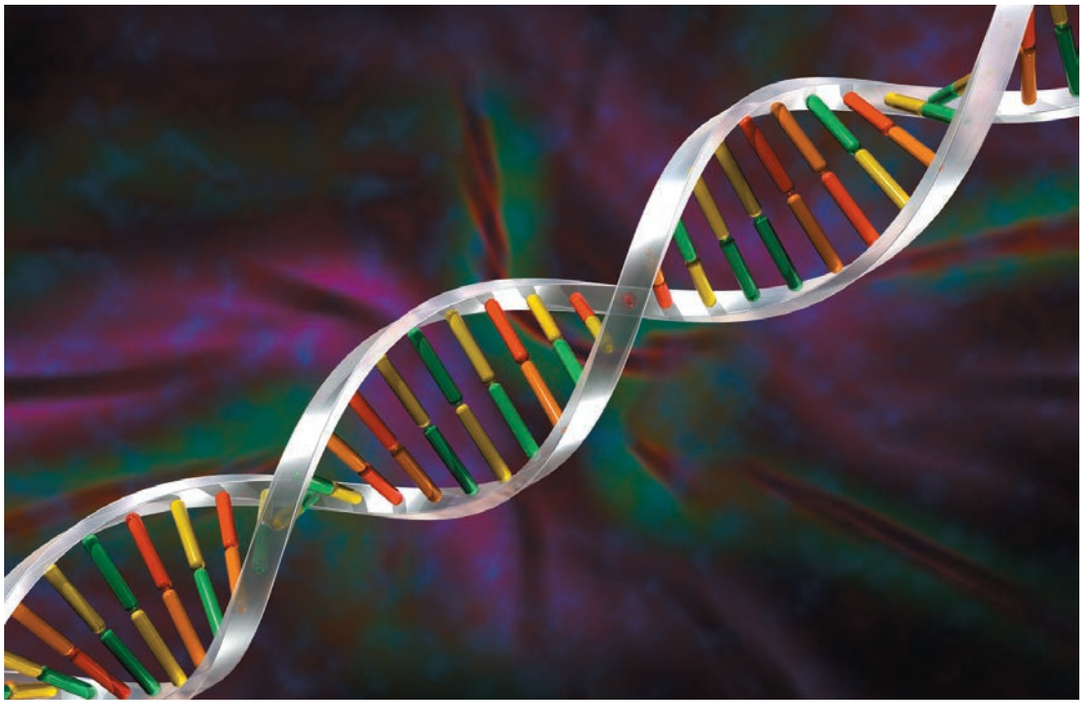
26 The Chemistry of the Nucleic Acids



a double helix

We have studied two of the three major kinds of biopolymers: polysaccharides in Chapter 20 and proteins in Chapter 21. Now we will look at the third kind of biopolymer—nucleic acids. There are two types of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA encodes an organism's entire hereditary information and controls the growth and division of cells. In all organisms (except certain viruses), the genetic information stored in DNA is transcribed into RNA. This information can then be translated for the synthesis of all the proteins needed for cellular structure and function.

DNA was first isolated in 1869 from the nuclei of white blood cells. Because it was found in the nucleus and was acidic, it was called *nucleic acid.* Eventually, scientists found that the nuclei of all cells contain DNA. The fact that DNA is the carrier of genetic information was not known until 1944, when it was found that DNA could be transferred from one species to another, along with inheritable traits. In 1953, James Watson and Francis Crick described the three-dimensional structure of DNA—the famed double helix.

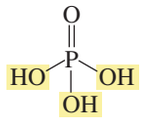
26.1 NUCLEOSIDES AND NUCLEOTIDES

Nucleic acids are chains of five-membered-ring sugars linked by phosphate groups. Notice that the linkages are **phosphodiesters** (Figure 26.1 on the next page).

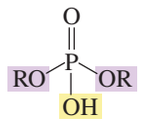
- In RNA, the five-membered-ring sugar is D-ribose.
- In DNA, the five-membered-ring sugar is 2′-deoxy-D-ribose (D-ribose without an OH group in the 2′-position).

The Bases in DNA and RNA

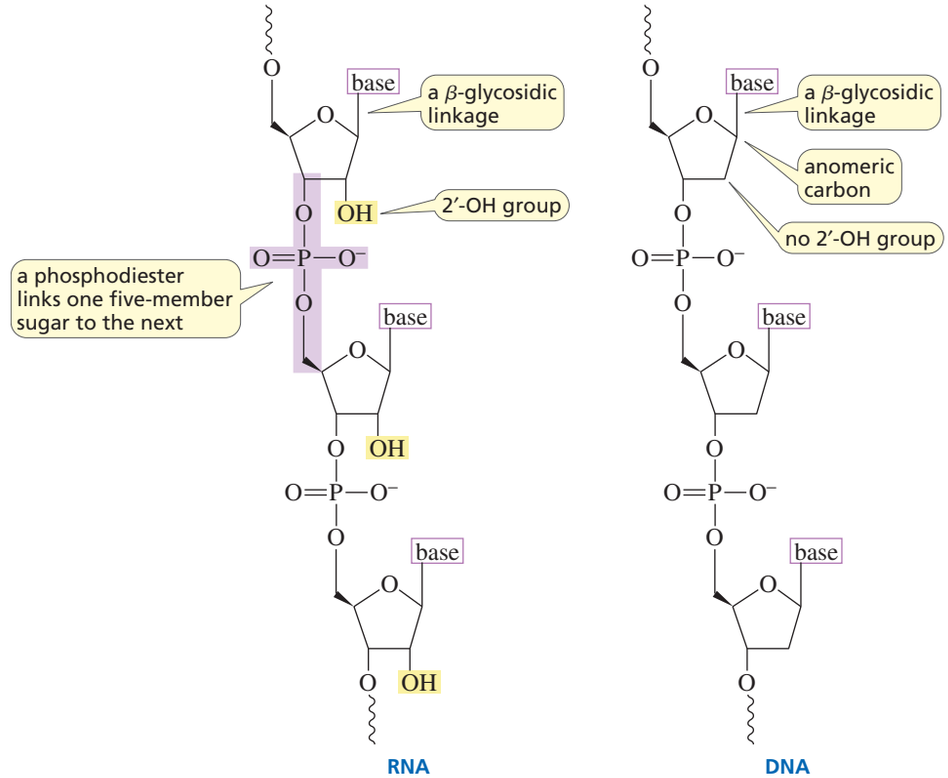
The anomeric carbon of each sugar is bonded to a nitrogen of a heterocyclic compound in a β-glycosidic linkage. (Recall from Section 20.10 that a β-linkage is one in which the substituents at C-1 and C-4 are on the same side of the furanose ring.) Because the heterocyclic compounds are amines, they are commonly referred to as **bases.**



phosphoric acid

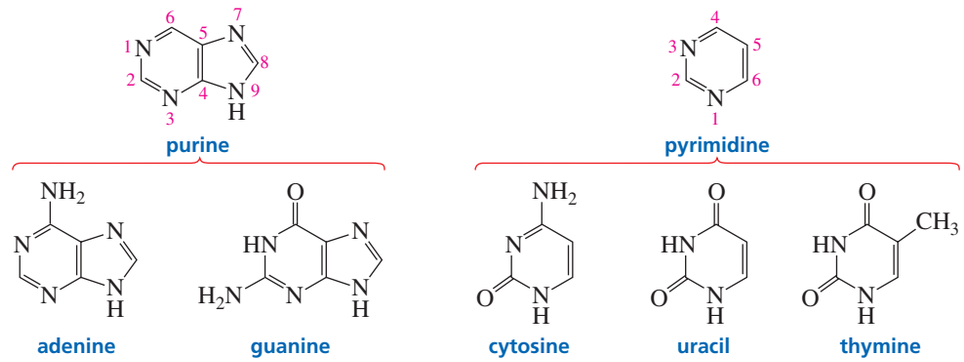


a phosphodiester



▶ **Figure 26.1**
Nucleic acids consist of a chain of five-membered-ring sugars linked by phosphate groups. The anomeric carbon of each sugar (D-ribose in RNA, 2′-deoxy-D-ribose in DNA) is bonded to a heterocyclic amine (a base) in a β-glycosidic linkage.

The vast differences in heredity between different species and between different members of the same species are determined by the sequence of the bases in DNA. Surprisingly, there are only four bases in DNA: two are substituted purines (adenine and guanine), and two are substituted pyrimidines (cytosine and thymine).

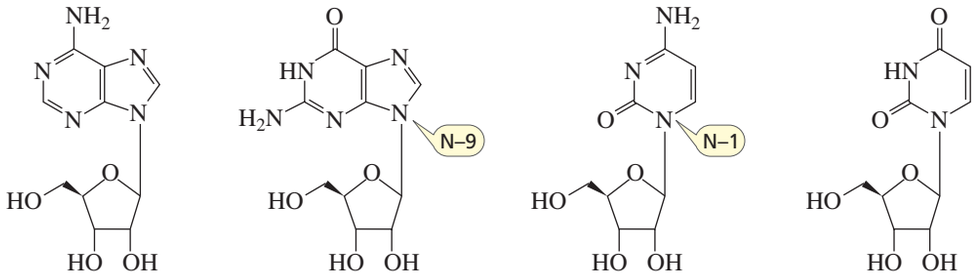


RNA also contains only four bases. Three (adenine, guanine, and cytosine) are the same as those in DNA, but the fourth base in RNA is uracil instead of thymine. Notice that thymine and uracil differ only by a methyl group. (Thymine is 5-methyluracil.) The reason DNA contains thymine instead of uracil is explained in Section 26.10.

Nucleosides

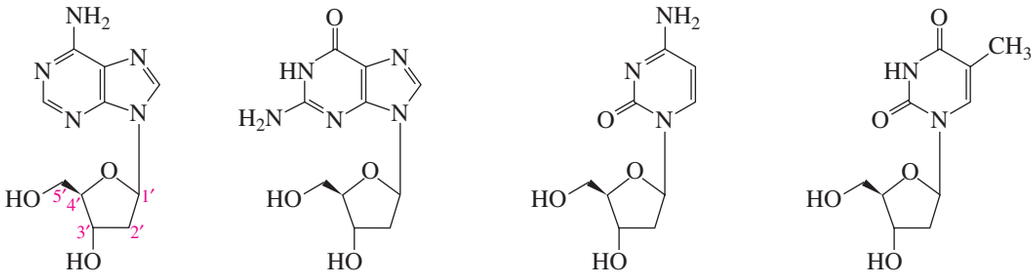
The anomeric carbon of the furanose ring is bonded to purines at N-9 and to pyrimidines at N-1. A compound containing a base bonded to D-ribose or to 2′-deoxy-D-ribose is called a **nucleoside.** The ring positions of the sugar component of a nucleoside are indicated by primed numbers to distinguish them from the ring positions of the base. This is why the sugar component of DNA is referred to as 2′-deoxy-D-ribose. The nucleosides of RNA—where the sugar is D-ribose— are more precisely called ribonucleosides, whereas the nucleosides of DNA—where the sugar is 2′-deoxy-D-ribose—are called deoxyribonucleosides.

nucleosides in RNA



adenosine guanosine cytidine uridine

nucleosides in DNA



2′-deoxyadenosine 2′-deoxyguanosine 2′-deoxycytidine thymidine

Notice the difference in the base names and their corresponding nucleoside names in Table 26.1. For example, adenine is the base, whereas adenosine is the nucleoside; similarly, cytosine is the base, whereas cytidine is the nucleoside, and so forth. Because uracil is found only in RNA, it is shown attached to D-ribose but not to 2′-deoxy-D-ribose; because thymine is found only in DNA, it is shown attached to 2′-deoxy-D-ribose but not to D-ribose.

Table 26.1 The Names of the Bases, the Nucleosides, and the Nucleotides

Base	Ribonucleoside	Deoxyribonucleoside	Ribonucleotide	Deoxyribonucleotide
Adenine	Adenosine	2′-Deoxyadenosine	Adenosine 5′-phosphate	2′-Deoxyadenosine 5′-phosphate
Guanine	Guanosine	2′-Deoxyguanosine	Guanosine 5′-phosphate	2′-Deoxyguanosine 5′-phosphate
Cytosine	Cytidine	2′-Deoxycytidine	Cytidine 5′-phosphate	2′-Deoxycytidine 5′-phosphate
Thymine	—	Thymidine	—	Thymidine 5′-phosphate
Uracil	Uridine	—	Uridine 5′-phosphate	

Nucleotides

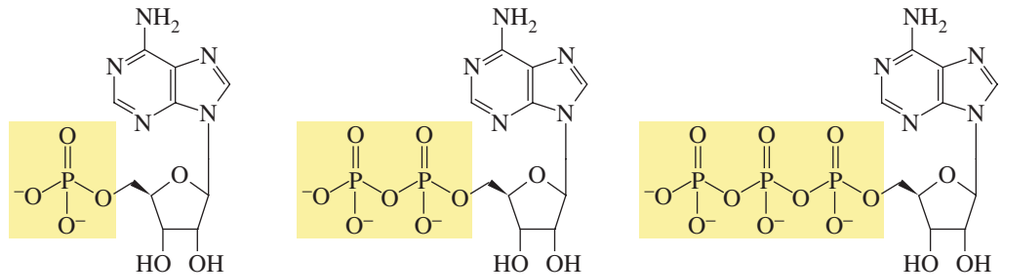
A **nucleotide** is a nucleoside with an OH group of the sugar bonded in an ester linkage to phosphoric acid. The nucleotides of RNA are more precisely called **ribonucleotides,** and those of DNA are called **deoxyribonucleotides.** The base names in nucleotides are the same as those in nucleosides.

Nucleoside = base + sugar

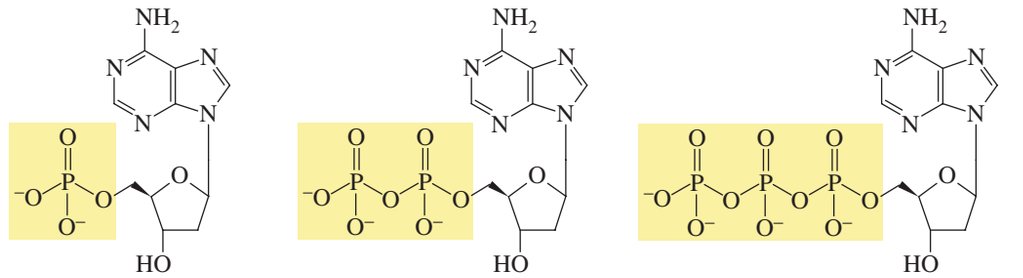
Nucleotide = base + sugar + phosphate

Because phosphoric acid can form an anhydride, nucleotides can exist as monophosphates, diphosphates, and triphosphates (Section 24.1). They are named by adding *monophosphate* or *diphosphate* or *triphosphate* to the name of the nucleoside.

nucleotides of adenosine



nucleotides of 2′-deoxyadenosine



The names of the nucleotides are abbreviated (A, G, C, T, U—followed by MP, DP, or TP, depending on whether it is a monophosphate, diphosphate, or triphosphate—with a d in front if it contains 2′-deoxy-D-ribose instead of D-ribose: for example, ATP, dATP).

The Structure of DNA: Watson, Crick, Franklin, and Wilkins

James D. Watson was born in Chicago in 1928. He graduated from the University of Chicago at the age of 19 and received a Ph.D. three years later from Indiana University. In 1951, as a postdoctoral fellow at Cambridge University, Watson worked on determining the three-dimensional structure of DNA.

Francis H. C. Crick (1916–2004) was born in Northampton, England. Originally trained as a physicist, Crick did research on radar during World War II. After the war, deciding that the most interesting problem in science was the physical basis of life, he entered Cambridge University to study the structure of biological molecules by X-ray analysis. He was a graduate student when he carried out his portion of the work that led to the proposal of the double helical structure of DNA. He received a Ph.D. in chemistry in 1953.

Francis Crick (*left*) and James Watson (*right*)

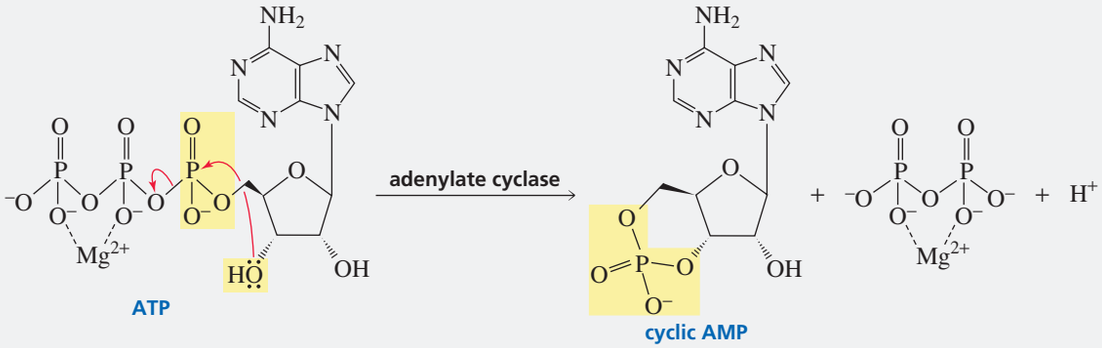
Rosalind Franklin (1920–1958) was born in London. She graduated from Cambridge University and studied X-ray diffraction techniques in Paris. In 1951 she returned to England and accepted a position to develop an X-ray diffraction unit in the biophysics department at King's College. Her X-ray studies showed that DNA was a helix with the sugars and phosphate groups on the outside of the molecule. Tragically, Franklin never protected herself from her X-ray source and died without knowing the role her work had played in determining the structure of DNA, and without being recognized for her contribution.

Rosalind Franklin

Watson and Crick shared the 1962 Nobel Prize in Physiology or Medicine with Maurice Wilkins for determining the double-helical structure of DNA. Wilkins (1916–2004), who contributed X-ray studies that confirmed the double-helical structure, was born in New Zealand to Irish immigrants and moved to England six years later with his parents. He received a Ph.D. from Birmingham University. During World War II he joined other British scientists who were working with American scientists on the development of the atomic bomb. He returned to England in 1945 and, having lost interest in physics, turned his attention to biology.

Cyclic AMP

Cyclic AMP (cAMP), which is present in all life forms, controls a wide variety of biological processes. This cyclic nucleotide is called a "second messenger" because it serves as a link between at least 20 different hormones (the first messengers) and enzymes that regulate cellular function.



For example, secretion of a hormone, such as adrenaline, activates adenylate cyclase, the enzyme responsible for the synthesis of cyclic AMP from ATP. Cyclic AMP then activates a regulatory enzyme (Section 24.13), generally by phosphorylating it.

PROBLEM 1

In acidic solutions, nucleosides are hydrolyzed to a sugar and a heterocyclic base. Propose a mechanism for this reaction.

PROBLEM 2

Draw the structure for each of the following:

a. dCDP
b. dTTP
c. dUMP
d. UDP
e. guanosine triphosphate
f. adenosine 5'-monophosphate

26.2 NUCLEIC ACIDS ARE COMPOSED OF NUCLEOTIDE SUBUNITS

Nucleic acids are composed of long strands of nucleotide subunits (Figure 26.1).

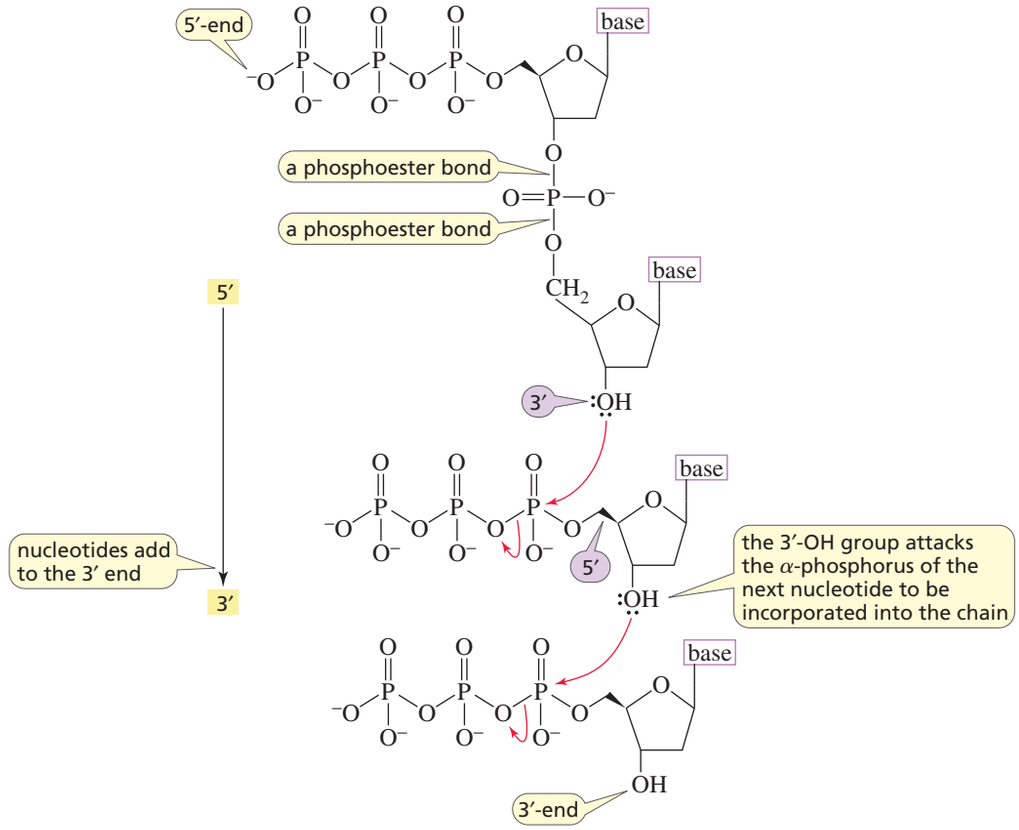
- A **dinucleotide** contains two nucleotide subunits.
- An **oligonucleotide** contains 3 to 10 nucleotide subunits.
- A **polynucleotide** contains many nucleotide subunits.

DNA and RNA are polynucleotides.

Biosynthesis of Nucleic Acids

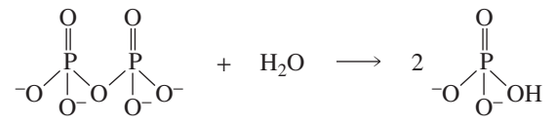
Nucleic acids are biosynthesized from nucleoside triphosphates, using enzymes called *DNA polymerases* (for the synthesis of DNA) or *RNA polymerases* (for the synthesis of RNA). The nucleotides are linked as a result of nucleophilic attack by a 3′-OH group of one nucleoside triphosphate on the α-phosphorus of another nucleoside triphosphate, breaking a phosphoanhydride bond and

eliminating pyrophosphate (Figure 26.2). Thus, the phosphodiester joins the 3′-OH group of one nucleotide and the 5′-OH group of the next nucleotide, and the growing polymer is synthesized in the 5′ → 3′ direction. In other words, new nucleotides are added to the 3′-end.



▶ Figure 26.2
Addition of nucleotides to a growing strand of DNA. Biosynthesis occurs in the 5′ → 3′ direction.

The pyrophosphate product is subsequently hydrolyzed, which makes the reaction that joins the nucleotides irreversible. Irreversibility is important if the genetic information in DNA is to be preserved (Section 24.1).



RNA strands are biosynthesized in the same way, using ribonucleoside triphosphates instead of 2′-deoxyribonucleoside triphosphates.

The Primary Structure of a Nucleic Acid

The **primary structure** of a nucleic acid is the sequence of bases in the strand. By convention, the sequence of bases is written in the 5′ → 3′ direction (the 5′-end is on the left). Remember that the nucleotide at the 5′-end of the strand has an unlinked 5′-triphosphate group, and the nucleotide at the 3′-end has an unlinked 3′-hydroxyl group.

DNA is synthesized in the 5′ → 3′ direction.

26.3 THE SECONDARY STRUCTURE OF DNA

Watson and Crick concluded, with the aid of Rosalind Franklin's X-ray data, that

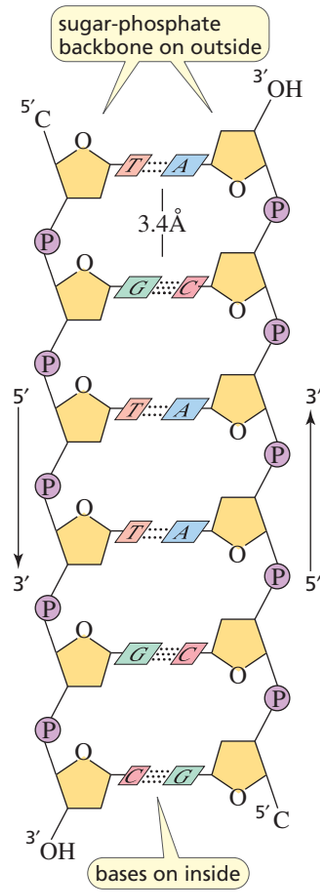
- DNA consists of two strands of nucleotides, with the sugar–phosphate backbone on the outside and the bases on the inside.
- the strands are antiparallel (they run in opposite directions).
- the strands are held together by hydrogen bonds between the bases on one strand and the bases on the other strand (Figure 26.3).

The DNA Strands are Complementary

Experiments carried out by Erwin Chargaff were critical to Watson and Crick's proposal for the structure of DNA. These experiments showed that the number of adenines in DNA equals the number of thymines, and the number of guanines equals the number of cytosines. Chargaff also noted that the number of adenines and thymines relative to the number of guanines and cytosines is characteristic of a given species but varies from species to species. In human DNA, for example, 60.4% of the bases are adenines and thymines, whereas 74.2% of the bases are adenines and thymines in the DNA of the bacterium *Sarcina lutea.*

Chargaff's data showing that [adenine] = [thymine] and [guanine] = [cytosine] could be explained if adenine (A) always paired with thymine (T), and guanine (G) always paired with cytosine (C). This means the two strands are *complementary:* where there is an A in one strand, there is a T in the opposing strand; and where there is a G in one strand, there is a C in the other strand (Figure 26.3).

Thus, if you know the sequence of bases in one strand,
you can figure out the sequence of bases in the other strand.

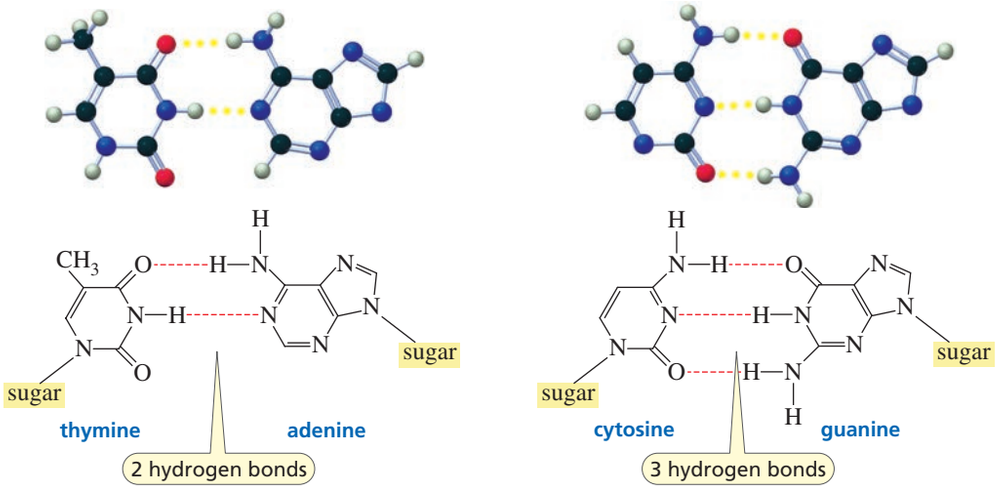


▲ **Figure 26.3**
The sugar–phosphate backbone of DNA is on the outside and the bases are on the inside. As pair with Ts and Gs pair with Cs. The two strands are antiparallel—that is, they run in opposite directions.

Hydrogen Bonding Dictates Base Pairing

Why does A pair with T? Why does G pair with C? First of all, the width of the double-stranded molecule is relatively constant, so a purine must pair with a pyrimidine. If the larger purines paired, the strands would bulge; if the smaller pyrimidines paired, the strands would have to pull in to bring the two pyrimidines close enough to form hydrogen bonds. But what causes A to pair with T rather than with C (the other pyrimidine)?

The base pairing is dictated by hydrogen bonding. Learning that the bases exist in the keto form and not the enol form (Section 17.2) allowed Watson to explain the pairing.* Adenine forms two hydrogen bonds with thymine but would form only one hydrogen bond with cytosine. Guanine forms three hydrogen bonds with cytosine but would form only one hydrogen bond with thymine (Figure 26.4).



LEARN THE STRATEGY

◀ **Figure 26.4**
Base pairing in DNA: adenine and thymine form two hydrogen bonds; cytosine and guanine form three hydrogen bonds.

* Watson was having difficulty understanding the base pairing in DNA because he thought the bases existed in the enol form (see Problem 4). When Jerry Donohue, an American crystallographer, informed him that the bases more likely existed in the keto form, Chargaff's data could easily be explained by hydrogen bonding between adenine and thymine and between guanine and cytosine.

USE THE STRATEGY

PROBLEM 3

Indicate whether each functional group of the five heterocyclic bases in nucleic acids is a hydrogen bond acceptor (A), a hydrogen bond donor (D), or both (D/A).

PROBLEM 4

Using the D, A, and D/A designations in Problem 3, indicate how base pairing would be affected if the bases existed in the enol form.

PROBLEM 5 ♦

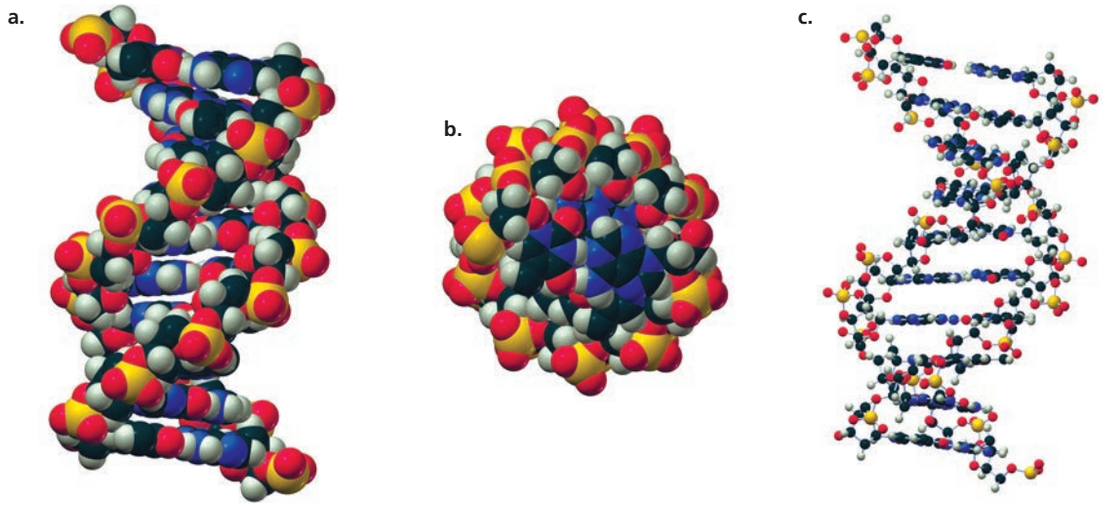
If one of the strands of DNA has the following sequence of bases running in the 5′ → 3′ direction,

5′—G—G—A—C—A—A—T—C—T—G—C—3′

a. what is the sequence of bases in the complementary strand?
b. what base is closest to the 5′-end in the complementary strand?

The Double Helix

The two antiparallel DNA strands are not linear but are twisted into a helix around a common axis (Figure 26.5a). The base pairs are planar and parallel to each other on the inside of the helix (Figure 26.5c). The secondary structure is, therefore, known as a **double helix.** The double helix resembles a circular staircase: the base pairs are the rungs, and the sugar–phosphate backbones are the handrails. The OH group of the phosphodiester linkages has a pK_a of about 2, so it is in its basic form (negatively charged) at physiological pH (Figure 26.2). The negatively charged phosphates repels nucleophiles, thereby preventing cleavage of the phosphodiester bonds.



▲ **Figure 26.5**

a. The DNA double helix.

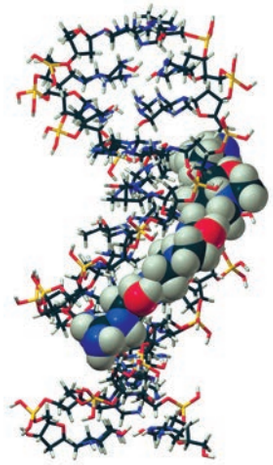
b. The view looking down the long axis of the helix. (The blue nitrogen atoms of the bases are on the inside; the yellow phosphorus atoms are on the outside.)

c. The bases are planar and parallel on the inside of the helix.

Hydrogen bonding between base pairs is just one of the forces holding the two strands of the DNA double helix together. The bases are planar aromatic molecules that stack on top of one another, each pair slightly rotated with respect to the next pair, like a partially spread-out hand of cards. In this arrangement, there are favorable interactions between the mutually induced dipoles of adjacent pairs of bases. These interactions, known as **stacking interactions,** are weak attractive forces, but when added together they contribute significantly to the stability of the double helix.

Confinement of the bases to the inside of the helix has an additional stabilizing effect—it reduces the surface area of the relatively nonpolar residues that is exposed to water, which increases the entropy of the surrounding water molecules (Section 21.15). Stacking interactions are strongest between two purines and weakest between two pyrimidines.

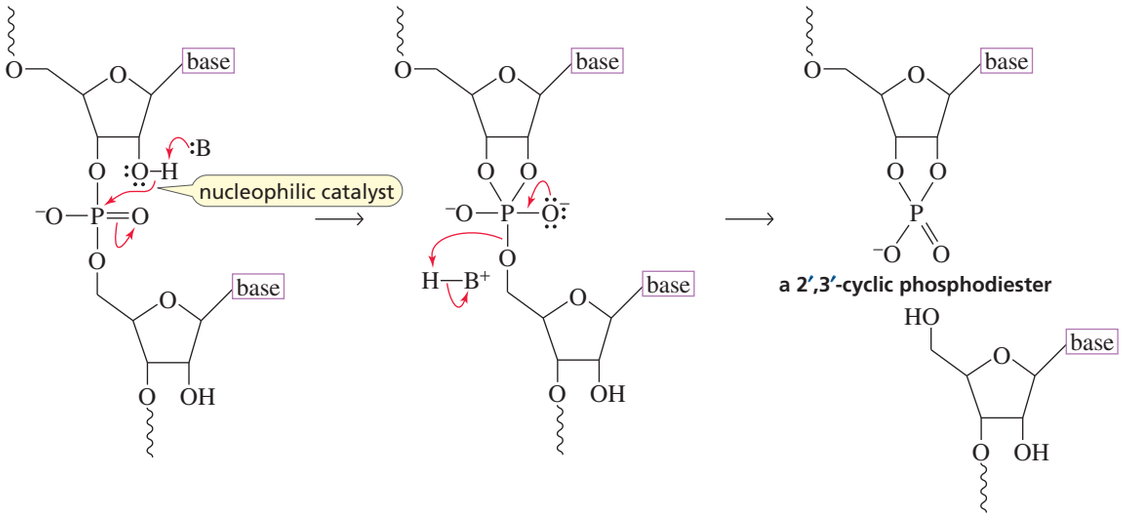
There are two different alternating grooves in a DNA double helix; a **major groove** and a narrower **minor groove.** Proteins and other molecules can bind to the grooves. The hydrogen-bonding properties of the functional groups facing into each groove determine what kind of molecules will bind to the groove. For example, netropsin is an antibiotic that works by binding to the minor groove of DNA (Figure 26.6).



▲ **Figure 26.6**
The antibiotic netropsin bound in the minor groove of DNA.

26.4 WHY DNA DOES NOT HAVE A 2′-OH GROUP

Unlike DNA, RNA is not stable because the 2′-OH group of ribose acts as a nucleophilic catalyst for the cleavage of RNA (Figure 26.7). This explains why the 2′-OH group is absent in DNA. DNA must remain intact throughout the life span of a cell in order to preserve the genetic information. Easy cleavage of DNA would have disastrous consequences for the cell and for life itself. RNA, in contrast, is synthesized as it is needed and is degraded once it has served its purpose.



▲ **Figure 26.7**
Catalysis of RNA cleavage by the 2′-OH group. RNA undergoes cleavage 3 billion times faster than DNA.

PROBLEM 6

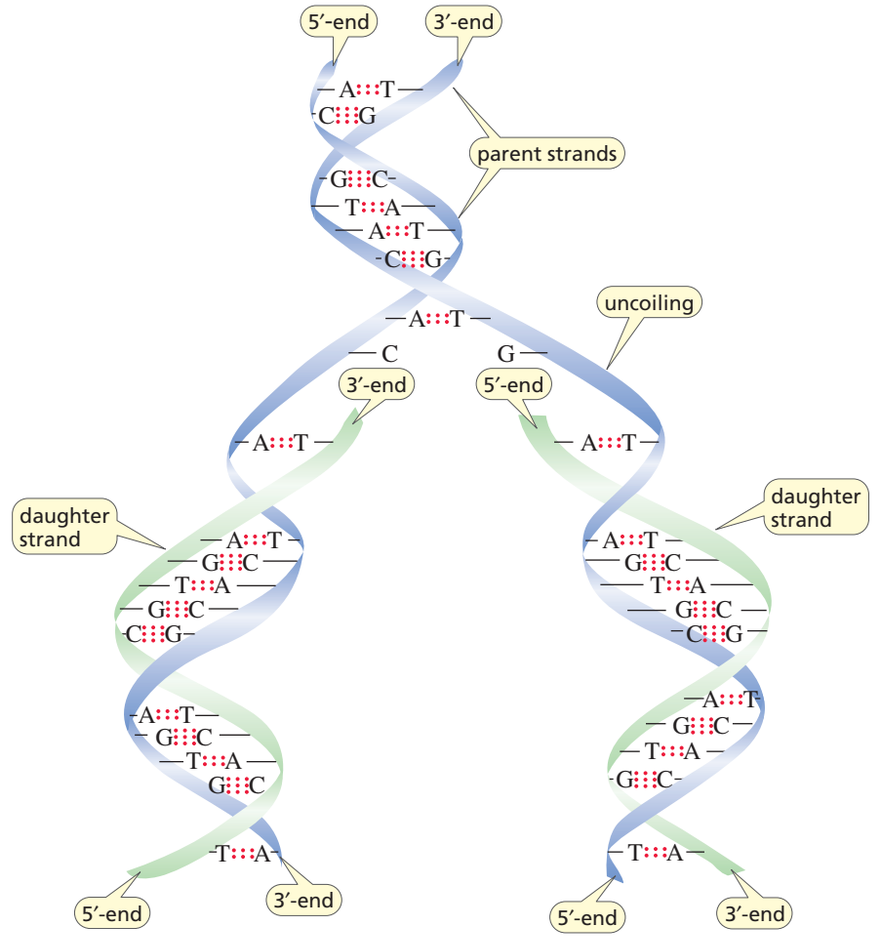
The 2′,3′-cyclic phosphodiester that is formed (Figure 26.7) when RNA is cleaved, forms a mixture of nucleotide 2′- and 3′-phosphates when it reacts with water. Propose a mechanism for this reaction.

26.5 THE BIOSYNTHESIS OF DNA IS CALLED REPLICATION

The genetic information of a human cell is contained in 23 pairs of chromosomes. Each chromosome is composed of thousands of **genes** (segments of DNA). The total DNA from a human cell—the **human genome**—contains 3.1 billion base pairs.

Part of the excitement created by Watson and Crick's proposed structure for DNA was that the structure immediately suggested how DNA is able to pass on genetic information to succeeding generations. Because the two strands are complementary, both carry the same genetic information. Thus, when organisms reproduce, DNA molecules can be copied using the same base-pairing principle that is fundamental to their structure—that is, each strand can serve as the template for the synthesis of a complementary new strand (Figure 26.8 on the next page). The new (daughter) DNA molecules are identical to the original (parent) molecule, so they contain all the original genetic information. The synthesis of identical copies of DNA is called **replication.**

DNA



▲ **Figure 26.8**
Replication of DNA. The green daughter strand on the left is synthesized continuously in the 5′ → 3′ direction; the green daughter strand on the right is synthesized discontinuously in the 5′ → 3′ direction.

All the reactions involved in nucleic acid synthesis are catalyzed by enzymes. The synthesis of DNA takes place in a region of the molecule where the strands have started to separate. Because a nucleic acid can be synthesized only in the 5′ → 3′ direction, only the daughter strand on the left in Figure 26.8 is synthesized continuously in a single piece (because it is synthesized in the 5′ → 3′ direction).

LEARN THE STRATEGY

The other daughter strand needs to grow in a 3′ → 5′ direction, so it is synthesized discontinuously in small pieces. Each piece is synthesized in the 5′ → 3′ direction, and the fragments are joined together by an enzyme called DNA ligase (see Figure 21.10 on page 1021). Each of the two new molecules of DNA—called daughter molecules—contains one of the original parent strands (blue strand in Figure 26.8) plus a newly synthesized strand (green strand). This process is called **semiconservative replication.**

USE THE STRATEGY

PROBLEM 7

Using a dark line for the original parental DNA and a wavy line for DNA synthesized from parental DNA, show what the population of DNA molecules would look like in the fourth generation. (Parental DNA is the first generation.)

26.6 DNA AND HEREDITY

If DNA contains hereditary information, there must be a method to decode that information. The decoding occurs in two steps.

1. The sequence of bases in DNA provides a blueprint for the synthesis of RNA; the synthesis of RNA from a DNA blueprint is called **transcription** (Section 26.7).