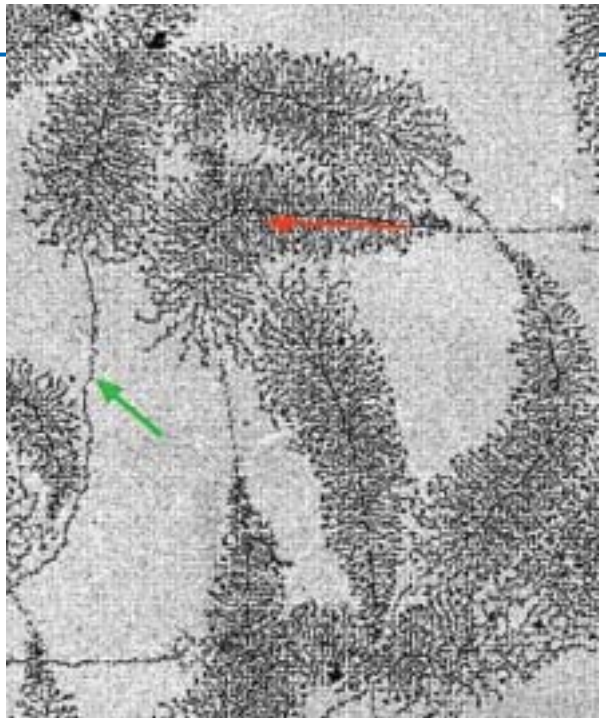
4

BASIC MOLECULAR GENETIC MECHANISMS



Electron micrograph of DNA (green arrow) being transcribed into RNA (red arrow). [O. L. Miller, Jr., and Barbara R. Beatty, Oak Ridge National Laboratory.]

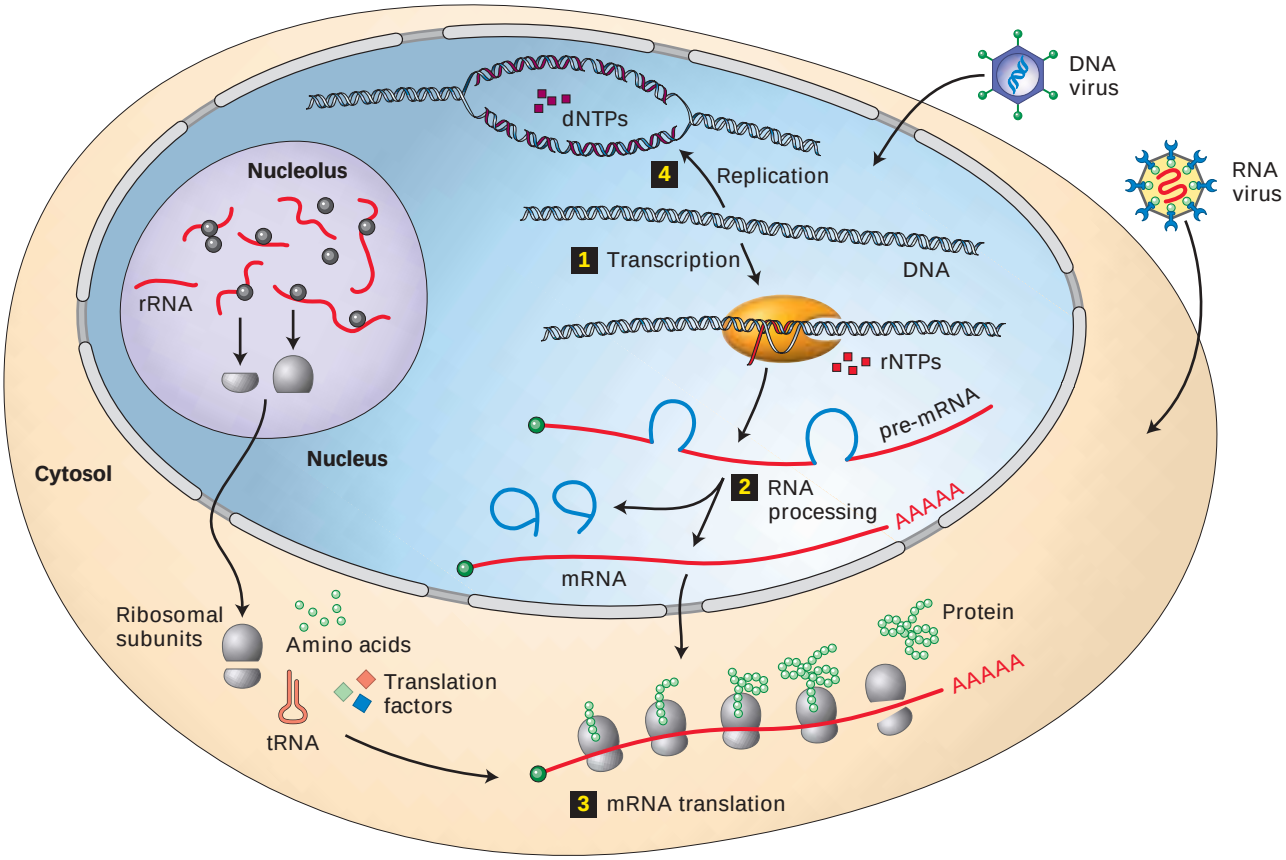
The extraordinary versatility of proteins as molecular machines and switches, cellular catalysts, and components of cellular structures was described in Chapter 3. In this chapter we consider the **nucleic acids.** These macromolecules (1) contain the information for determining the amino acid sequence and hence the structure and function of all the proteins of a cell, (2) are part of the cellular structures that select and align amino acids in the correct order as a polypeptide chain is being synthesized, and (3) catalyze a number of fundamental chemical reactions in cells, including formation of peptide bonds between amino acids during protein synthesis.

Deoxyribonucleic acid (DNA) contains all the information required to build the cells and tissues of an organism. The exact replication of this information in any species assures its genetic continuity from generation to generation and is critical to the normal development of an individual. The information stored in DNA is arranged in hereditary units, now known as **genes,** that control identifiable traits of an organism. In the process of **transcription,** the information stored in DNA is copied into **ribonucleic acid (RNA),** which has three distinct roles in protein synthesis.

Messenger RNA (mRNA) carries the instructions from DNA that specify the correct order of amino acids during protein synthesis. The remarkably accurate, stepwise assembly of amino acids into proteins occurs by **translation** of mRNA. In this process, the information in mRNA is interpreted by a second type of RNA called transfer RNA (tRNA) with the aid of a third type of RNA, ribosomal RNA (rRNA), and its associated proteins. As the correct amino acids are brought into sequence by tRNAs, they are linked by peptide bonds to make proteins.

Discovery of the structure of DNA in 1953 and subsequent elucidation of how DNA directs synthesis of RNA, which then directs assembly of proteins—the so-called *central dogma*—were monumental achievements marking the early days of molecular biology. However, the simplified representation of the central dogma as DNA→RNA→protein does not reflect the role of proteins in the synthesis of nucleic acids. Moreover, as discussed in later chapters, proteins are largely responsible for *regulating* **gene expression,** the entire process whereby the information encoded in DNA is decoded into the proteins that characterize various cell types.

OUTLINE



▲ FIGURE 4-1 **Overview of four basic molecular genetic processes.** In this chapter we cover the three processes that lead to production of proteins (1–3) and the process for replicating DNA (4). Because viruses utilize host-cell machinery, they have been important models for studying these processes. During transcription of a protein-coding gene by RNA polymerase (1), the four-base DNA code specifying the amino acid sequence of a protein is copied into a precursor messenger RNA (pre-mRNA) by the polymerization of ribonucleoside triphosphate monomers (rNTPs). Removal of extraneous sequences and other modifications to the pre-mRNA (2), collectively known as *RNA processing*, produce a functional mRNA, which is transported to the cytoplasm. During translation (3), the four-base code of the mRNA is decoded into the 20–amino acid "language" of proteins. Ribosomes, the macromolecular machines that translate the mRNA code, are composed of two subunits assembled in the nucleolus from ribosomal RNAs (rRNAs) and multiple proteins (*left*). After transport to the cytoplasm, ribosomal subunits associate with an mRNA and carry out protein synthesis with the help of transfer RNAs (tRNAs) and various translation factors. During DNA replication (4), which occurs only in cells preparing to divide, deoxyribonucleoside triphosphate monomers (dNTPs) are polymerized to yield two identical copies of each chromosomal DNA molecule. Each daughter cell receives one of the identical copies.

In this chapter, we first review the basic structures and properties of DNA and RNA. In the next several sections we discuss the basic processes summarized in Figure 4-1: transcription of DNA into RNA precursors, processing of these precursors to make functional RNA molecules, translation of mRNAs into proteins, and the replication of DNA. Along the way we compare gene structure in prokaryotes and eukaryotes and describe how bacteria control transcription, setting the stage for the more complex eukaryotic transcription-control mechanisms discussed in Chapter 11. After outlining the individual roles of mRNA, tRNA, and rRNA in protein synthesis, we present a detailed description of the components and biochemical steps in translation. We also consider the molecular problems involved in DNA replication and the complex cellular machinery for ensuring accurate copying of the genetic material. The final section of the chapter presents basic information about viruses, which are important model organisms for studying macromolecular synthesis and other cellular processes.

4.1 Structure of Nucleic Acids

DNA and RNA are chemically very similar. The primary structures of both are linear **polymers** composed of **monomers** called **nucleotides.** Cellular RNAs range in length from less than one hundred to many thousands of nucleotides. Cellular DNA molecules can be as long as several

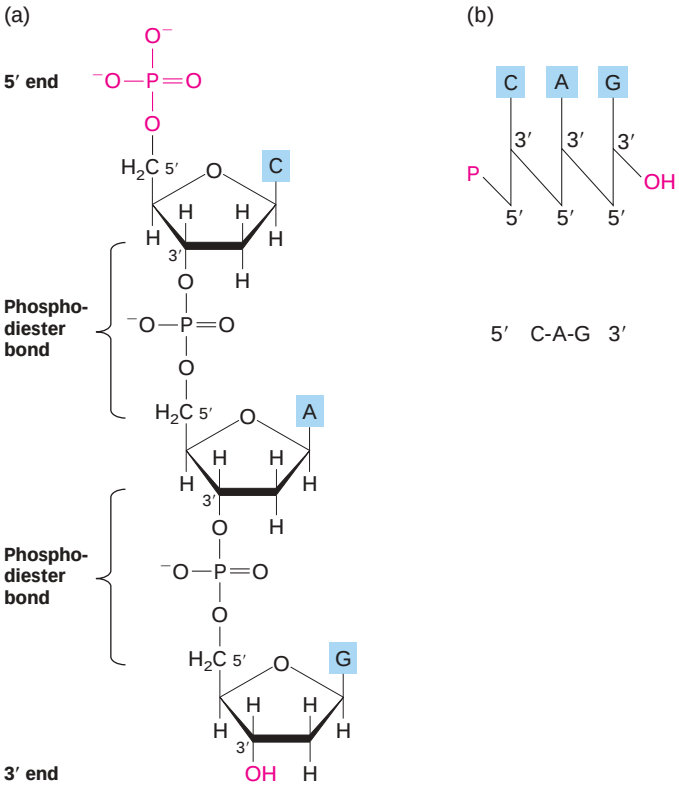
hundred million nucleotides. These large DNA units in association with proteins can be stained with dyes and visualized in the light microscope as chromosomes, so named because of their stainability.

A Nucleic Acid Strand Is a Linear Polymer with End-to-End Directionality

DNA and RNA each consist of only four different nucleotides. Recall from Chapter 2 that all nucleotides consist of an organic base linked to a five-carbon sugar that has a phosphate group attached to carbon 5. In RNA, the sugar is ribose; in DNA, deoxyribose (see Figure 2-14). The nucleotides used in synthesis of DNA and RNA contain five different bases. The bases *adenine* (A) and *guanine* (G) are **purines,** which contain a pair of fused rings; the bases *cytosine* (C), *thymine* (T), and *uracil* (U) are **pyrimidines,** which contain a single ring (see Figure 2-15). Both DNA and RNA contain three of these bases—A, G, and C; however, T is found only in DNA, and U only in RNA. (Note that the single-letter abbreviations for these bases are also commonly used to denote the entire nucleotides in nucleic acid polymers.)

A single nucleic acid strand has a *backbone* composed of repeating pentose-phosphate units from which the purine and pyrimidine bases extend as side groups. Like a polypeptide, a nucleic acid strand has an end-to-end chemical orientation: the *5′ end* has a hydroxyl or phosphate group on the 5′ carbon of its terminal sugar; the *3′ end* usually has a hydroxyl group on the 3′ carbon of its terminal sugar (Figure 4-2). This directionality, plus the fact that synthesis proceeds 5′ to 3′, has given rise to the convention that polynucleotide sequences are written and read in the 5′→3′ direction (from left to right); for example, the sequence AUG is assumed to be (5′)AUG(3′). As we will see, the 5′→3′ directionality of a nucleic acid strand is an important property of the molecule. The chemical linkage between adjacent nucleotides, commonly called a **phosphodiester bond,** actually consists of two phosphoester bonds, one on the 5′ side of the phosphate and another on the 3′ side.

The linear sequence of nucleotides linked by phosphodiester bonds constitutes the primary structure of nucleic acids. Like polypeptides, polynucleotides can twist and fold into three-dimensional conformations stabilized by noncovalent bonds. Although the primary structures of DNA and RNA are generally similar, their three-dimensional conformations are quite different. These structural differences are critical to the different functions of the two types of nucleic acids.

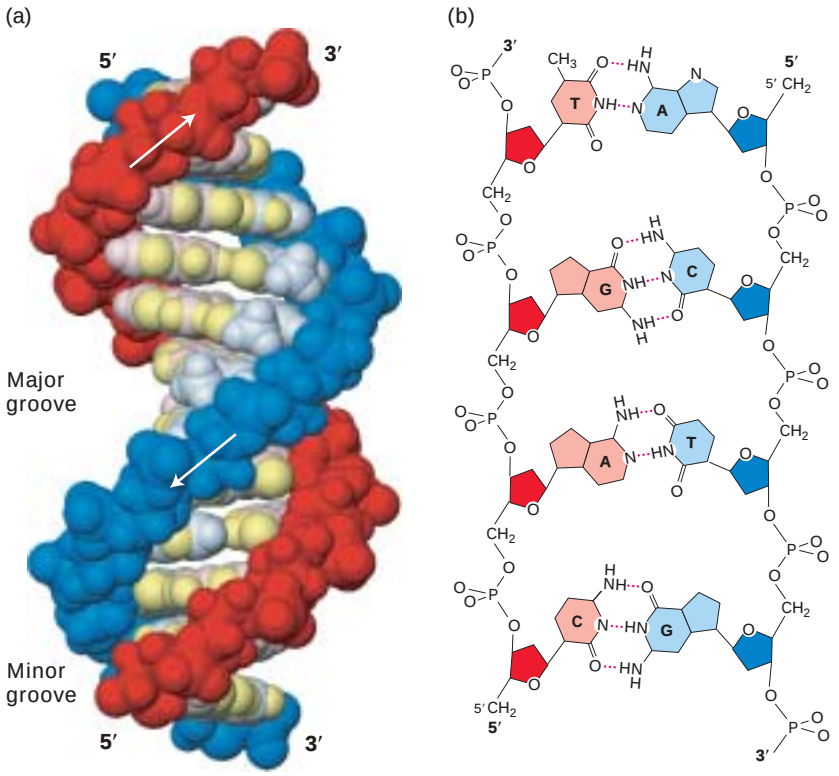


▲ FIGURE 4-2 **Alternative representations of a nucleic acid strand illustrating its chemical directionality.** Shown here is a single strand of DNA containing only three bases: cytosine (C), adenine (A), and guanine (G). (a) The chemical structure shows a hydroxyl group at the 3′ end and a phosphate group at the 5′ end. Note also that two phosphoester bonds link adjacent nucleotides; this two-bond linkage commonly is referred to as a *phosphodiester bond*. (b) In the "stick" diagram (*top*), the sugars are indicated as vertical lines and the phosphodiester bonds as slanting lines; the bases are denoted by their single-letter abbreviations. In the simplest representation (*bottom*), only the bases are indicated. By convention, a polynucleotide sequence is always written in the 5′→3′ direction (left to right) unless otherwise indicated.

Native DNA Is a Double Helix of Complementary Antiparallel Strands

The modern era of molecular biology began in 1953 when James D. Watson and Francis H. C. Crick proposed that DNA has a double-helical structure. Their proposal, based on analysis of x-ray diffraction patterns coupled with careful model building, proved correct and paved the way for our modern understanding of how DNA functions as the genetic material.

DNA consists of two associated polynucleotide strands that wind together to form a **double helix.** The two sugar-phosphate backbones are on the outside of the double helix, and the bases project into the interior. The adjoining bases in each strand stack on top of one another in parallel planes (Figure 4-3a). The orientation of the two strands is *antiparallel;* that is, their 5′→3′ directions are opposite. The strands are held in precise register by formation of **base pairs** between the two strands: A is paired with T through two hydrogen bonds; G is paired with C through three hydrogen bonds (Figure 4-3b). This base-pair complementarity is a consequence of the size, shape, and chemical composition of the bases. The presence of thousands of such hydrogen bonds in a DNA molecule contributes greatly to the stability



▲ FIGURE 4-3 **The DNA double helix.** (a) Space-filling model of B DNA, the most common form of DNA in cells. The bases (light shades) project inward from the sugar-phosphate backbones (dark red and blue) of each strand, but their edges are accessible through major and minor grooves. Arrows indicate the 5'→3' direction of each strand. Hydrogen bonds between the bases are in the center of the structure. The major and minor grooves are lined by potential hydrogen bond donors and acceptors (highlighted in yellow). (b) Chemical structure of DNA double helix. This extended schematic shows the two sugar-phosphate backbones and hydrogen bonding between the Watson-Crick base pairs, A·T and G·C. [Part (a) from R. Wing et al., 1980, *Nature* **287**:755; part (b) from R. E. Dickerson, 1983, *Sci. Am.* **249**:94.]

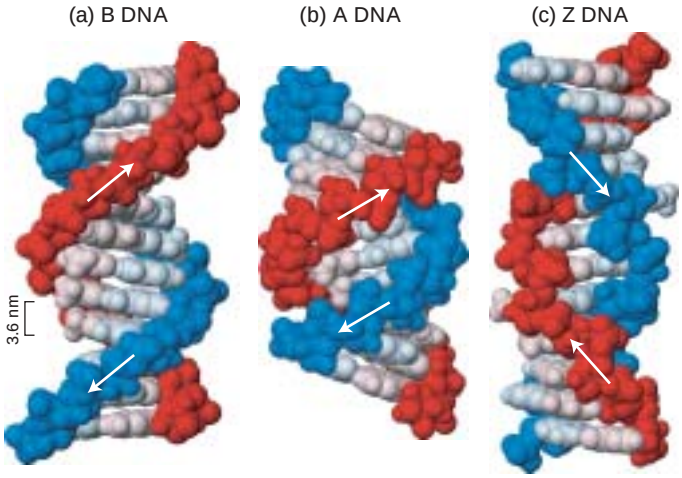
of the double helix. Hydrophobic and van der Waals interactions between the stacked adjacent base pairs further stabilize the double-helical structure.

In natural DNA, A always hydrogen bonds with T and G with C, forming A·T and G·C base pairs as shown in Figure 4-3b. These associations between a larger purine and smaller pyrimidine are often called *Watson-Crick base pairs.* Two polynucleotide strands, or regions thereof, in which all the nucleotides form such base pairs are said to be **complementary.** However, in theory and in synthetic DNAs other base pairs can form. For example, a guanine (a purine) could theoretically form hydrogen bonds with a thymine (a pyrimidine), causing only a minor distortion in the helix. The space available in the helix also would allow pairing between the two pyrimidines cytosine and thymine. Although the nonstandard G·T and C·T base pairs are normally not found in DNA, G·U base pairs are quite common in double-helical regions that form within otherwise single-stranded RNA.

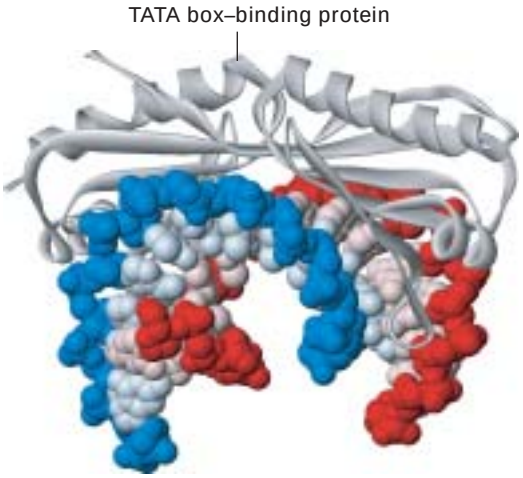
Most DNA in cells is a *right-handed* helix. The x-ray diffraction pattern of DNA indicates that the stacked bases are regularly spaced 0.36 nm apart along the helix axis. The helix makes a complete turn every 3.6 nm; thus there are about 10.5 pairs per turn. This is referred to as the *B form* of DNA, the normal form present in most DNA stretches in cells. On the outside of B-form DNA, the spaces between the intertwined strands form two helical grooves of different widths described as the *major* groove and the *minor* groove (see Figure 4-3a). As a consequence, the atoms on the edges of each base within these grooves are accessible from outside the helix, forming two types of binding surfaces. DNA-binding proteins can "read" the sequence of bases in duplex DNA by contacting atoms in either the major or the minor grooves.

In addition to the major B form, three additional DNA structures have been described. Two of these are compared to B DNA in Figure 4-4. In very low humidity, the crystallographic structure of B DNA changes to the *A form;* RNA-DNA and RNA-RNA helices exist in this form in cells and in vitro. Short DNA molecules composed of alternating purine-pyrimidine nucleotides (especially Gs and Cs) adopt an alternative left-handed configuration instead of the normal right-handed helix. This structure is called *Z DNA* because

the bases seem to zigzag when viewed from the side. Some evidence suggests that Z DNA may occur in cells, although its function is unknown. Finally, a triple-stranded DNA structure is formed when synthetic polymers of poly(A) and polydeoxy(U) are mixed in the test tube. In addition, homopolymeric stretches of DNA composed of C and T residues in one strand and A and G residues in the other can form a triple-stranded structure by binding matching lengths of synthetic poly(C+T). Such structures probably do not occur naturally in cells but may prove useful as therapeutic agents.



▲ FIGURE 4-4 **Models of various known DNA structures.** The sugar-phosphate backbones of the two strands, which are on the outside in all structures, are shown in red and blue; the bases (lighter shades) are oriented inward. (a) The B form of DNA has ≈10.5 base pairs per helical turn. Adjacent stacked base pairs are 0.36 nm apart. (b) The more compact A form of DNA has 11 base pairs per turn and exhibits a large tilt of the base pairs with respect to the helix axis. (c) Z DNA is a left-handed double helix.



▲ FIGURE 4-5 **Bending of DNA resulting from protein binding.** The conserved C-terminal domain of the TATA box–binding protein (TBP) binds to the minor groove of specific DNA sequences rich in A and T, untwisting and sharply bending the double helix. Transcription of most eukaryotic genes requires participation of TBP. [Adapted from D. B. Nikolov and S. K. Burley, 1997, *Proc. Nat'l. Acad. Sci. USA* **94**:15.]

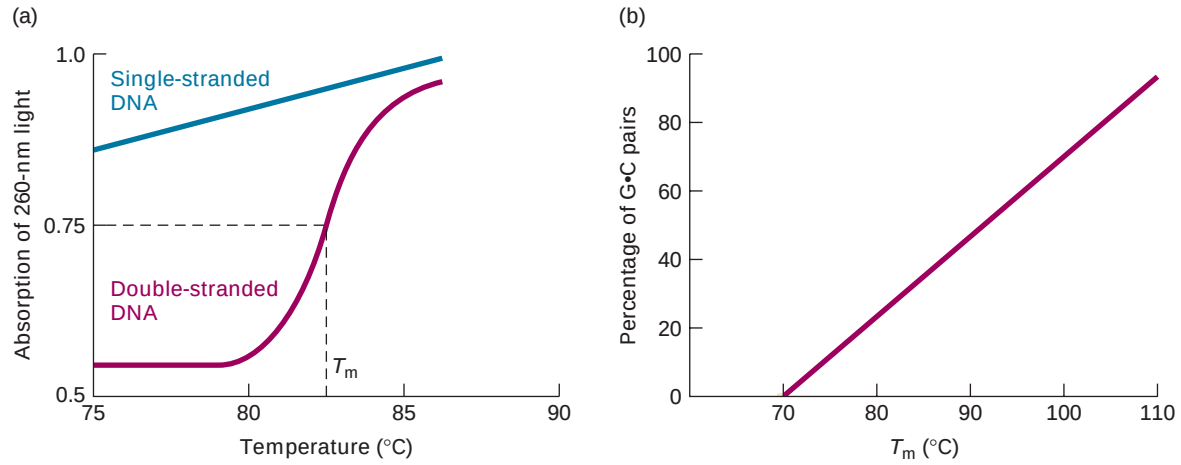
By far the most important modifications in the structure of standard B-form DNA come about as a result of protein binding to specific DNA sequences. Although the multitude of hydrogen and hydrophobic bonds between the bases provide stability to DNA, the double helix is flexible about its long axis. Unlike the α helix in proteins (see Figure 3-3), there are no hydrogen bonds parallel to the axis of the DNA helix. This property allows DNA to bend when complexed with a DNA-binding protein (Figure 4-5). Bending of DNA is critical to the dense packing of DNA in chromatin, the protein-DNA complex in which nuclear DNA occurs in eukaryotic cells (Chapter 10).

DNA Can Undergo Reversible Strand Separation

During replication and transcription of DNA, the strands of the double helix must separate to allow the internal edges of the bases to pair with the bases of the nucleotides to be polymerized into new polynucleotide chains. In later sections, we describe the cellular mechanisms that separate and subsequently reassociate DNA strands during replication and transcription. Here we discuss factors influencing the in vitro separation and reassociation of DNA strands.

The unwinding and separation of DNA strands, referred to as **denaturation,** or "melting," can be induced experimentally by increasing the temperature of a solution of DNA. As the thermal energy increases, the resulting increase in molecular motion eventually breaks the hydrogen bonds and other forces that stabilize the double helix; the strands then separate, driven apart by the electrostatic repulsion of the negatively charged deoxyribose-phosphate backbone of each strand. Near the denaturation temperature, a small increase in temperature causes a rapid, near simultaneous loss of the multiple weak interactions holding the strands together along the entire length of the DNA molecules, leading to an abrupt change in the absorption of ultraviolet (UV) light (Figure 4-6a).

The *melting temperature* T_m at which DNA strands will separate depends on several factors. Molecules that contain a greater proportion of G·C pairs require higher temperatures to denature because the three hydrogen bonds in G·C pairs make these base pairs more stable than A·T pairs, which have only two hydrogen bonds. Indeed, the percentage of G·C base pairs in a DNA sample can be estimated from its T_m (Figure 4-6b). The ion concentration also influences the T_m because the negatively charged phosphate groups in the



▲ EXPERIMENTAL FIGURE 4-6 **The temperature at which DNA denatures increases with the proportion of G·C pairs.** (a) Melting of doubled-stranded DNA can be monitored by the absorption of ultraviolet light at 260 nm. As regions of double-stranded DNA unpair, the absorption of light by those regions increases almost twofold. The temperature at which half the bases in a double-stranded DNA sample have denatured is denoted T_m (for temperature of melting). Light absorption by single-stranded DNA changes much less as the temperature is increased. (b) The T_m is a function of the G·C content of the DNA; the higher the G+C percentage, the greater the T_m.

two strands are shielded by positively charged ions. When the ion concentration is low, this shielding is decreased, thus increasing the repulsive forces between the strands and reducing the T_m. Agents that destabilize hydrogen bonds, such as formamide or urea, also lower the T_m. Finally, extremes of pH denature DNA at low temperature. At low (acid) pH, the bases become protonated and thus positively charged, repelling each other. At high (alkaline) pH, the bases lose protons and become negatively charged, again repelling each other because of the similar charge.

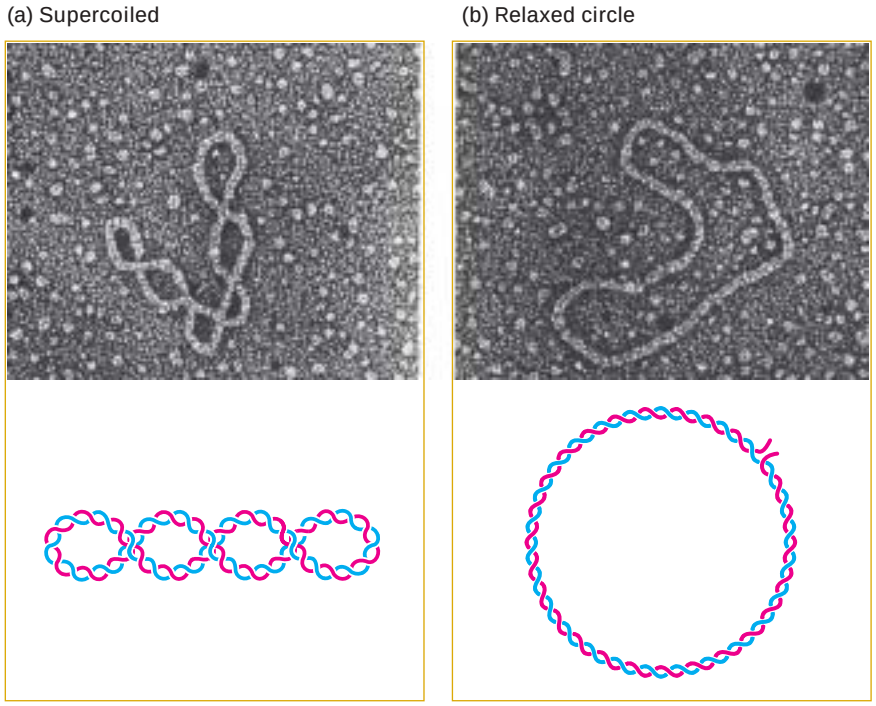
The single-stranded DNA molecules that result from denaturation form random coils without an organized structure. Lowering the temperature, increasing the ion concentration, or neutralizing the pH causes the two complementary strands to reassociate into a perfect double helix. The extent of such *renaturation* is dependent on time, the DNA concentration, and the ionic concentration. Two DNA strands not related in sequence will remain as random coils and will not renature; most importantly, they will not inhibit complementary DNA partner strands from finding each other and renaturing. Denaturation and renaturation of DNA are the basis of nucleic acid **hybridization,** a powerful technique used to study the relatedness of two DNA samples and to detect and isolate specific DNA molecules in a mixture containing numerous different DNA sequences (see Figure 9-16).

Many DNA Molecules Are Circular

Many prokaryotic genomic DNAs and many viral DNAs are circular molecules. Circular DNA molecules also occur in mitochondria, which are present in almost all eukaryotic cells, and in chloroplasts, which are present in plants and some unicellular eukaryotes.

Each of the two strands in a circular DNA molecule forms a closed structure without free ends. Localized unwinding of a circular DNA molecule, which occurs during DNA replication, induces torsional stress into the remaining portion of the molecule because the ends of the strands are not free to rotate. As a result, the DNA molecule twists back on itself, like a twisted rubber band, forming *supercoils* (Figure 4-7b). In other words, when part of the DNA helix is underwound, the remainder of the molecule becomes overwound. Bacterial and eukaryotic cells, however, contain *topoisomerase I,* which can relieve any torsional stress that develops in cellular DNA molecules during replication or other processes. This enzyme binds to DNA at random sites and breaks a phosphodiester bond in one strand. Such a one-strand break in DNA is called a *nick*. The broken end then winds around the uncut strand, leading to loss of supercoils (Figure 4-7a). Finally, the same enzyme joins (ligates) the two ends of the broken strand. Another type of enzyme, *topoisomerase II,* makes breaks in both strands of a double-stranded DNA and then religates them. As a result, topoisomerase II can both relieve torsional stress and link together two circular DNA molecules as in the links of a chain.

Although eukaryotic nuclear DNA is linear, long loops of DNA are fixed in place within chromosomes (Chapter 10). Thus torsional stress and the consequent formation of supercoils also could occur during replication of nuclear DNA. As in bacterial cells, abundant topoisomerase I in eukaryotic nuclei relieves any torsional stress in nuclear DNA that would develop in the absence of this enzyme.



◀ EXPERIMENTAL FIGURE 4-7 **DNA supercoils can be removed by cleavage of one strand.** (a) Electron micrograph of SV40 viral DNA. When the circular DNA of the SV40 virus is isolated and separated from its associated protein, the DNA duplex is underwound and assumes the supercoiled configuration. (b) If a supercoiled DNA is nicked (i.e., one strand cleaved), the strands can rewind, leading to loss of a supercoil. Topoisomerase I catalyzes this reaction and also reseals the broken ends. All the supercoils in isolated SV40 DNA can be removed by the sequential action of this enzyme, producing the relaxed-circle conformation. For clarity, the shapes of the molecules at the bottom have been simplified.

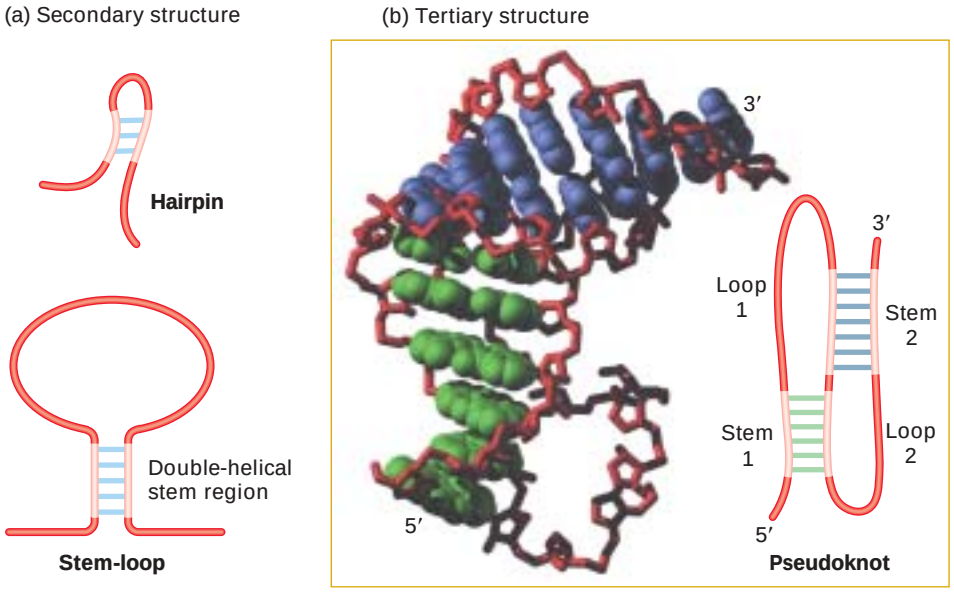
Different Types of RNA Exhibit Various Conformations Related to Their Functions

As noted earlier, the primary structure of RNA is generally similar to that of DNA with two exceptions: the sugar component of RNA, ribose, has a hydroxyl group at the 2′ position (see Figure 2-14b), and thymine in DNA is replaced by uracil in RNA. The hydroxyl group on C_2 of ribose makes RNA more chemically labile than DNA and provides a chemically reactive group that takes part in RNA-mediated catalysis. As a result of this lability, RNA is cleaved into mononucleotides by alkaline solution, whereas DNA is not. Like DNA, RNA is a long polynucleotide that can be double-stranded or single-stranded, linear or circular. It can also participate in a hybrid helix composed of one RNA strand and one DNA strand. As noted above, RNA-RNA and RNA-DNA double helices have a compact conformation like the A form of DNA (see Figure 4-4b).

Unlike DNA, which exists primarily as a very long double helix, most cellular RNAs are single-stranded and exhibit a variety of conformations (Figure 4-8). Differences in the sizes and conformations of the various types of RNA permit them to carry out specific functions in a cell. The simplest secondary structures in single-stranded RNAs are formed by pairing of complementary bases. "Hairpins" are formed by pairing of bases within ≈5–10 nucleotides of each other, and "stem-loops" by pairing of bases that are separated by >10 to several hundred nucleotides. These simple folds can cooperate to form more complicated tertiary structures, one of which is termed a "pseudoknot."

As discussed in detail later, tRNA molecules adopt a well-defined three-dimensional architecture in solution that is crucial in protein synthesis. Larger rRNA molecules also have locally well-defined three-dimensional structures, with more flexible links in between. Secondary and tertiary structures also have been recognized in mRNA, particularly near the ends of molecules. Clearly, then, RNA molecules are like proteins in that they have structured domains connected by less structured, flexible stretches.

The folded domains of RNA molecules not only are structurally analogous to the α helices and β strands found in proteins, but in some cases also have catalytic capacities. Such catalytic RNAs are called **ribozymes.** Although ribozymes usually are associated with proteins that stabilize the ribozyme structure, it is the RNA that acts as a catalyst. Some ribozymes can catalyze splicing, a remarkable process in which an internal RNA sequence is cut and removed, and the two resulting chains then ligated. This process occurs during formation of the majority of functional mRNA molecules in eukaryotic cells, and also occurs in bacteria and archaea. Remarkably, some RNAs carry out *self-splicing,* with the catalytic activity residing in the sequence that is removed. The mechanisms of splicing and self-splicing are discussed in detail in Chapter 12. As noted later in this chapter, rRNA



▶ FIGURE 4-8 **RNA secondary and tertiary structures.** (a) Stem-loops, hairpins, and other secondary structures can form by base pairing between distant complementary segments of an RNA molecule. In stem-loops, the single-stranded loop between the base-paired helical stem may be hundreds or even thousands of nucleotides long, whereas in hairpins, the short turn may contain as few as four nucleotides. (b) Pseudoknots, one type of RNA tertiary structure, are formed by interaction of secondary loops through base pairing between complementary bases (green and blue). Only base-paired bases are shown. A secondary structure diagram is shown at right. [Part (b) adapted from P. J. A. Michiels et al., 2001, *J. Mol. Biol.* **310**:1109.]

plays a catalytic role in the formation of peptide bonds during protein synthesis.

In this chapter, we focus on the functions of mRNA, tRNA, and rRNA in gene expression. In later chapters we will encounter other RNAs, often associated with proteins, that participate in other cell functions.

KEY CONCEPTS OF SECTION 4.1

Structure of Nucleic Acids

- Deoxyribonucleic acid (DNA), the genetic material, carries information to specify the amino acid sequences of proteins. It is transcribed into several types of ribonucleic acid (RNA), including messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA), which function in protein synthesis (see Figure 4-1).
- Both DNA and RNA are long, unbranched polymers of nucleotides, which consist of a phosphorylated pentose linked to an organic base, either a purine or pyrimidine.
- The purines adenine (A) and guanine (G) and the pyrimidine cytosine (C) are present in both DNA and RNA. The pyrimidine thymine (T) present in DNA is replaced by the pyrimidine uracil (U) in RNA.
- Adjacent nucleotides in a polynucleotide are linked by phosphodiester bonds. The entire strand has a chemical directionality: the 5′ end with a free hydroxyl or phosphate group on the 5′ carbon of the sugar, and the 3′ end with a free hydroxyl group on the 3′ carbon of the sugar (see Figure 4-2).
- Natural DNA (B DNA) contains two complementary antiparallel polynucleotide strands wound together into a regular right-handed double helix with the bases on the inside and the two sugar-phosphate backbones on the outside (see Figure 4-3). Base pairing between the strands and hydrophobic interactions between adjacent bases in the same strand stabilize this native structure.
- The bases in nucleic acids can interact via hydrogen bonds. The standard Watson-Crick base pairs are G·C, A·T (in DNA), and A·U (in RNA). Base pairing stabilizes the native three-dimensional structures of DNA and RNA.
- Binding of protein to DNA can deform its helical structure, causing local bending or unwinding of the DNA molecule.
- Heat causes the DNA strands to separate (denature). The melting temperature T_m of DNA increases with the percentage of G·C base pairs. Under suitable conditions, separated complementary nucleic acid strands will renature.
- Circular DNA molecules can be twisted on themselves, forming supercoils (see Figure 4-7). Enzymes called *topoisomerases* can relieve torsional stress and remove supercoils from circular DNA molecules.
- Cellular RNAs are single-stranded polynucleotides, some of which form well-defined secondary and tertiary structures (see Figure 4-8). Some RNAs, called *ribozymes*, have catalytic activity.

4.2 Transcription of Protein-Coding Genes and Formation of Functional mRNA

The simplest definition of a gene is a "unit of DNA that contains the information to specify synthesis of a single polypeptide chain or functional RNA (such as a tRNA)." The vast

majority of genes carry information to build protein molecules, and it is the RNA copies of such *protein-coding genes* that constitute the mRNA molecules of cells. The DNA molecules of small viruses contain only a few genes, whereas the single DNA molecule in each of the chromosomes of higher animals and plants may contain several thousand genes.

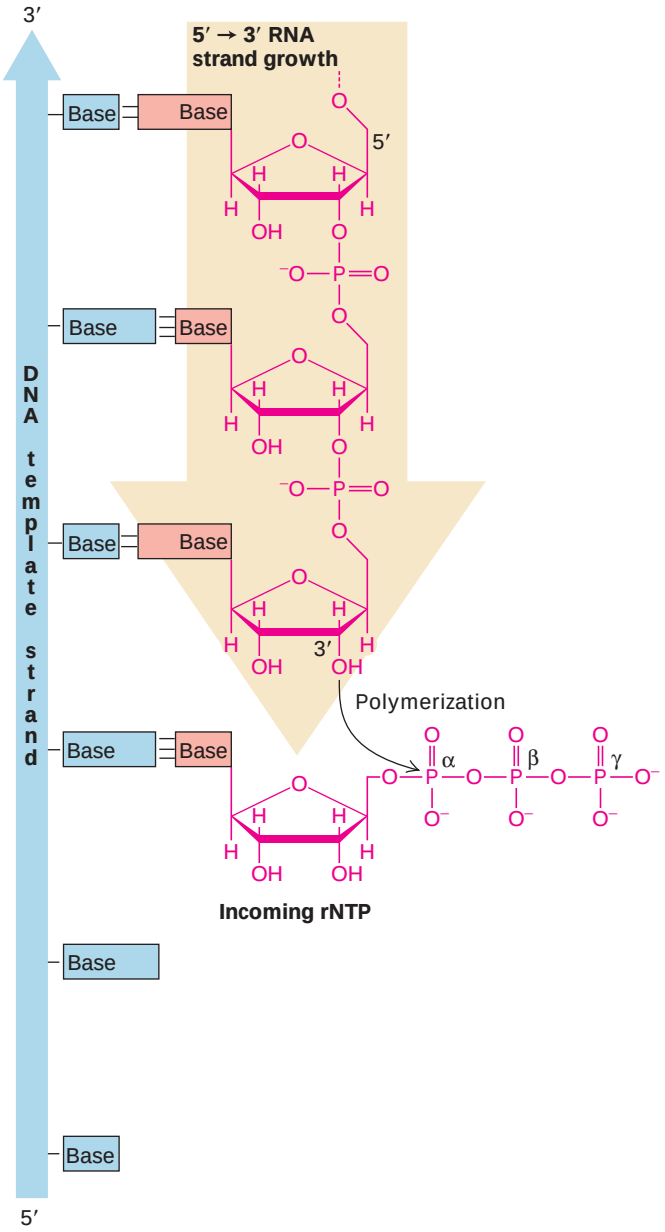
During synthesis of RNA, the four-base language of DNA containing A, G, C, and T is simply copied, or *transcribed*, into the four-base language of RNA, which is identical except that U replaces T. In contrast, during protein synthesis the four-base language of DNA and RNA is *translated* into the 20–amino acid language of proteins. In this section we focus on formation of functional mRNAs from protein-coding genes (see Figure 4-1, step 1). A similar process yields the precursors of rRNAs and tRNAs encoded by rRNA and tRNA genes; these precursors are then further modified to yield functional rRNAs and tRNAs (Chapter 12).

A Template DNA Strand Is Transcribed into a Complementary RNA Chain by RNA Polymerase

During transcription of DNA, one DNA strand acts as a *template*, determining the order in which ribonucleoside triphosphate (rNTP) monomers are polymerized to form a complementary RNA chain. Bases in the template DNA strand base-pair with complementary incoming rNTPs, which then are joined in a polymerization reaction catalyzed by **RNA polymerase.** Polymerization involves a nucleophilic attack by the 3′ oxygen in the growing RNA chain on the α phosphate of the next nucleotide precursor to be added, resulting in formation of a phosphodiester bond and release of pyrophosphate (PP_i). As a consequence of this mechanism, RNA molecules are always synthesized in the 5′→3′ direction (Figure 4-9).

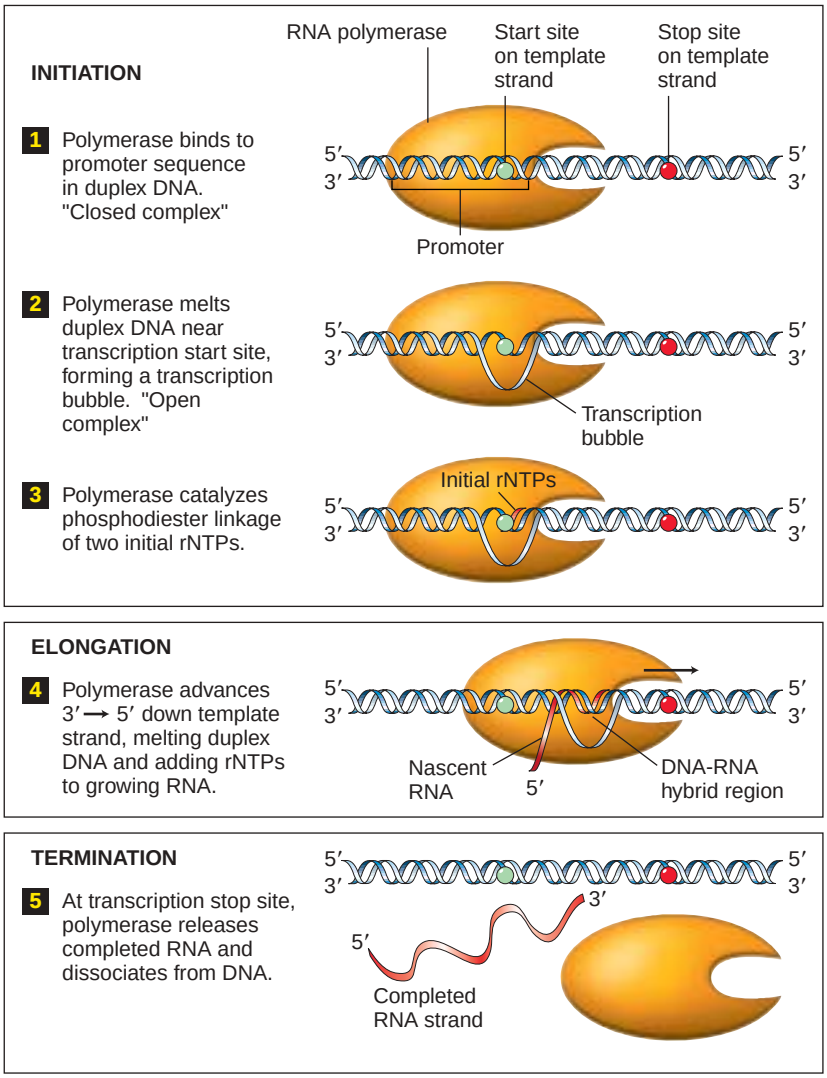
The energetics of the polymerization reaction strongly favors addition of ribonucleotides to the growing RNA chain because the high-energy bond between the α and β phosphate of rNTP monomers is replaced by the lower-energy phosphodiester bond between nucleotides. The equilibrium for the reaction is driven further toward chain elongation by pyrophosphatase, an enzyme that catalyzes cleavage of the released PP_i into two molecules of inorganic phosphate. Like the two strands in DNA, the template DNA strand and the growing RNA strand that is base-paired to it have opposite 5′→3′ directionality.

By convention, the site at which RNA polymerase begins transcription is numbered +1. **Downstream** denotes the direction in which a template DNA strand is transcribed (or mRNA translated); thus a downstream sequence is toward the 3′ end relative to the start site, considering the DNA strand with the same polarity as the transcribed RNA. **Upstream** denotes the opposite direction. Nucleotide positions in the DNA sequence downstream from a start site are indicated by a positive (+) sign; those upstream, by a negative (−) sign.



▲ FIGURE 4-9 **Polymerization of ribonucleotides by RNA polymerase during transcription.** The ribonucleotide to be added at the 3′ end of a growing RNA strand is specified by base pairing between the next base in the template DNA strand and the complementary incoming ribonucleoside triphosphate (rNTP). A phosphodiester bond is formed when RNA polymerase catalyzes a reaction between the 3′ O of the growing strand and the α phosphate of a correctly base-paired rNTP. RNA strands always are synthesized in the 5′→3′ direction and are opposite in polarity to their template DNA strands.

Stages in Transcription To carry out transcription, RNA polymerase performs several distinct functions, as depicted in Figure 4-10. During transcription *initiation*, RNA polymerase recognizes and binds to a specific site, called a **promoter,** in double-stranded DNA (step 1). Nuclear RNA



◀ FIGURE 4-10 **Three stages in transcription.** During initiation of transcription, RNA polymerase forms a transcription bubble and begins polymerization of ribonucleotides (rNTPs) at the start site, which is located within the promoter region. Once a DNA region has been transcribed, the separated strands reassociate into a double helix, displacing the nascent RNA except at its 3′ end. The 5′ end of the RNA strand exits the RNA polymerase through a channel in the enzyme. Termination occurs when the polymerase encounters a specific termination sequence (stop site). See the text for details.

MEDIA CONNECTIONS
Focus Animation: Basic Transcriptional Mechanism

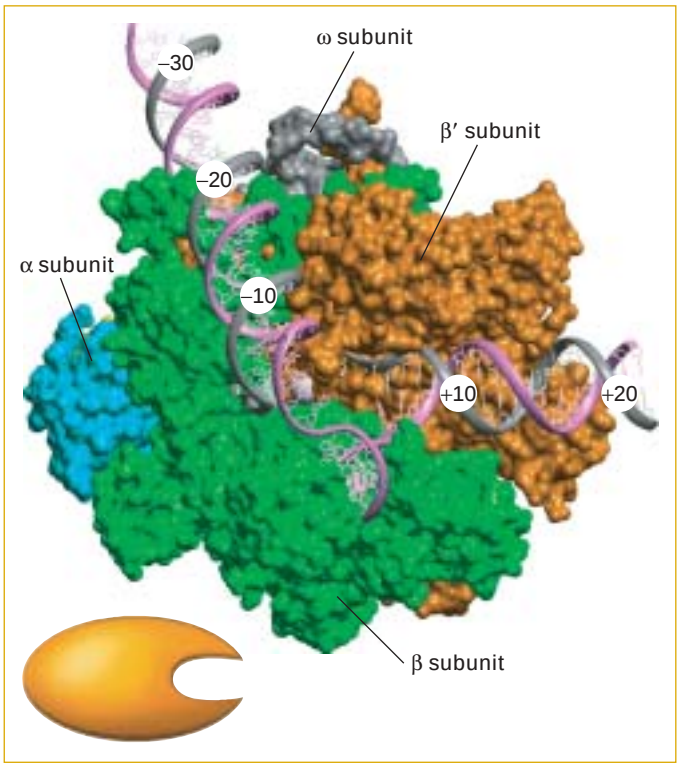
polymerases require various protein factors, called general **transcription factors,** to help them locate promoters and initiate transcription. After binding to a promoter, RNA polymerase melts the DNA strands in order to make the bases in the template strand available for base pairing with the bases of the ribonucleoside triphosphates that it will polymerize together. Cellular RNA polymerases melt approximately 14 base pairs of DNA around the transcription start site, which is located on the template strand within the promoter region (step 2). Transcription initiation is considered complete when the first two ribonucleotides of an RNA chain are linked by a phosphodiester bond (step 3).

After several ribonucleotides have been polymerized, RNA polymerase dissociates from the promoter DNA and general transcription factors. During the stage of *strand elongation*, RNA polymerase moves along the template DNA one base at a time, opening the double-stranded DNA in front of its direction of movement and hybridizing the strands behind it (Figure 4-10, step 4). One ribonucleotide at a time is added to the 3′ end of the growing (*nascent*) RNA chain during strand elongation by the polymerase. The enzyme maintains a melted region of approximately 14 base pairs, called the *transcription bubble*. Approximately eight nucleotides at the 3′ end of the growing RNA strand remain base-paired to the template DNA strand in the transcription bubble. The elongation complex, comprising RNA polymerase, template DNA, and the growing (nascent) RNA strand, is extraordinarily stable. For example, RNA polymerase transcribes the longest known mammalian genes, containing $\approx 2 \times 10^6$ base pairs, without dissociating from the DNA template or releasing the nascent RNA. Since RNA synthesis occurs at a rate of about 1000 nucleotides per minute at 37 °C, the elongation complex must remain intact for more than 24 hours to assure continuous RNA synthesis.

During transcription *termination*, the final stage in RNA synthesis, the completed RNA molecule, or **primary transcript,**

is released from the RNA polymerase and the polymerase dissociates from the template DNA (Figure 4-10, step 5). Specific sequences in the template DNA signal the bound RNA polymerase to terminate transcription. Once released, an RNA polymerase is free to transcribe the same gene again or another gene.

Structure of RNA Polymerases The RNA polymerases of bacteria, archaea, and eukaryotic cells are fundamentally similar in structure and function. Bacterial RNA polymerases are composed of two related large subunits (β′ and β), two copies of a smaller subunit (α), and one copy of a fifth subunit (ω) that is not essential for transcription or cell viability but stabilizes the enzyme and assists in the assembly of its subunits. Archaeal and eukaryotic RNA polymerases have several additional small subunits associated with this core complex, which we describe in Chapter 11. Schematic diagrams of the transcription process generally show RNA polymerase bound to an unbent DNA molecule, as in Figure 4-10. However, according to a current model of the interaction between bacterial RNA polymerase and promoter DNA, the DNA bends sharply following its entry into the enzyme (Figure 4-11).



▲ FIGURE 4-11 **Current model of bacterial RNA polymerase bound to a promoter.** This structure corresponds to the polymerase molecule as schematically shown in step 2 of Figure 4-10. The β′ subunit is in orange; β is in green. Part of one of the two α subunits can be seen in light blue; the ω subunit is in gray. The DNA template and nontemplate strands are shown, respectively, as gray and pink ribbons. A Mg^{2+} ion at the active center is shown as a gray sphere. Numbers indicate positions in the DNA sequence relative to the transcription start site, with positive (+) numbers in the direction of transcription and negative (−) numbers in the opposite direction. [Courtesy of R. H. Ebright, Waksman Institute.]

Organization of Genes Differs in Prokaryotic and Eukaryotic DNA

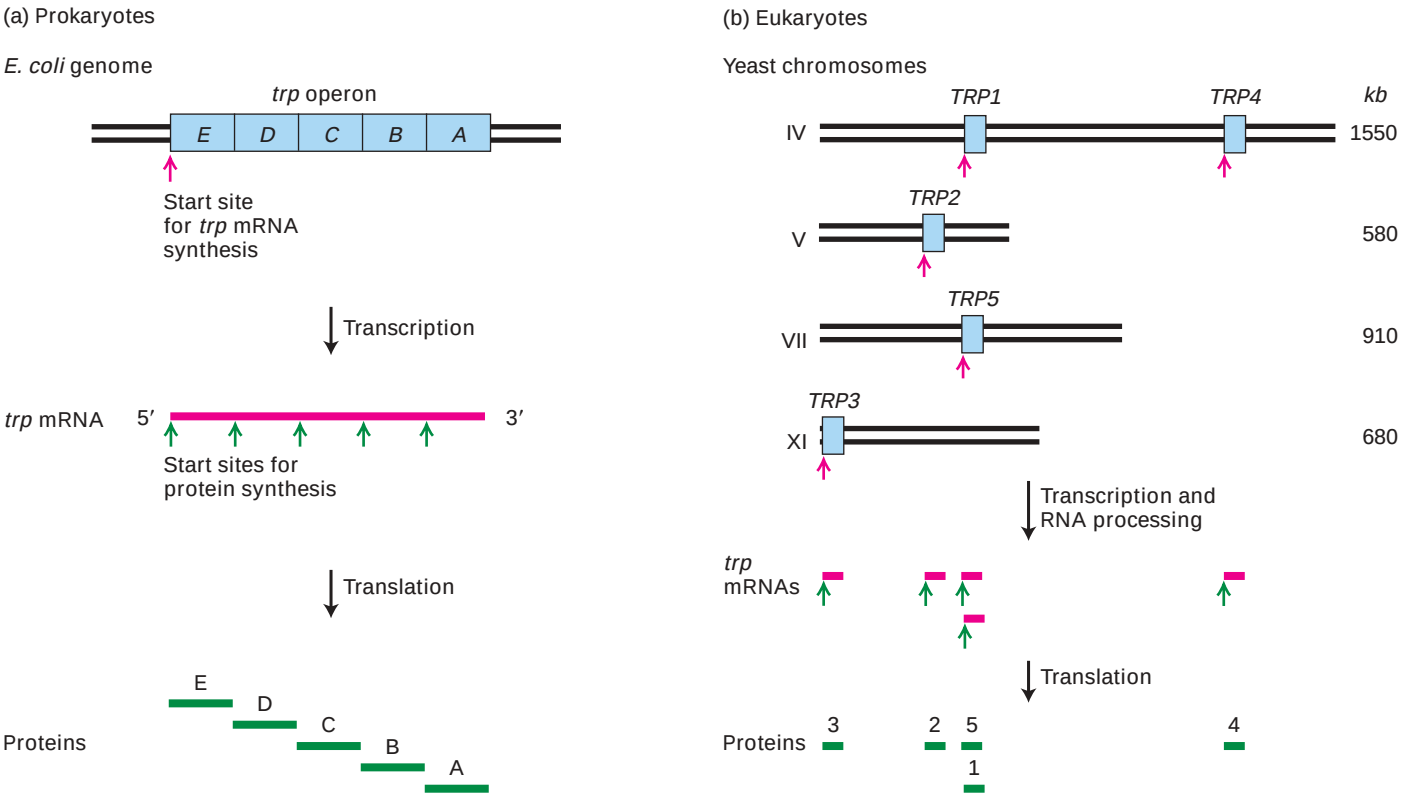
Having outlined the process of transcription, we now briefly consider the large-scale arrangement of information in DNA and how this arrangement dictates the requirements for RNA synthesis so that information transfer goes smoothly. In recent years, sequencing of the entire **genomes** from several organisms has revealed not only large variations in the number of protein-coding genes but also differences in their organization in prokaryotes and eukaryotes.

The most common arrangement of protein-coding genes in all prokaryotes has a powerful and appealing logic: genes devoted to a single metabolic goal, say, the synthesis of the amino acid tryptophan, are most often found in a contiguous array in the DNA. Such an arrangement of genes in a functional group is called an **operon,** because it operates as a unit from a single promoter. Transcription of an operon produces a continuous strand of mRNA that carries the message for a related series of proteins (Figure 4-12a). Each section of the mRNA represents the unit (or gene) that encodes one of the proteins in the series. In prokaryotic DNA the genes are closely packed with very few noncoding gaps, and the DNA is transcribed directly into colinear mRNA, which then is translated into protein.

This economic clustering of genes devoted to a single metabolic function does not occur in eukaryotes, even simple ones like yeasts, which can be metabolically similar to bacteria. Rather, eukaryotic genes devoted to a single pathway are most often physically separated in the DNA; indeed such genes usually are located on different chromosomes. Each gene is transcribed from its own promoter, producing one mRNA, which generally is translated to yield a single polypeptide (Figure 4-12b).

When researchers first compared the nucleotide sequences of eukaryotic mRNAs from multicellular organisms with the DNA sequences encoding them, they were surprised to find that the uninterrupted protein-coding sequence of a given mRNA was broken up (discontinuous) in its corresponding section of DNA. They concluded that the eukaryotic gene existed in pieces of coding sequence, the **exons,** separated by non-protein-coding segments, the **introns.** This astonishing finding implied that the long initial primary transcript—the RNA copy of the entire transcribed DNA sequence—had to be clipped apart to remove the introns and then carefully stitched back together to produce many eukaryotic mRNAs.

Although introns are common in multicellular eukaryotes, they are extremely rare in bacteria and archaea and



▲ FIGURE 4-12 **Comparison of gene organization, transcription, and translation in prokaryotes and eukaryotes.** (a) The tryptophan (*trp*) operon is a continuous segment of the *E. coli* chromosome, containing five genes (blue) that encode the enzymes necessary for the stepwise synthesis of tryptophan. The entire operon is transcribed from one promoter into one long continuous *trp* mRNA (red). Translation of this mRNA begins at five different start sites, yielding five proteins (green). The order of the genes in the bacterial genome parallels the sequential function of the encoded proteins in the tryptophan pathway. (b) The five genes encoding the enzymes required for tryptophan synthesis in yeast (*Saccharomyces cerevisiae*) are carried on four different chromosomes. Each gene is transcribed from its own promoter to yield a primary transcript that is processed into a functional mRNA encoding a single protein. The lengths of the yeast chromosomes are given in kilobases (10^3 bases).

uncommon in many unicellular eukaryotes such as baker's yeast. However, introns are present in the DNA of viruses that infect eukaryotic cells. Indeed, the presence of introns was first discovered in such viruses, whose DNA is transcribed by host-cell enzymes.

Eukaryotic Precursor mRNAs Are Processed to Form Functional mRNAs

In prokaryotic cells, which have no nuclei, translation of an mRNA into protein can begin from the 5′ end of the mRNA even while the 3′ end is still being synthesized by RNA polymerase. In other words, transcription and translation can occur concurrently in prokaryotes. In eukaryotic cells, however, not only is the nucleus separated from the cytoplasm where translation occurs, but also the primary transcripts of protein-coding genes are precursor mRNAs (**pre-mRNAs**) that must undergo several modifications, collectively termed *RNA processing*, to yield a functional mRNA (see Figure 4-1, step 2). This mRNA then must be exported to the cytoplasm before it can be translated into protein. Thus transcription and translation cannot occur concurrently in eukaryotic cells.

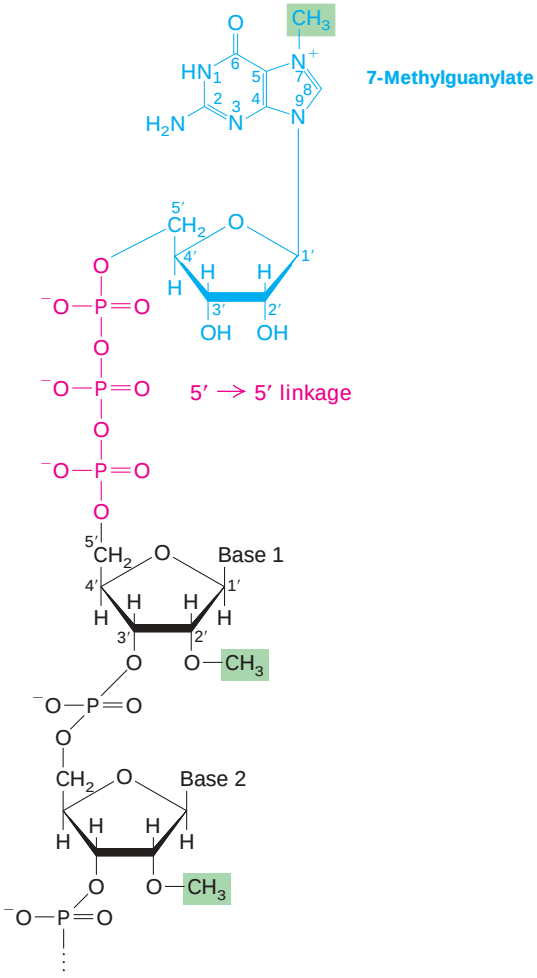
All eukaryotic pre-mRNAs initially are modified at the two ends, and these modifications are retained in mRNAs. As the 5′ end of a nascent RNA chain emerges from the surface of RNA polymerase II, it is immediately acted on by several enzymes that together synthesize the *5′ cap*, a 7-methylguanylate that is connected to the terminal nucleotide of the RNA by an unusual 5′,5′ triphosphate linkage (Figure 4-13). The cap protects an mRNA from enzymatic degradation and assists in its export to the cytoplasm. The cap also is bound by a protein factor required to begin translation in the cytoplasm.

Processing at the 3′ end of a pre-mRNA involves cleavage by an endonuclease to yield a free 3′-hydroxyl group to which a string of adenylic acid residues is added one at a time by an enzyme called *poly(A) polymerase*. The resulting *poly(A) tail* contains 100–250 bases, being shorter in yeasts and invertebrates than in vertebrates. Poly(A) polymerase is

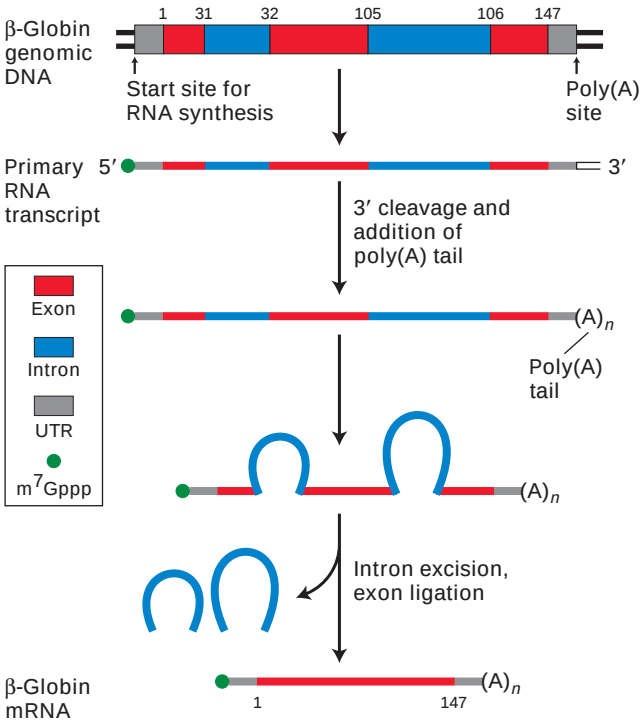
part of a complex of proteins that can locate and cleave a transcript at a specific site and then add the correct number of A residues, in a process that does not require a template.

The final step in the processing of many different eukaryotic mRNA molecules is **RNA splicing**: the internal cleavage of a transcript to excise the introns, followed by ligation of the coding exons. Figure 4-14 summarizes the basic steps in eukaryotic mRNA processing, using the β-globin gene as an example. We examine the cellular machinery for carrying out processing of mRNA, as well as tRNA and rRNA, in Chapter 12.

The functional eukaryotic mRNAs produced by RNA processing retain noncoding regions, referred to as *5′* and *3′*



▲ FIGURE 4-13 **Structure of the 5′ methylated cap of eukaryotic mRNA.** The distinguishing chemical features are the 5′→5′ linkage of 7-methylguanylate to the initial nucleotide of the mRNA molecule and the methyl group on the 2′ hydroxyl of the ribose of the first nucleotide (base 1). Both these features occur in all animal cells and in cells of higher plants; yeasts lack the methyl group on nucleotide 1. The ribose of the second nucleotide (base 2) also is methylated in vertebrates. [See A. J. Shatkin, 1976, *Cell* **9**:645.]



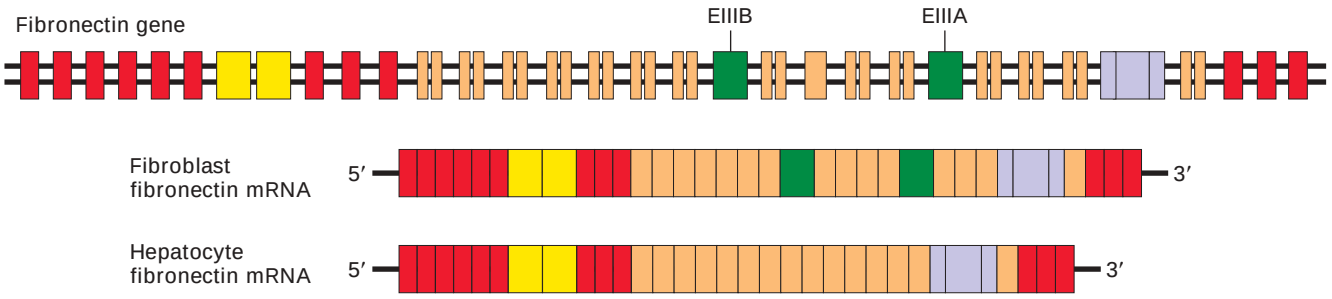
▲ FIGURE 4-14 **Overview of RNA processing to produce functional mRNA in eukaryotes.** The β-globin gene contains three protein-coding exons (coding region, red) and two intervening noncoding introns (blue). The introns interrupt the protein-coding sequence between the codons for amino acids 31 and 32 and 105 and 106. Transcription of eukaryotic protein-coding genes starts before the sequence that encodes the first amino acid and extends beyond the sequence encoding the last amino acid, resulting in noncoding regions (gray) at the ends of the primary transcript. These untranslated regions (UTRs) are retained during processing. The 5′ cap (m^7Gppp) is added during formation of the primary RNA transcript, which extends beyond the poly(A) site. After cleavage at the poly(A) site and addition of multiple A residues to the 3′ end, splicing removes the introns and joins the exons. The small numbers refer to positions in the 147–amino acid sequence of β-globin.

MEDIA CONNECTIONS
Overview Animation: Life Cycle of an mRNA

untranslated regions (UTRs), at each end. In mammalian mRNAs, the 5′ UTR may be a hundred or more nucleotides long, and the 3′ UTR may be several kilobases in length. Prokaryotic mRNAs also usually have 5′ and 3′ UTRs, but these are much shorter than those in eukaryotic mRNAs, generally containing fewer than 10 nucleotides.

Alternative RNA Splicing Increases the Number of Proteins Expressed from a Single Eukaryotic Gene

In contrast to bacterial and archaeal genes, the vast majority of genes in higher, multicellular eukaryotes contain multiple introns. As noted in Chapter 3, many proteins from



▲ FIGURE 4-15 **Cell type–specific splicing of fibronectin pre-mRNA in fibroblasts and hepatocytes.** The ≈75-kb fibronectin gene *(top)* contains multiple exons. The EIIIB and EIIIA exons (green) encode binding domains for specific proteins on the surface of fibroblasts. The fibronectin mRNA produced in fibroblasts includes the EIIIA and EIIIB exons, whereas these exons are spliced out of fibronectin mRNA in hepatocytes. In this diagram, introns (black lines) are not drawn to scale; most of them are much longer than any of the exons.

higher eukaryotes have a multidomain tertiary structure (see Figure 3-8). Individual repeated protein domains often are encoded by one exon or a small number of exons that code for identical or nearly identical amino acid sequences. Such repeated exons are thought to have evolved by the accidental multiple duplication of a length of DNA lying between two sites in adjacent introns, resulting in insertion of a string of repeated exons, separated by introns, between the original two introns. The presence of multiple introns in many eukaryotic genes permits expression of multiple, related proteins from a single gene by means of **alternative splicing.** In higher eukaryotes, alternative splicing is an important mechanism for production of different forms of a protein, called **isoforms,** by different types of cells.

Fibronectin, a multidomain extracellular adhesive protein found in mammals, provides a good example of alternative splicing (Figure 4-15). The fibronectin gene contains numerous exons, grouped into several regions corresponding to specific domains of the protein. Fibroblasts produce fibronectin mRNAs that contain exons EIIIA and EIIIB; these exons encode amino acid sequences that bind tightly to proteins in the fibroblast plasma membrane. Consequently, this fibronectin isoform adheres fibroblasts to the extracellular matrix. Alternative splicing of the fibronectin primary transcript in hepatocytes, the major type of cell in the liver, yields mRNAs that lack the EIIIA and EIIIB exons. As a result, the fibronectin secreted by hepatocytes into the blood does not adhere tightly to fibroblasts or most other cell types, allowing it to circulate. During formation of blood clots, however, the fibrin-binding domains of hepatocyte fibronectin binds to fibrin, one of the principal constituents of clots. The bound fibronectin then interacts with integrins on the membranes of passing, activated platelets, thereby expanding the clot by addition of platelets.

More than 20 different isoforms of fibronectin have been identified, each encoded by a different, alternatively spliced mRNA composed of a unique combination of fibronectin gene exons. Recent sequencing of large numbers of mRNAs isolated from various tissues and comparison of their sequences with genomic DNA has revealed that nearly 60 percent of all human genes are expressed as alternatively spliced mRNAs. Clearly, alternative RNA splicing greatly expands the number of proteins encoded by the genomes of higher, multicellular organisms.

KEY CONCEPTS OF SECTION 4.2

Transcription of Protein-Coding Genes and Formation of Functional mRNA

- Transcription of DNA is carried out by RNA polymerase, which adds one ribonucleotide at a time to the 3′ end of a growing RNA chain (see Figure 4-10). The sequence of the template DNA strand determines the order in which ribonucleotides are polymerized to form an RNA chain.
- During transcription initiation, RNA polymerase binds to a specific site in DNA (the promoter), locally melts the double-stranded DNA to reveal the unpaired template strand, and polymerizes the first two nucleotides.
- During strand elongation, RNA polymerase moves along the DNA, melting sequential segments of the DNA and adding nucleotides to the growing RNA strand.
- When RNA polymerase reaches a termination sequence in the DNA, the enzyme stops transcription, leading to release of the completed RNA and dissociation of the enzyme from the template DNA.
- In prokaryotic DNA, several protein-coding genes commonly are clustered into a functional region, an operon, which is transcribed from a single promoter into one mRNA encoding multiple proteins with related functions (see Figure 4-12a). Translation of a bacterial mRNA can begin before synthesis of the mRNA is complete.
- In eukaryotic DNA, each protein-coding gene is transcribed from its own promoter. The initial primary tran-

script very often contains noncoding regions (introns) interspersed among coding regions (exons).

- Eukaryotic primary transcripts must undergo RNA processing to yield functional RNAs. During processing, the ends of nearly all primary transcripts from protein-coding genes are modified by addition of a 5′ cap and 3′ poly(A) tail. Transcripts from genes containing introns undergo splicing, the removal of the introns and joining of the exons (see Figure 4-14).
- The individual domains of multidomain proteins found in higher eukaryotes are often encoded by individual exons or a small number of exons. Distinct isoforms of such proteins often are expressed in specific cell types as the result of alternative splicing of exons.

4.3 Control of Gene Expression in Prokaryotes

Since the structure and function of a cell are determined by the proteins it contains, the control of gene expression is a fundamental aspect of molecular cell biology. Most commonly, the "decision" to initiate transcription of the gene encoding a particular protein is the major mechanism for controlling production of the encoded protein in a cell. By controlling transcription initiation, a cell can regulate which proteins it produces and how rapidly. When transcription of a gene is *repressed*, the corresponding mRNA and encoded protein or proteins are synthesized at low rates. Conversely, when transcription of a gene is *activated*, both the mRNA and encoded protein or proteins are produced at much higher rates.

In most bacteria and other single-celled organisms, gene expression is highly regulated in order to adjust the cell's enzymatic machinery and structural components to changes in the nutritional and physical environment. Thus, at any given time, a bacterial cell normally synthesizes only those proteins of its entire proteome required for survival under the particular conditions. In multicellular organisms, control of gene expression is largely directed toward assuring that the right gene is expressed in the right cell at the right time during embryological development and tissue differentiation. Here we describe the basic features of transcription control in bacteria, using the *lac* operon in *E. coli* as our primary example. Many of the same processes, as well as others, are involved in eukaryotic transcription control, which is discussed in Chapter 11.

In *E. coli*, about half the genes are clustered into operons each of which encodes enzymes involved in a particular metabolic pathway or proteins that interact to form one multisubunit protein. For instance, the *trp* operon mentioned earlier encodes five enzymes needed in the biosynthesis of tryptophan (see Figure 4-12). Similarly, the *lac* operon encodes three enzymes required for the metabolism of lactose, a sugar present in milk. Since a bacterial operon is transcribed from one start site into a single mRNA, all the genes within an operon are coordinately regulated; that is, they are all activated or repressed to the same extent.

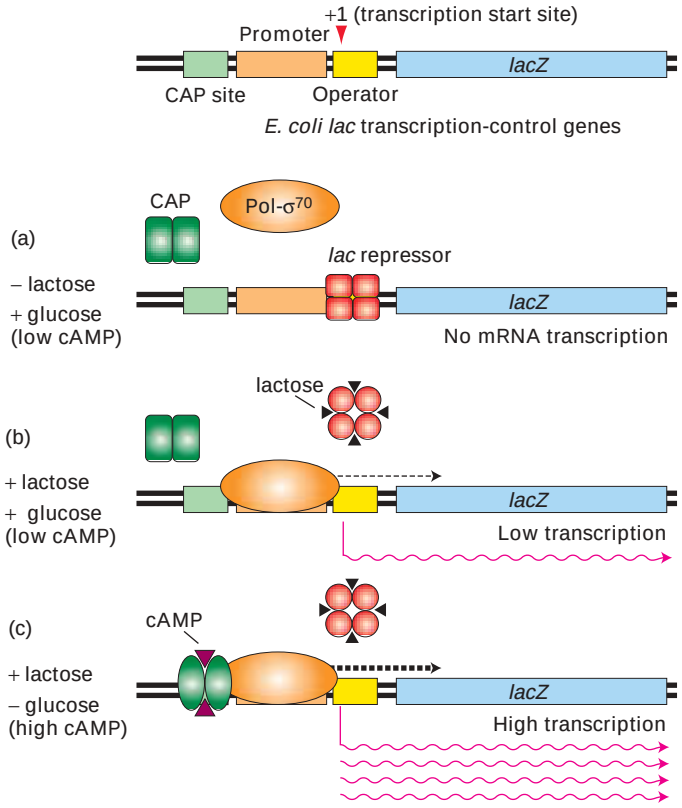
Transcription of operons, as well as of isolated genes, is controlled by an interplay between RNA polymerase and specific *repressor* and *activator* proteins. In order to initiate transcription, however, *E. coli* RNA polymerase must be associated with one of a small number of σ *(sigma) factors*, which function as initiation factors. The most common one in bacterial cells is σ^{70}.

Initiation of *lac* Operon Transcription Can Be Repressed and Activated

When *E. coli* is in an environment that lacks lactose, synthesis of *lac* mRNA is repressed, so that cellular energy is not wasted synthesizing enzymes the cells cannot use. In an environment containing both lactose and glucose, *E. coli* cells preferentially metabolize glucose, the central molecule of carbohydrate metabolism. Lactose is metabolized at a high rate only when lactose is present and glucose is largely depleted from the medium. This metabolic adjustment is achieved by repressing transcription of the *lac* operon until lactose is present, and synthesis of only low levels of *lac* mRNA until the cytosolic concentration of glucose falls to low levels. Transcription of the *lac* operon under different conditions is controlled by *lac* repressor and catabolite activator protein (CAP), each of which binds to a specific DNA sequence in the *lac* transcription-control region (Figure 4-16, *top*).

For transcription of the *lac* operon to begin, the σ^{70} subunit of the RNA polymerase must bind to the *lac* promoter, which lies just upstream of the start site. When no lactose is present, binding of the *lac* repressor to a sequence called the *lac* **operator,** which overlaps the transcription start site, blocks transcription initiation by the polymerase (Figure 4-16a). When lactose is present, it binds to specific binding sites in each subunit of the tetrameric *lac* repressor, causing a conformational change in the protein that makes it dissociate from the *lac* operator. As a result, the polymerase can initiate transcription of the *lac* operon. However, when glucose also is present, the rate of transcription initiation (i.e., the number of times per minute different polymerase molecules initiate transcription) is very low, resulting in synthesis of only low levels of *lac* mRNA and the proteins encoded in the *lac* operon (Figure 4-16b).

Once glucose is depleted from the media and the intracellular glucose concentration falls, *E. coli* cells respond by synthesizing cyclic AMP, cAMP (see Figure 3-27b). As the concentration of cAMP increases, it binds to a site in each subunit of the dimeric CAP protein, causing a conformational change that allows the protein to bind to the CAP site in the *lac* transcription-control region. The bound CAP-cAMP complex interacts with the polymerase bound to the promoter, greatly stimulating the rate of transcription initiation. This activation leads to synthesis of high levels of *lac*



▲ FIGURE 4-16 **Regulation of transcription from the *lac* operon of *E. coli*.** (*Top*) The transcription-control region, composed of ≈100 base pairs, includes three protein-binding regions: the CAP site, which binds catabolite activator protein; the *lac* promoter, which binds the RNA polymerase–σ^{70} complex; and the *lac* operator, which binds *lac* repressor. The *lacZ* gene, the first of three genes in the operon, is shown to the right. (a) In the absence of lactose, very little *lac* mRNA is produced because the *lac* repressor binds to the operator, inhibiting transcription initiation by RNA polymerase–σ^{70}. (b) In the presence of glucose and lactose, *lac* repressor binds lactose and dissociates from the operator, allowing RNA polymerase–σ^{70} to initiate transcription at a low rate. (c) Maximal transcription of the *lac* operon occurs in the presence of lactose and absence of glucose. In this situation, cAMP increases in response to the low glucose concentration and forms the CAP-cAMP complex, which binds to the CAP site, where it interacts with RNA polymerase to stimulate the rate of transcription initiation.

mRNA and subsequently of the enzymes encoded by the *lac* operon (Figure 4-16c).

Although the promoters for different *E. coli* genes exhibit considerable homology, their exact sequences differ. The promoter sequence determines the intrinsic rate at which an RNA polymerase–σ complex initiates transcription of a gene in the absence of a repressor or activator protein. Promoters that support a high rate of transcription initiation are called *strong promoters*. Those that support a low rate of transcription initiation are called *weak promoters*. The *lac* operon, for instance, has a weak promoter; its low intrinsic rate of initiation is further reduced by the *lac* repressor and substantially increased by the cAMP-CAP activator.

Small Molecules Regulate Expression of Many Bacterial Genes via DNA-Binding Repressors

Transcription of most *E. coli* genes is regulated by processes similar to those described for the *lac* operon. The general mechanism involves a specific repressor that binds to the operator region of a gene or operon, thereby blocking transcription initiation. A small molecule (or molecules), called an *inducer*, binds to the repressor, controlling its DNA-binding activity and consequently the rate of transcription as appropriate for the needs of the cell.

For example, when the tryptophan concentration in the medium and cytosol is high, the cell does not synthesize the several enzymes encoded in the *trp* operon. Binding of tryptophan to the *trp* repressor causes a conformational change that allows the protein to bind to the *trp* operator, thereby repressing expression of the enzymes that synthesize tryptophan. Conversely, when the tryptophan concentration in the medium and cytosol is low, tryptophan dissociates from the *trp* repressor, causing a conformational change in the protein that causes it to dissociate from the *trp* operator, allowing transcription of the *trp* operon. In the case of the *lac* operon, binding of the inducer lactose to the *lac* repressor reduces binding of the repressor to the operator, thereby promoting transcription.

Specific activator proteins, such as CAP in the lac operon, also control transcription of some but not all bacterial genes. These activators bind to DNA together with the RNA polymerase, stimulating transcription from a specific promoter. The DNA-binding activity of an activator is modulated in response to cellular needs by the binding of specific small molecules (e.g., cAMP) that alter the conformation of the activator.

Transcription by σ^{54}-RNA Polymerase Is Controlled by Activators That Bind Far from the Promoter

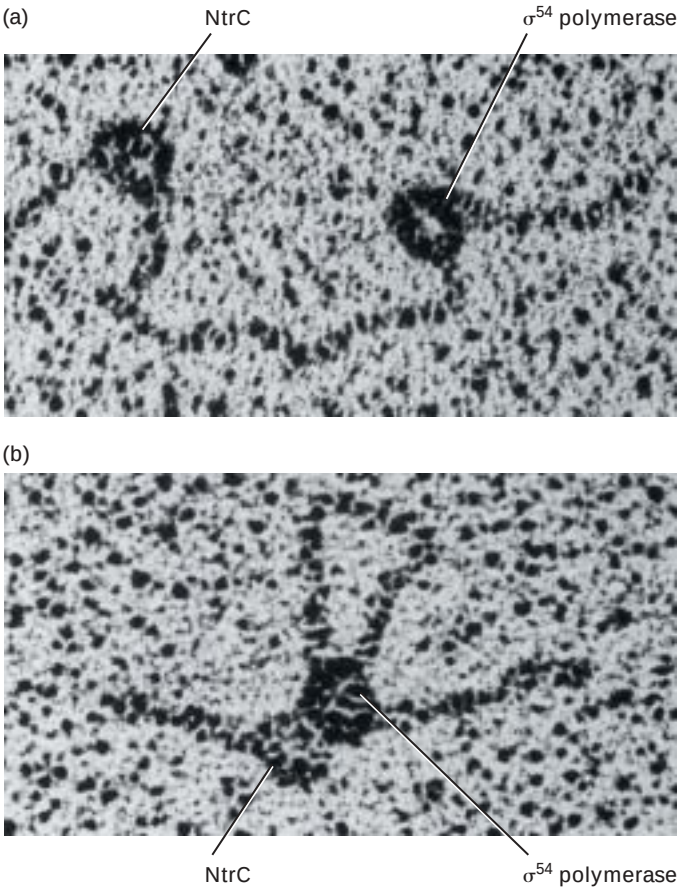
Most *E. coli* promoters interact with σ^{70}-RNA polymerase, the major form of the bacterial enzyme. Transcription of certain groups of genes, however, is carried out by *E. coli* RNA polymerases containing one of several alternative sigma factors that recognize different consensus promoter sequences than σ^{70} does. All but one of these are related to σ^{70} in sequence. Transcription initiation by RNA polymerases containing these σ^{70}-like factors is regulated by repressors and activators that bind to DNA near the region where the polymerase binds, similar to initiation by σ^{70}-RNA polymerase itself.

The sequence of one *E. coli* sigma factor, σ^{54}, is distinctly different from that of all the σ^{70}-like factors. Transcription of genes by RNA polymerases containing σ^{54} is regulated

solely by activators whose binding sites in DNA, referred to as **enhancers,** generally are located 80–160 base pairs upstream from the start site. Even when enhancers are moved more than a kilobase away from a start site, σ^{54}-activators can activate transcription.

The best-characterized σ^{54}-activator—the NtrC protein (nitrogen regulatory protein C)—stimulates transcription from the promoter of the *glnA* gene. This gene encodes the enzyme glutamine synthetase, which synthesizes the amino acid glutamine from glutamic acid and ammonia. The σ^{54}-RNA polymerase binds to the *glnA* promoter but does not melt the DNA strands and initiate transcription until it is activated by NtrC, a dimeric protein. NtrC, in turn, is regulated by a protein kinase called NtrB. In response to low levels of glutamine, NtrB phosphorylates dimeric NtrC, which then binds to an enhancer upstream of the *glnA* promoter. Enhancer-bound phosphorylated NtrC then stimulates the σ^{54}-polymerase bound at the promoter to separate the DNA strands and initiate transcription. Electron microscopy studies have shown that phosphorylated NtrC bound at enhancers and σ^{54}-polymerase bound at the promoter directly interact, forming a loop in the DNA between the binding sites (Figure 4-17). As discussed in Chapter 11, this activation mechanism is somewhat similar to the predominant mechanism of transcriptional activation in eukaryotes.

NtrC has ATPase activity, and ATP hydrolysis is required for activation of bound σ^{54}-polymerase by phosphorylated NtrC. Evidence for this is that mutants with an NtrC defective in ATP hydrolysis are invariably defective in stimulating the σ^{54}-polymerase to melt the DNA strands at the transcription start site. It is postulated that ATP hydrolysis supplies the energy required for melting the DNA strands. In contrast, the σ^{70}-polymerase does not require ATP hydrolysis to separate the strands at a start site.



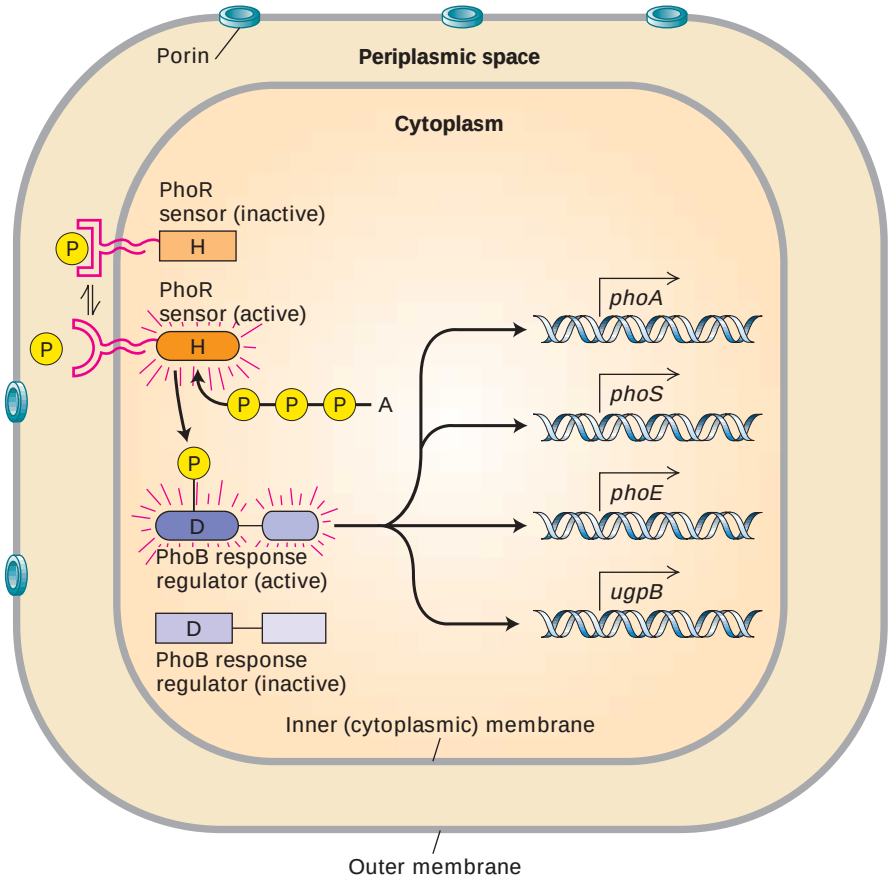
▲ EXPERIMENTAL FIGURE 4-17 **DNA looping permits interaction of bound NtrC and σ^{54}-polymerase.** (a) Electron micrograph of DNA restriction fragment with phosphorylated NtrC dimer binding to the enhancer region near one end and σ^{54} –RNA polymerase bound to the *glnA* promoter near the other end. (b) Electron micrograph of the same fragment preparation showing NtrC dimers and σ^{54}-polymerase binding to each other with the intervening DNA forming a loop between them. [From W. Su et al., 1990, *Proc. Nat'l. Acad. Sci. USA* **87:**5505; courtesy of S. Kustu.]

Many Bacterial Responses Are Controlled by Two-Component Regulatory Systems

As we've just seen, control of the *E. coli glnA* gene depends on two proteins, NtrC and NtrB. Such two-component regulatory systems control many responses of bacteria to changes in their environment. Another example involves the *E. coli* proteins PhoR and PhoB, which regulate transcription in response to the concentration of free phosphate. PhoR is a transmembrane protein, located in the inner (plasma) membrane, whose periplasmic domain binds phosphate with moderate affinity and whose cytosolic domain has protein kinase activity; PhoB is a cytosolic protein.

Large protein pores in the *E. coli* outer membrane allow ions to diffuse freely between the external environment and the periplasmic space. Consequently, when the phosphate concentration in the environment falls, it also falls in the periplasmic space, causing phosphate to dissociate from the PhoR periplasmic domain, as depicted in Figure 4-18. This causes a conformational change in the PhoR cytoplasmic domain that activates its protein kinase activity. The activated PhoR initially transfers a γ-phosphate from ATP to a histidine side chain in the PhoR kinase domain itself. The same phosphate is then transferred to a specific aspartic acid side chain in PhoB, converting PhoB from an inactive to an active transcriptional activator. Phosphorylated, active PhoB then induces transcription from several genes that help the cell cope with low phosphate conditions.

Many other bacterial responses are regulated by two proteins with homology to PhoR and PhoB. In each of these regulatory systems, one protein, called a *sensor,* contains a transmitter domain homologous to the PhoR protein kinase domain. The transmitter domain of the sensor protein is regulated by a second unique protein domain (e.g., the periplasmic domain of PhoR) that senses environmental changes. The second protein, called a *response regulator,* contains a



▶ FIGURE 4-18 **The PhoR/PhoB two-component regulatory system in *E. coli*.** In response to low phosphate concentrations in the environment and periplasmic space, a phosphate ion dissociates from the periplasmic domain of the inactive sensor protein PhoR. This causes a conformational change that activates a protein kinase transmitter domain in the cytosolic region of PhoR. The activated transmitter domain transfers an ATP γ phosphate to a conserved histidine in the transmitter domain. This phosphate is then transferred to an aspartic acid in the receiver domain of the response regulator PhoB. Several PhoB proteins can be phosphorylated by one activated PhoR. Phosphorylated PhoB proteins then activate transcription from genes encoding proteins that help the cell to respond to low phosphate, including *phoA, phoS, phoE,* and *ugpB.*

receiver domain homologous to the region of PhoB that is phosphorylated by activated PhoR. The receiver domain of the response regulator is associated with a second domain that determines the protein's function. The activity of this second functional domain is regulated by phosphorylation of the receiver domain. Although all transmitter domains are homologous (as are receiver domains), the transmitter domain of a specific sensor protein will phosphorylate only specific receiver domains of specific response regulators, allowing specific responses to different environmental changes. Note that NtrB and NtrC, discussed above, function as sensor and response regulator proteins, respectively, in the two-component regulatory system that controls transcription of *glnA*. Similar two-component histidyl-aspartyl phosphorelay regulatory systems are also found in plants.

KEY CONCEPTS OF SECTION 4.3

Control of Gene Expression in Prokaryotes

- Gene expression in both prokaryotes and eukaryotes is regulated primarily by mechanisms that control the initiation of transcription.
- Binding of the σ subunit in an RNA polymerase to a promoter region is the first step in the initiation of transcription in *E. coli.*
- The nucleotide sequence of a promoter determines its strength, that is, how frequently different RNA polymerase molecules can bind and initiate transcription per minute.
- Repressors are proteins that bind to operator sequences, which overlap or lie adjacent to promoters. Binding of a repressor to an operator inhibits transcription initiation.
- The DNA-binding activity of most bacterial repressors is modulated by small effector molecules (inducers). This allows bacterial cells to regulate transcription of specific genes in response to changes in the concentration of various nutrients in the environment.
- The *lac* operon and some other bacterial genes also are regulated by activator proteins that bind next to promoters and increase the rate of transcription initiation by RNA polymerase.
- The major sigma factor in *E. coli* is σ^{70}, but several other less abundant sigma factors are also found, each recognizing different consensus promoter sequences.
- Transcription initiation by all *E. coli* RNA polymerases, except those containing σ^{54}, can be regulated by repressors and activators that bind near the transcription start site (see Figure 4-16).
- Genes transcribed by σ^{54}–RNA polymerase are regulated by activators that bind to enhancers located ≈100 base

pairs upstream from the start site. When the activator and σ^{54}–RNA polymerase interact, the DNA between their binding sites forms a loop (see Figure 4-17).

■ In two-component regulatory systems, one protein acts as a sensor, monitoring the level of nutrients or other components in the environment. Under appropriate conditions, the γ-phosphate of an ATP is transferred first to a histidine in the sensor protein and then to an aspartic acid in a second protein, the response regulator. The phosphorylated response regulator then binds to DNA regulatory sequences, thereby stimulating or repressing transcription of specific genes (see Figure 4-18).

4.4 The Three Roles of RNA in Translation

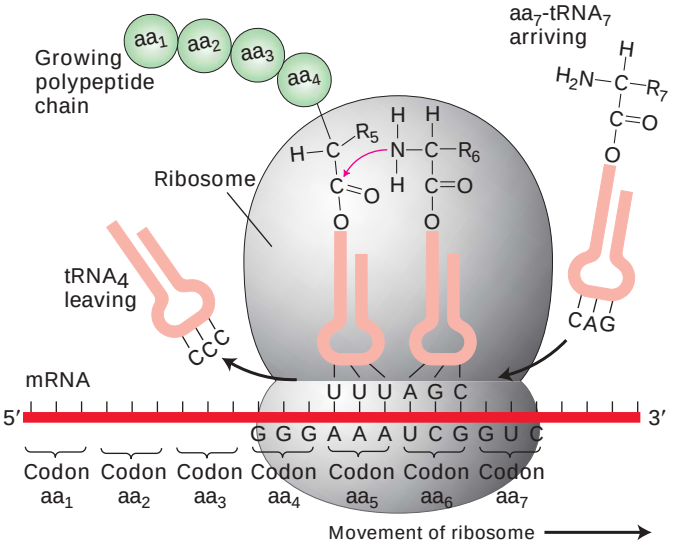
Although DNA stores the information for protein synthesis and mRNA conveys the instructions encoded in DNA, most biological activities are carried out by proteins. As we saw in Chapter 3, the linear order of amino acids in each protein determines its three-dimensional structure and activity. For this reason, assembly of amino acids in their correct order, as encoded in DNA, is critical to production of functional proteins and hence the proper functioning of cells and organisms.

Translation is the whole process by which the nucleotide sequence of an mRNA is used to order and to join the amino acids in a polypeptide chain (see Figure 4-1, step 3). In eukaryotic cells, protein synthesis occurs in the cytoplasm, where three types of RNA molecules come together to perform different but cooperative functions (Figure 4-19):

1. Messenger RNA (mRNA) carries the genetic information transcribed from DNA in the form of a series of three-nucleotide sequences, called **codons,** each of which specifies a particular amino acid.

2. Transfer RNA (tRNA) is the key to deciphering the codons in mRNA. Each type of amino acid has its own subset of tRNAs, which bind the amino acid and carry it to the growing end of a polypeptide chain if the next codon in the mRNA calls for it. The correct tRNA with its attached amino acid is selected at each step because each specific tRNA molecule contains a three-nucleotide sequence, an **anticodon,** that can base-pair with its complementary codon in the mRNA.

3. Ribosomal RNA (rRNA) associates with a set of proteins to form **ribosomes.** These complex structures, which physically move along an mRNA molecule, catalyze the assembly of amino acids into polypeptide chains. They also bind tRNAs and various accessory proteins necessary for protein synthesis. Ribosomes are composed of a large and a small subunit, each of which contains its own rRNA molecule or molecules.



▲ FIGURE 4-19 **The three roles of RNA in protein synthesis.** Messenger RNA (mRNA) is translated into protein by the joint action of transfer RNA (tRNA) and the ribosome, which is composed of numerous proteins and two major ribosomal RNA (rRNA) molecules (not shown). Note the base pairing between tRNA anticodons and complementary codons in the mRNA. Formation of a peptide bond between the amino group N on the incoming aa-tRNA and the carboxyl-terminal C on the growing protein chain (purple) is catalyzed by one of the rRNAs. aa = amino acid; R = side group. [Adapted from A. J. F. Griffiths et al., 1999, *Modern Genetic Analysis,* W. H. Freeman and Company.]

These three types of RNA participate in translation in all cells. Indeed, development of three functionally distinct RNAs was probably the molecular key to the origin of life. How the structure of each RNA relates to its specific task is described in this section; how the three types work together, along with required protein factors, to synthesize proteins is detailed in the following section. Since translation is essential for protein synthesis, the two processes commonly are referred to interchangeably. However, the polypeptide chains resulting from translation undergo post-translational folding and often other changes (e.g., chemical modifications, association with other chains) that are required for production of mature, functional proteins (Chapter 3).

Messenger RNA Carries Information from DNA in a Three-Letter Genetic Code

As noted above, the **genetic code** used by cells is a *triplet* code, with every three-nucleotide sequence, or codon, being "read" from a specified starting point in the mRNA. Of the 64 possible codons in the genetic code, 61 specify individual amino acids and three are stop codons. Table 4-1 shows that most amino acids are encoded by more than one codon. Only two—methionine and tryptophan—have a single

TABLE 4-1 The Genetic Code (RNA to Amino Acids)*

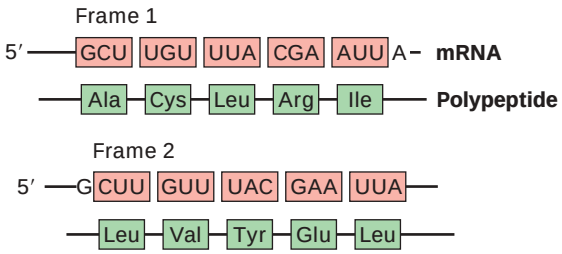
First Position (5′ end)	Second Position U	Second Position C	Second Position A	Second Position G	Third Position (3′ end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu (Met)*	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met (start)	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val (Met)*	Ala	Glu	Gly	G

*AUG is the most common initiator codon; GUG usually codes for valine, and CUG for leucine, but, rarely, these codons can also code for methionine to initiate a protein chain.

codon; at the other extreme, leucine, serine, and arginine are each specified by six different codons. The different codons for a given amino acid are said to be synonymous. The code itself is termed *degenerate,* meaning that more than one codon can specify the same amino acid.

Synthesis of all polypeptide chains in prokaryotic and eukaryotic cells begins with the amino acid methionine. In most mRNAs, the *start (initiator) codon* specifying this amino-terminal methionine is AUG. In a few bacterial mRNAs, GUG is used as the initiator codon, and CUG occasionally is used as an initiator codon for methionine in eukaryotes. The three codons UAA, UGA, and UAG do not specify amino acids but constitute *stop (termination) codons* that mark the carboxyl terminus of polypeptide chains in almost all cells. The sequence of codons that runs from a specific start codon to a stop codon is called a **reading frame.** This precise linear array of ribonucleotides in groups of three in mRNA specifies the precise linear sequence of amino acids in a polypeptide chain and also signals where synthesis of the chain starts and stops.

Because the genetic code is a comma-less, non-overlapping triplet code, a particular mRNA theoretically could be translated in three different reading frames. Indeed some mRNAs have been shown to contain overlapping information that can be translated in different reading frames, yielding different polypeptides (Figure 4-20). The vast majority of mRNAs, however, can be read in only one frame because stop codons encountered in the other two possible reading frames terminate translation before a functional protein is produced. Another unusual coding arrangement occurs because of *frame-*



▲ FIGURE 4-20 **Example of how the genetic code—a non-overlapping, comma-less triplet code—can be read in different frames.** If translation of the mRNA sequence shown begins at two different upstream start sites (not shown), then two overlapping reading frames are possible. In this example, the codons are shifted one base to the right in the lower frame. As a result, the same nucleotide sequence specifies different amino acids during translation. Although they are rare, many instances of such overlaps have been discovered in viral and cellular genes of prokaryotes and eukaryotes. It is theoretically possible for the mRNA to have a third reading frame.

shifting. In this case the protein-synthesizing machinery may read four nucleotides as one amino acid and then continue reading triplets, or it may back up one base and read all succeeding triplets in the new frame until termination of the chain occurs. These frameshifts are not common events, but a few dozen such instances are known.

The meaning of each codon is the same in most known organisms—a strong argument that life on earth evolved only once. However, the genetic code has been found to differ for a few codons in many mitochondria, in ciliated protozoans, and in *Acetabularia*, a single-celled plant. As shown in Table 4-2, most of these changes involve reading of normal stop codons as amino acids, not an exchange of one amino acid for another. These exceptions to the general code probably were later evolutionary developments; that is, at no single time was the code immutably fixed, although massive changes were not tolerated once a general code began to function early in evolution.

The Folded Structure of tRNA Promotes Its Decoding Functions

Translation, or decoding, of the four-nucleotide language of DNA and mRNA into the 20–amino acid language of proteins requires tRNAs and enzymes called *aminoacyl-tRNA synthetases*. To participate in protein synthesis, a tRNA molecule must become chemically linked to a particular amino acid via a high-energy bond, forming an **aminoacyl-tRNA;** the anticodon in the tRNA then base-pairs with a codon in mRNA so that the activated amino acid can be added to the growing polypeptide chain (Figure 4-21).

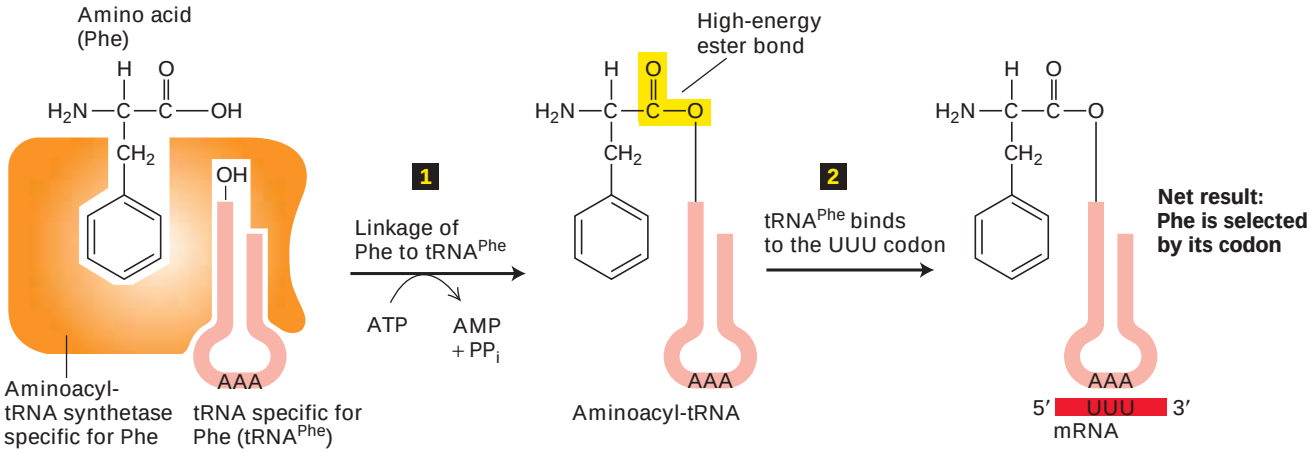
Some 30–40 different tRNAs have been identified in bacterial cells and as many as 50–100 in animal and plant cells. Thus the number of tRNAs in most cells is more than the number of amino acids used in protein synthesis (20) and also differs from the number of amino acid codons in the genetic code (61). Consequently, many amino acids have more than one tRNA to which they can attach (explaining how there can be more tRNAs than amino acids); in addition, many tRNAs can pair with more than one codon (explaining how there can be more codons than tRNAs).

The function of tRNA molecules, which are 70–80 nucleotides long, depends on their precise three-dimensional structures. In solution, all tRNA molecules fold into a similar stem-loop arrangement that resembles a cloverleaf when drawn in two dimensions (Figure 4-22a). The four stems are short double helices stabilized by Watson-Crick base pairing; three of the four stems have loops containing seven or eight bases at their ends, while the remaining, unlooped stem contains the free 3′ and 5′ ends of the chain. The three nucleotides composing the anticodon are located at the center of the middle loop, in an accessible position that facilitates codon-anticodon base pairing. In all tRNAs, the 3′ end of the unlooped amino acid *acceptor stem* has the sequence CCA, which in most cases is added after synthesis and processing of the tRNA are complete. Several bases in most tRNAs also are modified after synthesis. Viewed in three

TABLE 4-2 Known Deviations from the Universal Genetic Code

Codon	Universal Code	Unusual Code*	Occurrence
UGA	Stop	Trp	*Mycoplasma, Spiroplasma,* mitochondria of many species
CUG	Leu	Thr	Mitochondria in yeasts
UAA, UAG	Stop	Gln	*Acetabularia, Tetrahymena,* Paramecium, etc.
UGA	Stop	Cys	*Euplotes*

*"Unusual code" is used in nuclear genes of the listed organisms and in mitochondrial genes as indicated.
SOURCE: S. Osawa et al., 1992, *Microbiol. Rev.* **56**:229.



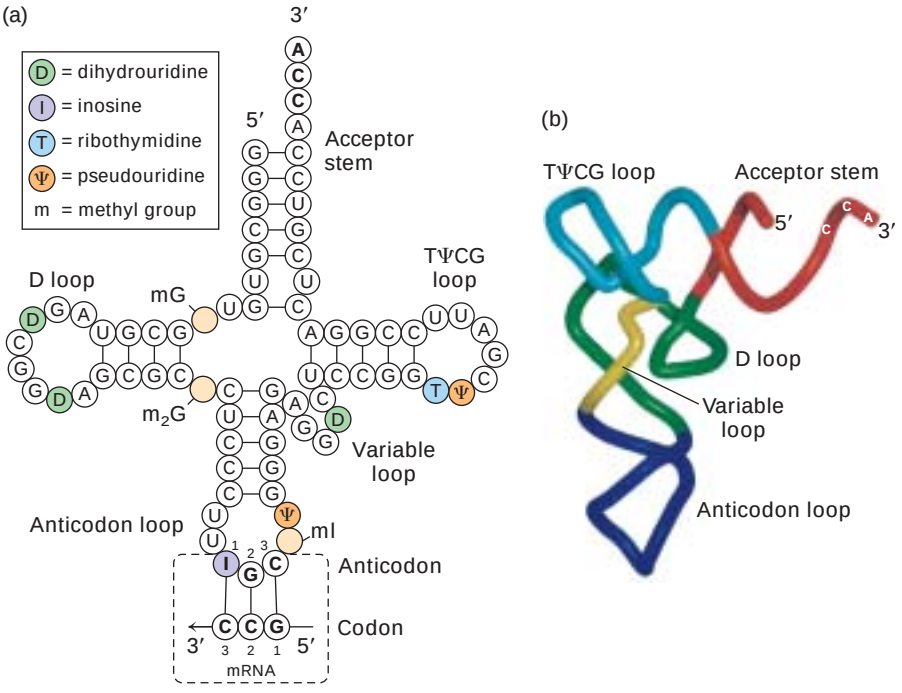
▲ FIGURE 4-21 **Two-step decoding process for translating nucleic acid sequences in mRNA into amino acid sequences in proteins.** Step 1: An aminoacyl-tRNA synthetase first couples a specific amino acid, via a high-energy ester bond (yellow), to either the 2′ or 3′ hydroxyl of the terminal adenosine in the corresponding tRNA. Step 2: A three-base sequence in the tRNA (the anticodon) then base-pairs with a codon in the mRNA specifying the attached amino acid. If an error occurs in either step, the wrong amino acid may be incorporated into a polypeptide chain. Phe = phenylalanine.

dimensions, the folded tRNA molecule has an L shape with the anticodon loop and acceptor stem forming the ends of the two arms (Figure 4-22b).

Nonstandard Base Pairing Often Occurs Between Codons and Anticodons

If perfect Watson-Crick base pairing were demanded between codons and anticodons, cells would have to contain exactly 61 different tRNA species, one for each codon that specifies an amino acid. As noted above, however, many cells contain fewer than 61 tRNAs. The explanation for the smaller number lies in the capability of a single tRNA anticodon to recognize more than one, but not necessarily every, codon corresponding to a given amino acid. This broader recognition can occur because of nonstandard pairing between bases in the so-called *wobble* position: that is, the third (3′) base in an mRNA codon and the corresponding first (5′) base in its tRNA anticodon.

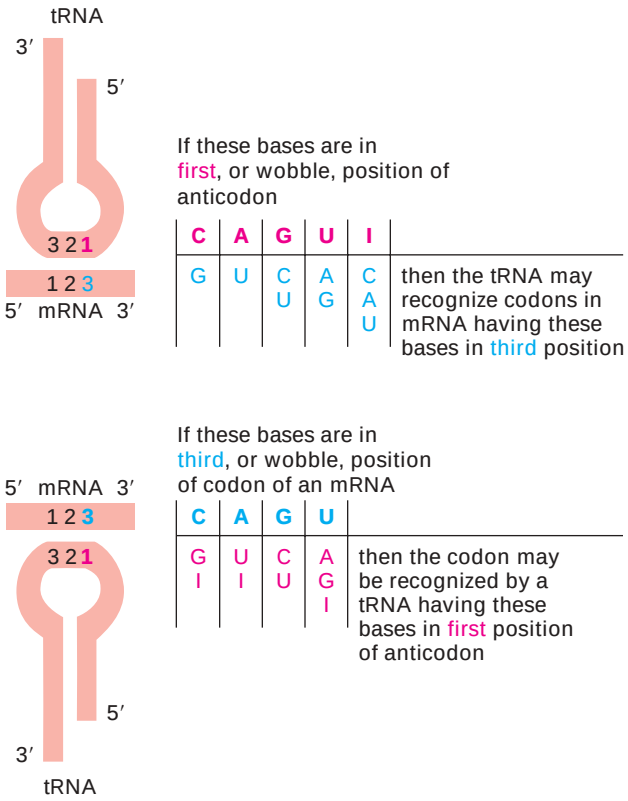
The first and second bases of a codon almost always form standard Watson-Crick base pairs with the third and



▶ FIGURE 4-22 **Structure of tRNAs.** (a) Although the exact nucleotide sequence varies among tRNAs, they all fold into four base-paired stems and three loops. The CCA sequence at the 3′ end also is found in all tRNAs. Attachment of an amino acid to the 3′ A yields an aminoacyl-tRNA. Some of the A, C, G, and U residues are modified in most tRNAs (see key). Dihydrouridine (D) is nearly always present in the D loop; likewise, ribothymidine (T) and pseudouridine (Ψ) are almost always present in the TΨCG loop. Yeast alanine tRNA, represented here, also contains other modified bases. The triplet at the tip of the anticodon loop base-pairs with the corresponding codon in mRNA. (b) Three-dimensional model of the generalized backbone of all tRNAs. Note the L shape of the molecule. [Part (a) see R. W. Holly et al., 1965, *Science* **147**:1462; part (b) from J. G. Arnez and D. Moras, 1997, *Trends Biochem. Sci.* **22**:211.]

second bases, respectively, of the corresponding anticodon, but four nonstandard interactions can occur between bases in the wobble position. Particularly important is the G·U base pair, which structurally fits almost as well as the standard G·C pair. Thus, a given anticodon in tRNA with G in the first (wobble) position can base-pair with the two corresponding codons that have either pyrimidine (C or U) in the third position (Figure 4-23). For example, the phenylalanine codons UUU and UUC (5′→3′) are both recognized by the tRNA that has GAA (5′→3′) as the anticodon. In fact, any two codons of the type NNPyr (N = any base; Pyr = pyrimidine) encode a single amino acid and are decoded by a single tRNA with G in the first (wobble) position of the anticodon.

Although adenine rarely is found in the anticodon wobble position, many tRNAs in plants and animals contain inosine (I), a deaminated product of adenine, at this position. Inosine can form nonstandard base pairs with A, C, and U. A tRNA with inosine in the wobble position thus can recognize the corresponding mRNA codons with A, C, or U in the third (wobble) position (see Figure 4-23). For this reason, inosine-containing tRNAs are heavily employed in translation of the synonymous codons that specify a single amino acid. For example, four of the six codons for leucine (CUA, CUC, CUU, and UUA) are all recognized by the same tRNA with the anticodon 3′-GAI-5′; the inosine in the wobble position forms nonstandard base pairs with the third base in the four codons. In the case of the UUA codon, a nonstandard G·U pair also forms between position 3 of the anticodon and position 1 of the codon.



▲ FIGURE 4-23 **Nonstandard codon-anticodon base pairing at the wobble position.** The base in the third (or wobble) position of an mRNA codon often forms a nonstandard base pair with the base in the first (or wobble) position of a tRNA anticodon. Wobble pairing allows a tRNA to recognize more than one mRNA codon *(top)*; conversely, it allows a codon to be recognized by more than one kind of tRNA *(bottom),* although each tRNA will bear the same amino acid. Note that a tRNA with I (inosine) in the wobble position can "read" (become paired with) three different codons, and a tRNA with G or U in the wobble position can read two codons. Although A is theoretically possible in the wobble position of the anticodon, it is almost never found in nature.

Aminoacyl-tRNA Synthetases Activate Amino Acids by Covalently Linking Them to tRNAs

Recognition of the codon or codons specifying a given amino acid by a particular tRNA is actually the second step in decoding the genetic message. The first step, attachment of the appropriate amino acid to a tRNA, is catalyzed by a specific aminoacyl-tRNA synthetase. Each of the 20 different synthetases recognizes *one* amino acid and *all* its compatible, or *cognate*, tRNAs. These coupling enzymes link an amino acid to the free 2′ or 3′ hydroxyl of the adenosine at the 3′ terminus of tRNA molecules by an ATP-requiring reaction. In this reaction, the amino acid is linked to the tRNA by a high-energy bond and thus is said to be *activated*. The energy of this bond subsequently drives formation of the peptide bonds linking adjacent amino acids in a growing polypeptide chain. The equilibrium of the aminoacylation reaction is driven further toward activation of the amino acid by hydrolysis of the high-energy phosphoanhydride bond in the released pyrophosphate (see Figure 4-21).

Because some amino acids are so similar structurally, aminoacyl-tRNA synthetases sometimes make mistakes. These are corrected, however, by the enzymes themselves, which have a *proofreading* activity that checks the fit in their amino acid–binding pocket. If the wrong amino acid becomes attached to a tRNA, the bound synthetase catalyzes removal of the amino acid from the tRNA. This crucial function helps guarantee that a tRNA delivers the correct amino acid to the protein-synthesizing machinery. The overall error rate for translation in *E. coli* is very low, approximately 1 per 50,000 codons, evidence of the importance of proofreading by aminoacyl-tRNA synthetases.

Ribosomes Are Protein-Synthesizing Machines

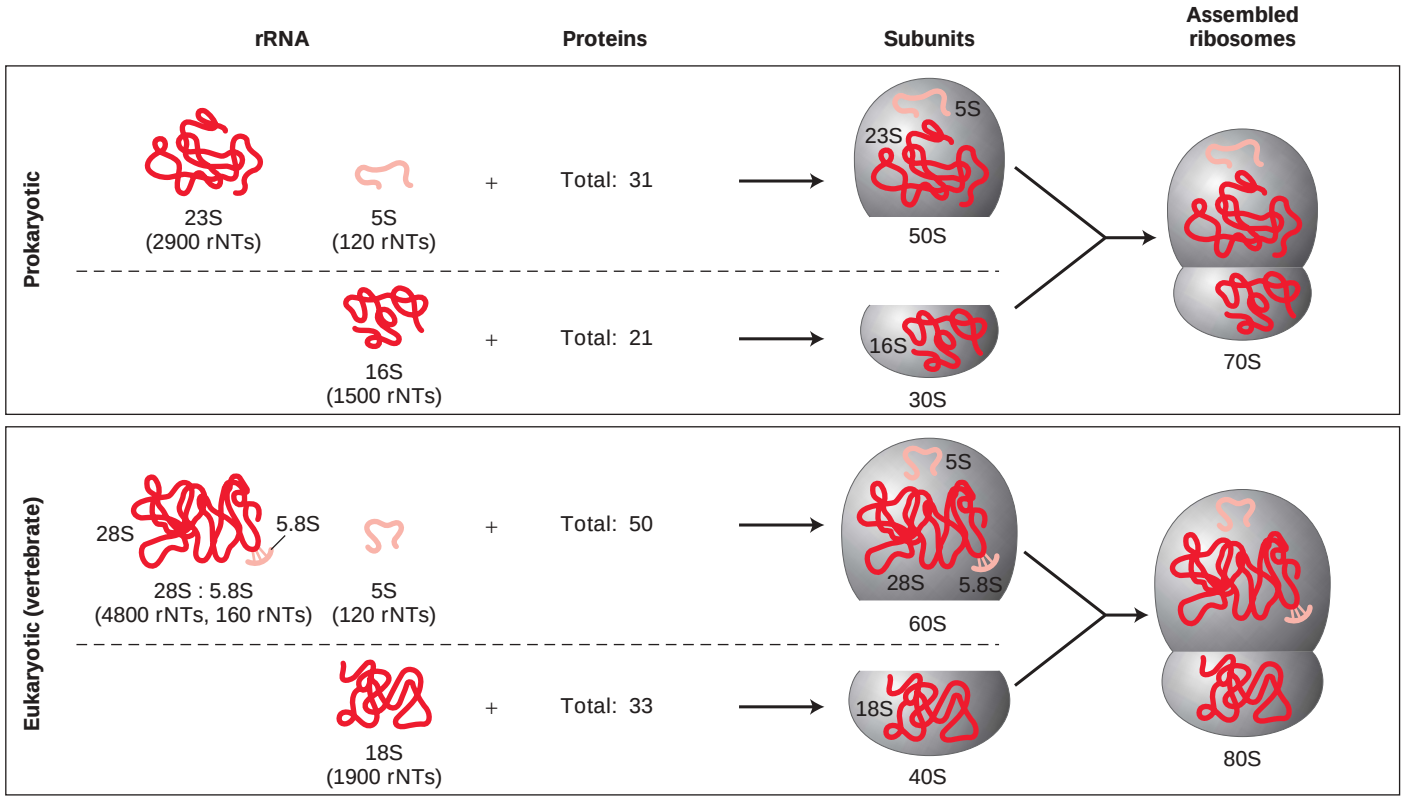
If the many components that participate in translating mRNA had to interact in free solution, the likelihood of simultaneous collisions occurring would be so low that the rate of amino acid polymerization would be very slow. The efficiency of translation is greatly increased by the binding of the mRNA and the individual aminoacyl-tRNAs to the most

abundant RNA-protein complex in the cell, the ribosome, which directs elongation of a polypeptide at a rate of three to five amino acids added per second. Small proteins of 100–200 amino acids are therefore made in a minute or less. On the other hand, it takes 2–3 hours to make the largest known protein, titin, which is found in muscle and contains about 30,000 amino acid residues. The cellular machine that accomplishes this task must be precise and persistent.

With the aid of the electron microscope, ribosomes were first discovered as small, discrete, RNA-rich particles in cells that secrete large amounts of protein. However, their role in protein synthesis was not recognized until reasonably pure ribosome preparations were obtained. In vitro radiolabeling experiments with such preparations showed that radioactive amino acids first were incorporated into growing polypeptide chains that were associated with ribosomes before appearing in finished chains.

A ribosome is composed of three (in bacteria) or four (in eukaryotes) different rRNA molecules and as many as 83 proteins, organized into a large subunit and a small subunit (Figure 4-24). The ribosomal subunits and the rRNA molecules are commonly designated in Svedberg units (S), a measure of the sedimentation rate of suspended particles centrifuged under standard conditions. The small ribosomal subunit contains a single rRNA molecule, referred to as *small rRNA*. The large subunit contains a molecule of *large rRNA* and one molecule of 5S rRNA, plus an additional molecule of 5.8S rRNA in vertebrates. The lengths of the rRNA molecules, the quantity of proteins in each subunit, and consequently the sizes of the subunits differ in bacterial and eukaryotic cells. The assembled ribosome is 70S in bacteria and 80S in vertebrates. But more interesting than these differences are the great structural and functional similarities between ribosomes from all species. This consistency is another reflection of the common evolutionary origin of the most basic constituents of living cells.

The sequences of the small and large rRNAs from several thousand organisms are now known. Although the primary nucleotide sequences of these rRNAs vary considerably, the same parts of each type of rRNA theoretically can form base-paired stem-loops, which would generate a similar three-dimensional structure for each rRNA in all organisms. The actual three-dimensional structures of bacterial rRNAs from *Thermus thermopolis* recently have been determined by x-ray crystallography of the 70S ribosome. The multiple, much smaller ribosomal proteins for the most part are associated



▲ FIGURE 4-24 **The general structure of ribosomes in prokaryotes and eukaryotes.** In all cells, each ribosome consists of a large and a small subunit. The two subunits contain rRNAs (red) of different lengths, as well as a different set of proteins. All ribosomes contain two major rRNA molecules (23S and 16S rRNA in bacteria; 28S and 18S rRNA in vertebrates) and a 5S rRNA. The large subunit of vertebrate ribosomes also contains a 5.8S rRNA base-paired to the 28S rRNA. The number of ribonucleotides (rNTs) in each rRNA type is indicated.

with the surface of the rRNAs. Although the number of protein molecules in ribosomes greatly exceeds the number of RNA molecules, RNA constitutes about 60 percent of the mass of a ribosome.

During translation, a ribosome moves along an mRNA chain, interacting with various protein factors and tRNAs and very likely undergoing large conformational changes. Despite the complexity of the ribosome, great progress has been made in determining the overall structure of bacterial ribosomes and in identifying various reactive sites. X-ray crystallographic studies on the *T. thermophilus* 70S ribosome, for instance, not only have revealed the dimensions and overall shape of the ribosomal subunits but also have localized the positions of tRNAs bound to the ribosome during elongation of a growing protein chain. In addition, powerful chemical techniques such as **footprinting,** which is described in Chapter 11, have been used to identify specific nucleotide sequences in rRNAs that bind to protein or another RNA. Some 40 years after the initial discovery of ribosomes, their overall structure and functioning during protein synthesis are finally becoming clear, as we describe in the next section.

KEY CONCEPTS OF SECTION 4.4

The Three Roles of RNA in Translation

- Genetic information is transcribed from DNA into mRNA in the form of a comma-less, overlapping, degenerate triplet code.
- Each amino acid is encoded by one or more three-nucleotide sequences (codons) in mRNA. Each codon specifies one amino acid, but most amino acids are encoded by multiple codons (see Table 4-1).
- The AUG codon for methionine is the most common start codon, specifying the amino acid at the NH_2-terminus of a protein chain. Three codons (UAA, UAG, UGA) function as stop codons and specify no amino acids.
- A reading frame, the uninterrupted sequence of codons in mRNA from a specific start codon to a stop codon, is translated into the linear sequence of amino acids in a polypeptide chain.
- Decoding of the nucleotide sequence in mRNA into the amino acid sequence of proteins depends on tRNAs and aminoacyl-tRNA synthetases.
- All tRNAs have a similar three-dimensional structure that includes an acceptor arm for attachment of a specific amino acid and a stem-loop with a three-base anticodon sequence at its ends (see Figure 4-22). The anticodon can base-pair with its corresponding codon in mRNA.
- Because of nonstandard interactions, a tRNA may base-pair with more than one mRNA codon; conversely, a particular codon may base-pair with multiple tRNAs. In each case, however, only the proper amino acid is inserted into a growing polypeptide chain.
- Each of the 20 aminoacyl-tRNA synthetases recognizes a single amino acid and covalently links it to a cognate tRNA, forming an aminoacyl-tRNA (see Figure 4-21). This reaction activates the amino acid, so it can participate in peptide bond formation.
- Both prokaryotic and eukaryotic ribosomes—the large ribonucleoprotein complexes on which translation occurs—consist of a small and a large subunit (see Figure 4-24). Each subunit contains numerous different proteins and one major rRNA molecule (small or large). The large subunit also contains one accessory 5S rRNA in bacteria and two accessory rRNAs in eukaryotes (5S and 5.8S in vertebrates).
- Analogous rRNAs from many different species fold into quite similar three-dimensional structures containing numerous stem-loops and binding sites for proteins, mRNA, and tRNAs. Much smaller ribosomal proteins are associated with the periphery of the rRNAs.

4.5 Stepwise Synthesis of Proteins on Ribosomes

The previous sections have introduced the major participants in protein synthesis—mRNA, aminoacylated tRNAs, and ribosomes containing large and small rRNAs. We now take a detailed look at how these components are brought together to carry out the biochemical events leading to formation of polypeptide chains on ribosomes. Similar to transcription, the complex process of translation can be divided into three stages—initiation, elongation, and termination—which we consider in order. We focus our description on translation in eukaryotic cells, but the mechanism of translation is fundamentally the same in all cells.

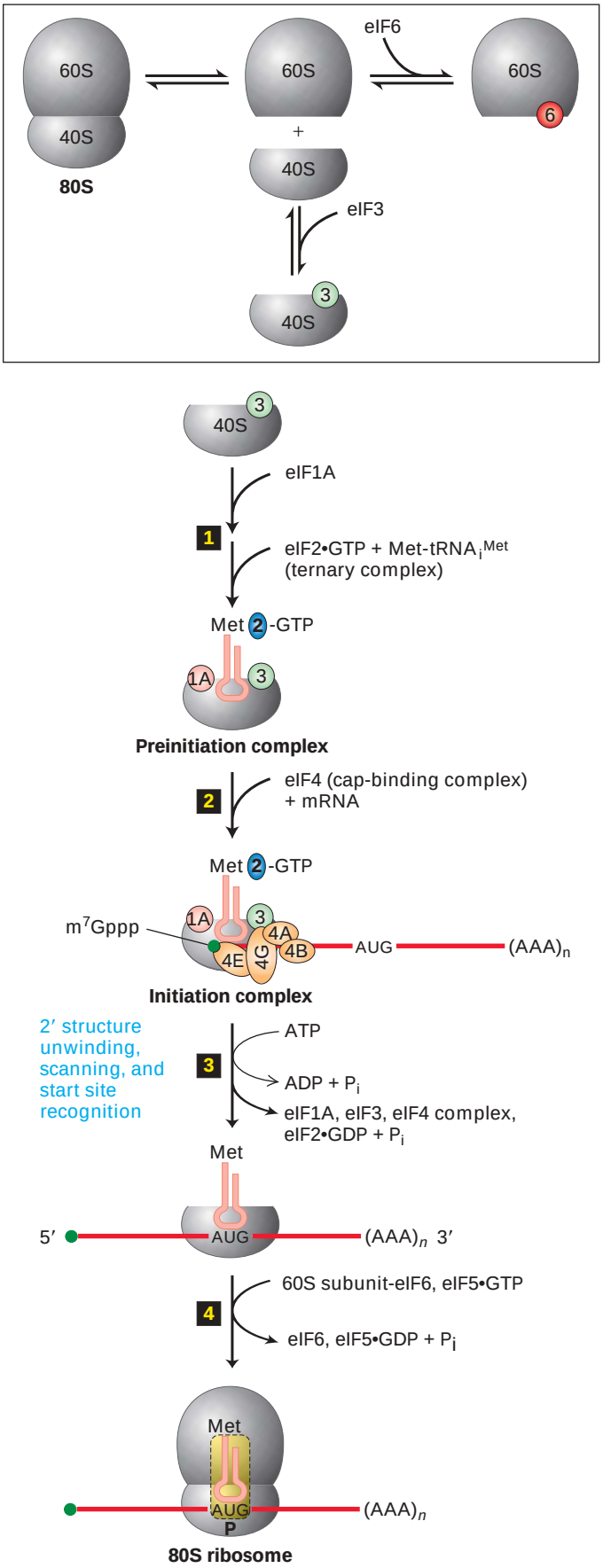
Methionyl-$tRNA_i^{Met}$ Recognizes the AUG Start Codon

As noted earlier, the AUG codon for methionine functions as the start codon in the vast majority of mRNAs. A critical aspect of translation initiation is to begin protein synthesis at the start codon, thereby establishing the correct reading frame for the entire mRNA. Both prokaryotes and eukaryotes contain two different methionine tRNAs: $tRNA_i^{Met}$ can initiate protein synthesis, and $tRNA^{Met}$ can incorporate methionine only into a growing protein chain. The same aminoacyl-tRNA synthetase (MetRS) charges both tRNAs with methionine. But *only* Met-$tRNA_i^{Met}$ (i.e., activated methionine attached to $tRNA_i^{Met}$) can bind at the appropriate site on the small ribosomal subunit, the *P site,* to begin synthesis of a polypeptide chain. The regular Met-$tRNA^{Met}$ and all other charged tRNAs bind only to another ribosomal site, the *A site,* as described later.

Translation Initiation Usually Occurs Near the First AUG Closest to the 5′ End of an mRNA

During the first stage of translation, a ribosome assembles, complexed with an mRNA and an activated initiator tRNA, which is correctly positioned at the start codon. Large and small ribosomal subunits not actively engaged in translation are kept apart by binding of two **initiation factors,** designated eIF3 and eIF6 in eukaryotes. A translation *preinitiation complex* is formed when the 40S subunit–eIF3 complex is bound by eIF1A and a ternary complex of the Met-tRNA$_i^{Met}$, eIF2, and GTP (Figure 4-25, step 1). Cells can regulate protein synthesis by phosphorylating a serine residue on the eIF2 bound to GDP; the phosphorylated complex is unable to exchange the bound GDP for GTP and cannot bind Met-tRNA$_i^{Met}$, thus inhibiting protein synthesis.

During translation initiation, the 5′ cap of an mRNA to be translated is bound by the eIF4E subunit of the eIF4 cap-binding complex. The mRNA-eIF4 complex then associates with the preinitiation complex through an interaction of the eIF4G subunit and eIF3, forming the *initiation complex* (Figure 4-25, step 2). The initiation complex then probably slides along, or *scans,* the associated mRNA as the **helicase** activity of eIF4A uses energy from ATP hydrolysis to unwind the RNA secondary structure. Scanning stops when the tRNA$_i^{Met}$ anticodon recognizes the start codon, which is the first AUG downstream from the 5′ end in most eukaryotic mRNAs (step 3). Recognition of the start codon leads to hydrolysis of the GTP associated with eIF2, an irreversible step that prevents further scanning. Selection of the initiating AUG is facilitated by specific surrounding nucleotides called the *Kozak sequence,* for Marilyn Kozak, who defined it: (5′) **A**CCAUG**G** (3′). The A preceding the AUG (underlined) and the G immediately following it are the most important nucleotides affecting translation initiation efficiency. Once the small ribosomal subunit with its bound Met-tRNA$_i^{Met}$ is correctly positioned at the start codon, union with the large (60S) ribosomal subunit completes formation of an 80S ribosome. This requires the action of another factor (eIF5) and hydrolysis



▶ FIGURE 4-25 **Initiation of translation in eukaryotes.** *(Inset)* When a ribosome dissociates at the termination of translation, the 40S and 60S subunits associate with initiation factors eIF3 and eIF6, forming complexes that can initiate another round of translation. Steps 1 and 2: Sequential addition of the indicated components to the 40S subunit–eIF3 complex forms the initiation complex. Step 3: Scanning of the mRNA by the associated initiation complex leads to positioning of the small subunit and bound Met-tRNA$_i^{Met}$ at the start codon. Step 4: Association of the large subunit (60S) forms an 80S ribosome ready to translate the mRNA. Two initiation factors, eIF2 (step 1) and eIF5 (step 4) are GTP-binding proteins, whose bound GTP is hydrolyzed during translation initiation. The precise time at which particular initiation factors are released is not yet well characterized. See the text for details. [Adapted from R. Mendez and J. D. Richter, 2001, *Nature Rev. Mol. Cell Biol.* **2**:521]

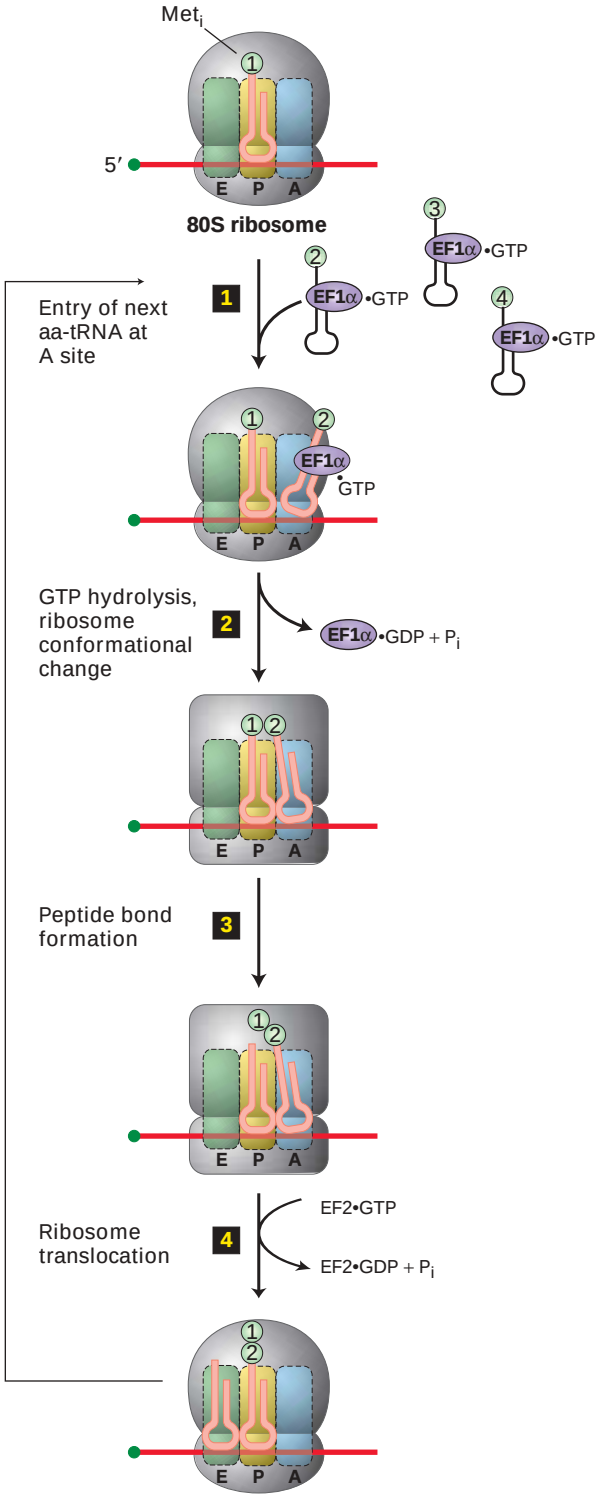
of a GTP associated with it (step 4). Coupling the joining reaction to GTP hydrolysis makes this an irreversible step, so that the ribosomal subunits do not dissociate until the entire mRNA is translated and protein synthesis is terminated. As discussed later, during chain elongation, the growing polypeptide remains attached to the tRNA at this P site in the ribosome.

The eukaryotic protein-synthesizing machinery begins translation of most cellular mRNAs within about 100 nucleotides of the 5′ capped end as just described. However, some cellular mRNAs contain an internal ribosome entry site (IRES) located far downstream of the 5′ end. In addition, translation of some viral mRNAs, which lack a 5′ cap, is initiated at IRESs by the host-cell machinery of infected eukaryotic cells. Some of the same translation initiation factors that assist in ribosome scanning from a 5′ cap are required for locating an internal AUG start codon, but exactly how an IRES is recognized is less clear. Recent results indicate that some IRESs fold into an RNA structure that binds to a third site on the ribosome, the *E site*, thereby positioning a nearby internal AUG start codon in the P site.

During Chain Elongation Each Incoming Aminoacyl-tRNA Moves Through Three Ribosomal Sites

The correctly positioned eukaryotic 80S ribosome–Met-tRNA$_i^{Met}$ complex is now ready to begin the task of stepwise addition of amino acids by the in-frame translation of the mRNA. As is the case with initiation, a set of special proteins, termed **elongation factors (EFs),** are required to carry out this process of chain elongation. The key steps in elongation are entry of each succeeding aminoacyl-tRNA, formation of a peptide bond, and the movement, or *translocation*, of the ribosome one codon at a time along the mRNA.

At the completion of translation initiation, as noted already, Met-tRNA$_i^{Met}$ is bound to the P site on the assembled 80S ribosome (Figure 4-26, *top*). This region of the ribosome is called the *P* site because the tRNA chemically linked to the growing polypeptide chain is located here. The second



► FIGURE 4-26 **Cycle of peptidyl chain elongation during translation in eukaryotes.** Once the 80S ribosome with Met-tRNA$_i^{Met}$ in the ribosome P site is assembled *(top)*, a ternary complex bearing the second amino acid (aa$_2$) coded by the mRNA binds to the A site (step 1). Following a conformational change in the ribosome induced by hydrolysis of GTP in EF1α·GTP (step 2), the large rRNA catalyzes peptide bond formation between Met$_i$ and aa$_2$ (step 3). Hydrolysis of GTP in EF2·GTP causes another conformational change in the ribosome that results in its translocation one codon along the mRNA and shifts the unacylated tRNA$_i^{Met}$ to the E site and the tRNA with the bound peptide to the P site (step 4). The cycle can begin again with binding of a ternary complex bearing aa$_3$ to the now-open A site. In the second and subsequent elongation cycles, the tRNA at the E site is ejected during step 2 as a result of the conformational change induced by hydrolysis of GTP in EF1α·GTP. See the text for details. [Adapted from K. H. Nierhaus et al., 2000, in R. A. Garrett et al., eds., *The Ribosome: Structure, Function, Antibiotics, and Cellular Interactions,* ASM Press, p. 319.]

MEDIA CONNECTIONS
Focus Animation: Protein Synthesis

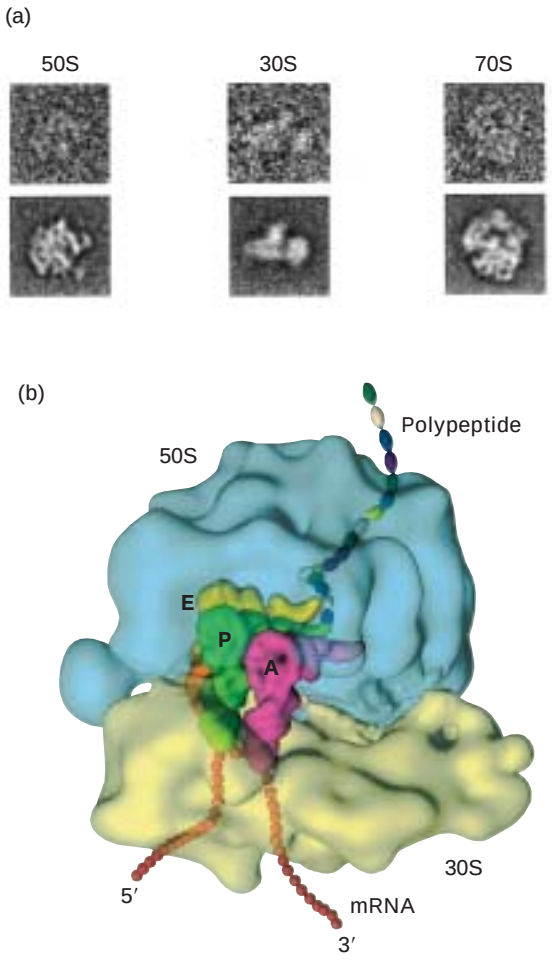
aminoacyl-tRNA is brought into the ribosome as a ternary complex in association with EF1α·GTP and becomes bound to the *A* site, so named because it is where *a*minoacylated tRNAs bind (step 1). If the anticodon of the incoming (second) aminoacyl-tRNA correctly base-pairs with the second codon of the mRNA, the GTP in the associated EF1α·GTP is hydrolyzed. The hydrolysis of GTP promotes a conformational change in the ribosome that leads to tight binding of the aminoacyl-tRNA in the A site and release of the resulting EF1α·GDP complex (step 2). This conformational change also positions the aminoacylated 3′ end of the tRNA in the A site in close proximity to the 3′ end of the Met-$tRNA_i^{Met}$ in the P site. GTP hydrolysis, and hence tight binding, does not occur if the anticodon of the incoming aminoacyl-tRNA cannot base-pair with the codon at the A site. In this case, the ternary complex diffuses away, leaving an empty A site that can associate with other aminoacyltRNA–EF1α·GTP complexes until a correctly base-paired tRNA is bound. This phenomenon contributes to the fidelity with which the correct aminoacyl-tRNA is loaded into the A site.

With the initiating Met-$tRNA_i^{Met}$ at the P site and the second aminoacyl-tRNA tightly bound at the A site, the α amino group of the second amino acid reacts with the "activated" (ester-linked) methionine on the initiator tRNA, forming a peptide bond (Figure 4-26, step 3; see Figures 4-19 and 4-21). This *peptidyltransferase reaction* is catalyzed by the large rRNA, which precisely orients the interacting atoms, permitting the reaction to proceed. The catalytic ability of the large rRNA in bacteria has been demonstrated by carefully removing the vast majority of the protein from large ribosomal subunits. The nearly pure bacterial 23S rRNA can catalyze a peptidyltransferase reaction between analogs of aminoacylated-tRNA and peptidyl-tRNA. Further support for the catalytic role of large rRNA in protein synthesis comes from crystallographic studies showing that no proteins lie near the site of peptide bond synthesis in the crystal structure of the bacterial large subunit.

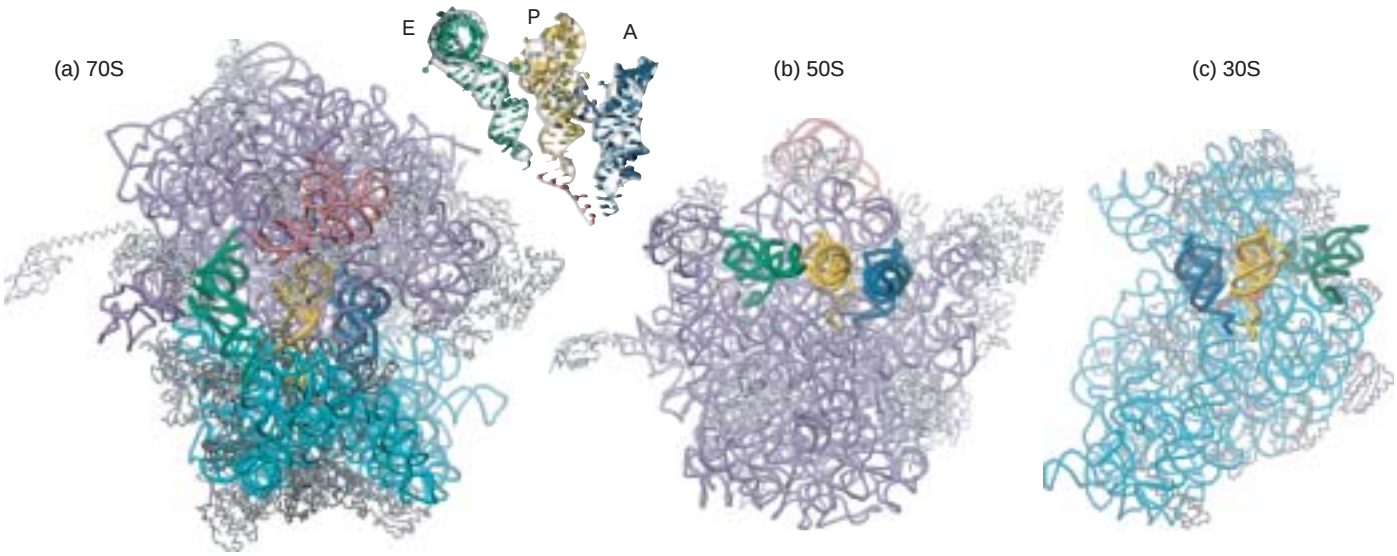
Following peptide bond synthesis, the ribosome is translocated along the mRNA a distance equal to one codon. This translocation step is promoted by hydrolysis of the GTP in eukaryotic EF2·GTP. As a result of translocation, $tRNA_i^{Met}$, now without its activated methionine, is moved to the *E* (*e*xit) site on the ribosome; concurrently, the second tRNA, now covalently bound to a dipeptide (a peptidyl-tRNA), is moved to the P site (Figure 4-26, step 4). Translocation thus returns the ribosome conformation to a state in which the A site is open and able to accept another aminoacylated tRNA complexed with EF1α·GTP, beginning another cycle of chain elongation.

Repetition of the elongation cycle depicted in Figure 4-26 adds amino acids one at a time to the C-terminus of the growing polypeptide as directed by the mRNA sequence until a stop codon is encountered. In subsequent cycles, the conformational change that occurs in step 2 ejects the unacylated tRNA from the E site. As the nascent polypeptide chain becomes longer, it threads through a channel in the large ribosomal subunit, exiting at a position opposite the side that interacts with the small subunit (Figure 4-27).

The locations of tRNAs bound at the A, P, and E sites are visible in the recently determined crystal structure of the bacterial ribosome (Figure 4-28). Base pairing is also apparent between the tRNAs in the A and P sites with their respective codons in mRNA (see Figure 4-28, *inset*). An RNA-RNA hybrid of only three base pairs is not stable under physio-



▲ FIGURE 4-27 **Low-resolution model of *E. coli* 70S ribosome.** (a) Top panels show cryoelectron microscopic images of *E. coli* 70S ribosomes and 50S and 30S subunits. Bottom panels show computer-derived averages of many dozens of images in the same orientation. (b) Model of a 70S ribosome based on the computer-derived images and on chemical cross-linking studies. Three tRNAs are superimposed on the A (pink), P (green), and E (yellow) sites. The nascent polypeptide chain is buried in a tunnel in the large ribosomal subunit that begins close to the acceptor stem of the tRNA in the P site. [See I. S. Gabashvili et al., 2000, *Cell* **100**:537; courtesy of J. Frank.]



▲ FIGURE 4-28 **Structure of *T. thermophilus* 70S ribosome as determined by x-ray crystallography.** (a) Model of the entire ribosome viewed from the side diagrammed in Figure 4-26 with large subunit on top and small subunit below. The tRNAs positioned at the A (blue), P (yellow), and E (green) sites are visible in the interface between the subunits with their anticodon loops pointing down into the small subunit. 16S rRNA is cyan; 23S rRNA, purple; 5S rRNA, pink; mRNA, red; small ribosomal proteins, dark gray; and large ribosomal proteins, light gray. Note that the ribosomal proteins are located primarily on the surface of the ribosome and the rRNAs on the inside. (b) View of the large subunit rotated 90° about the horizontal from the view in (a) showing the face that interacts with the small subunit. The tRNA anticodon loops point out of the page. In the intact ribosome, these extend into the small subunit where the anticodons of the tRNAs in the A and P sites base-pair with codons in the mRNA. (c) View of the face of the small subunit that interacts with the large subunit in (b). Here the tRNA anticodon loops point into the page. The TΨCG loops and acceptor stems extend out of the page and the 3′ CCA ends of the tRNAs in the A and P sites point downward. Note the close opposition of the acceptor stems of tRNAs in the A and P sites, which allows the amino group of the acylated tRNA in the A site to react with the carboxyl-terminal C of the peptidyl-tRNA in the P site (see Figure 4-19). In the intact ribosome, these are located at the peptidyltransferase active site of the large subunit. [Adapted from M. M. Yusupov et al., 2001, *Science* **292**:883.]

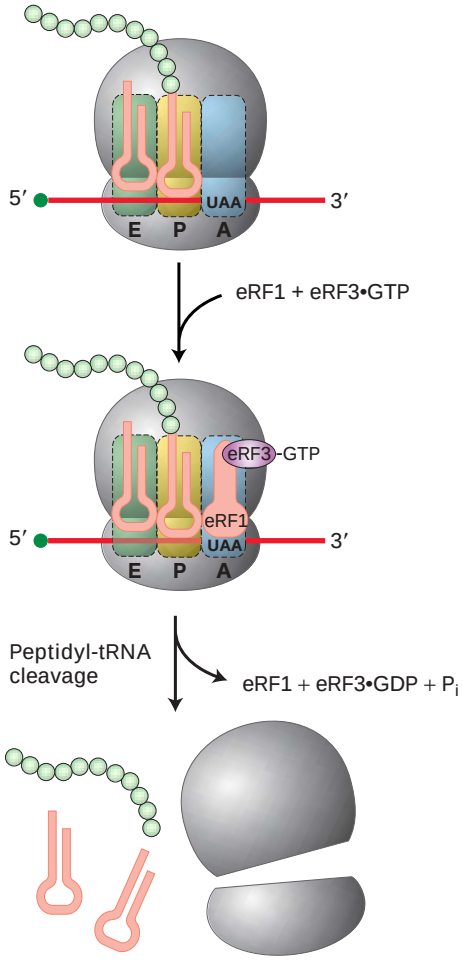
logical conditions. However, multiple interactions between the large and small rRNAs and general domains of tRNAs (e.g., the D and TΨCG loops) stabilize the tRNAs in the A and P sites, while other RNA-RNA interactions sense correct codon-anticodon base pairing, assuring that the genetic code is read properly.

Translation Is Terminated by Release Factors When a Stop Codon Is Reached

The final stage of translation, like initiation and elongation, requires highly specific molecular signals that decide the fate of the mRNA–ribosome–tRNA-peptidyl complex. Two types of specific protein **release factors** (RFs) have been discovered. Eukaryotic eRF1, whose shape is similar to that of tRNAs, apparently acts by binding to the ribosomal A site and recognizing stop codons directly. Like some of the initiation and elongation factors discussed previously, the second eukaryotic release factor, eRF3, is a GTP-binding protein. The eRF3·GTP acts in concert with eRF1 to promote cleavage of the peptidyl-tRNA, thus releasing the completed protein chain (Figure 4-29). Bacteria have two release factors (RF1 and RF2) that are functionally analogous to eRF1 and a GTP-binding factor (RF3) that is analogous to eRF3.

After its release from the ribosome, a newly synthesized protein folds into its native three-dimensional conformation, a process facilitated by other proteins called **chaperones** (Chapter 3). Additional release factors then promote dissociation of the ribosome, freeing the subunits, mRNA, and terminal tRNA for another round of translation.

We can now see that one or more GTP-binding proteins participate in each stage of translation. These proteins belong to the **GTPase superfamily** of switch proteins that cycle between a GTP-bound active form and GDP-bound inactive form (see Figure 3-29). Hydrolysis of the bound GTP is thought to cause conformational changes in the GTPase itself or other associated proteins that are critical to various complex molecular processes. In translation initiation, for instance, hydrolysis of eIF2·GTP to eIF2·GDP prevents further scanning of the mRNA once the start site is encountered and allows binding of the large ribosomal subunit to the small subunit (see Figure 4-25, step 3). Similarly, hydrolysis of



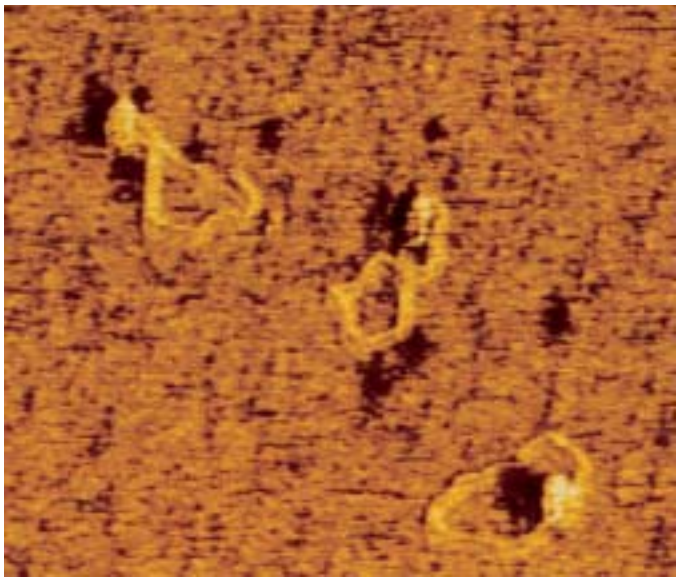
▲ FIGURE 4-29 **Termination of translation in eukaryotes.** When a ribosome bearing a nascent protein chain reaches a stop codon (UAA, UGA, UAG), release factor eRF1 enters the ribosomal complex, probably at or near the A site together with eRF3·GTP. Hydrolysis of the bound GTP is accompanied by cleavage of the peptide chain from the tRNA in the P site and release of the tRNAs and the two ribosomal subunits.

EF2·GTP to EF2·GDP during chain elongation leads to translocation of the ribosome along the mRNA (see Figure 4-26, step 4).

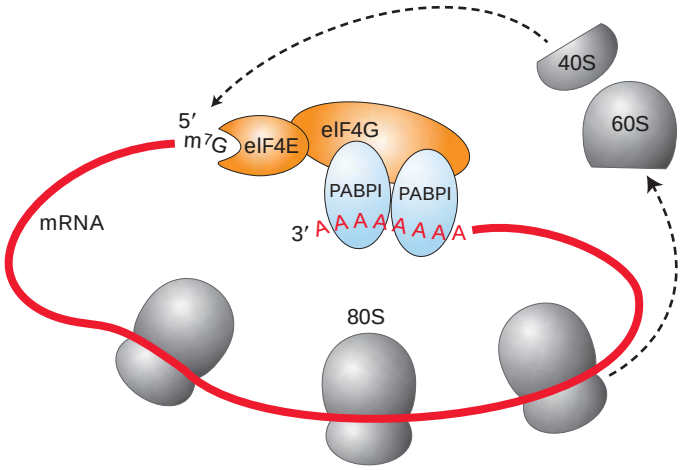
Polysomes and Rapid Ribosome Recycling Increase the Efficiency of Translation

As noted earlier, translation of a single eukaryotic mRNA molecule to yield a typical-sized protein takes 30–60 seconds. Two phenomena significantly increase the overall rate at which cells can synthesize a protein: the simultaneous translation of a single mRNA molecule by multiple ribosomes and rapid recycling of ribosomal subunits after they disengage from the 3′ end of an mRNA. Simultaneous translation of an mRNA by multiple ribosomes is readily observable in electron micrographs and by sedimentation analysis, revealing mRNA attached to multiple ribosomes bearing nascent growing polypeptide chains. These structures, referred to as **polyribosomes** or *polysomes*, were seen to be circular in electron micrographs of some tissues. Subsequent studies with yeast cells explained the circular shape of polyribosomes and suggested the mode by which ribosomes recycle efficiently.

These studies revealed that multiple copies of a cytosolic protein found in all eukaryotic cells, *poly(A)-binding protein* (PABPI), can interact with both an mRNA poly(A) tail and the 4G subunit of yeast eIF4. Moreover, the 4E subunit of yeast eIF4 binds to the 5′ end of an mRNA. As a result of these interactions, the two ends of an mRNA molecule can be bridged by the intervening proteins, forming a "circular" mRNA (Figure 4-30). Because the two ends of a polysome are relatively close together, ribosomal subunits that disengage from the 3′ end are positioned near the 5′ end, facilitating re-initiation by the interaction of the 40S subunit with eIF4 bound to the 5′ cap. The circular pathway depicted in Figure 4-31, which may operate in many eukaryotic cells, would enhance ribosome recycling and thus increase the efficiency of protein synthesis.



▲ EXPERIMENTAL FIGURE 4-30 **Eukaryotic mRNA forms a circular structure owing to interactions of three proteins.** In the presence of purified poly(A)-binding protein I (PABPI), eIF4E, and eIF4G, eukaryotic mRNAs form circular structures, visible in this force-field electron micrograph. In these structures, protein-protein and protein-mRNA interactions form a bridge between the 5′ and 3′ ends of the mRNA as diagrammed in Figure 4-31. [Courtesy of A. Sachs.]



◀ FIGURE 4-31 **Model of protein synthesis on circular polysomes and recycling of ribosomal subunits.** Multiple individual ribosomes can simultaneously translate a eukaryotic mRNA, shown here in circular form stabilized by interactions between proteins bound at the 3′ and 5′ ends. When a ribosome completes translation and dissociates from the 3′ end, the separated subunits can rapidly find the nearby 5′ cap (m^7G) and initiate another round of synthesis.

MEDIA CONNECTIONS
Overview Animation: Life Cycle of an mRNA

KEY CONCEPTS OF SECTION 4.5

Stepwise Synthesis of Proteins on Ribosomes

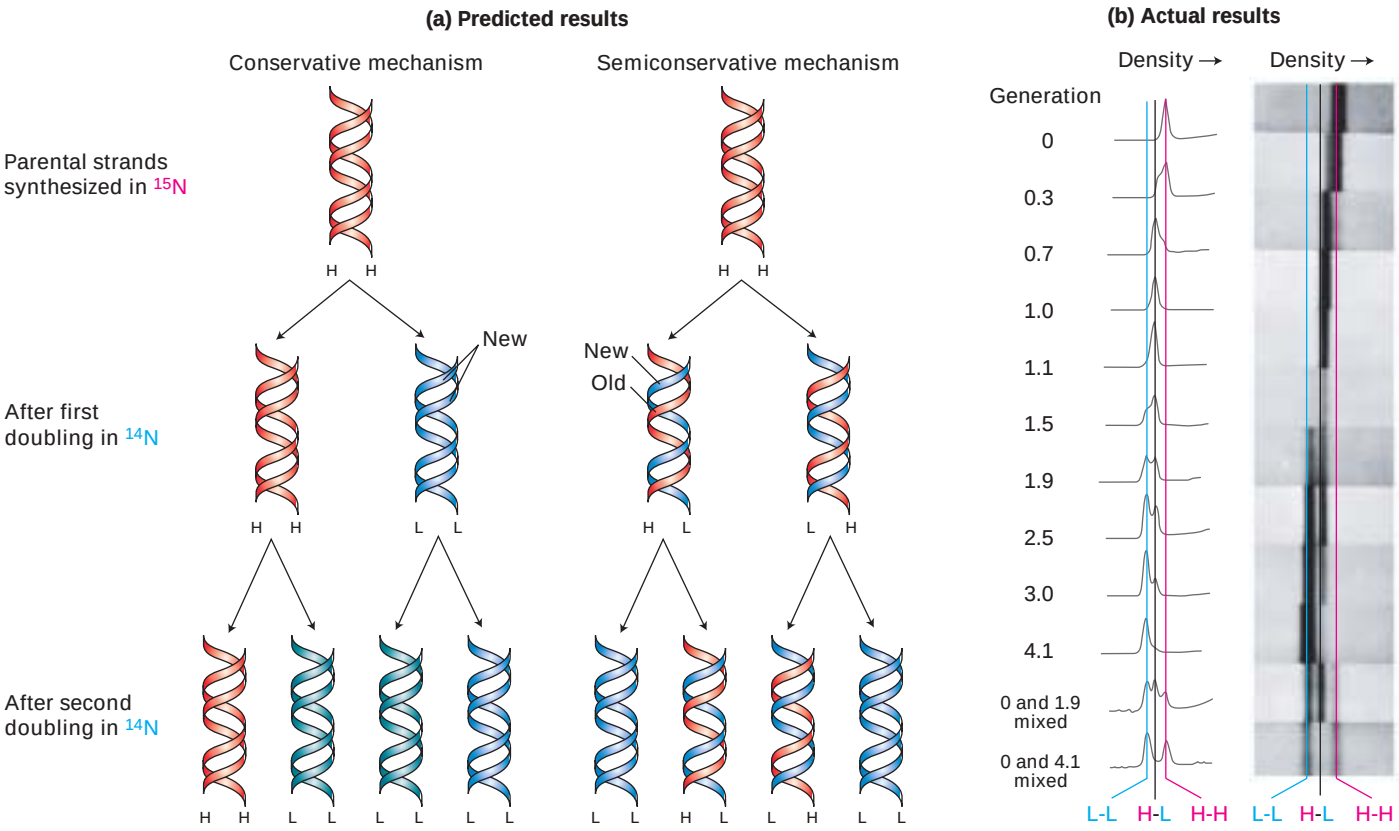
- Of the two methionine tRNAs found in all cells, only one ($tRNA_i^{Met}$) functions in initiation of translation.
- Each stage of translation—initiation, chain elongation, and termination—requires specific protein factors including GTP-binding proteins that hydrolyze their bound GTP to GDP when a step has been completed successfully.
- During initiation, the ribosomal subunits assemble near the translation start site in an mRNA molecule with the tRNA carrying the amino-terminal methionine (Met-$tRNA_i^{Met}$) base-paired with the start codon (Figure 4-25).
- Chain elongation entails a repetitive four-step cycle: loose binding of an incoming aminoacyl-tRNA to the A site on the ribosome; tight binding of the correct aminoacyl-tRNA to the A site accompanied by release of the previously used tRNA from the E site; transfer of the growing peptidyl chain to the incoming amino acid catalyzed by large rRNA; and translocation of the ribosome to the next codon, thereby moving the peptidyl-tRNA in the A site to the P site and the now unacylated tRNA in the P site to the E site (see Figure 4-26).
- In each cycle of chain elongation, the ribosome undergoes two conformational changes monitored by GTP-binding proteins. The first permits tight binding of the incoming aminoacyl-tRNA to the A site and ejection of a tRNA from the E site, and the second leads to translocation.
- Termination of translation is carried out by two types of termination factors: those that recognize stop codons and those that promote hydrolysis of peptidyl-tRNA (see Figure 4-29).
- The efficiency of protein synthesis is increased by the simultaneous translation of a single mRNA by multiple ribosomes. In eukaryotic cells, protein-mediated interactions bring the two ends of a polyribosome close together, thereby promoting the rapid recycling of ribosomal subunits, which further increases the efficiency of protein synthesis (see Figure 4-31).

4.6 DNA Replication

Now that we have seen how genetic information encoded in the nucleotide sequences of DNA is translated into the structures of proteins that perform most cell functions, we can appreciate the necessity of the precise copying of DNA sequences during DNA replication (see Figure 4-1, step 4). The regular pairing of bases in the double-helical DNA structure suggested to Watson and Crick that new DNA strands are synthesized by using the existing (*parental*) strands as **templates** in the formation of new, *daughter* strands complementary to the parental strands.

This base-pairing template model theoretically could proceed either by a *conservative* or a *semiconservative* mechanism. In a conservative mechanism, the two daughter strands would form a new double-stranded (*duplex*) DNA molecule and the parental duplex would remain intact. In a semiconservative mechanism, the parental strands are permanently separated and each forms a duplex molecule with the daughter strand base-paired to it. Definitive evidence that duplex DNA is replicated by a semiconservative mechanism came from a now classic experiment conducted by M. Meselson and W. F. Stahl, outlined in Figure 4-32.

Copying of a DNA template strand into a complementary strand thus is a common feature of DNA replication and transcription of DNA into RNA. In both cases, the information in the template is preserved. In some viruses, single-stranded RNA molecules function as templates for synthesis of complementary RNA or DNA strands. However, the vast preponderance of RNA and DNA in cells is synthesized from preexisting duplex DNA.

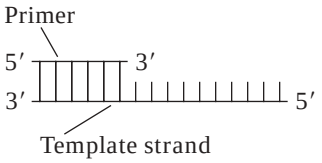


▲ EXPERIMENTAL FIGURE 4-32 **The Meselson-Stahl experiment showed that DNA replicates by a semiconservative mechanism.** In this experiment, *E. coli* cells initially were grown in a medium containing ammonium salts prepared with "heavy" nitrogen (^{15}N) until all the cellular DNA was labeled. After the cells were transferred to a medium containing the normal "light" isotope (^{14}N), samples were removed periodically from the cultures and the DNA in each sample was analyzed by equilibrium density-gradient centrifugation (see Figure 5-37). This technique can separate heavy-heavy (H-H), light-light (L-L), and heavy-light (H-L) duplexes into distinct bands. (a) Expected composition of daughter duplex molecules synthesized from ^{15}N-labeled DNA after *E. coli* cells are shifted to ^{14}N-containing medium if DNA replication occurs by a conservative or semiconservative mechanism. Parental heavy (H) strands are in red; light (L) strands synthesized after shift to ^{14}N-containing medium are in blue. Note that the conservative mechanism never generates H-L DNA and that the semiconservative mechanism never generates H-H DNA but does generate H-L DNA during the second and subsequent doublings. With additional replication cycles, the ^{15}N-labeled (H) strands from the original DNA are diluted, so that the vast bulk of the DNA would consist of L-L duplexes with either mechanism. (b) Actual banding patterns of DNA subjected to equilibrium density-gradient centrifugation before and after shifting ^{15}N-labeled *E. coli* cells to ^{14}N-containing medium. DNA bands were visualized under UV light and photographed. The traces on the left are a measure of the density of the photographic signal, and hence the DNA concentration, along the length of the centrifuge cells from left to right. The number of generations (far left) following the shift to ^{14}N-containing medium was determined by counting the concentration of *E. coli* cells in the culture. This value corresponds to the number of DNA replication cycles that had occurred at the time each sample was taken. After one generation of growth, all the extracted DNA had the density of H-L DNA. After 1.9 generations, approximately half the DNA had the density of H-L DNA; the other half had the density of L-L DNA. With additional generations, a larger and larger fraction of the extracted DNA consisted of L-L duplexes; H-H duplexes never appeared. These results match the predicted pattern for the semiconservative replication mechanism depicted in (a). The bottom two centrifuge cells contained mixtures of H-H DNA and DNA isolated at 1.9 and 4.1 generations in order to clearly show the positions of H-H, H-L, and L-L DNA in the density gradient. [Part (b) from M. Meselson and F. W. Stahl, 1958, *Proc. Nat'l. Acad. Sci. USA* **44**:671.]

DNA Polymerases Require a Primer to Initiate Replication

Analogous to RNA, DNA is synthesized from deoxynucleoside 5′-triphosphate precursors (dNTPs). Also like RNA synthesis, DNA synthesis always proceeds in the 5′→3′ direction because chain growth results from formation of a phosphoester bond between the 3′ oxygen of a growing strand and the α phosphate of a dNTP (see Figure 4-9). As discussed earlier, an RNA polymerase can find an appropriate transcription start site on duplex DNA and initiate the

synthesis of an RNA complementary to the template DNA strand (see Figure 4-10). In contrast, **DNA polymerases** cannot initiate chain synthesis de novo; instead, they require a short, preexisting RNA or DNA strand, called a **primer,** to begin chain growth. With a primer base-paired to the template strand, a DNA polymerase adds deoxynucleotides to the free hydroxyl group at the 3′ end of the primer as directed by the sequence of the template strand:



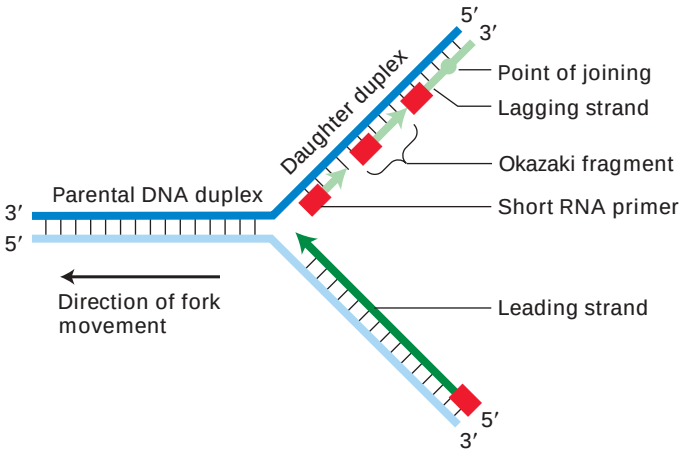
When RNA is the primer, the daughter strand that is formed is RNA at the 5′ end and DNA at the 3′ end.

Duplex DNA Is Unwound, and Daughter Strands Are Formed at the DNA Replication Fork

In order for duplex DNA to function as a template during replication, the two intertwined strands must be unwound, or melted, to make the bases available for base pairing with the bases of the dNTPs that are polymerized into the newly synthesized daughter strands. This unwinding of the parental DNA strands is by specific **helicases,** beginning at unique segments in a DNA molecule called *replication origins,* or simply *origins*. The nucleotide sequences of origins from different organisms vary greatly, although they usually contain A·T-rich sequences. Once helicases have unwound the parental DNA at an origin, a specialized RNA polymerase called **primase** forms a short RNA primer complementary to the unwound template strands. The primer, still base-paired to its complementary DNA strand, is then elongated by a DNA polymerase, thereby forming a new daughter strand.

The DNA region at which all these proteins come together to carry out synthesis of daughter strands is called the **replication fork,** or growing fork. As replication proceeds, the growing fork and associated proteins move away from the origin. As noted earlier, local unwinding of duplex DNA produces torsional stress, which is relieved by topoisomerase I. In order for DNA polymerases to move along and copy a duplex DNA, helicase must sequentially unwind the duplex and topoisomerase must remove the supercoils that form.

A major complication in the operation of a DNA replication fork arises from two properties: the two strands of the parental DNA duplex are antiparallel, and DNA polymerases (like RNA polymerases) can add nucleotides to the growing new strands only in the 5′→3′ direction. Synthesis of one daughter strand, called the **leading strand,** can proceed continuously from a single RNA primer in the 5′→3′ direction, *the same direction as movement of the replication fork* (Figure 4-33). The problem comes in synthesis of the other daughter strand, called the **lagging strand.**

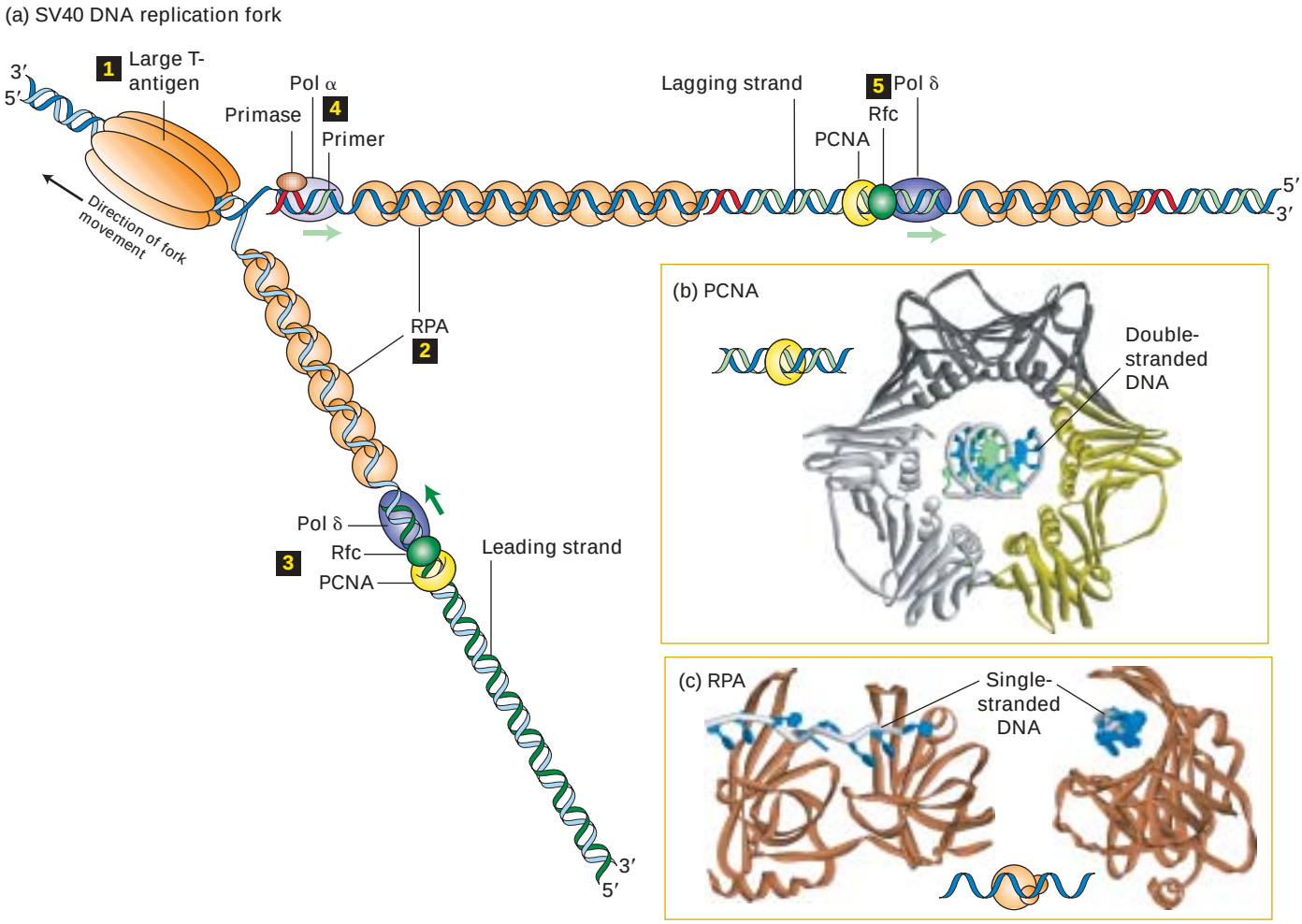


▲ FIGURE 4-33 **Schematic diagram of leading-strand and lagging-strand DNA synthesis at a replication fork.** Nucleotides are added by a DNA polymerase to each growing daughter strand in the 5′→3′ direction (indicated by arrowheads). The leading strand is synthesized continuously from a single RNA primer (red) at its 5′ end. The lagging strand is synthesized discontinuously from multiple RNA primers that are formed periodically as each new region of the parental duplex is unwound. Elongation of these primers initially produces Okazaki fragments. As each growing fragment approaches the previous primer, the primer is removed and the fragments are ligated. Repetition of this process eventually results in synthesis of the entire lagging strand.

Because growth of the lagging strand must occur in the 5′→3′ direction, copying of its template strand must somehow occur in the *opposite* direction from the movement of the replication fork. A cell accomplishes this feat by synthesizing a new primer every few hundred bases or so on the second parental strand, as more of the strand is exposed by unwinding. Each of these primers, base-paired to their template strand, is elongated in the 5′→3′ direction, forming discontinuous segments called **Okazaki fragments** after their discoverer Reiji Okazaki (see Figure 4-33). The RNA primer of each Okazaki fragment is removed and replaced by DNA chain growth from the neighboring Okazaki fragment; finally an enzyme called *DNA ligase* joins the adjacent fragments.

Helicase, Primase, DNA Polymerases, and Other Proteins Participate in DNA Replication

Detailed understanding of the eukaryotic proteins that participate in DNA replication has come largely from studies with small viral DNAs, particularly SV40 DNA, the circular genome of a small virus that infects monkeys. Figure 4-34 depicts the multiple proteins that coordinate copying of SV40 DNA at a replication fork. The assembled proteins at a replication fork further illustrate the concept of molecular machines introduced in Chapter 3. These multicomponent



▲ FIGURE 4-34 **Model of an SV40 DNA replication fork and assembled proteins.** (a) A hexamer of large T-antigen (1), a viral protein, functions as a helicase to unwind the parental DNA strands. Single-strand regions of the parental template unwound by large T-antigen are bound by multiple copies of the heterotrimeric protein RPA (2). The leading strand is synthesized by a complex of DNA polymerase δ (Pol δ), PCNA, and Rfc (3). Primers for lagging-strand synthesis (red, RNA; light blue, DNA) are synthesized by a complex of DNA polymerase α (Pol α) and primase (4). The 3′ end of each primer synthesized by Pol α–primase is then bound by a PCNA-Rfc–Pol δ complex, which proceeds to extend the primer and synthesize most of each Okazaki fragment (5). See the text for details. (b) The three subunits of PCNA, shown in different colors, form a circular structure with a central hole through which double-stranded DNA passes. A diagram of DNA is shown in the center of a ribbon model of the PCNA trimer. (c) The large subunit of RPA contains two domains that bind single-stranded DNA. On the left, the two DNA-binding domains of RPA are shown perpendicular to the DNA backbone (white backbone with blue bases). Note that the single DNA strand is extended with the bases exposed, an optimal conformation for replication by a DNA polymerase. On the right, the view is down the length of the single DNA strand, revealing how RPA β strands wrap around the DNA. [Part (a) adapted from S. J. Flint et al., 2000, *Virology: Molecular Biology, Pathogenesis, and Control,* ASM Press; part (b) after J. M. Gulbis et al., 1996, *Cell* **87**:297; and part (c) after A. Bochkarev et al., 1997, *Nature* **385**:176.]

complexes permit the cell to carry out an ordered sequence of events that accomplish essential cell functions.

In the molecular machine that replicates SV40 DNA, a hexamer of a viral protein called *large T-antigen* unwinds the parental strands at a replication fork. All other proteins involved in SV40 DNA replication are provided by the host cell. Primers for leading and lagging daughter-strand DNA are synthesized by a complex of primase, which synthesizes a short RNA primer, and *DNA polymerase* α (Pol α), which extends the RNA primer with deoxynucleotides, forming a mixed RNA-DNA primer.

The primer is extended into daughter-strand DNA by *DNA polymerase* δ (Pol δ), which is less likely to make errors during copying of the template strand than is Pol α. Pol δ forms a complex with *Rfc* (*r*eplication *f*actor *C*) and *PCNA* (*p*roliferating *c*ell *n*uclear *a*ntigen), which displaces

the primase–Pol α complex following primer synthesis. As illustrated in Figure 4-34b, PCNA is a homotrimeric protein that has a central hole through which the daughter duplex DNA passes, thereby preventing the PCNA-Rfc–Pol δ complex from dissociating from the template.

After parental DNA is separated into single-stranded templates at the replication fork, it is bound by multiple copies of RPA (*r*eplication *p*rotein *A*), a heterotrimeric protein (Figure 4-34c). Binding of RPA maintains the template in a uniform conformation optimal for copying by DNA polymerases. Bound RPA proteins are dislodged from the parental strands by Pol α and Pol δ as they synthesize the complementary strands base-paired with the parental strands.

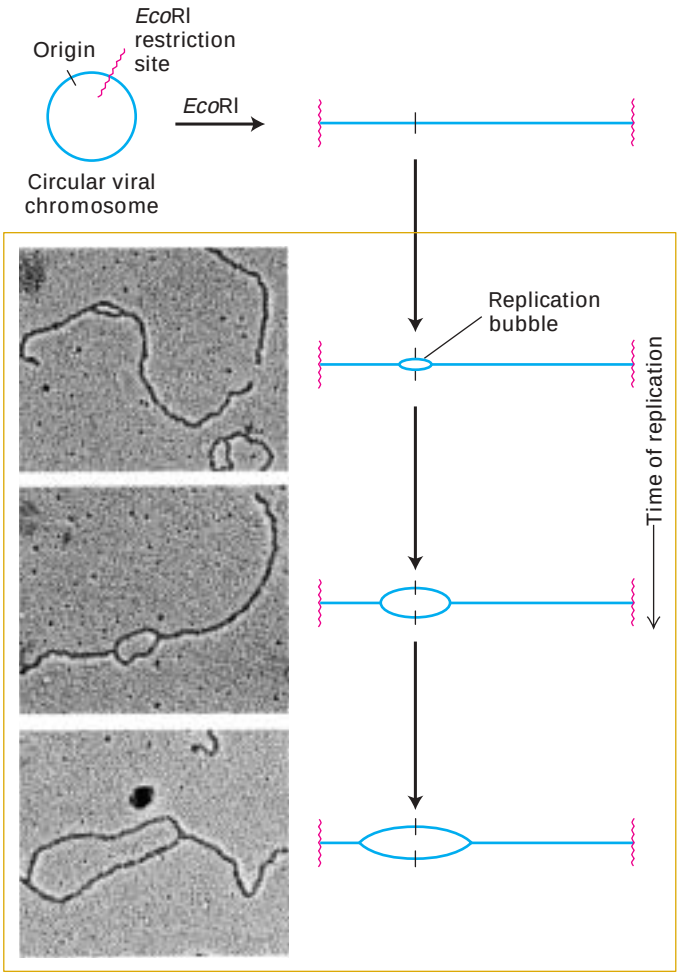
Several eukaryotic proteins that function in DNA replication are not depicted in Figure 4-34. A topoisomerase associates with the parental DNA ahead of the helicase to remove torsional stress introduced by the unwinding of the parental strands. Ribonuclease H and FEN I remove the ribonucleotides at the 5′ ends of Okazaki fragments; these are replaced by deoxynucleotides added by DNA polymerase δ as it extends the upstream Okazaki fragment. Successive Okazaki fragments are coupled by DNA ligase through standard 5′→3′ phosphoester bonds.

DNA Replication Generally Occurs Bidirectionally from Each Origin

As indicated in Figures 4-33 and 4-34, both parental DNA strands that are exposed by local unwinding at a replication fork are copied into a daughter strand. In theory, DNA replication from a single origin could involve one replication fork that moves in one direction. Alternatively, two replication forks might assemble at a single origin and then move in opposite directions, leading to *bidirectional growth* of both daughter strands. Several types of experiments, including the one shown in Figure 4-35, provided early evidence in support of bidirectional strand growth.

The general consensus is that all prokaryotic and eukaryotic cells employ a bidirectional mechanism of DNA replication. In the case of SV40 DNA, replication is initiated by binding of two large T-antigen hexameric helicases to the single SV40 origin and assembly of other proteins to form two replication forks. These then move away from the SV40 origin in opposite directions with leading- and lagging-strand synthesis occurring at both forks. As shown in Figure 4-36, the left replication fork extends DNA synthesis in the leftward direction; similarly, the right replication fork extends DNA synthesis in the rightward direction.

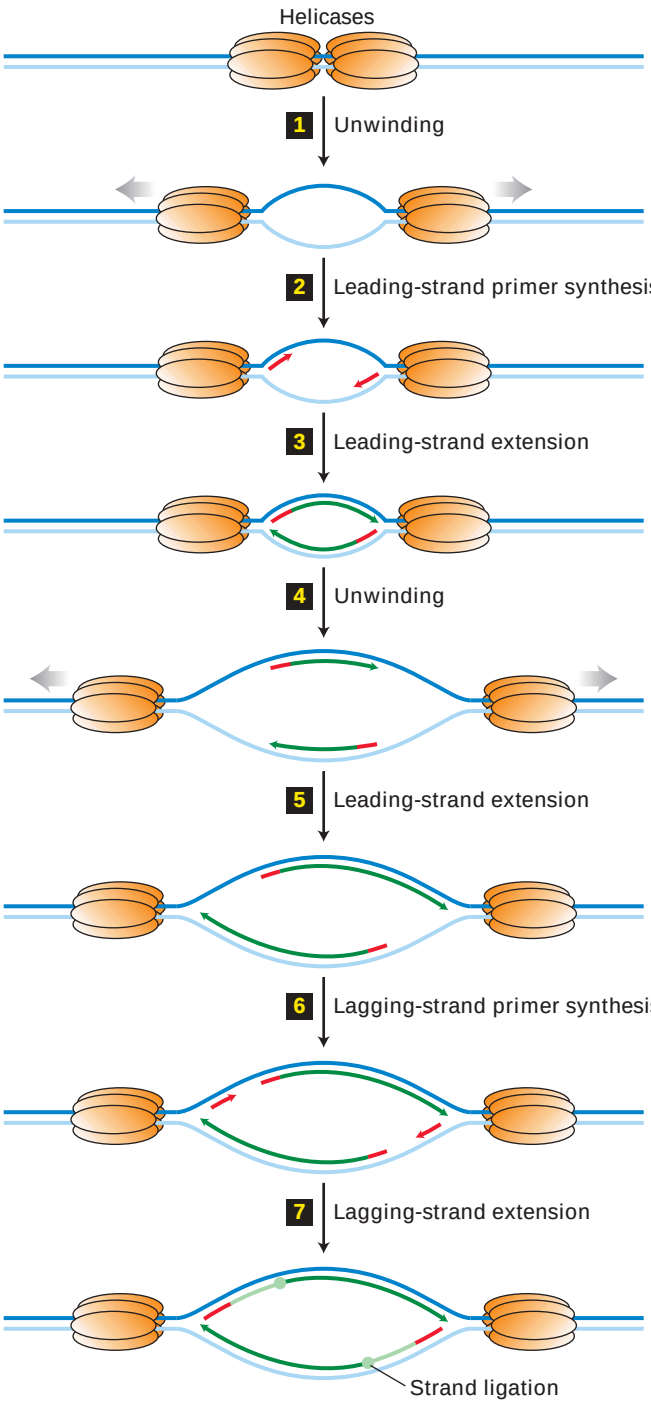
Unlike SV40 DNA, eukaryotic chromosomal DNA molecules contain multiple replication origins separated by tens to hundreds of kilobases. A six-subunit protein called *ORC*, for *o*rigin *r*ecognition *c*omplex, binds to each origin and associates with other proteins required to load cellular hexameric helicases composed of six homologous *MCM* proteins.



▲ EXPERIMENTAL FIGURE 4-35 **Electron microscopy of replicating SV40 DNA indicates bidirectional growth of DNA strands from an origin.** The replicating viral DNA from SV40-infected cells was cut by the restriction enzyme *Eco*RI, which recognizes one site in the circular DNA. Electron micrographs of treated samples showed a collection of cut molecules with increasingly longer replication "bubbles," whose centers are a constant distance from each end of the cut molecules. This finding is consistent with chain growth in two directions from a common origin located at the center of a bubble, as illustrated in the corresponding diagrams. [See G. C. Fareed et al., 1972, *J. Virol.* **10**:484; photographs courtesy of N. P. Salzman.]

Two opposed MCM helicases separate the parental strands at an origin, with RPA proteins binding to the resulting single-stranded DNA. Synthesis of primers and subsequent steps in replication of cellular DNA are thought to be analogous to those in SV40 DNA replication (see Figures 4-34 and 4-36).

Replication of cellular DNA and other events leading to proliferation of cells are tightly regulated, so that the appropriate numbers of cells constituting each tissue are produced during development and throughout the life of an organism. As in transcription of most genes, control of the initiation



◀ FIGURE 4-36 **Bidirectional mechanism of DNA replication.** The left replication fork here is comparable to the replication fork diagrammed in Figure 4-34, which also shows proteins other than large T-antigen. (*Top*) Two large T-antigen hexameric helicases first bind at the replication origin in opposite orientations. Step 1: Using energy provided from ATP hydrolysis, the helicases move in opposite directions, unwinding the parental DNA and generating single-strand templates that are bound by RPA proteins. Step 2: Primase–Pol α complexes synthesize short primers base-paired to each of the separated parental strands. Step 3: PCNA-Rfc–Pol δ complexes replace the primase–Pol α complexes and extend the short primers, generating the leading strands (dark green) at each replication fork. Step 4: The helicases further unwind the parental strands, and RPA proteins bind to the newly exposed single-strand regions. Step 5: PCNA-Rfc–Pol δ complexes extend the leading strands further. Step 6: Primase–Pol α complexes synthesize primers for lagging-strand synthesis at each replication fork. Step 7: PCNA-Rfc–Pol δ complexes displace the primase–Pol α complexes and extend the lagging-strand Okazaki fragments (light green), which eventually are ligated to the 5′ ends of the leading strands. The position where ligation occurs is represented by a circle. Replication continues by further unwinding of the parental strands and synthesis of leading and lagging strands as in steps 4–7. Although depicted as individual steps for clarity, unwinding and synthesis of leading and lagging strands occur concurrently.

step is the primary mechanism for regulating cellular DNA replication. Activation of MCM helicase activity, which is required to initiate cellular DNA replication, is regulated by specific protein kinases called S-phase **cyclin-dependent kinases.** Other cyclin-dependent kinases regulate additional aspects of cell proliferation, including the complex process of mitosis by which a eukaryotic cell divides into two daughter cells. We discuss the various regulatory mechanisms that determine the rate of cell division in Chapter 21.

KEY CONCEPTS OF SECTION 4.6

DNA Replication

- Each strand in a parental duplex DNA acts as a template for synthesis of a daughter strand and remains base-paired to the new strand, forming a daughter duplex (semiconservative mechanism). New strands are formed in the 5′→3′ direction.
- Replication begins at a sequence called an *origin*. Each eukaryotic chromosomal DNA molecule contains multiple replication origins.
- DNA polymerases, unlike RNA polymerases, cannot unwind the strands of duplex DNA and cannot initiate synthesis of new strands complementary to the template strands.
- At a replication fork, one daughter strand (the leading strand) is elongated continuously. The other daughter strand (the lagging strand) is formed as a series of discontinuous Okazaki fragments from primers synthesized every few hundred nucleotides (Figure 4-33).
- The ribonucleotides at the 5′ end of each Okazaki fragment are removed and replaced by elongation of the 3′ end of the next Okazaki fragment. Finally, adjacent Okazaki fragments are joined by DNA ligase.
- Helicases use energy from ATP hydrolysis to separate the parental (template) DNA strands. Primase synthesizes

a short RNA primer, which remains base-paired to the template DNA. This initially is extended at the 3′ end by DNA polymerase α (Pol α), resulting in a short (5′)RNA-(3′)DNA daughter strand.

- Most of the DNA in eukaryotic cells is synthesized by Pol δ, which takes over from Pol α and continues elongation of the daughter strand in the 5′→3′ direction. Pol δ remains stably associated with the template by binding to Rfc protein, which in turn binds to PCNA, a trimeric protein that encircles the daughter duplex DNA (see Figure 4-34).
- DNA replication generally occurs by a bidirectional mechanism in which two replication forks form at an origin and move in opposite directions, with both template strands being copied at each fork (see Figure 4-36).
- Synthesis of eukaryotic DNA in vivo is regulated by controlling the activity of the MCM helicases that initiate DNA replication at multiple origins spaced along chromosomal DNA.

4.7 Viruses: Parasites of the Cellular Genetic System

Viruses cannot reproduce by themselves and must commandeer a host cell's machinery to synthesize viral proteins and in some cases to replicate the viral genome. RNA viruses, which usually replicate in the host-cell cytoplasm, have an RNA genome, and DNA viruses, which commonly replicate in the host-cell nucleus, have a DNA genome (see Figure 4-1). Viral genomes may be single- or double-stranded, depending on the specific type of virus. The entire infectious virus particle, called a **virion,** consists of the nucleic acid and an outer shell of protein. The simplest viruses contain only enough RNA or DNA to encode four proteins; the most complex can encode 100–200 proteins. In addition to their obvious importance as causes of disease, viruses are extremely useful as research tools in the study of basic biological processes.

Most Viral Host Ranges Are Narrow

The surface of a virion contains many copies of one type of protein that binds specifically to multiple copies of a receptor protein on a host cell. This interaction determines the *host range*—the group of cell types that a virus can infect—and begins the infection process. Most viruses have a rather limited host range.

A virus that infects only bacteria is called a **bacteriophage,** or simply a *phage*. Viruses that infect animal or plant cells are referred to generally as animal viruses or plant viruses. A few viruses can grow in both plants and the insects that feed on them. The highly mobile insects serve as vectors for transferring such viruses between susceptible plant hosts. Wide host ranges are also characteristic of some strictly animal viruses, such as vesicular stomatitis virus, which grows in insect vectors and in many different types of mammals. Most animal viruses, however, do not cross phyla, and some (e.g., poliovirus) infect only closely related species such as primates. The host-cell range of some animal viruses is further restricted to a limited number of cell types because only these cells have appropriate surface receptors to which the virions can attach.

Viral Capsids Are Regular Arrays of One or a Few Types of Protein

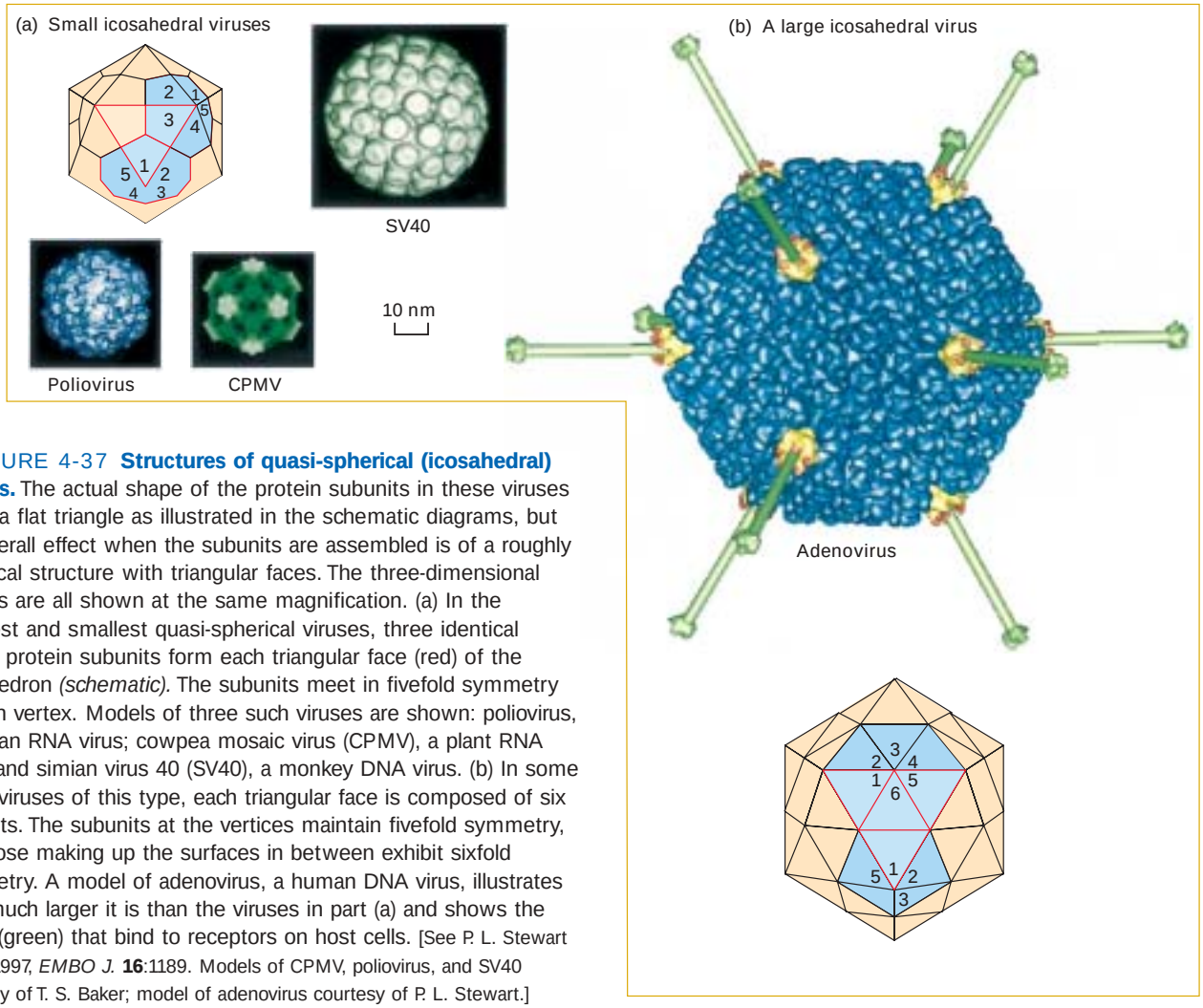
The nucleic acid of a virion is enclosed within a protein coat, or **capsid,** composed of multiple copies of one protein or a few different proteins, each of which is encoded by a single viral gene. Because of this structure, a virus is able to encode all the information for making a relatively large capsid in a small number of genes. This efficient use of genetic information is important, since only a limited amount of RNA or DNA, and therefore a limited number of genes, can fit into a virion capsid. A capsid plus the enclosed nucleic acid is called a **nucleocapsid.**

Nature has found two basic ways of arranging the multiple capsid protein subunits and the viral genome into a nucleocapsid. In some viruses, multiple copies of a single coat protein form a *helical* structure that encloses and protects the viral RNA or DNA, which runs in a helical groove within the protein tube. Viruses with such a helical nucleocapsid, such as tobacco mosaic virus, have a rodlike shape. The other major structural type is based on the *icosahedron*, a solid, approximately spherical object built of 20 identical faces, each of which is an equilateral triangle.

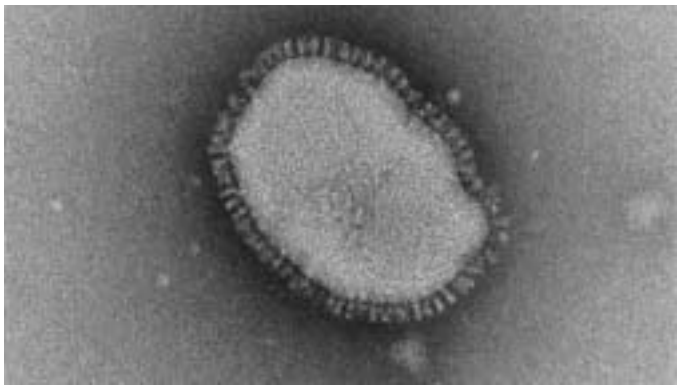
The number and arrangement of coat proteins in icosahedral, or quasi-spherical, viruses differ somewhat depending on their size. In small viruses of this type, each of the 20 triangular faces is constructed of three identical capsid protein subunits, making a total of 60 subunits per capsid. All the protein subunits are in *equivalent* contact with one another (Figure 4-37a). In large quasi-spherical viruses, each face of the icosahedron is composed of more than three subunits. As a result, the contacts between subunits not at the vertices are *quasi-equivalent* (Figure 4-37b). Models of several quasi-spherical viruses, based on cryoelectron microscopy, are shown in Figure 4-37. In the smaller viruses (e.g., poliovirus), clefts that encircle each of the vertices of the icosahedral structure interact with receptors on the surface of host cells during infection. In the larger viruses (e.g., adenovirus), long fiberlike proteins extending from the nucleocapsid interact with cell-surface receptors on host cells.

In many DNA bacteriophages, the viral DNA is located within an icosahedral "head" that is attached to a rodlike "tail." During infection, viral proteins at the tip of the tail bind to host-cell receptors, and then the viral DNA passes down the tail into the cytoplasm of the host cell.

In some viruses, the symmetrically arranged nucleocapsid is covered by an external membrane, or **envelope,** which



▲ FIGURE 4-37 **Structures of quasi-spherical (icosahedral) viruses.** The actual shape of the protein subunits in these viruses is not a flat triangle as illustrated in the schematic diagrams, but the overall effect when the subunits are assembled is of a roughly spherical structure with triangular faces. The three-dimensional models are all shown at the same magnification. (a) In the simplest and smallest quasi-spherical viruses, three identical capsid protein subunits form each triangular face (red) of the icosahedron *(schematic).* The subunits meet in fivefold symmetry at each vertex. Models of three such viruses are shown: poliovirus, a human RNA virus; cowpea mosaic virus (CPMV), a plant RNA virus; and simian virus 40 (SV40), a monkey DNA virus. (b) In some larger viruses of this type, each triangular face is composed of six subunits. The subunits at the vertices maintain fivefold symmetry, but those making up the surfaces in between exhibit sixfold symmetry. A model of adenovirus, a human DNA virus, illustrates how much larger it is than the viruses in part (a) and shows the fibers (green) that bind to receptors on host cells. [See P. L. Stewart et al., 1997, *EMBO J.* **16**:1189. Models of CPMV, poliovirus, and SV40 courtesy of T. S. Baker; model of adenovirus courtesy of P. L. Stewart.]

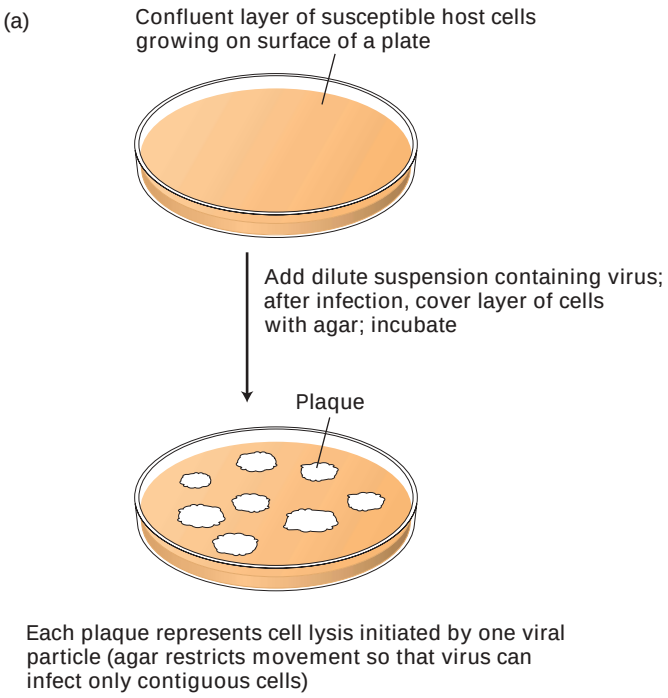


▲ EXPERIMENTAL FIGURE 4-38 **Viral protein spikes protrude from the surface of an influenza virus virion.** Influenza viruses are surrounded by an envelope consisting of a phospholipid bilayer and embedded viral proteins. The large spikes seen in this electron micrograph of a negatively stained influenza virion are composed of neuraminidase, a tetrameric protein, or hemagglutinin, a trimeric protein (see Figure 3-7). Inside is the helical nucleocapsid. [Courtesy of A. Helenius and J. White.]

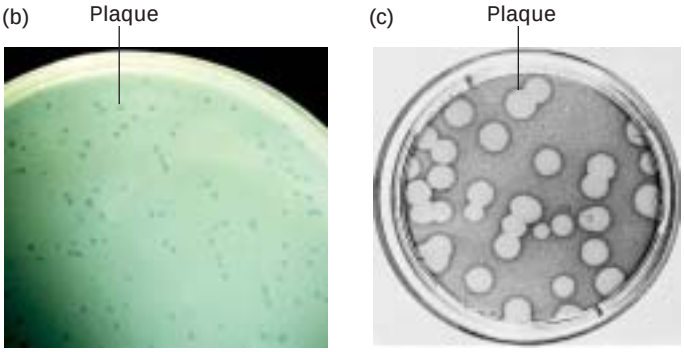
consists mainly of a phospholipid bilayer but also contains one or two types of virus-encoded glycoproteins (Figure 4-38). The phospholipids in the viral envelope are similar to those in the plasma membrane of an infected host cell. The viral envelope is, in fact, derived by budding from that membrane, but contains mainly viral glycoproteins, as we discuss shortly.

Viruses Can Be Cloned and Counted in Plaque Assays

The number of infectious viral particles in a sample can be quantified by a **plaque assay.** This assay is performed by culturing a dilute sample of viral particles on a plate covered with host cells and then counting the number of local lesions, called *plaques,* that develop (Figure 4-39). A plaque develops on the plate wherever a single virion initially infects a single cell. The virus replicates in this initial host cell and then lyses (ruptures) the cell, releasing many progeny virions that infect the neighboring cells on the plate. After a few such cycles of infection, enough cells are lysed to pro-



◀ EXPERIMENTAL FIGURE 4-39 **Plaque assay determines the number of infectious particles in a viral suspension.** (a) Each lesion, or plaque, which develops where a single virion initially infected a single cell, constitutes a pure viral clone. (b) Plate illuminated from behind shows plaques formed by bacteriophage λ plated on *E. coli.* (c) Plate showing plaques produced by poliovirus plated on HeLa cells. [Part (b) courtesy of Barbara Morris; part (c) from S. E. Luria et al., 1978, *General Virology,* 3d ed., Wiley, p. 26.]



duce a visible clear area, or plaque, in the layer of remaining uninfected cells.

Since all the progeny virions in a plaque are derived from a single parental virus, they constitute a virus **clone.** This type of plaque assay is in standard use for bacterial and animal viruses. Plant viruses can be assayed similarly by counting local lesions on plant leaves inoculated with viruses. Analysis of viral mutants, which are commonly isolated by plaque assays, has contributed extensively to current understanding of molecular cellular processes. The plaque assay also is critical in isolating bacteriophage λ clones carrying segments of cellular DNA, as discussed in Chapter 9.

Lytic Viral Growth Cycles Lead to Death of Host Cells

Although details vary among different types of viruses, those that exhibit a *lytic cycle* of growth proceed through the following general stages:

1. *Adsorption*—Virion interacts with a host cell by binding of multiple copies of capsid protein to specific receptors on the cell surface.

2. *Penetration*—Viral genome crosses the plasma membrane. For animal and plant viruses, viral proteins also enter the host cell.

3. *Replication*—Viral mRNAs are produced with the aid of the host-cell transcription machinery (DNA viruses) or by viral enzymes (RNA viruses). For both types of viruses, viral mRNAs are translated by the host-cell translation machinery. Production of multiple copies of the viral genome is carried out either by viral proteins alone or with the help of host-cell proteins.

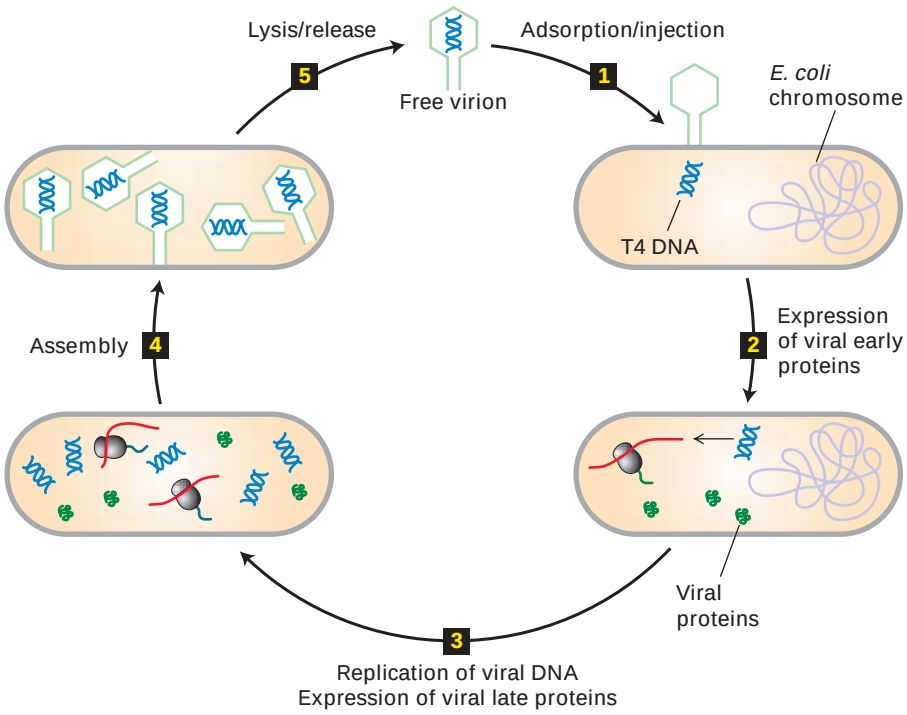
4. *Assembly*—Viral proteins and replicated genomes associate to form progeny virions.

5. *Release*—Infected cell either ruptures suddenly (**lysis**), releasing all the newly formed virions at once, or disintegrates gradually, with slow release of virions.

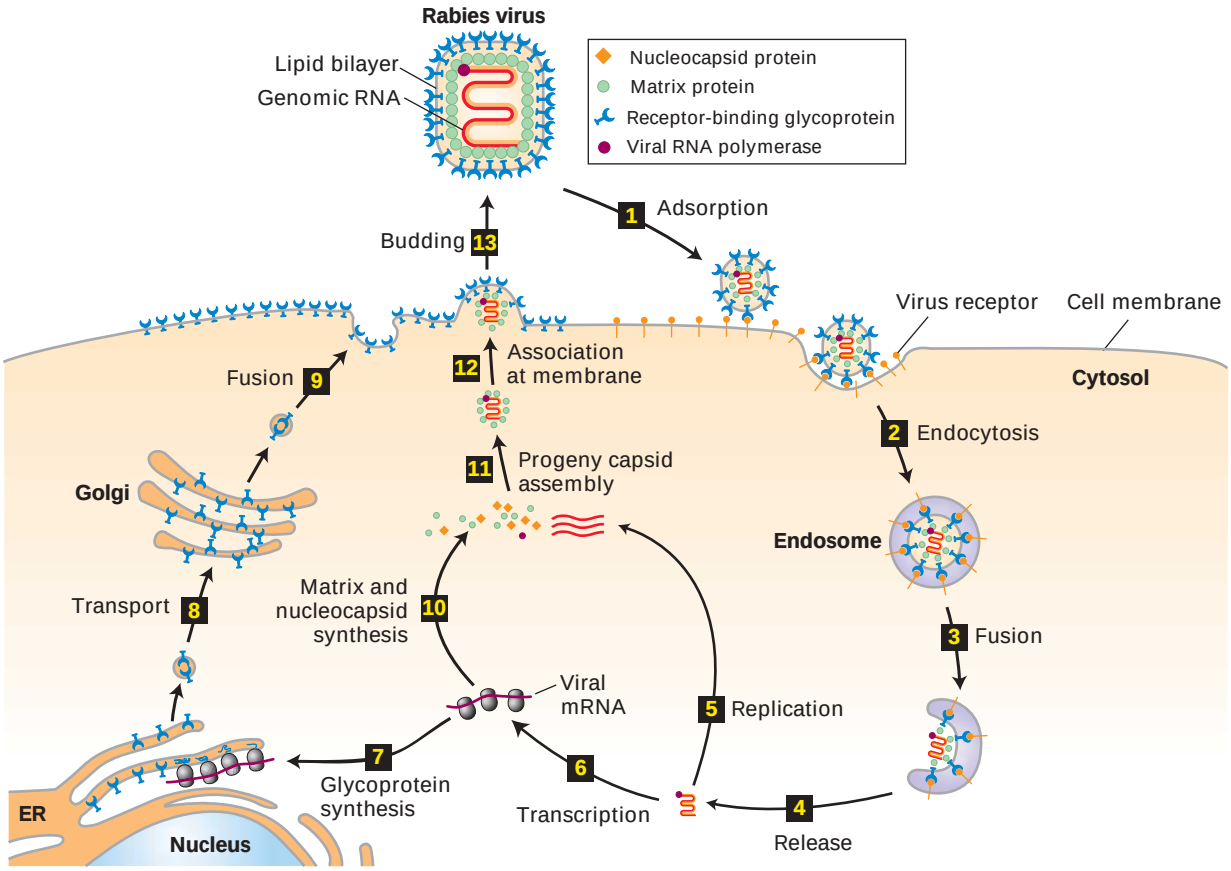
Figure 4-40 illustrates the lytic cycle for T4 bacteriophage, a nonenveloped DNA virus that infects *E. coli.* Viral capsid proteins generally are made in large amounts because many copies of them are required for the assembly of each progeny virion. In each infected cell, about 100–200 T4 progeny virions are produced and released by lysis.

The lytic cycle is somewhat more complicated for DNA viruses that infect eukaryotic cells. In most such viruses, the DNA genome is transported (with some associated proteins) into the cell nucleus. Once inside the nucleus, the viral DNA is transcribed into RNA by the host's transcription machinery. Processing of the viral RNA primary transcript by host-cell enzymes yields viral mRNA, which is transported to the cytoplasm and translated into viral proteins by host-cell ribosomes, tRNA, and translation factors. The viral proteins are then transported back into the nucleus, where some of them either replicate the viral DNA directly or direct cellular proteins to replicate the viral DNA, as in the case of SV40 discussed in the last section. Assembly of the capsid proteins with the newly replicated viral DNA occurs in the nucleus, yielding hundreds to thousands of progeny virions.

Most plant and animal viruses with an RNA genome do not require nuclear functions for lytic replication. In some



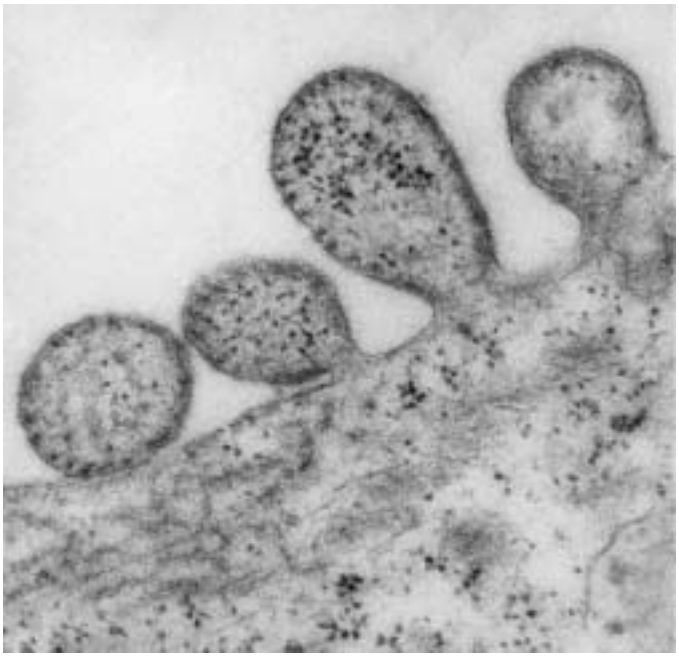
▶ FIGURE 4-40 **Lytic replication cycle of *E. coli* bacteriophage T4, a nonenveloped virus with a double-stranded DNA genome.** After viral coat proteins at the tip of the tail in T4 interact with specific receptor proteins on the exterior of the host cell, the viral genome is injected into the host (step 1). Host-cell enzymes then transcribe viral "early" genes into mRNAs and subsequently translate these into viral "early" proteins (step 2). The early proteins replicate the viral DNA and induce expression of viral "late" proteins by host-cell enzymes (step 3). The viral late proteins include capsid and assembly proteins and enzymes that degrade the host-cell DNA, supplying nucleotides for synthesis of more viral DNA. Progeny virions are assembled in the cell (step 4) and released (step 5) when viral proteins lyse the cell. Newly liberated viruses initiate another cycle of infection in other host cells.



of these viruses, a virus-encoded enzyme that enters the host during penetration transcribes the genomic RNA into mRNAs in the cell cytoplasm. The mRNA is directly translated into viral proteins by the host-cell translation machinery. One or more of these proteins then produces additional copies of the viral RNA genome. Finally, progeny genomes are assembled with newly synthesized capsid proteins into progeny virions in the cytoplasm.

After the synthesis of hundreds to thousands of new virions has been completed, most infected bacterial cells and some infected plant and animal cells are lysed, releasing all the virions at once. In many plant and animal viral infections, however, no discrete lytic event occurs; rather, the dead host cell releases the virions as it gradually disintegrates.

As noted previously, enveloped animal viruses are surrounded by an outer phospholipid layer derived from the plasma membrane of host cells and containing abundant viral glycoproteins. The processes of adsorption and release of enveloped viruses differ substantially from these processes in nonenveloped viruses. To illustrate lytic replication of enveloped viruses, we consider the rabies virus, whose nucleocapsid consists of a single-stranded RNA genome surrounded by multiple copies of nucleocapsid protein. Like other lytic RNA viruses, rabies virions are replicated in the cytoplasm and do not require host-cell nuclear enzymes. As shown in Figure 4-41, a rabies virion is adsorbed by endocytosis, and release of progeny virions occurs by *budding* from the host-cell plasma membrane. Budding virions are clearly visible in electron micrographs of infected cells, as illustrated in Figure 4-42. Many tens of thousands of progeny virions bud from an infected host cell before it dies.

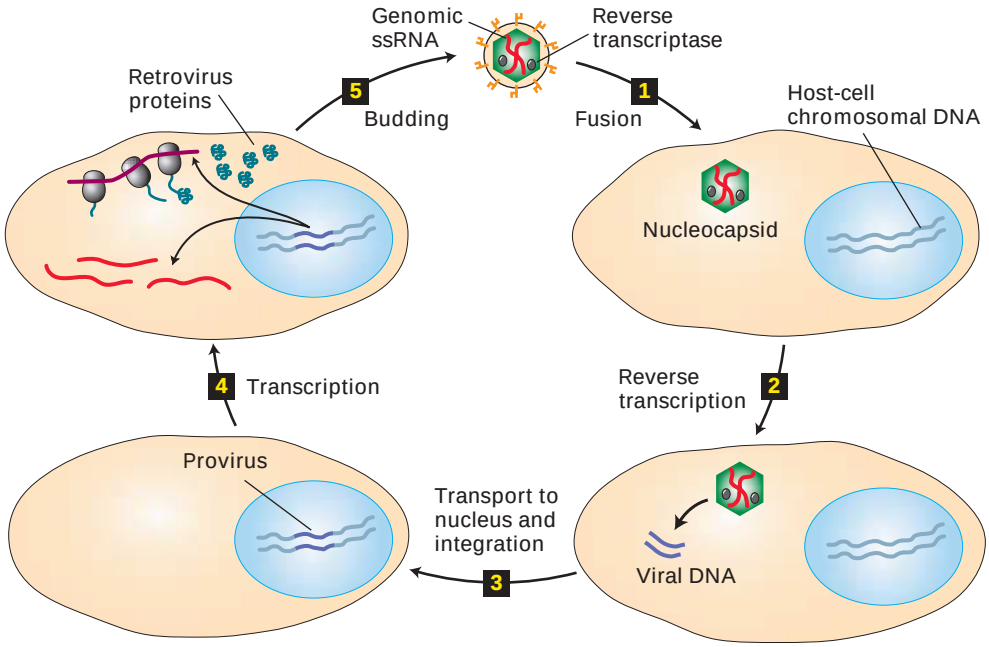


▲ EXPERIMENTAL FIGURE 4-42 **Progeny virions of enveloped viruses are released by budding from infected cells.** In this transmission electron micrograph of a cell infected with measles virus, virion buds are clearly visible protruding from the cell surface. Measles virus is an enveloped RNA virus with a helical nucleocapsid, like rabies virus, and replicates as illustrated in Figure 4-41. [From A. Levine, 1991, *Viruses,* Scientific American Library, p. 22.]

◀ FIGURE 4-41 **Lytic replication cycle of rabies virus, an enveloped virus with a single-stranded RNA genome.** The structural components of this virus are depicted at the top. Note that the nucleocapsid is helical rather than icosahedral. After a virion adsorbs to multiple copies of a specific host membrane protein (step 1), the cell engulfs it in an endosome (step 2). A cellular protein in the endosome membrane pumps H^+ ions from the cytosol into the endosome interior. The resulting decrease in endosomal pH induces a conformational change in the viral glycoprotein, leading to fusion of the viral envelope with the endosomal lipid bilayer membrane and release of the nucleocapsid into the cytosol (steps 3 and 4). Viral RNA polymerase uses ribonucleoside triphosphates in the cytosol to replicate the viral RNA genome (step 5) and to synthesize viral mRNAs (step 6). One of the viral mRNAs encodes the viral transmembrane glycoprotein, which is inserted into the membrane of the endoplasmic reticulum (ER) as it is synthesized on ER-bound ribosomes (step 7). Carbohydrate is added to the large folded domain inside the ER lumen and is modified as the membrane and the associated glycoprotein pass through the Golgi apparatus (step 8). Vesicles with mature glycoprotein fuse with the host plasma membrane, depositing viral glycoprotein on the cell surface with the large receptor-binding domain outside the cell (step 9). Meanwhile, other viral mRNAs are translated on host-cell ribosomes into nucleocapsid protein, matrix protein, and viral RNA polymerase (step 10). These proteins are assembled with replicated viral genomic RNA (bright red) into progeny nucleocapsids (step 11), which then associate with the cytosolic domain of viral transmembrane glycoproteins in the plasma membrane (step 12). The plasma membrane is folded around the nucleocapsid, forming a "bud" that eventually is released (step 13).

Viral DNA Is Integrated into the Host-Cell Genome in Some Nonlytic Viral Growth Cycles

Some bacterial viruses, called *temperate phages*, can establish a nonlytic association with their host cells that does not kill the cell. For example, when λ bacteriophage infects *E. coli*, the viral DNA may be integrated into the host-cell chromosome rather than being replicated. The integrated viral DNA, called a *prophage*, is replicated as part of the cell's DNA from one host-cell generation to the next. This phenomenon is referred to as **lysogeny.** Under certain conditions, the prophage DNA is activated, leading to its excision from the host-cell chromosome, entrance into the lytic cycle, and subsequent production and release of progeny virions.



▲ FIGURE 4-43 **Retroviral life cycle.** Retroviruses have a genome of two identical copies of single-stranded RNA and an outer envelope. Step 1: After viral glycoproteins in the envelope interact with a specific host-cell membrane protein, the retroviral envelope fuses directly with the plasma membrane, allowing entry of the nucleocapsid into the cytoplasm of the cell. Step 2: Viral reverse transcriptase and other proteins copy the viral ssRNA genome into a double-stranded DNA. Step 3: The viral dsDNA is transported into the nucleus and integrated into one of many possible sites in the host-cell chromosomal DNA. For simplicity, only one host-cell chromosome is depicted. Step 4: The integrated viral DNA (provirus) is transcribed by the host-cell RNA polymerase, generating mRNAs (dark red) and genomic RNA molecules (bright red). The host-cell machinery translates the viral mRNAs into glycoproteins and nucleocapsid proteins. Step 5: Progeny virions then assemble and are released by budding as illustrated in Figure 4-41.

MEDIA CONNECTIONS
Overview Animation: Life Cycle of a Retrovirus

The genomes of a number of animal viruses also can integrate into the host-cell genome. Probably the most important are the **retroviruses,** which are enveloped viruses with a genome consisting of two identical strands of RNA. These viruses are known as *retroviruses* because their RNA genome acts as a template for formation of a DNA molecule—the opposite flow of genetic information compared with the more common transcription of DNA into RNA. In the retroviral life cycle (Figure 4-43), a viral enzyme called **reverse transcriptase** initially copies the viral RNA genome into single-stranded DNA complementary to the virion RNA; the same enzyme then catalyzes synthesis of a complementary DNA strand. (This complex reaction is detailed in Chapter 10 when we consider closely related intracellular parasites called retrotransposons.) The resulting double-stranded DNA is integrated into the chromosomal DNA of the infected cell. Finally, the integrated DNA, called a *provirus*, is transcribed by the cell's own machinery into RNA, which either is translated into viral proteins or is packaged within virion coat proteins to form progeny virions that are released by budding from the host-cell membrane. Because most retroviruses do not kill their host cells, infected cells can replicate, producing daughter cells with integrated proviral DNA. These daughter cells continue to transcribe the proviral DNA and bud progeny virions.

MEDICINE

Some retroviruses contain cancer-causing genes (**oncogenes**), and cells infected by such retroviruses are oncogenically transformed into tumor cells. Studies of oncogenic retroviruses (mostly viruses of birds and mice) have revealed a great deal about the processes that lead to transformation of a normal cell into a cancer cell (Chapter 23).

Among the known human retroviruses are human T-cell lymphotrophic virus (HTLV), which causes a form of leukemia, and human immunodeficiency virus (HIV), which causes acquired immune deficiency syndrome (AIDS). Both of these viruses can infect only specific cell types, primarily certain cells of the immune system and, in the case of HIV, some central nervous system neurons and glial cells. Only these cells have cell-surface receptors that interact with viral envelope proteins, accounting for the host-cell specificity of these viruses. Unlike most other retroviruses, HIV eventually kills its host cells. The eventual death of large numbers of

immune-system cells results in the defective immune response characteristic of AIDS.

Some DNA viruses also can integrate into a host-cell chromosome. One example is the human papillomaviruses (HPVs), which most commonly cause warts and other benign skin lesions. The genomes of certain HPV serotypes, however, occasionally integrate into the chromosomal DNA of infected cervical epithelial cells, initiating development of cervical cancer. Routine Pap smears can detect cells in the early stages of the transformation process initiated by HPV integration, permitting effective treatment. ▮

KEY CONCEPTS OF SECTION 4.7

Viruses: Parasites of the Cellular Genetic System

- Viruses are small parasites that can replicate only in host cells. Viral genomes may be either DNA (DNA viruses) or RNA (RNA viruses) and either single- or double-stranded.
- The capsid, which surrounds the viral genome, is composed of multiple copies of one or a small number of virus-encoded proteins. Some viruses also have an outer envelope, which is similar to the plasma membrane but contains viral transmembrane proteins.
- Most animal and plant DNA viruses require host-cell nuclear enzymes to carry out transcription of the viral genome into mRNA and production of progeny genomes. In contrast, most RNA viruses encode enzymes that can transcribe the RNA genome into viral mRNA and produce new copies of the RNA genome.
- Host-cell ribosomes, tRNAs, and translation factors are used in the synthesis of all viral proteins in infected cells.
- Lytic viral infection entails adsorption, penetration, synthesis of viral proteins and progeny genomes (replication), assembly of progeny virions, and release of hundreds to thousands of virions, leading to death of the host cell (see Figure 4-40). Release of enveloped viruses occurs by budding through the host-cell plasma membrane (see Figure 4-41).
- Nonlytic infection occurs when the viral genome is integrated into the host-cell DNA and generally does not lead to cell death.
- Retroviruses are enveloped animal viruses containing a single-stranded RNA genome. After a host cell is penetrated, reverse transcriptase, a viral enzyme carried in the virion, converts the viral RNA genome into double-stranded DNA, which integrates into chromosomal DNA (see Figure 4-43).
- Unlike infection by other retroviruses, HIV infection eventually kills host cells, causing the defects in the immune response characteristic of AIDS.
- Tumor viruses, which contain oncogenes, may have an RNA genome (e.g., human T-cell lymphotrophic virus) or a DNA genome (e.g., human papillomaviruses). In the case of these viruses, integration of the viral genome into a host-cell chromosome can cause transformation of the cell into a tumor cell.

PERSPECTIVES FOR THE FUTURE

In this chapter we first reviewed the basic structure of DNA and RNA and then described fundamental aspects of the transcription of DNA by RNA polymerases. Eukaryotic RNA polymerases are discussed in greater detail in Chapter 11, along with additional factors required for transcription initiation in eukaryotic cells and interactions with regulatory transcription factors that control transcription initiation. Next, we discussed the genetic code and the participation of tRNA and the protein-synthesizing machine, the ribosome, in decoding the information in mRNA to allow accurate assembly of protein chains. Mechanisms that regulate protein synthesis are considered further in Chapter 12. Finally, we considered the molecular details underlying the accurate replication of DNA required for cell division. Chapter 21 covers the mechanisms that regulate when a cell replicates its DNA and that coordinate DNA replication with the complex process of mitosis that distributes the daughter DNA molecules equally to each daughter cell.

These basic cellular processes form the foundation of molecular cell biology. Our current understanding of these processes is grounded in a wealth of experimental results and is not likely to change. However, the depth of our understanding will continue to increase as additional details of the structures and interactions of the macromolecular machines involved are uncovered. The determination in recent years of the three-dimensional structures of RNA polymerases, ribosomal subunits, and DNA replication proteins has allowed researchers to design ever more penetrating experimental approaches for revealing how these macromolecules operate at the molecular level. The detailed level of understanding that results may allow the design of new and more effective drugs for treating human illnesses. For example, the recent high-resolution structures of ribosomes are providing insights into the mechanism by which antibiotics inhibit bacterial protein synthesis without affecting the function of mammalian ribosomes. This new knowledge may allow the design of even more effective antibiotics. Similarly, detailed understanding of the mechanisms regulating transcription of specific human genes may lead to therapeutic strategies that can reduce or prevent inappropriate immune responses that lead to multiple sclerosis and arthritis, the inappropriate cell division that is the hallmark of cancer, and other pathological processes.

Much of current biological research is focused on discovering how molecular interactions endow cells with decision-making capacity and their special properties. For this reason several of the following chapters describe current knowledge about how such interactions regulate transcription and protein synthesis in multicellular organisms and how such regulation endows cells with the capacity to become

specialized and grow into complicated organs. Other chapters deal with how protein-protein interactions underlie the construction of specialized organelles in cells, and how they determine cell shape and movement. The rapid advances in molecular cell biology in recent years hold promise that in the not too distant future we will understand how the regulation of specialized cell function, shape, and mobility coupled with regulated cell replication and cell death (apoptosis) lead to the growth of complex organisms like trees and human beings.

KEY TERMS

anticodon *119*
codons *119*
complementary *104*
DNA polymerases *133*
double helix *103*
envelope (viral) *137*
exons *111*
genetic code *119*
introns *111*
lagging strand *133*
leading strand *133*
messenger RNA (mRNA) *119*
Okazaki fragments *133*
operon *111*
phosphodiester bond *103*
plaque assay *138*
polyribosomes *130*
primary transcript *110*
primer *133*
promoter *109*
reading frame *120*
replication fork *133*
reverse transcriptase *142*
ribosomal RNA (rRNA) *119*
ribosomes *119*
RNA polymerase *109*
transcription *101*
transfer RNA (tRNA) *119*
translation *101*
Watson-Crick base pairs *103*

REVIEW THE CONCEPTS

1. What are Watson-Crick base pairs? Why are they important?

2. TATA box–binding protein binds to the minor groove of DNA, resulting in the bending of the DNA helix (see Figure 4-5). What property of DNA allows the TATA box–binding protein to recognize the DNA helix?

3. Preparing plasmid (double-stranded, circular) DNA for sequencing involves annealing a complementary, short, single-stranded oligonucleotide DNA primer to one strand of the plasmid template. This is routinely accomplished by heating the plasmid DNA and primer to 90 °C and then slowly bringing the temperature down to 25 °C. Why does this protocol work?

4. What difference between RNA and DNA helps to explain the greater stability of DNA? What implications does this have for the function of DNA?

5. What are the major differences in the synthesis and structure of prokaryotic and eukaryotic mRNAs?

6. While investigating the function of a specific growth factor receptor gene from humans, it was found that two types of proteins are synthesized from this gene. A larger protein containing a membrane-spanning domain functions to recognize growth factors at the cell surface, stimulating a specific downstream signaling pathway. In contrast, a related, smaller protein is secreted from the cell and functions to bind available growth factor circulating in the blood, thus inhibiting the downstream signaling pathway. Speculate on how the cell synthesizes these disparate proteins.

7. Describe the molecular events that occur at the *lac* operon when *E. coli* cells are shifted from a glucose-containing medium to a lactose-containing medium.

8. The concentration of free phosphate affects transcription of some *E. coli* genes. Describe the mechanism for this.

9. Contrast how selection of the translational start site occurs on bacterial, eukaryotic, and poliovirus mRNAs.

10. What is the evidence that the 23S rRNA in the large rRNA subunit has a peptidyl transferase activity?

11. How would a mutation in the poly(A)-binding protein I gene affect translation? How would an electron micrograph of polyribosomes from such a mutant differ from the normal pattern?

12. What characteristic of DNA results in the requirement that some DNA synthesis is discontinuous? How are Okazaki fragments and DNA ligase utilized by the cell?

13. What gene is unique to retroviruses? Why is the protein encoded by this gene absolutely necessary for maintaining the retroviral life cycle, but not that of other viruses?

ANALYZE THE DATA

NASA has identified a new microbe present on Mars and requests that you determine the genetic code of this organism. To accomplish this goal, you isolate an extract from this microbe that contains all the components necessary for protein synthesis except mRNA. Synthetic mRNAs are added to this extract and the resulting polypeptides are analyzed:

Synthetic mRNA	*Resulting Polypeptides*
AAAAAAAAAAAAAAAA	Lysine-Lysine-Lysine etc.
CACACACACACACACA	Threonine-Histidine-Threonine-Histidine etc.
AACAACAACAACAACA	Threonine-Threonine-Threonine etc.
	Glutamine-Glutamine-Glutamine etc.
	Asparagine-Asparagine-Asparagine etc.

From these data, what specifics can you conclude about the microbe's genetic code? What is the sequence of the anticodon loop of a tRNA carrying a threonine? If you found that this microbe contained 61 different tRNAs, what could you speculate about the fidelity of translation in this organism?

REFERENCES

Structure of Nucleic Acids

Dickerson, R. E. 1983. The DNA helix and how it is read. *Sci. Am.* **249**:94–111.

Doudna, J. A., and T. R. Cech. 2002. The chemical repertoire of natural ribozymes. *Nature* **418**:222–228.

Kornberg, A., and T. A. Baker. 1992. *DNA Replication*, 2d ed. W. H. Freeman and Company, chap. 1. A good summary of the principles of DNA structure.

Wang, J. C. 1980. Superhelical DNA. *Trends Biochem. Sci.* **5**:219–221.

Transcription of Protein-Coding Genes and Formation of Functional mRNA

Brenner, S., F. Jacob, and M. Meselson. 1961. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190:**576–581.

Young, B. A., T. M. Gruber, and C. A. Gross. 2002. Views of transcription initiation. *Cell* **109**:417–420.

Control of Gene Expression in Prokaryotes

Bell, C. E., and M. Lewis. 2001. The Lac repressor: a second generation of structural and functional studies. *Curr. Opin. Struc. Biol.* **11**:19–25.

Busby, S., and R. H. Ebright. 1999. Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.* **293**:199–213.

Darst, S. A. 2001. Bacterial RNA polymerase. *Curr. Opin. Struc. Biol.* **11**:155–162.

Muller-Hill, B. 1998. Some repressors of bacterial transcription. *Curr. Opin. Microbiol.* **1**:145–151.

The Three Roles of RNA in Translation

Alexander, R. W., and P. Schimmel. 2001. Domain-domain communication in aminoacyl-tRNA synthetases. *Prog. Nucleic Acid Res. Mol. Biol.* **69**:317–349.

Bjork, G. R., et al. 1987. Transfer RNA modification. *Ann. Rev. Biochem.* **56**:263–287.

Garrett, R. A., et al., eds. 2000. *The Ribosome: Structure, Function, Antibiotics, and Cellular Interactions.* ASM Press.

Hatfield, D. L., and V. N. Gladyshev. 2002. How selenium has altered our understanding of the genetic code. *Mol. Cell Biol.* **22**:3565–3576.

Hoagland, M. B., et al. 1958. A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.* **231**:241–257.

Holley, R. W., et al. 1965. Structure of a ribonucleic acid. *Science* **147**:1462–1465.

Ibba, M., and D. Soll. 2001.The renaissance of aminoacyl-tRNA synthesis. *EMBO Rep.* **2**:382–387.

Khorana, G. H., et al. 1966. Polynucleotide synthesis and the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **31**:39–49.

Maguire, B. A., and R. A. Zimmermann. 2001. The ribosome in focus. *Cell* **104**:813–816.

Nirenberg, M., et al. 1966. The RNA code in protein synthesis. *Cold Spring Harbor Symp. Quant. Biol.* **31**:11–24.

Ramakrishnan, V. 2002. Ribosome structure and the mechanism of translation. *Cell* **108**:557–572.

Rich, A., and S.-H. Kim. 1978. The three-dimensional structure of transfer RNA. *Sci. Am.* **240**(1):52–62 (offprint 1377).

Stepwise Synthesis of Proteins on Ribosomes

Gingras, A. C., R. Raught, and N. Sonenberg. 1999. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Ann. Rev. Biochem.* **68**:913–963.

Green, R. 2000. Ribosomal translocation: EF-G turns the crank. *Curr. Biol.* **10**:R369–R373.

Hellen, C. U., and P. Sarnow. 2001. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genet. Devel.* **15**:1593–1612.

Kisselev, L. L., and R. H. Buckingham. 2000. Translational termination comes of age. *Trends Biochem. Sci.* **25**:561–566.

Kozak, M. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**:187–208.

Noller, H. F., et al. 2002. Translocation of tRNA during protein synthesis. *FEBS Lett.* **514**:11–16.

Pestova, T. V., et al. 2001. Molecular mechanisms of translation initiation in eukaryotes. *Proc. Nat'l. Acad. Sci. USA* **98**:7029–7036.

Poole, E., and W. Tate. 2000. Release factors and their role as decoding proteins: specificity and fidelity for termination of protein synthesis. *Biochim. Biophys. Acta* **1493**:1–11.

Ramakrishnan, V. 2002. Ribosome structure and the mechanism of translation. *Cell* **108**:557–572.

Sonenberg, N., J. W. B. Hershey, and M. B. Mathews, eds. 2000. *Translational Control of Gene Expression*. Cold Spring Harbor Laboratory Press.

DNA Replication

Bullock, P. A. 1997. The initiation of simian virus 40 DNA replication in vitro. *Crit. Rev. Biochem. Mol. Biol.* **32**:503–568.

Kornberg, A., and T. A. Baker. 1992. *DNA Replication*, 2d ed. W. H. Freeman and Company

Waga, S., and B. Stillman. 1998. The DNA replication fork in eukaryotic cells. *Ann. Rev. Biochem.* **67**:721–751.

Viruses: Parasites of the Cellular Genetic System

Flint, S. J., et al. 2000. *Principles of Virology: Molecular Biology, Pathogenesis, and Control.* ASM Press.

Hull, R. 2002. *Mathews' Plant Virology*. Academic Press.

Knipe, D. M., and P. M. Howley, eds. 2001. *Fields Virology.* Lippincott Williams & Wilkins.

Kornberg, A., and T. A. Baker. 1992. *DNA Replication*, 2d ed. W. H. Freeman and Company. Good summary of bacteriophage molecular biology.