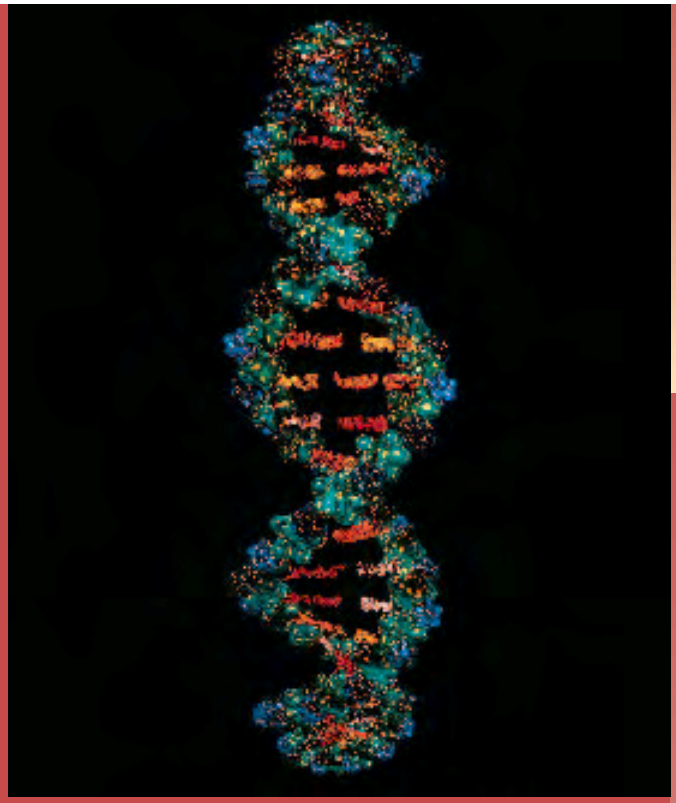
© The McGraw–Hill Companies, 2003



Computer-generated model of DNA.

24

Introduction to Molecular Genetics

Learning Goals

1. Draw the general structure of DNA and RNA nucleotides.
2. Describe the structure of DNA and compare it with RNA.
3. Explain DNA replication.
4. List three classes of RNA molecules and describe their functions.
5. Explain the process of transcription.
6. List and explain the three types of post-transcriptional modifications of eukaryotic mRNA.
7. Describe the essential elements of the genetic code, and develop a "feel" for its elegance.
8. Describe the process of translation.
9. Define mutation and understand how mutations cause cancer and cell death.
10. Describe the tools used in the study of DNA and in genetic engineering.
11. Describe the process of polymerase chain reaction and discuss potential uses of the process.
12. Discuss strategies for genome analysis and DNA sequencing.

BIOCHEMISTRY

Outline

Chemistry Connection: *Molecular Genetics and Detection of Human Genetic Disease*

24.1 The Structure of the Nucleotide
- Chemical Composition of DNA and RNA
- Nucleotide Structure

24.2 The Structure of DNA and RNA
- DNA Structure: The Double Helix
- Chromosomes

A Clinical Perspective: *Fooling the AIDS Virus with "Look-Alike" Nucleotides*
- RNA Structure

24.3 DNA Replication

24.4 Information Flow in Biological Systems
- Classes of RNA Molecules
- Transcription
- Post-transcriptional Processing of RNA

24.5 The Genetic Code

24.6 Protein Synthesis
- The Role of Transfer RNA
- The Process of Translation

24.7 Mutation, Ultraviolet Light, and DNA Repair
- The Nature of Mutations
- The Results of Mutations
- Mutagens and Carcinogens
- Ultraviolet Light Damage and DNA Repair

A Clinical Perspective: *The Ames Test for Carcinogens*
- Consequences of Defects in DNA Repair

24.8 Recombinant DNA
- Tools Used in the Study of DNA

A Human Perspective: *DNA Fingerprinting*
- Genetic Engineering

24.9 Polymerase Chain Reaction

A Medical Perspective: *A Genetic Approach to Familial Emphysema*

24.10 The Human Genome Project
- Genetic Strategies for Genome Analysis
- DNA Sequencing

Summary

Key Terms

Questions and Problems

Critical Thinking Problems

General, Organic, *and* Biochemistry

Fourth Edition

Denniston • Topping • Caret

CHEMISTRY CONNECTION

Molecular Genetics and Detection of Human Genetic Disease

It is estimated that 3–5% of the human population suffers from a serious genetic defect. That's 200 million people! But what if genetic disease could be detected and "cured"? Two new technologies, *gene therapy* and *preimplantation diagnosis*, may help us realize this dream.

For a couple with a history of a genetic disease in the family, pregnancy is a time of anxiety. Through *genetic counseling* these couples can learn the probability that their child has the disease. For about 200 genetic diseases the uncertainty can be eliminated. *Amniocentesis* (removal of 10–20 mL of fluid from the sac around the fetus) and *chorionic villus* sampling (removal of cells from a fetal membrane) are two procedures that are used to obtain fetal cells for genetic testing. Fetal cells are cultured and tested by enzyme assays and DNA tests to look for genetic diseases. If a genetic disease is diagnosed, the parents must make a difficult decision: to abort the fetus or to carry the child to term and deal with the effects of the genetic disease.

The power of modern molecular genetics is obvious in our ability to find a "bad" gene from just a few cells. But scientists have developed an even more impressive way to test for genetic disease before the embryo implants into the uterine lining. This technique, called *preimplantation diagnosis*, involves fertilizing a human egg and allowing the resulting conceptus to divide in a sterile petri dish. When the conceptus consists of 8–16 cells, *one* cell is removed for genetic testing. Only genetically normal embryos are implanted in the mother. Thus the genetic diseases that we can detect could be eliminated from the population by preimplantation diagnosis because only a conceptus with "good" genes is used.

Gene therapy is a second way in which genetic diseases may one day be eliminated. Foreign genes, including growth hormone, have been introduced into fertilized mouse eggs and the conceptuses implanted in female mice. The baby mice born with the foreign growth hormone gene were about three times larger than their normal littermates! One day, this kind of technology may be used to introduce normal genes into human fertilized eggs carrying a defective gene, thereby replacing the defective gene with a normal one.

In this chapter we will examine the molecules that carry and express our genetic information, DNA and RNA. Only by understanding the structure and function of these molecules have we been able to develop the amazing array of genetic tools that currently exists. We hope that as we continue to learn more about human genetics, we will be able to detect and one day correct most of the known genetic diseases.

Introduction

Look around at the students in your chemistry class. They all share many traits: upright stance, a head with two eyes, a nose, and a mouth facing forward, one ear on each side of the head, and so on. You would have no difficulty listing the similarities that define you and your classmates as *Homo sapiens*.

As you look more closely at the individuals you begin to notice many differences. Eye color, hair color, skin color, the shape of the nose, height, body build: all these traits, and many more, show amazing variety from one person to the next. Even within one family, in which the similarities may be more pronounced, each individual has a unique appearance. In fact, only identical twins look exactly alike—well, most of the time.

The molecule responsible for all these similarities and differences is deoxyribonucleic acid (DNA). Tightly wound up in structures called *chromosomes* in the nucleus of the cell, DNA carries the genetic code to produce the thousands of different proteins that make us who we are. These proteins include enzymes that are responsible for production of the pigment melanin. The more melanin we are genetically programmed to make, the darker our hair, eyes, and skin will be. Others are structural proteins. The gene for α-keratin that makes up hair determines whether our hair will be wavy, straight, or curly. Thousands of genes carry the genetic information for thousands of proteins that dictate our form and, some believe, our behavior.

Genetic traits are passed from one generation to the next. When a sperm fertilizes an ovum, a conceptus is created from a single set of maternal chromosomes and a single set of paternal chromosomes. As this fertilized egg divides, each daughter cell will receive one copy of each of these chromosomes. The genes on

these chromosomes will direct fetal development, from that fertilized cell to a newborn with all the characteristics we recognize as human.

In this chapter we will explore the structure of DNA and the molecular events that translate the genetic information of a gene into the structure of a protein.

24.1 The Structure of the Nucleotide

Chemical Composition of DNA and RNA

Learning Goal 1

DNA was discovered by Friedrich Meischer in 1869. However, it was not recognized as the genetic information until 1950. In 1953 James Watson and Francis Crick published a paper describing the structure of the DNA molecule. From that paper and many other studies we know that **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)** are long polymers of **nucleotides.** Every nucleotide is composed of three units: a five-carbon sugar, a nitrogenous base, and either one, two, or three phosphoryl groups. Nitrogenous bases are heterocyclic ring structures having backbones consisting of carbon and nitrogen atoms. There are two classes of nitrogenous bases. The **purine** nitrogenous bases consist of a six-member ring fused to a five-member ring. The **pyrimidine** nitrogenous bases consist of a single six-member ring. The structures are seen in Figure 24.1.

The five-carbon sugar in RNA is ribose, and the sugar in DNA is 2′-deoxyribose. The only difference between these two sugars is the absence of an hydroxyl group on the 2′ carbon of 2′-deoxyribose. The purines in both DNA and RNA are adenine and guanine. Both DNA and RNA contain the pyrimidine cytosine; however, the fourth base is thymine in DNA and uracil in RNA. The chemical compositions of DNA and RNA are summarized in Table 24.1.

The carbon atoms of the sugars found in nucleic acids are indicated with a prime: 1′, 2′, 3′, and so on. This is to distinguish them from the ring atoms of the nitrogenous bases.

In addition, the nucleotides that make up DNA and RNA contain phosphoryl groups. Nucleotides may be *mono*-, *di*-, or *tri*phosphates.

Nucleotide Structure

Nucleotides are produced by the combination of a sugar, a nitrogenous base, and at least one phosphoryl group, as shown in Figure 24.2. Because this large structure contains two cyclic molecules, the sugar and the base, we must have an easy way to describe the ring atoms of each. For this reason the ring atoms of the sugar are designated with a prime to distinguish them from atoms in the base. The covalent bond between the sugar and the phosphoryl group is a phosphoester linkage formed by a condensation reaction between the 5′-OH of the sugar and an —OH of the phosphoryl group. The bond between the base and the sugar is a β-*N*-glycosidic linkage, and it joins the 1′-carbon of the sugar and a nitrogen atom of the nitrogenous base (N-9 of purines and N-1 of pyrimidines).

To name a nucleotide, simply begin with the name of the nitrogenous base, and apply the following simple rules:

- Remove the *-ine* ending, and replace it with either *-osine* for purines or *-idine* for pyrimidines. Uracil is the one exception to this rule. In this case the *-acil* ending is replaced with *-idine,* producing the name *uridine.*
- Nucleotides with the sugar ribose are **ribonucleotides,** and those having the sugar 2′-deoxyribose are **deoxyribonucleotides.** For a deoxyribonucleotide, the prefix *deoxy-* is placed before the modified nitrogenous base name. No prefix is required for ribonucleotides, or for thymidine, which is found only in DNA.
- Add a prefix to indicate the number of phosphoryl groups that are attached. A *mono*phosphate carries one phosphoryl group; a *di*phosphate carries two phosphoryl groups; and a *tri*phosphate carries three phosphoryl groups.

In diphosphates and triphosphates the phosphoryl groups are bonded to one another through phosphoanhydride bonds (Section 15.4).

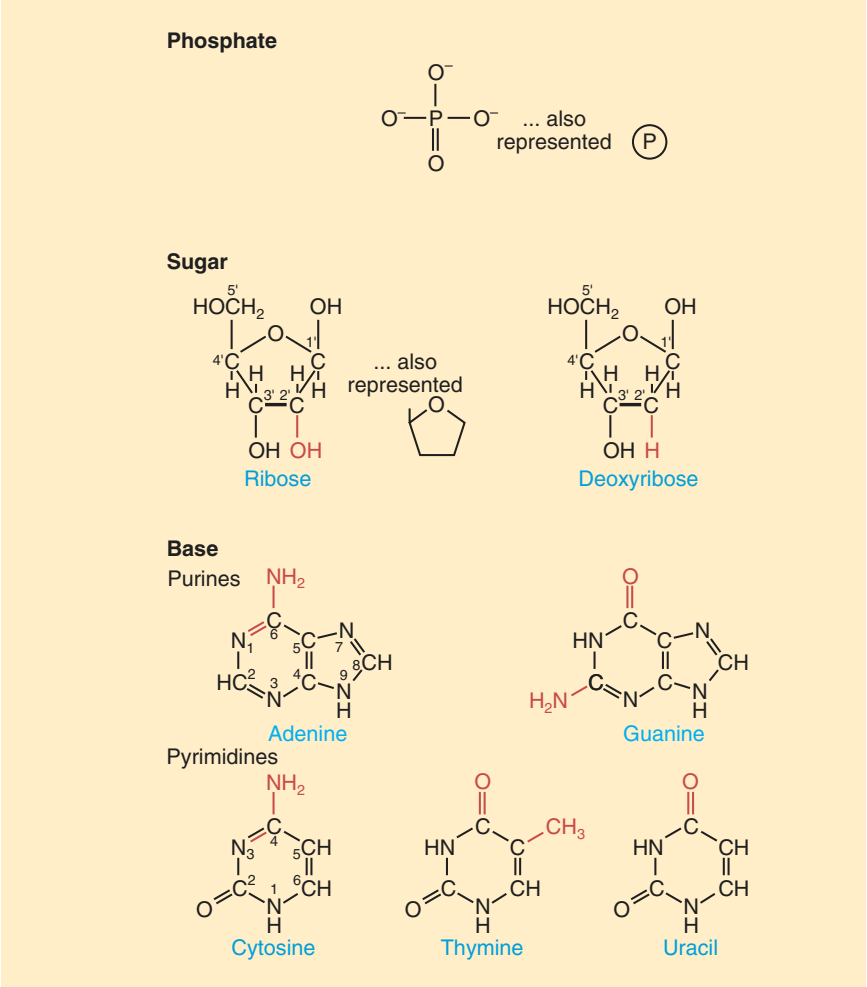


Figure 24.1
The components of nucleic acids include phosphate groups, the five-carbon sugars ribose and deoxyribose, and purine and pyrimidine nitrogenous bases. The ring positions of the sugars are designated with primes (′) to distinguish them from the ring positions of the bases.

Table 24.1 Chemical Composition of Nucleic Acids

	DNA	RNA
Sugar	2′-Deoxyribose	Ribose
Purine nitrogenous bases	Adenine (A)	Adenine (A)
	Guanine (G)	Guanine (G)
Pyrimidine nitrogenous bases	Cytosine (C)	Cytosine (C)
	Thymine (T)	Uracil (U)

Because the full names of the nucleotides are so cumbersome, a simple abbreviation is generally used. These abbreviations are summarized in Table 24.2.

Question 24.1

Referring to the structures in Figures 24.1 and 24.2, draw the structures and write the names for nucleotides consisting of the following units.

a. Ribose, adenine, two phosphoryl groups
b. 2′-Deoxyribose, guanine, three phosphoryl groups

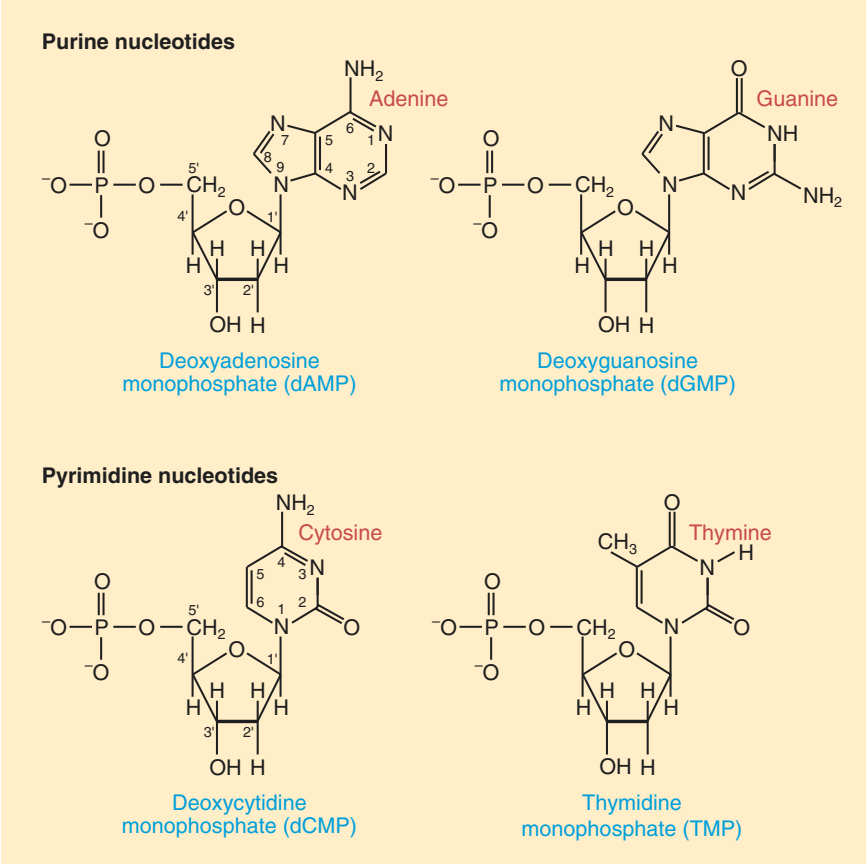


Figure 24.2
Structures, names, and common abbreviations of four deoxyribonucleotides or deoxyribonucleotide-5′-phosphates.

Table 24.2 **Names and Abbreviations of the Ribonucleotides and Deoxyribonucleotides Containing Adenine**

Nucleotide	Abbreviation
Deoxyadenosine monophosphate	dAMP
Deoxyadenosine diphosphate	dADP
Deoxyadenosine triphosphate	dATP
Adenosine monophosphate	AMP
Adenosine diphosphate	ADP
Adenosine triphosphate	ATP

Question 24.2

Referring to the structures in Figures 24.1 and 24.2, draw the structures and write the names for nucleotides consisting of the following units:

a. 2′-Deoxyribose, thymine, one phosphoryl group
b. Ribose, cytosine, three phosphoryl groups
c. Ribose, uracil, one phosphoryl group

Question 24.3

Write the names and abbreviations of the deoxyribonucleotides and ribonucleotides of guanine.

Question 24.4

Write the names and abbreviations of the deoxyribonucleotides and ribonucleotides of cytosine.

24.2 The Structure of DNA and RNA

Learning Goal 2

A single strand of DNA is a polymer of nucleotides bonded to one another by 3′–5′ phosphodiester bonds. The backbone of the polymer is called the *sugar-phosphate backbone* because it is composed of alternating units of the five-carbon sugar 2′-deoxyribose and phosphoryl groups in phosphodiester linkage. A nitrogenous base is bonded to each sugar by an *N*-glycosidic linkage (Figure 24.3).

DNA Structure: The Double Helix

James Watson and Francis Crick were the first to describe the three-dimensional structure of DNA in 1953. They deduced the structure by building models based on the experimental results of others. Irwin Chargaff observed that the amount of adenine in any DNA molecule is equal to the amount of thymine. Similarly, he found that the amounts of cytosine and guanine are also equal. The X-ray diffraction studies of Rosalind Franklin and Maurice Wilkens revealed several repeat

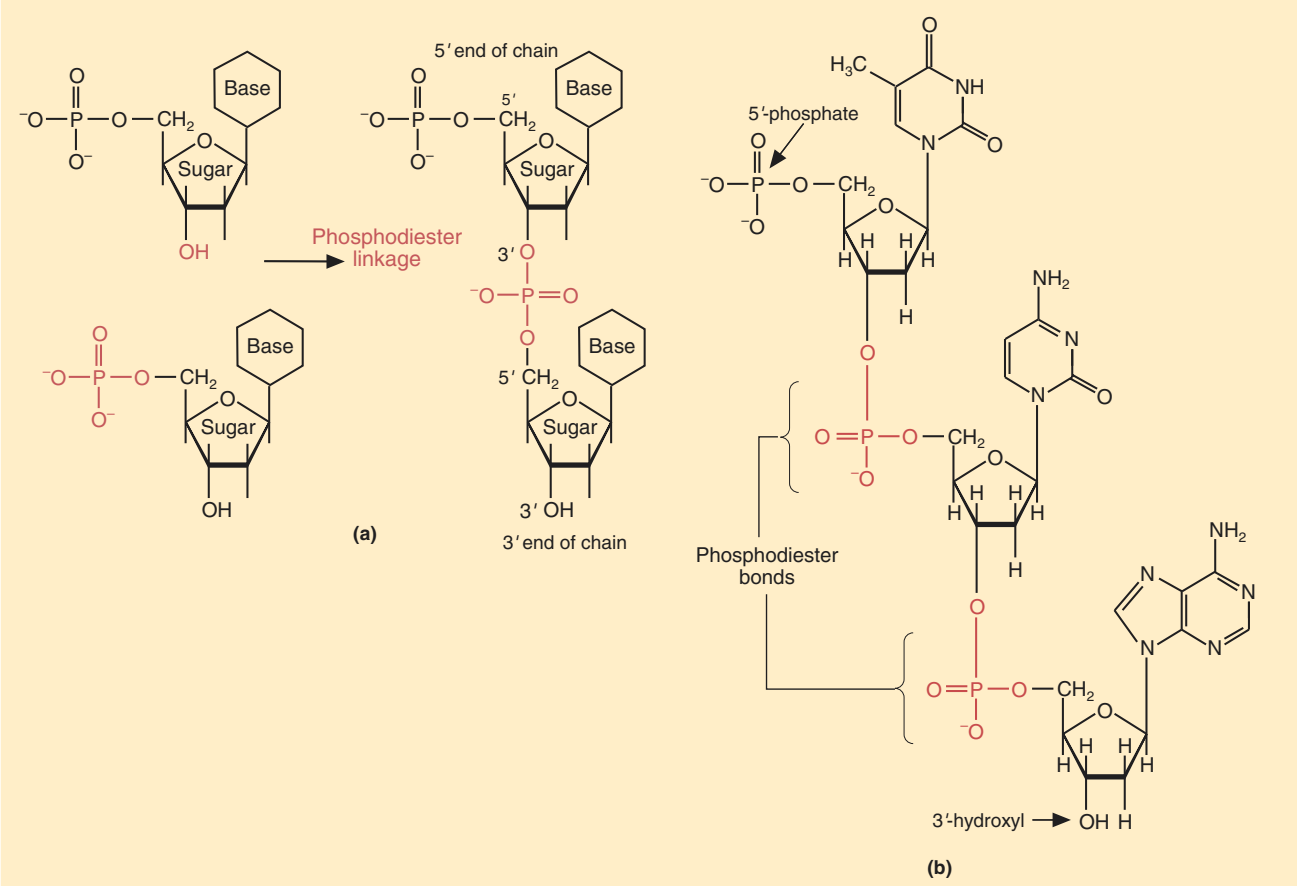


Figure 24.3
The covalent, primary structure of DNA. (a) The esterification reaction by which two nucleotides become linked by a phosphodiester bond. (b) A series of three covalently linked deoxyribonucleotides.

distances that characterize the structure of DNA: 0.34 nm, 3.4 nm, and 2 nm. (Look at the structure of DNA in Figure 24.4 to see the significance of these measurements.) With this information Watson and Crick concluded that DNA is a **double helix** of two strands of DNA wound around one another. It is useful to compare the structure of the double helix to a spiral staircase. The sugar-phosphate backbones of the two strands of DNA spiral around the outside of the helix like the handrails on a spiral staircase. The nitrogenous bases extend into the center at right angles to the axis of the helix. You can imagine the nitrogenous bases forming the steps of the staircase. The structure of this elegant molecule is shown in Figure 24.4.

The two strands of DNA are held together by hydrogen bonds between the nitrogenous bases in the center of the helix. Adenine forms two hydrogen bonds with

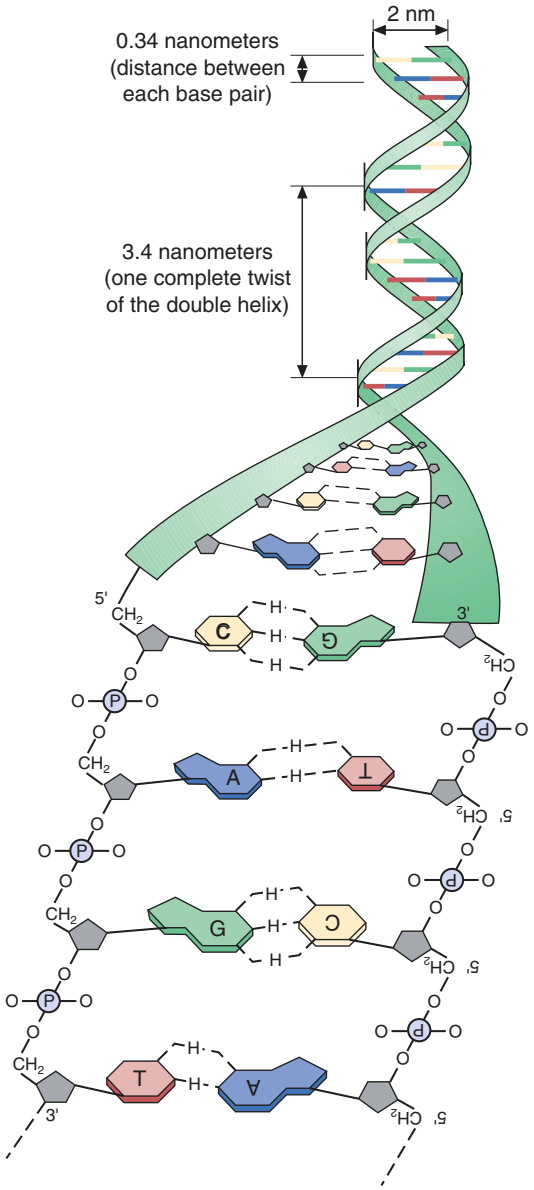
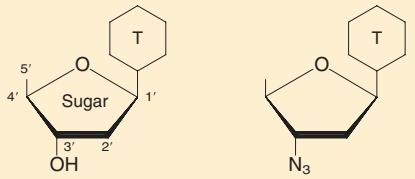


Figure 24.4
Schematic ribbon diagram of the DNA double helix showing the dimensions of the DNA molecule and the antiparallel orientation of the two strands.

A CLINICAL PERSPECTIVE

Fooling the AIDS Virus with "Look-Alike" Nucleotides

The virus that is responsible for the acquired immune deficiency syndrome (AIDS) is called the *human immunodeficiency virus,* or *HIV.* HIV is a member of a family of viruses called *retroviruses,* all of which have single-stranded RNA as their genetic material. The RNA is copied by a viral enzyme called *reverse transcriptase* into a double-stranded DNA molecule. This process is the opposite of the central dogma, which states that the flow of genetic information is from DNA to RNA. But these viruses reverse that flow, RNA to DNA. For this reason these viruses are called retroviruses, which literally means "backward viruses." The process of producing a DNA copy of the RNA is called *reverse transcription.*



2′-Deoxythymidine 3′-Azido-2′,3′-dideoxythymidine (AZT)

Comparison of the structures of the normal nucleoside, 2′-deoxythymidine, and the nucleoside analog, 3′-azido-2′, 3′-dideoxythymidine.

Because our genetic information is DNA and it is expressed by the classical DNA → RNA → protein pathway, our cells have no need for a reverse transcriptase enzyme. Thus the HIV reverse transcriptase is a good target for antiviral chemotherapy because inhibition of reverse transcription should kill the virus but have no effect on the human host. Many drugs have been tested for the ability to selectively inhibit HIV reverse transcription. Among these is the DNA chain terminator 3′-azido-2′, 3′-dideoxythymidine, commonly called *AZT* or *zidovudine.*

How does AZT work? It is one of many drugs that looks like one of the normal nucleosides. These are called *nucleoside analogs.* A nucleoside is just a nucleotide without any phosphate groups attached. The analog is phosphorylated by the cell and then tricks a polymerase, in this case viral reverse transcriptase, into incorporating it into the growing DNA chain in place of the normal phosphorylated nucleoside. AZT is a nucleoside analog that looks like the nucleoside thymidine except that in the 3′ position of the deoxyribose sugar there is an azido group (—N_3) rather than the 3′-OH group. Compare the structures of thymidine and AZT shown in the accompanying figure. The 3′-OH group is necessary for further DNA polymerization because it is there that the phosphoester linkage must be made between the growing DNA strand and the next nucleotide. If an azido group or some other group is present at the 3′ position, the nucleotide analog can be incorporated into the growing DNA strand, but further chain elongation is blocked, as shown in the following figure. If the viral RNA cannot be reverse transcribed into the DNA form, the virus will not be able to replicate and can be considered to be dead.

AZT is particularly effective because the HIV reverse transcriptase actually prefers it over the normal nucleotide, thymidine. Nonetheless, AZT is not a cure. At best it prolongs the life

thymine, and cytosine forms three hydrogen bonds with guanine (Figures 24.4 and 24.5). These are called **base pairs.** The two strands of DNA are **complementary strands** because the sequence of bases on one automatically determines the sequence of bases on the other. When there is an adenine on one strand, there will always be a thymine in the same location on the opposite strand.

The diameter of the double helix is 2.0 nm. This is dictated by the dimensions of the purine-pyrimidine base pairs. The helix completes one turn every ten base pairs. One complete turn is 3.4 nm. Thus each base pair advances the helix by 0.34 nm.

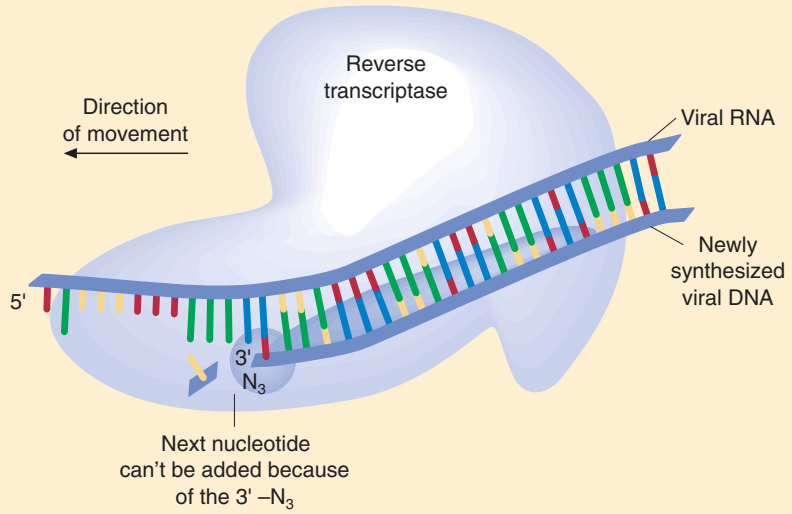
One last important feature of the DNA double helix is that the two strands are **antiparallel strands,** as this example shows:

```
5′  P—S—P—S—P—S—P—S—P—S—P—S—OH 3′
       |   |   |   |   |   |
       A   T   G   C   G   A
       :   :   :   :   :   :
       T   A   C   G   C   T
       |   |   |   |   |   |
3′ OH—S—P—S—P—S—P—S—P—S—P—S—P  5′
```

of a person with AIDS for a year or two. Eventually, however, AZT has a negative effect on the body. The cells of our bone marrow are constantly dividing to produce new blood cells: red blood cells to carry oxygen to the tissues, white blood cells of the immune system, and platelets for blood clotting. For cells to divide, they must replicate their DNA. The DNA polymerases of these dividing cells also accidentally incorporate AZT into the growing DNA chains with the result that cells of the bone marrow begin to die. This can result in anemia and even further depression of the immune response.

Another problem that has arisen with prolonged use of AZT is that AZT-resistant mutants of the virus appear. It is well known that HIV is a virus that mutates rapidly. Some of these mutant forms of the virus have an altered reverse transcriptase that will no longer use AZT. When these mutants appear, AZT is no longer useful in treating the infection.

It is hoped that research with other nucleoside analogs, alternative types of antiviral treatments, and combinations of drugs will provide a means of effectively treating HIV infection. Such a therapy must have fewer toxic side effects while stopping the replication of the virus and the progress of the disease.



The mechanism by which AZT inhibits HIV reverse transcriptase. Incorporation of AZT into the growing HIV DNA strand in place of deoxythymidine results in DNA chain termination; the azido group on the 3′ carbon of the sugar cannot react to produce the phosphoester linkage required to add the next nucleotide.

In other words, the two strands of the helix run in opposite directions (see Figure 24.4). Only when the two strands are antiparallel can the base pairs form the hydrogen bonds that hold the two strands together.

Chromosomes

Chromosomes are pieces of DNA that carry the genetic instructions, or genes, of an organism. Organisms such as the prokaryotes have only a single chromosome and its structure is relatively simple. Others, the eukaryotes, have many chromosomes, each of which has many different levels of structure. The complete set of genetic information in all the chromosomes of an organism is called the **genome.**

Prokaryotes are organisms with a simple cellular structure in which there is no true nucleus surrounded by a nuclear membrane and there are no true membrane-bound organelles. This group includes all of the bacteria. In these organisms the chromosome is a circular DNA molecule that is supercoiled, which means that the helix is coiled on itself. The supercoiled DNA molecule is attached to a complex of proteins at roughly forty sites along its length, forming a series of loops. This structure, called the nucleoid, can be seen in Figure 24.6.

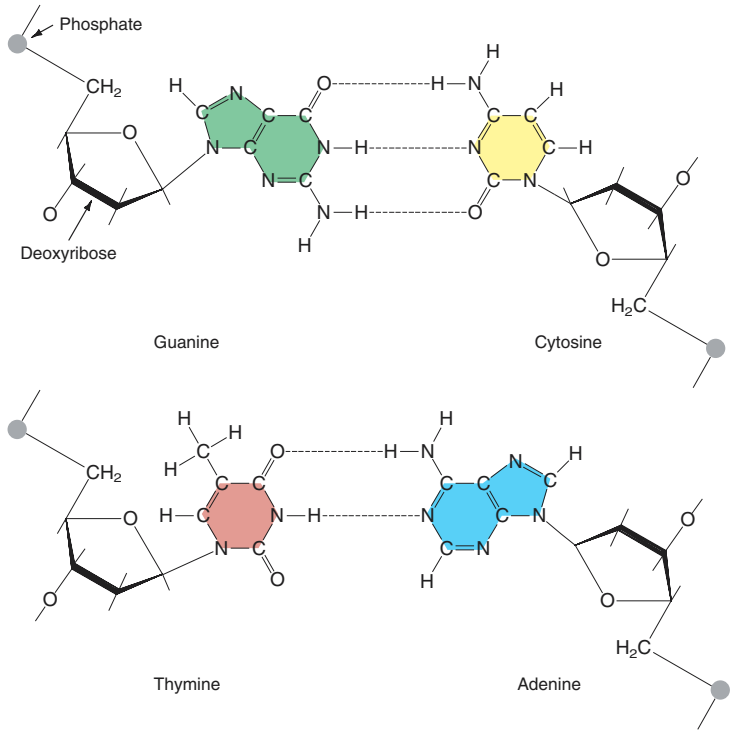


Figure 24.5
Base pairing in DNA. Adenine is always paired with thymine (A—T), and guanine is paired with cytosine (G—C).

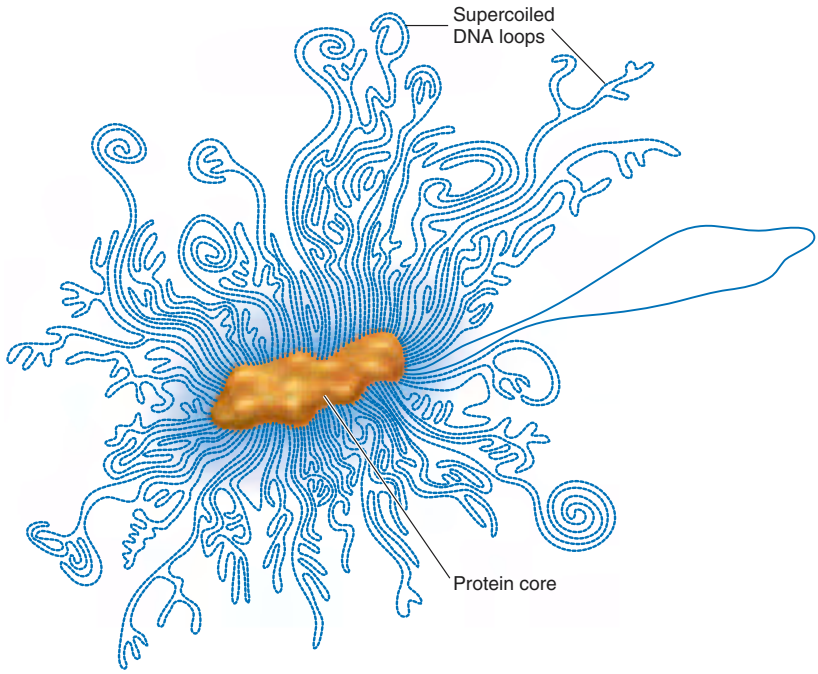


Figure 24.6
Structure of a bacterial nucleoid. The nucleoid is made up of the supercoiled, circular chromosome attached to a protein core.

Eukaryotes are organisms that have cells containing a true nucleus enclosed by a nuclear membrane. They also have a variety of membrane-bound organelles that segregate different cellular functions into different compartments. As an example, the processes of aerobic respiration are located within the mitochondria.

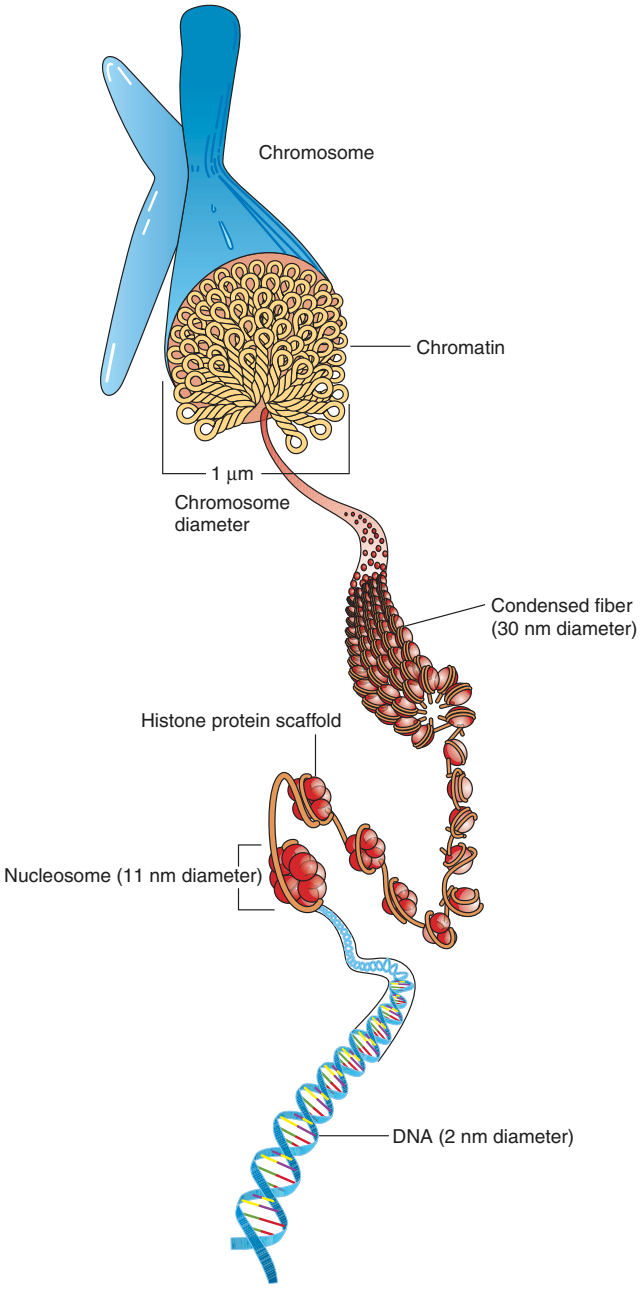


Figure 24.7
The eukaryotic chromosome has many levels of structure.

All animals, plants, and fungi are eukaryotes. The number and size of the chromosomes of eukaryotes vary from one species to the next. For instance, humans have 23 pairs of chromosomes, while the Adder's Tongue fern has 631 pairs of chromosomes. But the chromosome structure is the same for all those organisms that have been studied.

Eukaryotic chromosomes are very complex structures (Figure 24.7). The first level of structure is the **nucleosome,** which consists of a strand of DNA wrapped around a small disk made up of histone proteins. At this level the DNA looks like beads along a string. The string of beads then coils into a larger structure called the

30 nm fiber. These, in turn, are further coiled into a *200 nm fiber.* Other proteins are probably involved in the organization of the 200 nm fiber. The full complexities of the eukaryotic chromosome are not yet understood, but there are probably many such levels of coiled structures.

RNA Structure

The sugar-phosphate backbone of RNA consists of ribonucleotides, also linked by 3′–5′ phosphodiester bonds. These phosphodiester bonds are identical to those found in DNA. However, RNA molecules differ from DNA molecules in three basic properties.

- RNA molecules are usually single-stranded.
- The sugar-phosphate backbone of RNA consists of *ribonucleotides* linked by 3′–5′ phosphodiester bonds. Thus the sugar ribose is found in place of 2′-deoxyribose.
- The nitrogenous base uracil (U) replaces thymine (T).

Although RNA molecules are single-stranded, base pairing between uracil and adenine and between guanine and cytosine can still occur. We will show the importance of this property as we examine the way in which RNA molecules are involved in the expression of the genetic information in DNA.

24.3 DNA Replication

Learning Goal 3

DNA must be replicated before a cell divides so that each daughter cell inherits a copy of each gene. A cell that is missing a critical gene will die, just as an individual with a genetic disease, a defect in an important gene, may die early in life. Thus it is essential that the process of DNA replication produces an absolutely accurate copy of the original genetic information. If mistakes are made in critical genes, the result may be lethal mutations.

The first step in DNA replication is the separation of the strands of DNA. Proteins do this by breaking the hydrogen bonds between the base pairs. Then the enzyme **DNA polymerase** "reads" each parental strand, also called the *template,* and catalyzes the polymerization of a complementary daughter strand. Deoxyribonucleotide triphosphate molecules are the precursors for DNA replication. However, the last two phosphoryl groups are cleaved away in the process. This cleavage releases the energy needed by DNA polymerase to form the phosphoester linkage between the 3′-OH of 2′-deoxyribose and the 5′-phosphoryl group of the deoxyribonucleotide monophosphate to be added to the DNA chain.

Because DNA polymerase "reads" each parental strand and produces a new complementary daughter DNA strand, each new DNA molecule consists of one parental strand and one newly synthesized daughter strand. This mode of DNA replication is called **semiconservative replication** (Figure 24.8).

DNA polymerase can catalyze a reaction only between the 5′-phosphoryl group on a nucleotide to be added to the growing chain and the hydroxyl group on the growing daughter strand. Thus, the parental DNA strands can be copied in only one direction. This is a problem, since the two strands of a DNA molecule are antiparallel. As a result, only one daughter strand can be produced continuously from the **replication fork,** the point at which new nucleotides are added to the growing daughter strand.

The other strand must grow in the opposite direction of the moving replication fork. For this reason, it must be synthesized in short segments. Each of these is started at the replication fork as it moves in the opposite direction. The new pieces of DNA are covalently linked to one another by the enzyme DNA ligase (Figure 24.9).

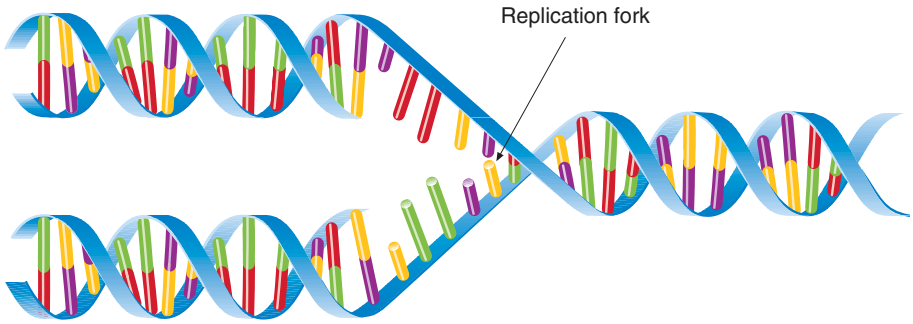


Figure 24.8
In semiconservative DNA replication, each parent strand serves as a template for the synthesis of a new daughter strand.

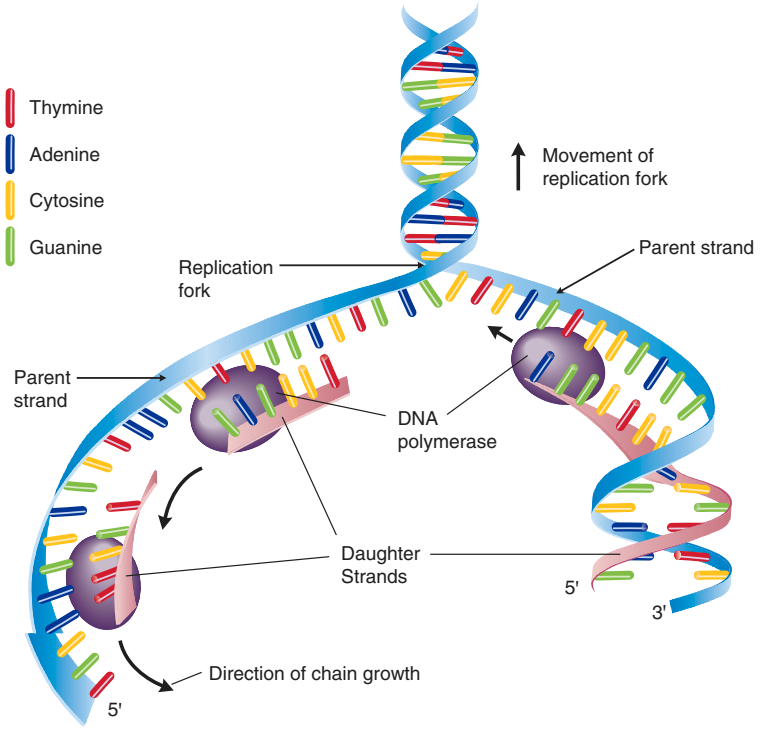


Figure 24.9
Because the two strands of DNA are antiparallel and DNA polymerase can only catalyze 5′ → 3′ replication, only one of the two DNA strands (on the right) can be read continuously to produce a daughter strand. The other must be synthesized in segments that are extended away from the direction of movement of the replication fork (on the left). These discontinuous segments are later covalently joined together by DNA ligase.

Because it is critical to produce an accurate copy of the parental DNA, it is very important to avoid errors in the replication process. In addition to catalyzing the replication of new DNA strands, DNA polymerase is able to *proofread* the newly synthesized strand. If the wrong nucleotide has been added to the growing DNA strand, it is removed and replaced with the correct one. In this way a faithful copy of the parental DNA is ensured.

All the genetic information of bacteria such as *E. coli* is contained on a single circular piece of DNA made up of about three million nucleotide pairs and called the *chromosome.* DNA replication in *E. coli* begins at a unique sequence on the circular chromosome known as the **replication origin.** Replication occurs bidirectionally at the rate of about five hundred new nucleotides every second! Because DNA synthesis occurs bidirectionally, there are two replication forks moving in opposite directions. Replication is complete when the two replication forks meet halfway around the circular chromosome.

DNA replication in eukaryotes is more complex. The human genome consists of approximately three billion nucleotide pairs. Just one chromosome may be nearly one hundred times longer than a bacterial chromosome. To accomplish this huge job, DNA replication begins at many replication origins and proceeds bidirectionally along each chromosome.

24.4 Information Flow in Biological Systems

The **central dogma** of molecular biology states that in cells the flow of genetic information contained in DNA is a one-way street that leads from DNA to RNA to protein. The process by which a single strand of DNA serves as a template for the synthesis of an RNA molecule is called **transcription.** The word *transcription* is derived from the Latin word *transcribere* and simply means "to make a copy." Thus in this process, part of the information in the DNA is copied into a strand of RNA. The process by which the message is converted into protein is called **translation.** Unlike transcription the process of translation involves converting the information from one language to another. In this case the genetic information in the linear sequence of nucleotides is being translated into a protein, a linear sequence of amino acids. The expression of the information contained in DNA is fundamental to the growth, development, and maintenance of all organisms.

Classes of RNA Molecules

Learning Goal 4

Three classes of RNA molecules are produced by transcription: messenger RNA, transfer RNA, and ribosomal RNA.

1. **Messenger RNA (mRNA)** carries the genetic information for a protein from DNA to the ribosomes. It is a complementary RNA copy of a gene on the DNA.
2. **Ribosomal RNA (rRNA)** is a structural and functional component of the ribosomes, which are "platforms" on which protein synthesis occurs. There are three types of rRNA molecules in bacterial ribosomes and four in the ribosomes of eukaryotes.
3. **Transfer RNA (tRNA)** translates the genetic code of the mRNA into the primary sequence of amino acids in the protein. In addition to the primary structure, tRNA molecules have a cloverleaf-shaped secondary structure resulting from base pair hydrogen bonding (A—U and G—C) and a roughly L-shaped tertiary structure (Figure 24.10). The sequence CCA is found at the 3′ end of the tRNA. The 3′–OH group of the terminal nucleotide, adenosine, can be covalently attached to an amino acid. Three nucleotides at the base of the cloverleaf structure form the **anticodon.** As we will discuss in more detail in Section 24.6, this triplet of bases forms hydrogen bonds to a **codon** (complementary sequence of bases) on a messenger RNA (mRNA) molecule on the surface of a ribosome during protein synthesis. This hydrogen bonding of codon and anticodon brings the correct amino acid to the site of protein synthesis at the appropriate location in the growing peptide chain (Figure 24.11).

Transcription

Learning Goal 5

Transcription, shown in Figure 24.12, is catalyzed by the enzyme **RNA polymerase.** The process occurs in three stages. The first, called *initiation,* involves binding of RNA polymerase to a specific nucleotide sequence, the **promoter,** at the beginning of a gene. This interaction of RNA polymerase with specific promoter DNA sequences allows RNA polymerase to recognize the start point for transcription. It also determines which DNA strand will be transcribed. Unlike DNA repli-

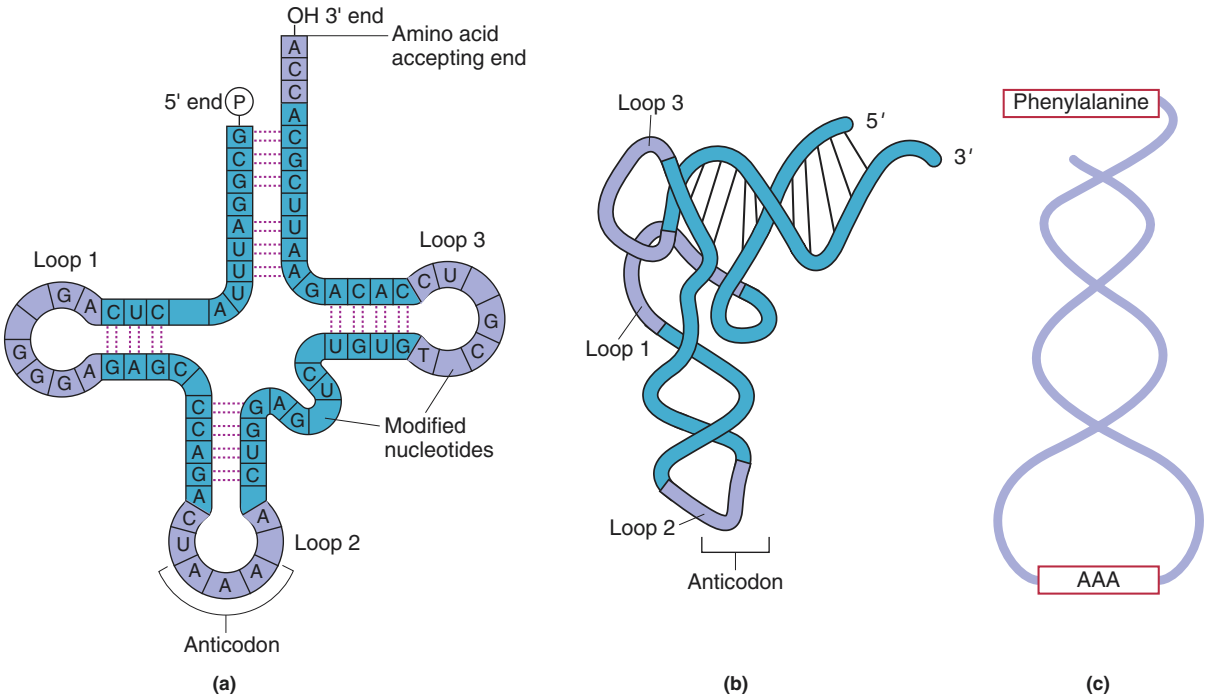


Figure 24.10
Structure of tRNA. (a) The primary structure of a tRNA is the linear sequence of ribonucleotides. Here we see the hydrogen-bonded secondary structure of a tRNA showing the three loops and the amino acid–accepting end. (b) The three-dimensional structure of a tRNA. (c) A schematic diagram that will be used to represent a tRNA throughout the chapter.

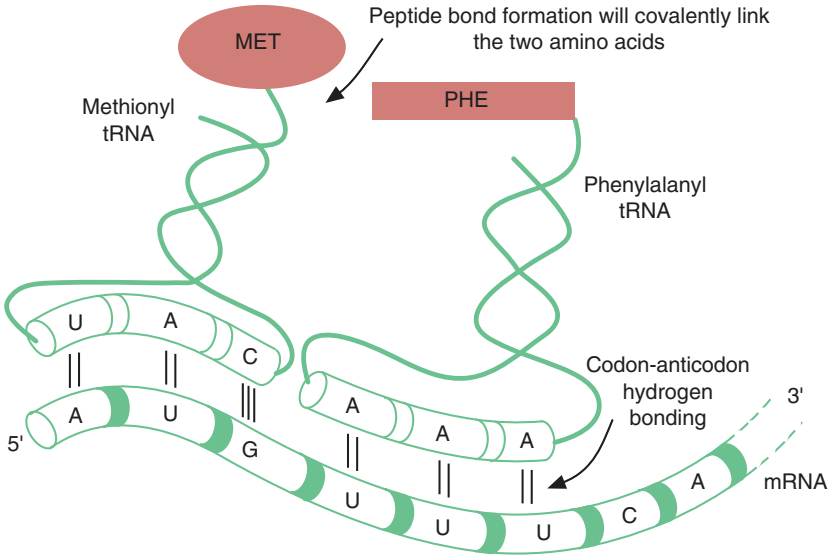


Figure 24.11
Codon-anticodon binding.

cation, transcription produces a complementary copy of only one of the two strands of DNA. As it binds to the DNA, RNA polymerase separates the two strands of DNA so that it can "read" the base sequence of the DNA.

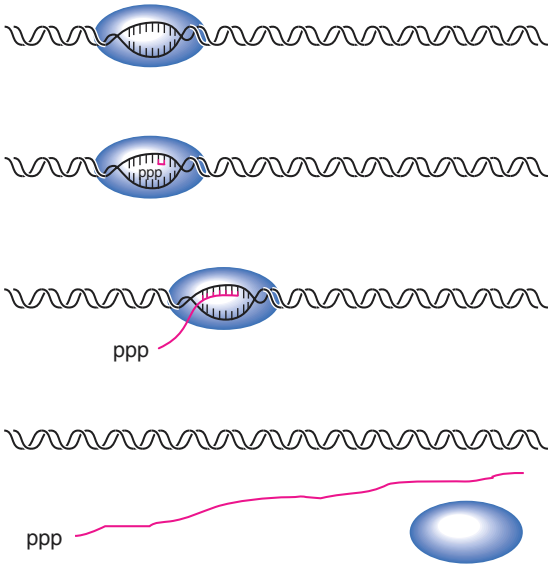


Figure 24.12
The stages of transcription.

The second stage, chain elongation, begins as the RNA polymerase "reads" the DNA template strand and catalyzes the polymerization of a complementary RNA copy. With each catalytic step, RNA polymerase transfers a complementary ribonucleotide to the end of the growing RNA chain and catalyzes the formation of a 3′–5′ phosphodiester bond between the 5′ phosphoryl group of the incoming ribonucleotide and the 3′ hydroxyl group of the last ribonucleotide of the growing RNA chain. This reaction is shown in Figure 24.13.

The final stage of transcription is termination. The RNA polymerase finds a termination sequence at the end of the gene and releases the newly formed RNA molecule.

Question 24.5 What is the function of RNA polymerase in the process of transcription?

Question 24.6 What is the function of the promoter sequence in the process of transcription?

Post-transcriptional Processing of RNA

Learning Goal 6

In bacteria, which are prokaryotes, termination releases a mature mRNA for translation. In fact, because prokaryotes have no nuclear membrane separating the DNA from the cytoplasm, translation begins long before the mRNA is completed. In eukaryotes, transcription produces a **primary transcript** that must undergo extensive **post-transcriptional modification** before it is exported out of the nucleus for translation in the cytoplasm.

Eukaryotic primary transcripts undergo three post-transcriptional modifications. These are the addition of a 5′ cap structure and a 3′ poly(A) tail, and RNA splicing.

In the first modification, a **cap structure** is enzymatically added to the 5′ end of the primary transcript. The cap structure (Figure 24.14) consists of 7-methylguanosine attached to the 5′ end of the RNA by a 5′–5′ triphosphate bridge. The first two nucleotides of the mRNA are also methylated. The cap structure is required for efficient translation of the final mature mRNA.

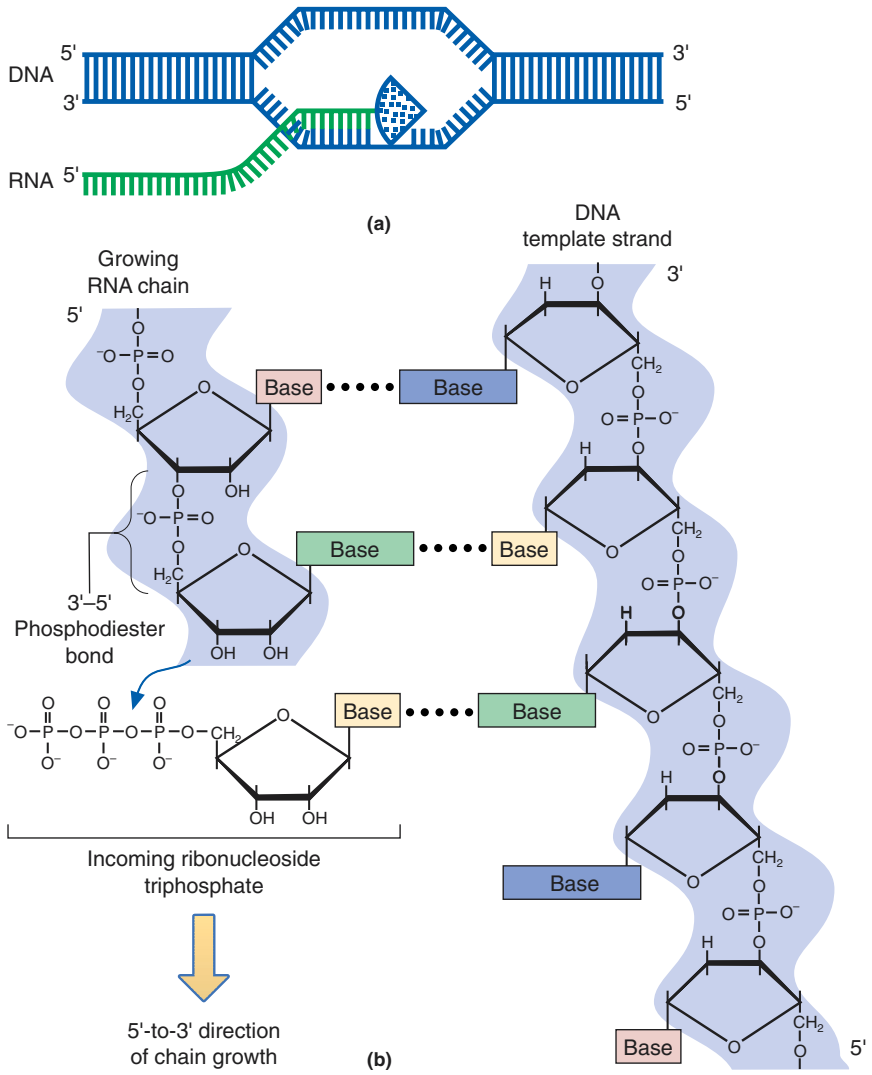


Figure 24.13
The reaction catalyzed by RNA polymerase. (a) RNA polymerase separates the two strands of DNA and produces an RNA copy of one of the two DNA strands. (b) Phosphodiester bond formation occurs as a nucleotide is added to the growing RNA chain.

The second modification is the enzymatic addition of a **poly(A) tail** to the 3′ end of the transcript. *Poly(A) polymerase* uses ATP and catalyzes the stepwise polymerization of 100–200 adenosine nucleotides on the 3′ end of the RNA. The poly(A) tail protects the 3′ end of the mRNA from enzymatic degradation and thus prolongs the lifetime of the mRNA.

The third modification, **RNA splicing,** involves the removal of portions of the primary transcript that are not protein coding. Bacterial genes are continuous; all the nucleotide sequences of the gene are found in the mRNA. However, study of the gene structure of eukaryotes revealed a fascinating difference. Eukaryotic genes are discontinuous; there are *extra* DNA sequences within these genes that do not encode any amino acid sequences for the protein. These sequences are called *intervening sequences* or **introns.** The primary transcript contains both the introns and the protein coding sequences, called **exons.** The presence of introns in the mRNA would make it impossible for the process of translation to synthesize the correct protein. Therefore they must be removed, which is done by the process of RNA splicing.

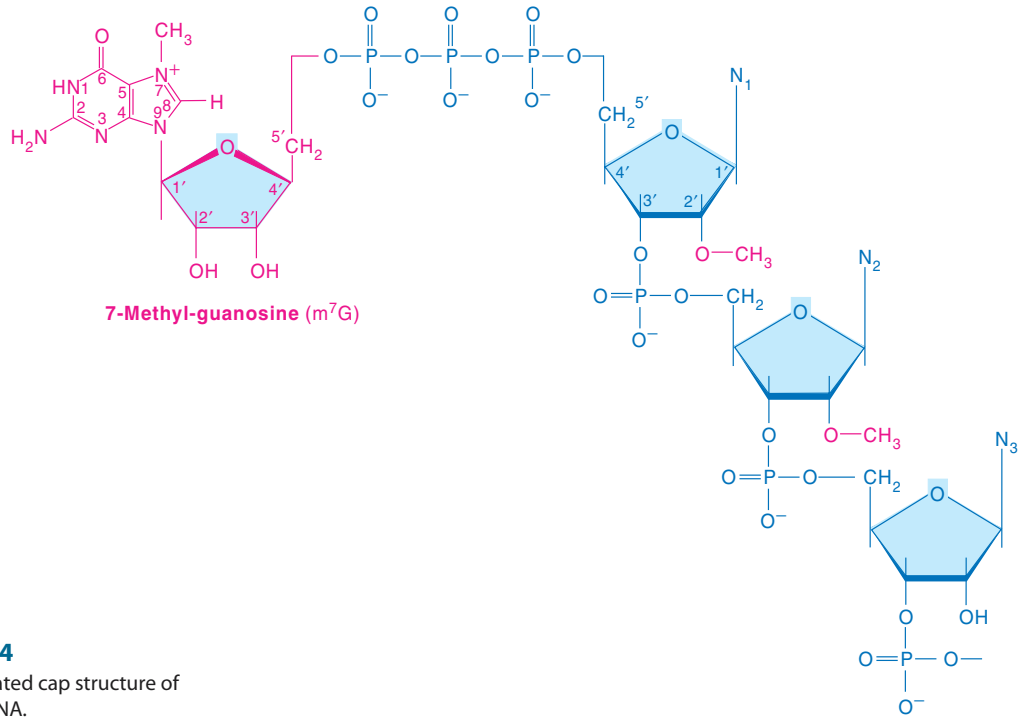


Figure 24.14
The 5′-methylated cap structure of eukaryotic mRNA.

As you can imagine, RNA splicing must be very precise. If too much, or too little, RNA is removed, the mRNA will not carry the correct code for the protein. Thus there are "signals" in the DNA to mark the boundaries of the introns. The sequence GpU is always found at the intron's 5′ boundary and the sequence ApG is found at the 3′ boundary.

Recognition of the splice boundaries and stabilization of the splicing complex requires the assistance of particles called *spliceosomes*. Spliceosomes are composed of a variety of *small nuclear ribonucleoproteins* (snRNPs, read "snurps"). Each snRNP consists of a small RNA and associated proteins. The RNA components of different snRNPs are complementary to different sequences involved in splicing. By hydrogen bonding to a splice boundary or intron sequences the snRNPs recognize and bring together the sequences involved in the splicing reactions.

One of the first eukaryotic genes shown to contain introns was the gene for the β subunit of adult hemoglobin (Figure 24.15). On the DNA the gene for β-hemoglobin is 1200 nucleotides long, but only 438 nucleotides carry the genetic information for protein. The remaining sequences are found in two introns of 116 and 646 nucleotides that are removed by splicing before translation. It is interesting that the larger intron is longer than the final β-hemoglobin mRNA! In the genes that have been studied, introns have been found to range in size from 50 to 20,000 nucleotides in length, and there may be many throughout a gene. Thus a typical human gene might be 10–30 times longer than the final mRNA.

24.5 The Genetic Code

Learning Goal 7

The mRNA carries the genetic code for a protein. But what is the nature of this code? In 1954, George Gamow proposed that because there are only four "letters" in the DNA alphabet (A, T, G, and C) and because there are twenty amino acids, the genetic code must contain words made of at least three letters taken from the