorption of ultraviolet light. A mixture of ^{14}N DNA and ^{15}N DNA molecules gave clearly separate bands because they differ in density by about 1% (Figure 4.20).

DNA was extracted from the bacteria at various times after they were transferred from a ^{15}N to a ^{14}N medium. Analysis of these samples by the density-gradient technique showed that there was a single band of DNA after one generation. The density of this band was precisely halfway between the densities of the ^{14}N DNA and ^{15}N DNA bands (Figure 4.21).

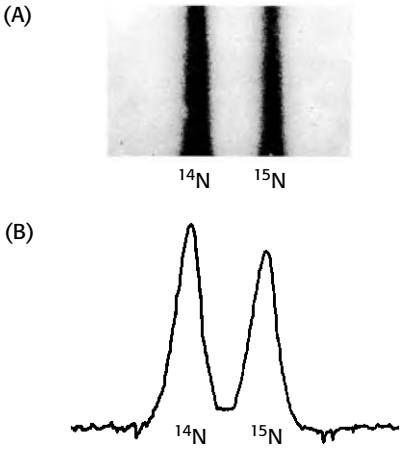


FIGURE 4.20 Resolution of ^{14}N DNA and ^{15}N DNA by density-gradient centrifugation. (A) Ultraviolet-absorption photograph of a centrifuged cell showing the two distinct bands of DNA. (B) Densitometric tracing of the absorption photograph. [Data from M. Meselson and F. W. Stahl. *Proc. Natl. Acad. Sci. U. S. A.* 44:671–682, 1958.]

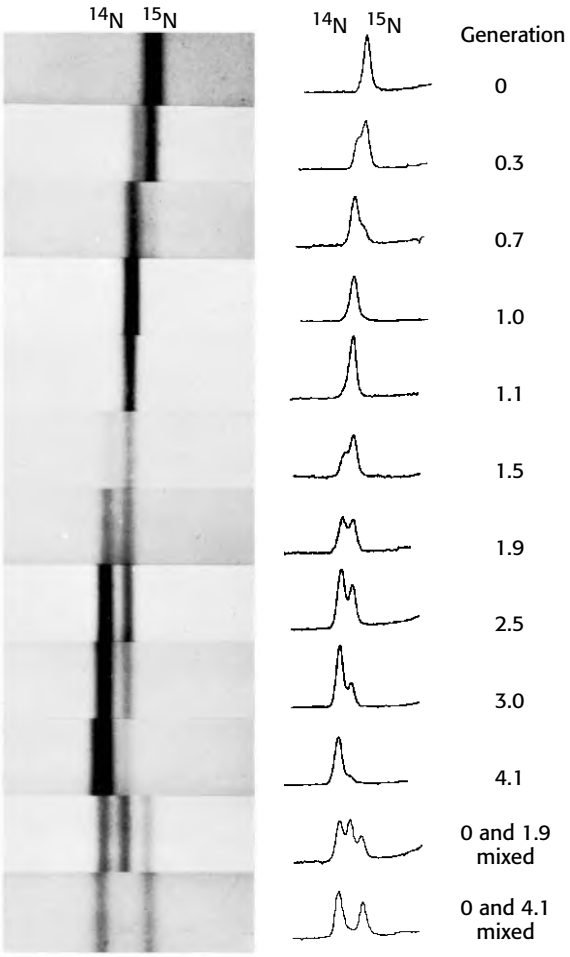


FIGURE 4.21 Detection of semiconservative replication of *E. coli* DNA by density-gradient centrifugation. The position of a band of DNA depends on its content of ^{14}N and ^{15}N. After 1.0 generation, all of the DNA molecules were hybrids containing equal amounts of ^{14}N and ^{15}N. [Data from M. Meselson and F. W. Stahl. *Proc. Natl. Acad. Sci. U. S. A.* 44:671–682, 1958.]

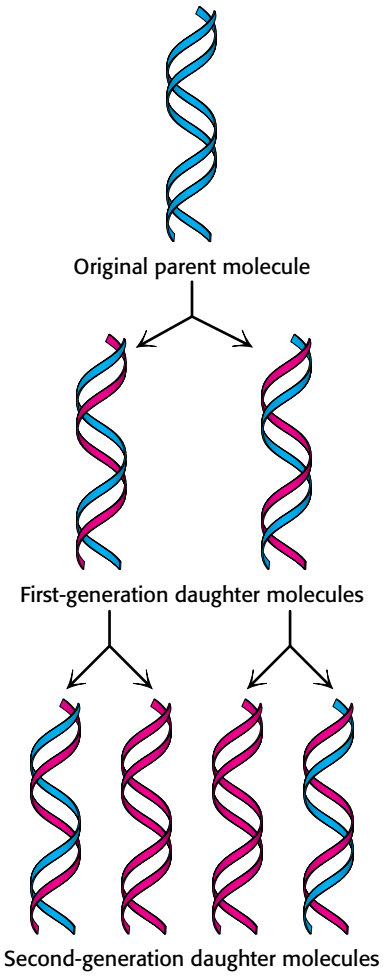


FIGURE 4.22 Diagram of semiconservative replication. Parental DNA is shown in blue and newly synthesized DNA in red. [Information from M. Meselson and F. W. Stahl. *Proc. Natl. Acad. Sci. U. S. A.* 44:671–682, 1958.]

The absence of ^{15}N DNA indicated that parental DNA was not preserved as an intact unit after replication. The absence of ^{14}N DNA indicated that all the daughter DNA derived some of their atoms from the parent DNA. This proportion had to be half because the density of the hybrid DNA band was halfway between the densities of the ^{14}N DNA and ^{15}N DNA bands.

After two generations, there were equal amounts of two bands of DNA. One was hybrid DNA, and the other was ^{14}N DNA. Meselson and Stahl concluded from these incisive experiments that replication was semiconservative, and so each new double helix contains a parent strand and a newly synthesized strand. Their results agreed perfectly with the Watson–Crick model for DNA replication (Figure 4.22).

The double helix can be reversibly melted

During DNA replication and transcription, the two strands of the double helix must be separated from each other, at least in a local region. The two strands of a DNA helix readily come apart when the hydrogen bonds between base pairs are disrupted. In the laboratory, the double helix can be disrupted by heating a solution of DNA or by adding acid or alkali to ionize its bases. The dissociation of the double helix is called *melting* because it occurs abruptly at a certain temperature. The *melting temperature* (T_m) of DNA is defined as the temperature at which half the helical structure is lost. Inside cells, however, the double helix is not melted by the addition of heat. Instead, proteins called *helicases* use chemical energy (from ATP) to disrupt the helix (Chapter 28).

Stacked bases in nucleic acids absorb less ultraviolet light than do unstacked bases, an effect called *hypochromism*. Thus, the melting of nucleic acids is readily monitored by measuring their absorption of light, which is maximal at a wavelength of 260 nm (Figure 4.23).

Separated complementary strands of nucleic acids spontaneously reassociate to form a double helix when the temperature is lowered below T_m. This renaturation process is sometimes called *annealing*. The facility with which double helices can be melted and then reassociated is crucial for the biological functions of nucleic acids.

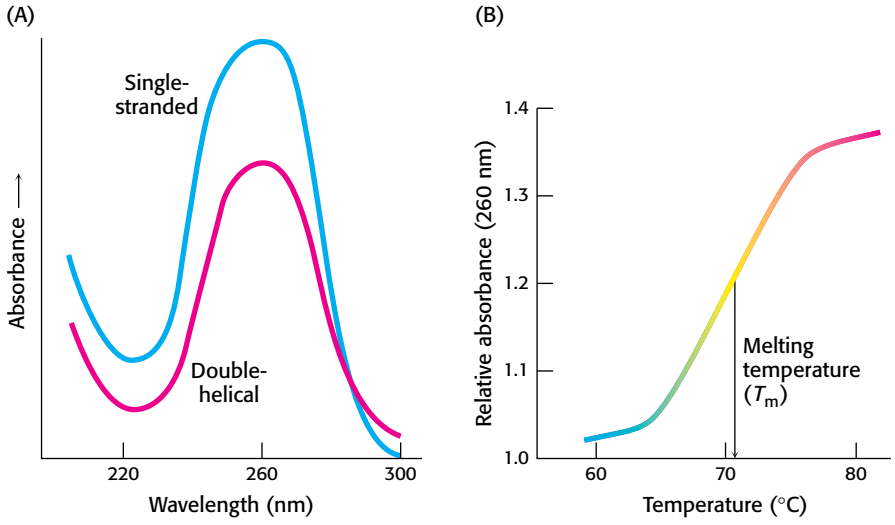


FIGURE 4.23 Hypochromism. (A) Single-stranded DNA absorbs light more effectively than does double-helical DNA. (B) The absorbance of a DNA solution at a wavelength of 260 nm increases when the double helix is melted into single strands.

The ability to melt and reanneal DNA reversibly in the laboratory provides a powerful tool for investigating sequence similarity. For instance, DNA molecules from two different organisms can be melted and allowed to reanneal, or *hybridize,* in the presence of each other. If the sequences are similar, hybrid DNA duplexes, with DNA from each organism contributing a strand of the double helix, can form. The degree of hybridization is an indication of the relatedness of the genomes and hence the organisms. Similar hybridization experiments with RNA and DNA can locate genes in a cell's DNA that correspond to a particular RNA. We will return to this important technique in Chapter 5.

4.4 DNA Is Replicated by Polymerases That Take Instructions from Templates

We now turn to the molecular mechanism of DNA replication. The full replication machinery in a cell comprises more than 20 proteins engaged in intricate and coordinated interplay. In 1958, Arthur Kornberg and his colleagues isolated from *E. coli* the first known of the enzymes, called *DNA polymerases,* that promote the formation of the bonds joining units of the DNA backbone. *E. coli* has a number of DNA polymerases, designated by roman numerals, that participate in DNA replication and repair (Chapter 28).

DNA polymerase catalyzes phosphodiester-bridge formation

DNA polymerases catalyze the step-by-step addition of deoxyribonucleotide units to a DNA strand (Figure 4.24). The reaction catalyzed, in its simplest form, is

$$(\text{DNA})_n + \text{dNTP} \rightleftharpoons (\text{DNA})_{n+1} + \text{PP}_\text{i}$$

where dNTP stands for any deoxyribonucleotide and PP_i is a pyrophosphate ion.

DNA synthesis has the following characteristics:

1. The reaction requires all four activated precursors—that is, *the deoxynucleoside 5′-triphosphates dATP, dGTP, dCTP, and TTP*—as well as Mg^{2+} ion.

2. *The new DNA strand is assembled directly on a preexisting DNA template.* DNA polymerases catalyze the formation of a phosphodiester linkage efficiently only if the base on the incoming nucleoside triphosphate is complementary to the base on the template strand. Thus, DNA polymerase is a *template-directed enzyme* that synthesizes a product with a base sequence complementary to that of the template.

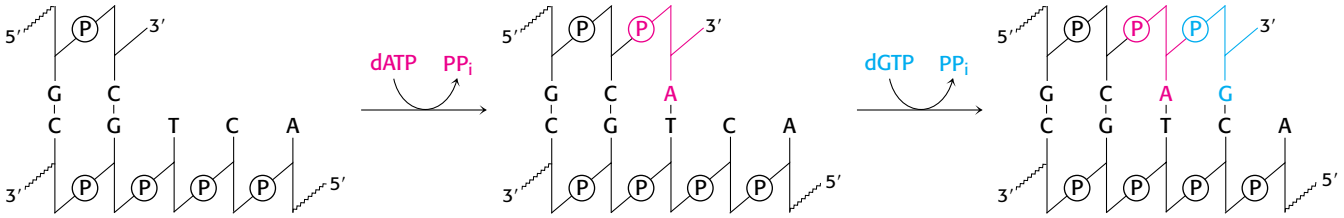


FIGURE 4.24 Polymerization reaction catalyzed by DNA polymerases.

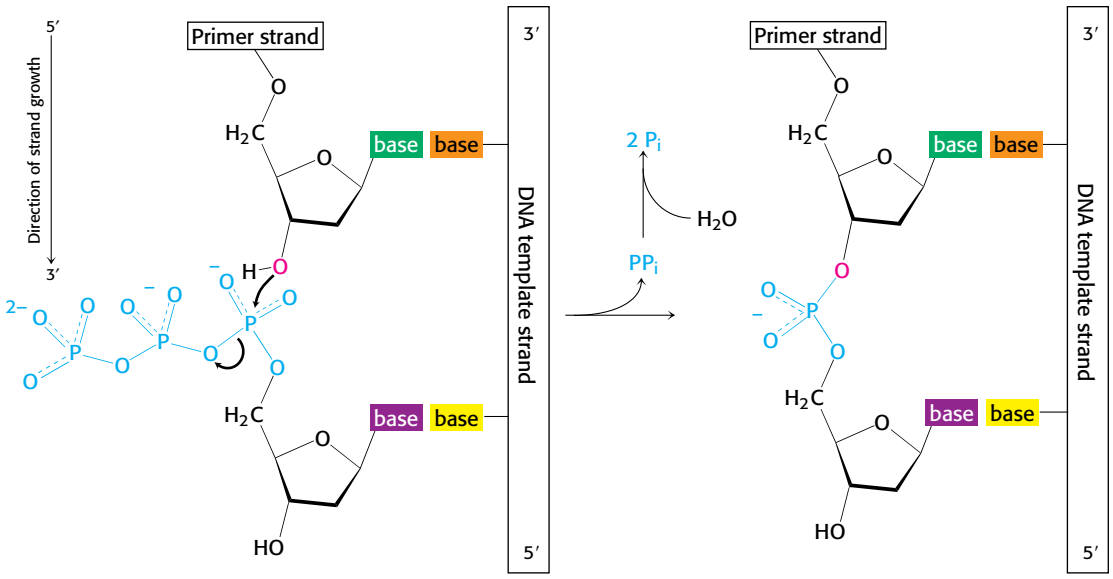


FIGURE 4.25 Strand-elongation reaction. DNA polymerases catalyze the formation of a phosphodiester bridge. [Source: J. L. Tymoczko, J. Berg, and L. Stryer, *Biochemistry: A Short Course*, 2nd ed. (W. H. Freeman and Company, 2013), Fig. 34.2.]

3. *DNA polymerases require a primer to begin synthesis.* A *primer* strand having a free 3′-OH group must be already bound to the template strand. The chain-elongation reaction catalyzed by DNA polymerases is a nucleophilic attack by the 3′-OH terminus of the growing strand on the innermost phosphorus atom of the deoxynucleoside triphosphate (Figure 4.25). A phosphodiester bridge is formed and pyrophosphate is released. The subsequent hydrolysis of pyrophosphate to yield two ions of orthophosphate (P_i) by *pyrophosphatase* helps drive the polymerization forward. *Elongation of the DNA chain proceeds in the 5′-to-3′ direction.*

4. *Many DNA polymerases are able to correct mistakes in DNA by removing mismatched nucleotides.* These polymerases have a distinct nuclease activity that allows them to excise incorrect bases by a separate reaction. This nuclease activity contributes to the remarkably high fidelity of DNA replication, which has an error rate of less than 10^{-8} per base pair.

The genes of some viruses are made of RNA

Genes in all cellular organisms are made of DNA. The same is true for some viruses but, for others, the genetic material is RNA. Viruses are genetic elements enclosed in protein coats that can move from one cell to another but are not capable of independent growth. A well-studied example of an RNA virus is the tobacco mosaic virus, which infects the leaves of tobacco plants. This virus consists of a single strand of RNA (6390 nucleotides) surrounded by a protein coat of 2130 identical subunits. An RNA polymerase that takes direction from an RNA template, called an *RNA-directed RNA polymerase*, copies the viral RNA. The infected cells die because of virus-instigated programmed cell death; in essence, the virus instructs the cell to commit suicide. Cell death results in discoloration in the tobacco leaf in a variegated pattern, hence the name mosaic virus.

Another important class of RNA virus comprises the *retroviruses,* so called because the genetic information flows from RNA to DNA rather than from DNA to RNA. This class includes human immunodeficiency virus 1 (HIV-1), the cause of acquired immunodeficiency syndrome (AIDS), as well as a number of RNA viruses that produce tumors in

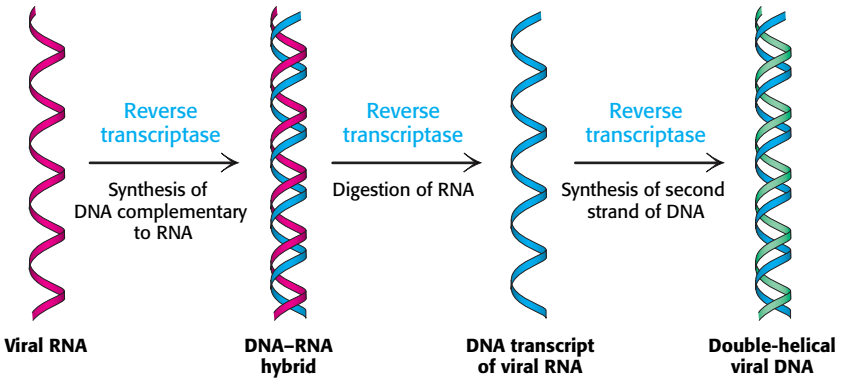


FIGURE 4.26 Flow of information from RNA to DNA in retroviruses. The RNA genome of a retrovirus is converted into DNA by reverse transcriptase, an enzyme brought into the cell by the infecting virus particle. Reverse transcriptase possesses several activities and catalyzes the synthesis of a complementary DNA strand, the digestion of the RNA, and the subsequent synthesis of the DNA strand.

susceptible animals. Retrovirus particles contain two copies of a single-stranded RNA molecule. On entering the cell, the RNA is copied into DNA through the action of a viral enzyme called *reverse transcriptase*, which acts as both a polymerase and an RNase (Figure 4.26). The resulting double-helical DNA version of the viral genome can become incorporated into the chromosomal DNA of the host and is replicated along with the normal cellular DNA. At a later time, the integrated viral genome is expressed to form viral RNA and viral proteins, which assemble into new virus particles.

4.5 Gene Expression Is the Transformation of DNA Information into Functional Molecules

The information stored as DNA becomes useful when it is expressed in the production of RNA and proteins. This rich and complex topic is the subject of several chapters later in this book, but here we introduce the basics of gene expression. DNA can be thought of as archival information, stored and manipulated judiciously to minimize damage (mutations). It is expressed in two steps. First, an RNA copy is made that encodes directions for protein synthesis. This messenger RNA can be thought of as a photocopy of the original information: it can be made in multiple copies, used, and then disposed of. Second, the information in messenger RNA is translated to synthesize functional proteins. Other types of RNA molecules exist to facilitate this translation.

Several kinds of RNA play key roles in gene expression

Scientists used to believe that RNA played a passive role in gene expression, as a mere conveyor of information. However, recent investigations have shown that RNA plays a variety of roles, from catalysis to regulation. Cells contain several kinds of RNA (Table 4.3):

TABLE 4.3 RNA molecules in *E. coli*

Type	Relative amount (%)	Sedimentation coefficient (*S*)	Mass (kDa)	Number of nucleotides
Ribosomal RNA (rRNA)	80	23	1.2×10^3	3700
		16	0.55×10^3	1700
		5	3.6×10^1	120
Transfer RNA (tRNA)	15	4	2.5×10^1	75
Messenger RNA (mRNA)	5		Heterogeneous	