8 Microbial Genetics

Virtually all the microbial traits you have read about in earlier chapters are controlled or influenced by heredity. The inherited traits of microbes include their shape and structural features, their metabolism, their ability to move or behave in various ways, and their ability to interact with other organisms—perhaps causing disease. Individual organisms transmit these characteristics to their offspring through genes.

Researchers are trying to solve the difficult medical problem of microbes developing antibiotic resistance. Microorganisms can become resistant to antibiotics in any of several ways, all of which depend on genetics. The emergence of vancomycin-resistant *Staphylococcus aureus (VRSA)* poses a serious threat to patient care. In this chapter you will see how VRSA acquired this trait.

Emergent diseases provide another example of the importance of understanding genetics. New diseases are the results of genetic changes in some existing organism; for example, *E. coli* O157:H7 acquired the genes for Shiga toxin from *Shigella*.

Currently, microbiologists are using genetics to discover relatedness among organisms and the origins of organisms such as HIV and West Nile virus and to study the potential for avian influenza viruses to infect humans.

UNDER THE MICROSCOPE

Bacterial Chromosome. The single chromosome, normally tightly packed inside a bacterial cell, has burst from an *E. coli* cell after the cell wall and plasma membrane were damaged.

STRUCTURE AND FUNCTION OF THE GENETIC MATERIAL

LEARNING OBJECTIVES

- Define *genetics, genome, chromosome, gene, genetic code, genotype, phenotype,* and *genomics.*
- Describe how DNA serves as genetic information.

Genetics is the science of heredity; it includes the study of what genes are, how they carry information, how they are replicated and passed to subsequent generations of cells or passed between organisms, and how the expression of their information within an organism determines the particular characteristics of that organism. The genetic information in a cell is called the **genome.** A cell's genome includes its chromosomes and plasmids. **Chromosomes** are structures containing DNA that physically carry hereditary information; the chromosomes contain the genes. **Genes** are segments of DNA (except in some viruses, in which they are made of RNA) that code for functional products. We saw in Chapter 2, on page 48, that DNA is a macromolecule composed of repeating units called *nucleotides*. Recall that each nucleotide consists of a nitrogenous base (adenine, thymine, cytosine, or guanine), deoxyribose (a pentose sugar), and a phosphate group (see Figure 2.16, page 48). The DNA within a cell exists as long strands of nucleotides twisted together in pairs to form a double helix. Each strand has a string of alternating sugar and phosphate groups (its *sugar-phosphate backbone*), and a nitrogenous base is attached to each sugar in the backbone. The two strands are held together by hydrogen bonds between their nitrogenous bases. The **base pairs** always occur in a specific way: adenine always pairs with thymine, and cytosine always pairs with guanine. Because of this specific base pairing, the base sequence of one DNA strand determines the base sequence of the other strand. The two strands of DNA are thus *complementary*. You can think of these complementary DNA sequences as being like a positive photograph and its negative.

The structure of DNA helps explain two primary features of biological information storage. First, the linear sequence of bases provides the actual information. Genetic information is encoded by the sequence of bases along a strand of DNA, in much the same way as our written language uses a linear sequence of letters to form words and sentences. The genetic language, however, uses an alphabet with only four letters—the four kinds of nitrogenous bases in DNA (or RNA). But 1000 of these four bases, the number contained in an average-sized gene, can be arranged in 4^{1000} different ways. This astronomically large number explains how genes can be varied enough to provide all the information a cell needs to grow and perform its functions. The **genetic code,** the set of rules that determines how a nucleotide sequence is converted into the amino acid sequence of a protein, is discussed in more detail later in the chapter.

Second, the complementary structure allows for the precise duplication of DNA during cell division. Again, think of the photograph analogy: if you have a negative, you can always make another copy of the positive print. Likewise with DNA: if you know the sequence of one strand, you also know the sequence of the complementary strand.

Much of cellular metabolism is concerned with translating the genetic message of genes into specific proteins. A gene usually codes for a messenger RNA (mRNA) molecule, which ultimately results in the formation of a protein. Alternatively, the gene product can be a ribosomal RNA (rRNA) or a transfer RNA (tRNA). As we will see, all of these types of RNA are involved in the process of protein synthesis. When the ultimate molecule for which a gene codes (a protein, for example) has been produced, we say that the gene has been *expressed*.

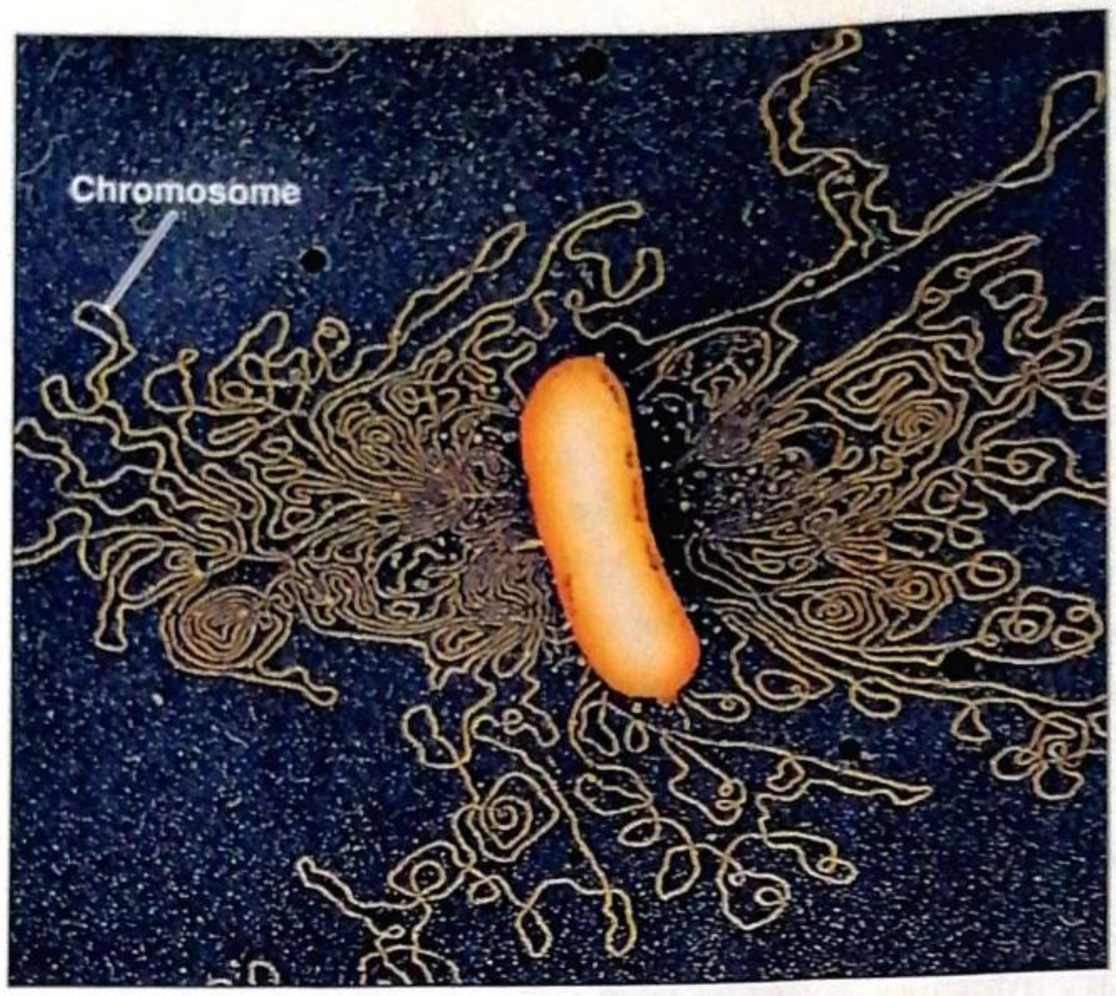
GENOTYPE AND PHENOTYPE

The **genotype** of an organism is its genetic makeup, the information that codes for all the particular characteristics of the organism. The genotype represents *potential* properties, but not the properties themselves. **Phenotype** refers to *actual, expressed* properties, such as the organism's ability to perform a particular chemical reaction. Phenotype, then, is the manifestation of genotype.

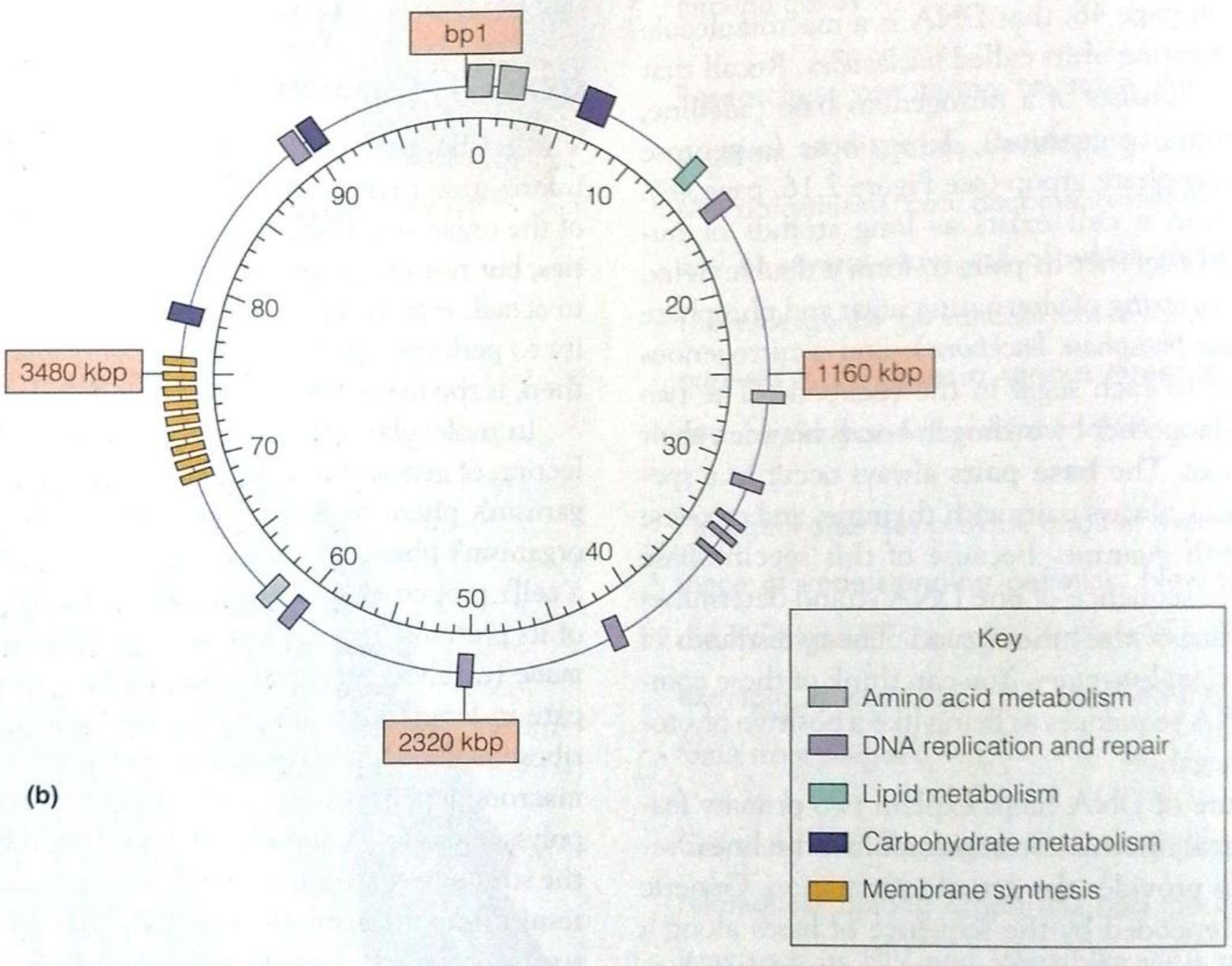
In molecular terms, an organism's genotype is its collection of genes, its entire DNA. What constitutes the organism's phenotype in molecular terms? In a sense, an organism's phenotype is its collection of proteins. Most of a cell's properties derive from the structures and functions of its proteins. In microbes, most proteins are either *enzymatic* (catalyze particular reactions) or *structural* (participate in large functional complexes such as membranes or ribosomes). Even phenotypes that depend on structural macromolecules other than proteins (such as lipids or polysaccharides) rely indirectly on proteins. For instance, the structure of a complex lipid or polysaccharide molecule results from the catalytic activities of enzymes that synthesize, process, and degrade those molecules. Thus, although it is not completely accurate to say that phenotypes are due only to proteins, it is a useful simplification.

DNA AND CHROMOSOMES

Bacteria typically have a single circular chromosome consisting of a single circular molecule of DNA with associated proteins. The chromosome is looped and folded (Figure 8.1a)



(a)



(b)

FIGURE 8.1 Chromosomes. (**a**) A prokaryotic chromosome. The tangled mass and looping strands of DNA emerging from this disrupted *E. coli* cell are part of its single chromosome. (**b**) A genetic map of the chromosome of *E. coli*. The numbers inside the circle are minutes based on the length of time it takes to transfer the genes during mating between two cells; the numbers in colored boxes are base pairs.

Q **What is a gene? What is an open-reading frame?**

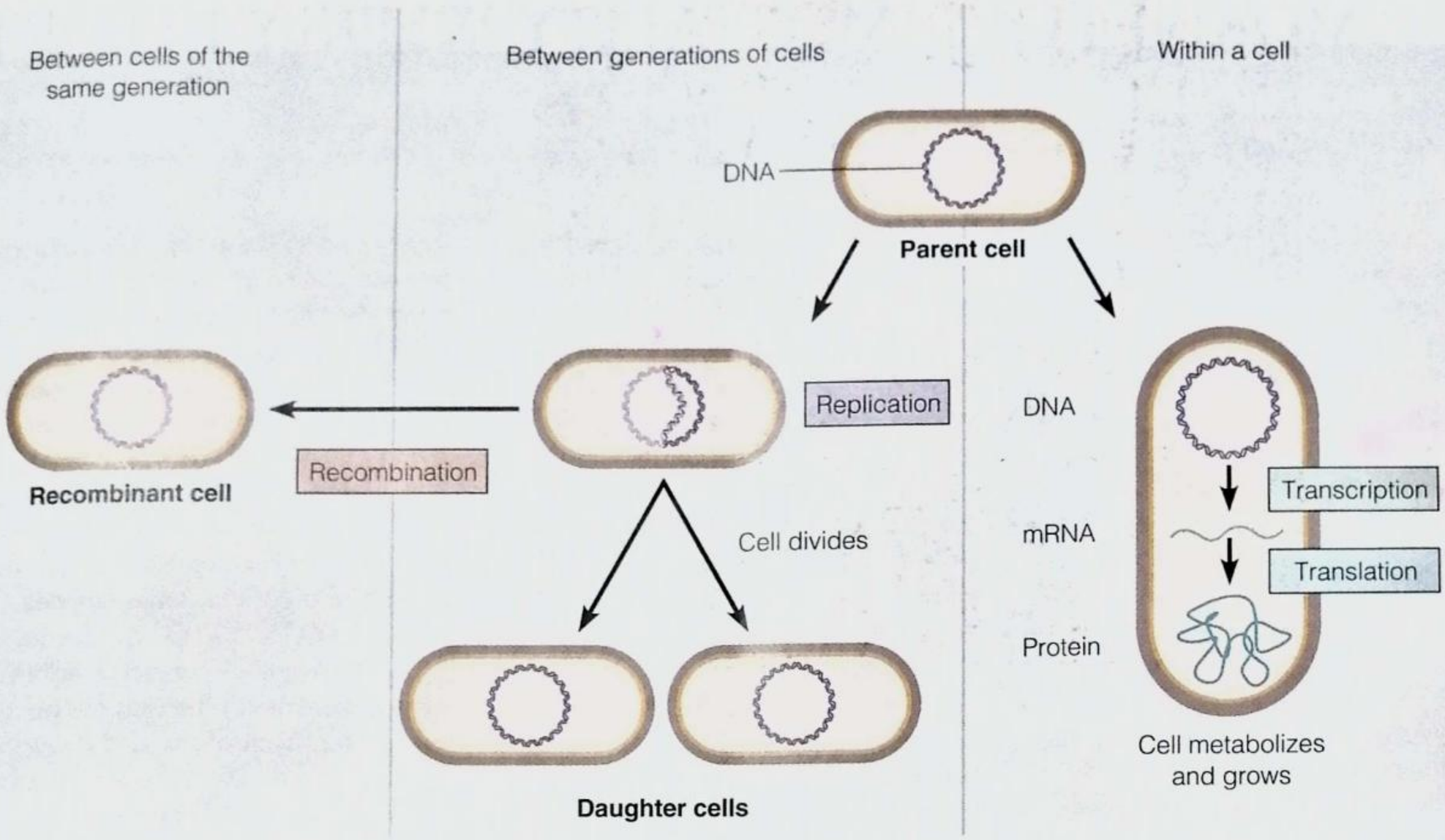


FIGURE 8.2 An overview of the flow of genetic information. Genetic information can be transferred between generations of cells, through DNA replication. Occasionally, genetic information can be transferred between cells of the same generation through recombination. Genetic information is also used within a cell to produce the proteins the cell needs to function (through transcription and translation). The cell represented here is a bacterium with a single circular chromosome. A small version of this figure will be included in figures throughout this chapter to indicate the relationships of different processes.

Q **All of these processes can occur at the same time in a bacterial cell. Which process results in reproduction?**

and attached at one or several points to the plasma membrane. The DNA of *E. coli*, the most-studied bacterial species, has about 4.6 million base pairs and is about 1 mm long—1000 times longer than the entire cell. However, the chromosome takes up only about 10% of the cell's volume because the DNA is twisted, or *supercoiled*—much like a telephone cord when you put the handset back on the receiver.

The location of genes on a bacterial chromosome can be determined by experiments on the transfer of genes from one cell to another. These processes will be discussed later in this chapter. The bacterial chromosome map that results is marked in minutes corresponding to when the genes are transferred from a donor cell to a recipient cell (Figure 8.1b).

In recent years, the complete base sequences of several bacterial chromosomes have been determined. Computers are used to search for *open-reading frames*, that is, regions of DNA that are likely to encode a protein. As you will see later, these are base sequences between start and stop codons. The sequencing and molecular characterization of genomes is called **genomics.** The use of genomics to track West Nile virus is described in the box on the following page.

THE FLOW OF GENETIC INFORMATION

DNA replication makes possible the flow of genetic information from one generation to the next. As shown in Figure 8.2, the DNA of a cell replicates before cell division so that each offspring cell receives a chromosome identical to the parent's. Within each metabolizing cell, the genetic information contained in DNA also flows in another way: it is transcribed into mRNA and then translated into protein. We describe the processes of transcription and translation later in this chapter.

DNA REPLICATION

LEARNING OBJECTIVE

- Describe the process of DNA replication.

In DNA replication, one "parental" double-stranded DNA molecule is converted to two identical "daughter" molecules. The complementary structure of the nitrogenous base sequences in the DNA molecule is the key to understanding DNA replication. Because the bases along the two strands of double-helical DNA are complementary,

MORBIDITY & MORTALITY WEEKLY REPORT

TRACKING WEST NILE VIRUS

On August 23, 1999, an infectious disease physician from a hospital in northern Queens contacted the New York City Department of Health (NYCDOH) to report two patients with encephalitis. On investigation, NYCDOH initially identified a cluster of six patients with encephalitis, five of whom had profound muscle weakness and required respiratory support. No bacteria were cultured from the patients' blood or cerebrospinal fluid. Viruses transmitted by mosquitoes are a likely cause of aseptic encephalitis during the summer months. These viruses are called arboviruses. Arboviruses, arthropod-borne, are viruses that are maintained in nature through biological transmission between susceptible vertebrate hosts by blood-feeding arthropods such as mosquitoes.

Testing of these initial cases for antibodies to the common North American arboviruses was positive for Saint Louis encephalitis virus (SLE) on September 3 at the CDC. SLE belongs to the family Flaviviridae.

Subsequent nucleic-acid sequencing of these isolates was performed at the CDC on September 23. Comparison of the nucleic acid sequences to databases indicated that the viruses were closely related to West Nile virus (WNV), which had never been isolated in the western hemisphere.

By 2004, WNV had been found in birds in all states except Alaska and Hawaii. The recognition of WNV in the Western Hemisphere in the summer of 1999 marked the first introduction in recent history of an Old World flavivirus into the New World. The United States is not alone, however, in reporting new or heightened activity in humans and other animals. In 2003, WNV caused encephalitis in horses in Mexico, and incursions of flaviviruses into new areas are likely to continue through increasing global commerce and travel.

West Nile virus was first isolated in 1937 in the West Nile district of Uganda. In the early 1950s, scientists recognized WNV encephalitis outbreaks in humans in Egypt and Israel. Initially considered a minor arbovirus, WNV has recently emerged as a major public health and veterinary concern in southern Europe, the Mediterranean basin, and North America.

Currently researchers are looking at the virus's genome for clues about its path around the world. The flavivirus genome consists of a positive, single-stranded RNA 11,000 to 12,000 nucleotides long. (Positive RNA can act as mRNA and be translated.) The virus has acquired several mutations, and researchers are looking for clues in these mutations to determine the virus's journey.

Using the portions of the genomes (shown below) that encode viral proteins, how similar are these viruses? Can you figure out its movement around the world?

Although genetically related groups or clades can be seen, the actual journey of the virus remains elusive.

SOURCE: Adapted from CDC data.

Portion of the nucleotide base region of the viral envelope protein. Although WNV is an RNA virus, the convention is to write genomes as DNA in the 5′ → 3′ direction.

Australia	A	C	C	C	C	G	T	C	C	A	C	C	C	T	T	T	C	A	A	T	T
Egypt	A	A	T	C	C	C	T	C	C	T	C	T	C	C	T	T	C	G	A	C	T
France	A	A	T	C	C	C	T	C	C	T	C	G	C	C	T	T	C	G	A	C	T
Israel	A	A	C	C	C	C	T	C	C	T	C	T	C	C	T	T	C	G	A	C	T
Italy	A	A	C	C	A	C	T	C	T	T	C	C	C	C	T	A	C	G	A	T	T
Kenya	A	A	C	C	A	C	T	C	T	T	C	C	C	C	T	A	C	G	A	T	T
Mexico	A	A	C	C	C	T	T	C	C	T	C	C	C	C	T	T	C	G	A	T	T
United States	A	A	C	C	C	C	T	C	C	T	C	C	C	C	T	T	C	G	A	T	T
Uganda	A	T	A	C	G	A	T	C	A	T	G	C	T	C	G	T	C	C	A	T	C

one strand can act as a template for the production of the other strand (Figure 8.3a).

DNA replication requires the presence of several cellular proteins that direct a particular sequence of events. Enzymes involved in DNA replication and other processes are listed in Table 8.1 on page 220. When replication begins, the supercoiling is relaxed by *topoisomerase* or *gyrase* and the two strands of parental DNA are unwound by *helicase* and separated from each other in one small DNA segment after another. Free nucleotides present in the

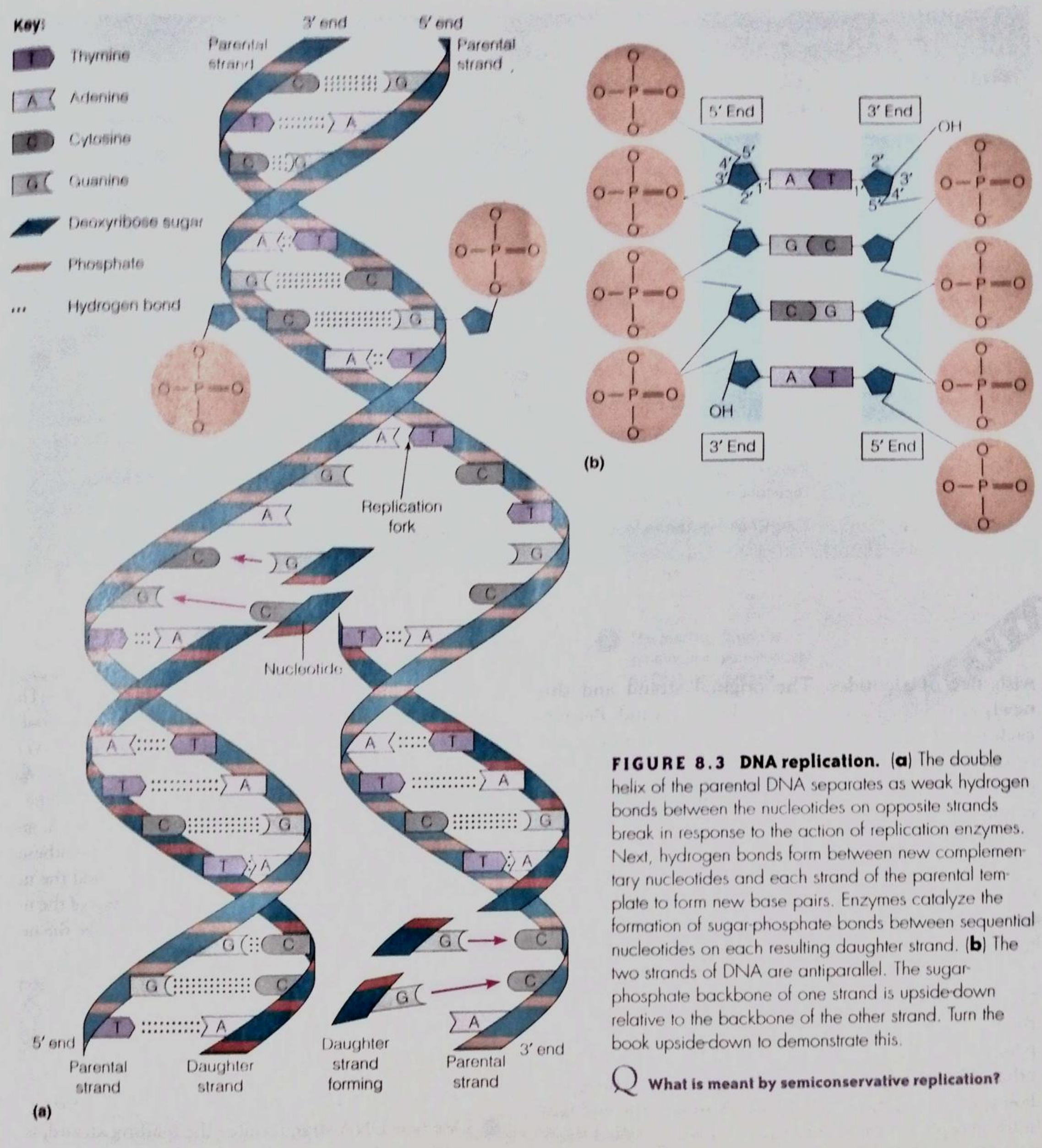


FIGURE 8.3 DNA replication. (**a**) The double helix of the parental DNA separates as weak hydrogen bonds between the nucleotides on opposite strands break in response to the action of replication enzymes. Next, hydrogen bonds form between new complementary nucleotides and each strand of the parental template to form new base pairs. Enzymes catalyze the formation of sugar-phosphate bonds between sequential nucleotides on each resulting daughter strand. (**b**) The two strands of DNA are antiparallel. The sugar-phosphate backbone of one strand is upside-down relative to the backbone of the other strand. Turn the book upside-down to demonstrate this.

Q **What is meant by semiconservative replication?**

cytoplasm of the cell are matched up to the exposed bases of the single-stranded parental DNA. Where thymine is present on the original strand, only adenine can fit into place on the new strand; where guanine is present on the original strand, only cytosine can fit into place, and so on. Any bases that are improperly base-paired are removed and replaced by replication enzymes. Once aligned, the newly added nucleotide is joined to the growing DNA strand by an enzyme called **DNA polymerase.** Then the parental DNA is unwound a bit further to allow the addition of the next nucleotides. The point at which replication occurs is called the *replication fork*.

As the replication fork moves along the parental DNA, each of the unwound single strands combines

TABLE 8.1 Important Enzymes in DNA Replication, Expression, and Repair

DNA gyrase	Relaxes supercoiling ahead of the replication fork.
DNA ligase	Makes covalent bonds to join DNA strands; joins Okazaki fragments and new segments in excision repair.
DNA polymerase	Synthesizes DNA; proofreads and repairs DNA.
Endonucleases	Cut DNA backbone in a strand of DNA; facilitate repair and insertions.
Exonucleases	Cut DNA from an exposed end of DNA; facilitate repair.
Helicase	Unwinds double-stranded DNA.
Methylase	Adds methyl group to selected bases in newly-made DNA.
Photolyases	Use visible light energy to separate UV-induced pyrimidine dimers.
Primase	Makes RNA primers from a DNA template.
Ribozyme	RNA enzyme that removes introns and splices exons together.
RNA polymerase	Copies RNA from a DNA template.
Topoisomerase	Relaxes supercoiling ahead of the replication fork; separates DNA circles at the end of DNA replication.
Transposase	Cuts DNA backbone leaving single-stranded "sticky ends."

with new nucleotides. The original strand and this newly synthesized daughter strand then rewind. Because each new double-stranded DNA molecule contains one original (conserved) strand and one new strand, the process of replication is referred to as **semiconservative replication.**

Before looking at DNA replication in more detail, let's take a closer look at the structure of DNA (see Figure 2.16, on page 48). It is important to understand the concept that the paired DNA strands are oriented in opposite directions relative to each other. Notice in Figure 2.16 that the carbon atoms of the sugar component of each nucleotide are numbered 1′ (pronounced "one prime") to 5′. In order for the paired bases to be next to each other, the sugar components in one strand are upside-down relative to the other. The end with the hydroxyl attached to the 3′ carbon is called the 3′ end of the DNA strand; the end having a phosphate attached to the 5′ carbon is called the 5′ end. The way in which the two strands fit together dictates that the 5′ → 3′ direction of one strand runs counter to the 5′ → 3′ direction of the other strand (see Figure 8.3b). This structure of DNA affects the replication process because DNA polymerases can add new nucleotides to the 3′ end only. Therefore, as the replication fork moves along the parental DNA, the two new strands must grow in different directions.

DNA replication requires a great deal of energy. The energy is supplied from the nucleotides, which are actually nucleoside triphosphates. You already know about ATP; the only difference between ATP and the adenine nucleotide in DNA is the sugar component. Deoxyribose is the sugar in the nucleosides used to synthesize DNA, and nucleoside triphosphates with ribose are used to synthesize RNA. Two phosphate groups are removed to add the nucleotide to a growing strand of DNA; hydrolysis of the nucleoside is exergonic and provides energy to make the new bonds in the DNA strand (Figure 8.4).

Figure 8.5 provides more detail about the many steps that go into this complex process.

❶–❷ Once the parental DNA is unwound and stabilized, the replication fork forms at a fixed site called the origin of replication.

❸ One new DNA strand, called the **leading strand,** is synthesized continuously as the DNA polymerase moves toward the replication fork making DNA in the 5′ → 3′ direction.

❹ Remember that DNA polymerase can only add new nucleotides to the 3′ end, so a short piece of RNA called an **RNA primer,** made by *primase*, starts synthesis. DNA polymerase can then add nucleotides to the 3′ end of the RNA. Consequently, the **lagging strand** of

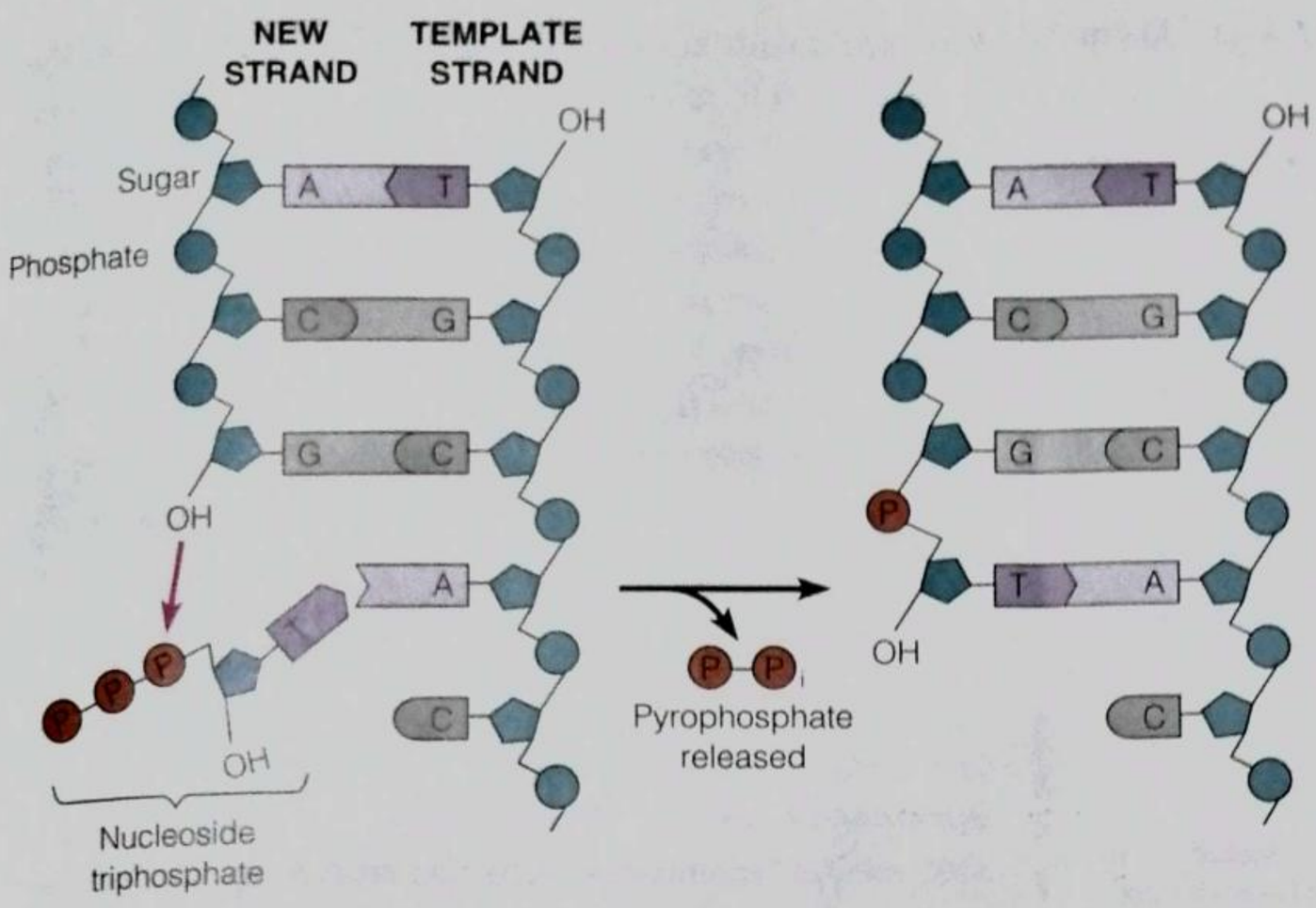


FIGURE 8.4 Adding a nucleotide to DNA. When a nucleoside triphosphate bonds to the sugar in a growing DNA strand, it loses two phosphates. Hydrolysis of the phosphate bonds provides the energy for the reaction.

Q **Why is one strand "upside-down" relative to the other strand? Why can't both strands "face" the same way?**

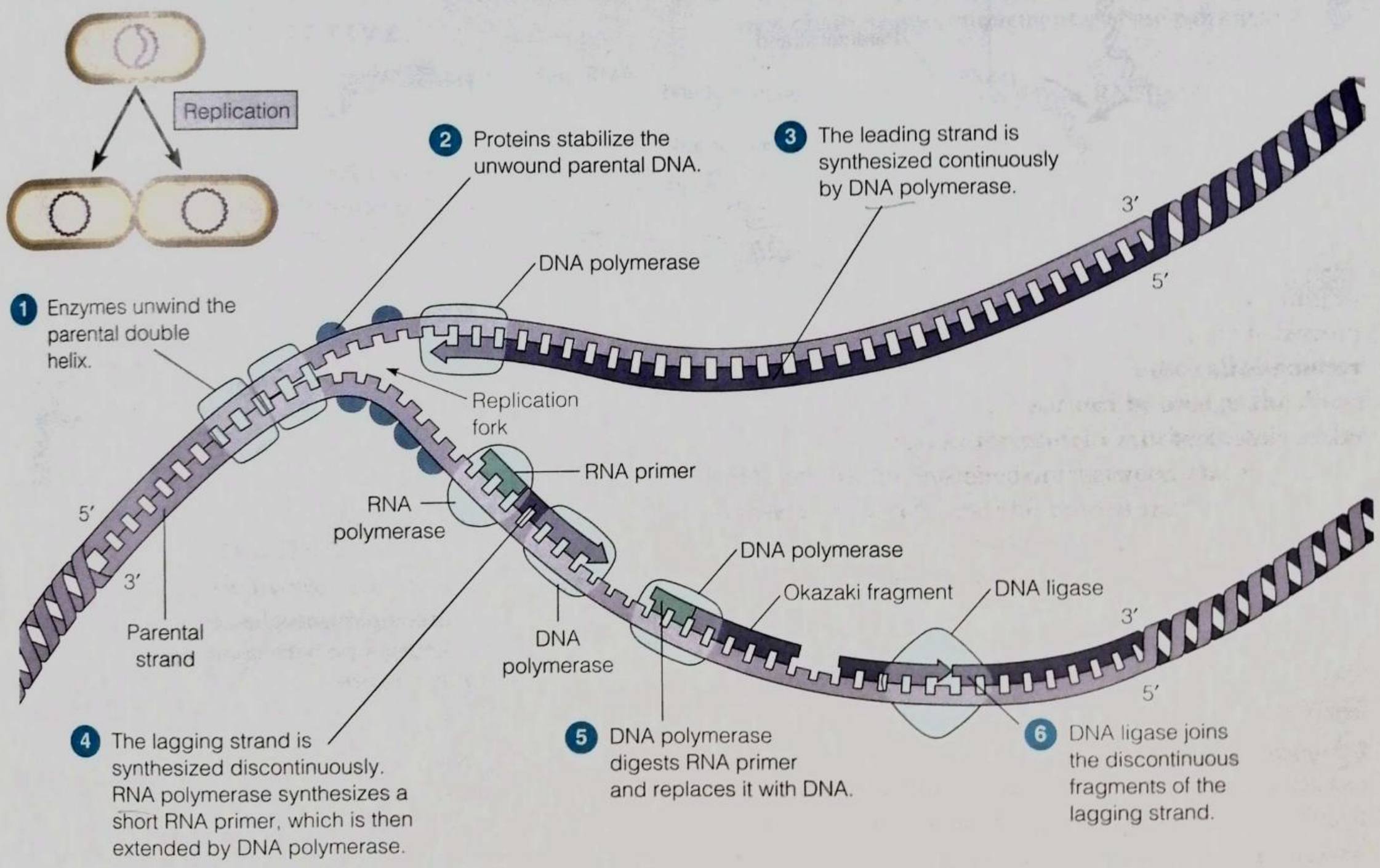


FIGURE 8.5 A summary of events at the DNA replication fork. Enzymes at the replication fork unwind the parental double helix. DNA polymerase synthesizes a continuous strand of new DNA, using one of the parental strands as a template. DNA polymerase also uses the other parental strand as a template, but because the orientation of the sugars is opposite, RNA polymerase starts the synthesis by adding a short stretch of RNA called an RNA primer. The DNA polymerase digests away the RNA as it makes a small piece of DNA. The small units are subsequently joined by DNA ligase.

Q **Why is one strand of DNA synthesized discontinuously?**

(a)

(b)

FIGURE 8.6 Replication of bacterial DNA. (**a**) An *E. coli* chromosome in the process of replicating. (In the corresponding diagram at right, the arrows show the direction in which the replication forks are moving.) The chromosome is about one-third replicated. Notice that one of the new helices is perpendicular to the other. (**b**) A diagram of the bidirectional replication of a circular bacterial DNA molecule.

Q What is the origin of replication?

new DNA is synthesized in pieces consisting of about 1000 nucleotides, called *Okazaki fragments*, as the DNA polymerase moves away from the replication fork.

5. DNA polymerase removes the RNA primer, and
6. the enzyme **DNA ligase** joins the newly made DNA fragments.

DNA replication by some bacteria, such as *E. coli*, goes *bidirectionally* around the chromosome (Figure 8.6). Two replication forks move in opposite directions away from the origin of replication. Because the bacterial chromosome is a closed loop, the replication forks eventually meet when replication is completed. The two loops must be separated by a topoisomerase. Much evidence shows an

association between the bacterial plasma membrane and the origin of replication. After duplication, if each copy of the origin binds to the membrane at opposite poles, then each daughter cell would receive one copy of the DNA molecule—that is, one complete chromosome.

DNA replication is an amazingly accurate process. Typically, mistakes are made at a rate of only 1 in every 10^{10} bases incorporated. Such accuracy is largely due to the *proofreading* capability of DNA polymerase. As each new base is added, the enzyme evaluates whether it forms the proper complementary base-pairing structure. If not, the enzyme excises the improper base and replaces it with the correct one. In this way, DNA replication can be performed very accurately, allowing each daughter chromosome to be virtually identical to the parental DNA. **Animation: Go to The Microbiology Place website or CD-ROM and click "Animations" to view DNA Replication.**

RNA AND PROTEIN SYNTHESIS

LEARNING OBJECTIVE

- Describe protein synthesis, including transcription, RNA processing, and translation.

How is the information in DNA used to make the proteins that control cell activities? In the process of *transcription*, genetic information in DNA is copied, or transcribed, into a complementary base sequence of RNA. The cell then uses the information encoded in this RNA to synthesize specific proteins through the process of *translation*. We now take a closer look at these two processes as they occur in a bacterial cell.

TRANSCRIPTION

Transcription is the synthesis of a complementary strand of RNA from a DNA template. We will discuss transcription in prokaryotic cells here. Transcription in eukaryotes is discussed on page 226. As mentioned earlier, there are three kinds of RNA in bacterial cells: messenger RNA, ribosomal RNA, and transfer RNA. Ribosomal RNA forms an integral part of ribosomes, the cellular machinery for protein synthesis. Transfer RNA is also involved in protein synthesis, as we will see. **Messenger RNA (mRNA)** carries the coded information for making specific proteins from DNA to ribosomes, where proteins are synthesized.

During transcription, a strand of mRNA is synthesized using a specific gene—a portion of the cell's DNA—as a template. In other words, the genetic information stored in the sequence of nitrogenous bases of DNA is rewritten so that the same information appears in the base sequence of mRNA. As in DNA replication, a G in the DNA template dictates a C in the mRNA being made, a C in the DNA template dictates a G in the mRNA, and a T in the DNA template dictates an A in the mRNA. However, an A in the DNA template dictates a uracil (U) in the mRNA because RNA contains U instead of T. (U has a chemical structure slightly different from T, but it base-pairs in the same way.) If, for example, the template portion of DNA has the base sequence 3′-ATGCAT, the newly synthesized mRNA strand will have the complementary base sequence 5′-UACGUA.

The process of transcription requires both an enzyme called *RNA polymerase* and a supply of RNA nucleotides (Figure 8.7). Transcription begins when

1. RNA polymerase binds to the DNA at a site called the **promoter.** Only one of the two DNA strands serves as the template for RNA synthesis for a given gene. Like DNA, RNA is synthesized in the 5′ → 3′ direction.
2. RNA polymerase assembles free nucleotides into a new chain, using complementary base pairing as a guide.
3. As the new RNA chain grows, RNA polymerase moves along the DNA.
4. RNA synthesis continues until RNA polymerase reaches a site on the DNA called the **terminator.**
5. When this happens, RNA polymerase and the newly formed, single-stranded mRNA are released from the DNA.

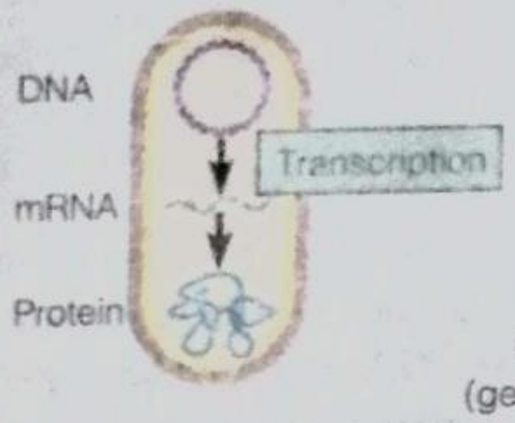
The process of transcription allows the cell to produce short-term copies of genes that can be used as the direct source of information for protein synthesis. Messenger RNA acts as an intermediate between the permanent storage form, DNA, and the process that uses the information, translation. **Animation: Go to The Microbiology Place website or CD-ROM and click "Animations" to view Transcription.**

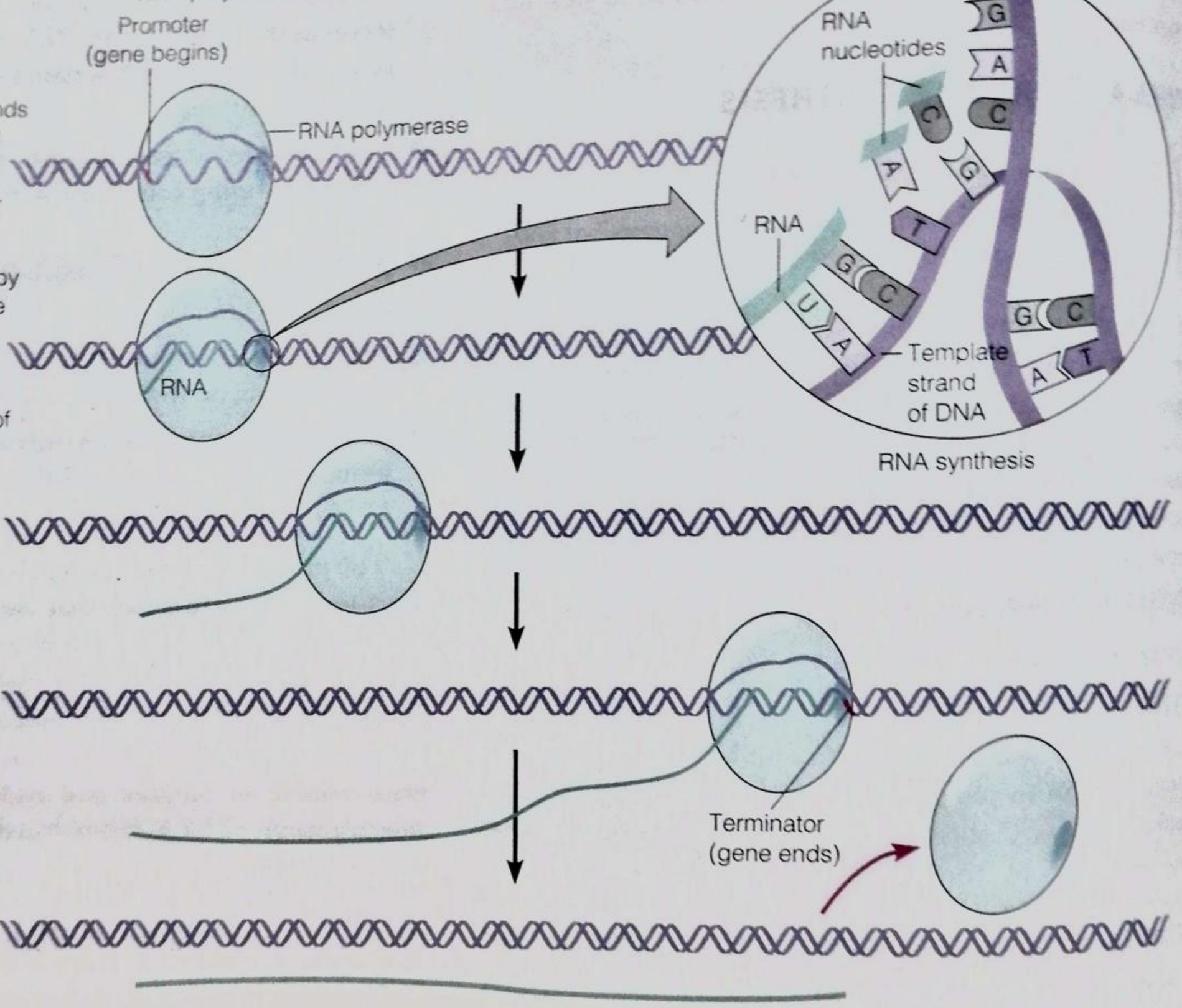
TRANSLATION

We have seen how the genetic information in DNA is transferred to mRNA during transcription. Now we will see how mRNA serves as the source of information for the synthesis of proteins. Protein synthesis is called **translation** because it involves decoding the "language" of nucleic acids and converting that information into the "language" of proteins.

The language of mRNA is in the form of **codons,** groups of three nucleotides, such as AUG, GGC, or AAA. The sequence of codons on an mRNA molecule determines the sequence of amino acids that will be in the protein being synthesized. Each codon "codes" for a

RNA polymerase bound to DNA





1. RNA polymerase binds to the promoter, and DNA unwinds at the beginning of a gene.
2. RNA is synthesized by complementary base pairing of free nucleotides with the nucleotide bases on the template strand of DNA.
3. The site of synthesis moves along DNA; DNA that has been transcribed rewinds.
4. Transcription reaches the terminator.
5. RNA and RNA polymerase are released and the DNA helix re-forms.

FIGURE 8.7 The process of transcription. The orienting diagram indicates the relationship of transcription to the overall flow of genetic information within a cell.

Q **In transcription, what is copied and what is made?**

particular amino acid. This is the genetic code (Figure 8.8).

Codons are written in terms of their base sequence in mRNA. Notice that there are 64 possible codons but only 20 amino acids. This means that most amino acids are signaled by several alternative codons, a situation referred to as the **degeneracy** of the code. For example, leucine has six codons and alanine has four codons.

Degeneracy allows for a certain amount of change, or mutation, in the DNA without affecting the protein ultimately produced.

Of the 64 codons, 61 are sense codons and 3 are nonsense codons. **Sense codons** code for amino acids, and **nonsense codons** (also called *stop codons*) do not. Rather, the nonsense codons—UAA, UAG, and UGA—signal the end of the protein molecule's synthesis. The start codon that initiates the synthesis of the protein molecule is AUG, which is also the codon for methionine. In bacteria, the start AUG codes for formylmethionine rather than the methionine found in other parts of the protein. The initiating methionine is often removed later, so not all proteins begin with methionine.

The codons of mRNA are converted into protein through the process of translation. The codons of an mRNA are "read" sequentially; and, in response to each codon, the appropriate amino acid is assembled into a growing chain. The site of translation is the ribosome, and **transfer RNA (tRNA)** molecules both recognize the specific codons and transport the required amino acids.

Each tRNA molecule has an **anticodon,** a sequence of three bases that is complementary to a codon. In this way, a tRNA molecule can base-pair with its associated codon. Each tRNA can also carry on its other end the amino acid coded for by the codon that the tRNA recognizes. The functions of the ribosome are to direct the orderly binding of tRNAs to codons and to assemble the amino acids brought there into a chain, ultimately producing a protein.

Figure 8.9 on pages 226–227 shows the details of translation.

1. The necessary components assemble: the two ribosomal subunits, a tRNA with the anticodon UAC, and the mRNA molecule to be translated, along with several additional protein factors. This sets up the initiator codon (AUG) in the proper position to allow translation to begin.
2. The first tRNA binds to the start codon, bringing with it the amino acid methionine.
3. When the tRNA that recognizes the second codon moves into position on the ribosome, the first amino acid is transferred by the ribosome.
4. After the ribosome joins the two amino acids with a peptide bond, the first tRNA molecule leaves the ribosome.
5. The ribosome then moves along the mRNA to the next codon.
6. As the proper amino acids are brought into line one by one, peptide bonds are formed between them, and a polypeptide chain results. (See Figure 2.14, page 45.)

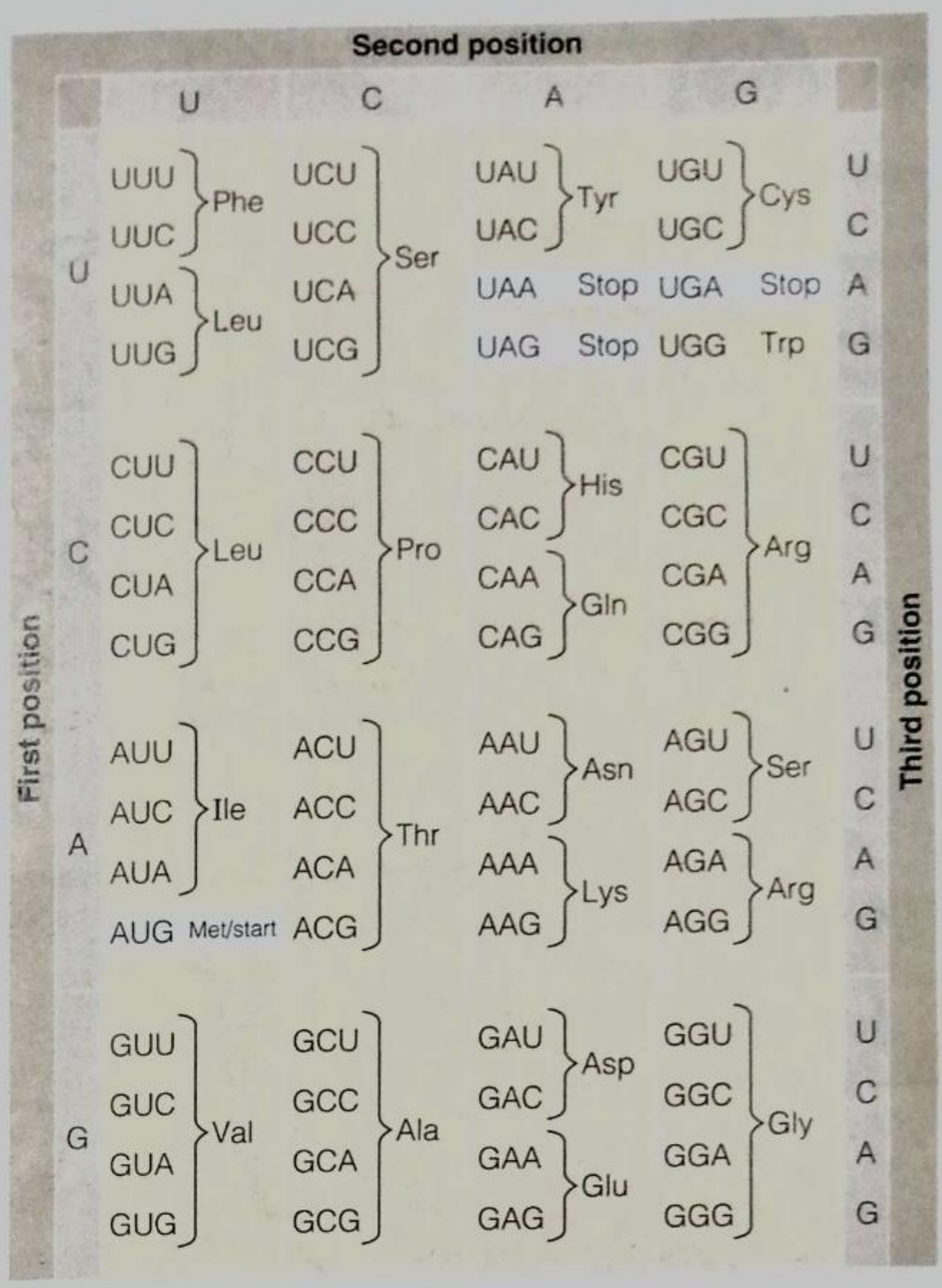
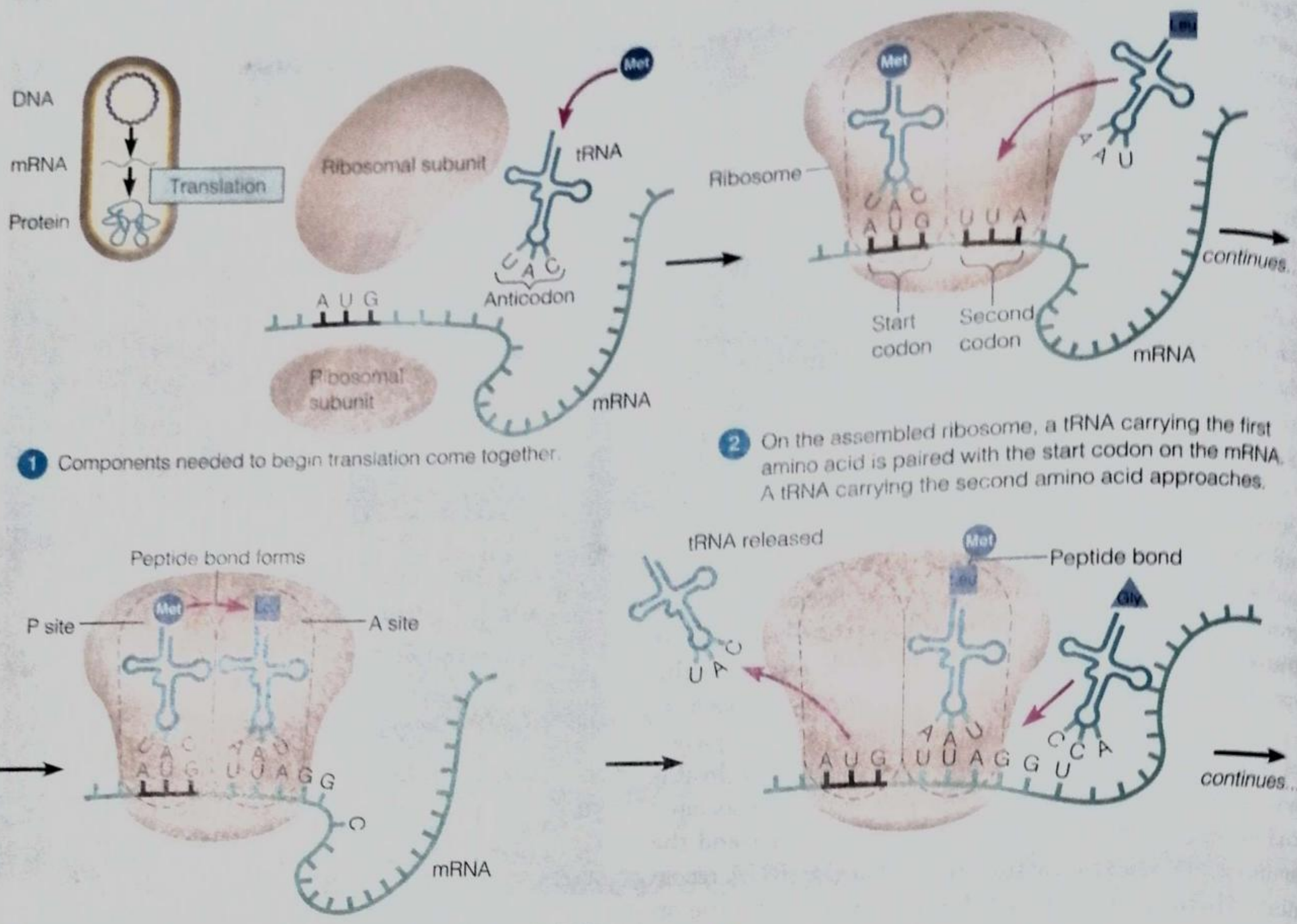


FIGURE 8.8 The genetic code. The three nucleotides in an mRNA codon are designated, respectively, as the first position, second position, and third position of the codon on the mRNA. Each set of three nucleotides specifies a particular amino acid, represented by a three-letter abbreviation (see Table 2.4, page 44). The codon AUG, which specifies the amino acid methionine, is also the start of protein synthesis. The word *Stop* identifies the nonsense codons that signal the termination of protein synthesis.

Q **Why is the genetic code described as degenerate?**

7. Translation ends when one of the three nonsense codons in the mRNA is reached.
8. When the ribosome arrives at this codon, it comes apart into its two subunits, and the mRNA and newly synthesized polypeptide chain are released. The ribosome, the mRNA, and the tRNAs are then available to be used again.

The ribosome moves along the mRNA in the 5′ → 3′ direction. As a ribosome moves along the mRNA, it will soon allow the start codon to be exposed. Additional ribosomes can then assemble and begin synthesizing protein. In this way, there are usually a number of ribosomes attached to a single mRNA, all at various stages of protein synthesis. In prokaryotic cells, the translation of mRNA into protein can begin even before transcription is complete



1 Components needed to begin translation come together.

2 On the assembled ribosome, a tRNA carrying the first amino acid is paired with the start codon on the mRNA. A tRNA carrying the second amino acid approaches.

3 The place on the ribosome where the first tRNA sits is called the P site. In the A site next to it, the second codon of the mRNA pairs with a tRNA carrying the second amino acid.

4 The first amino acid joins to the second by a peptide bond, and the first tRNA is released.

FIGURE 8.9 The process of translation. The overall goal of translation is to produce proteins using mRNAs as the source of biological information. The complex cycle of events illustrated here shows the primary role of tRNA and ribosomes in the decoding of this information. The ribosome acts as the site where the mRNA-encoded information is decoded, as well as the site where individual amino acids are connected into polypeptide chains. The tRNA molecules act as the actual "translators"—one end of each tRNA recognizes a specific mRNA codon, while the other end carries the amino acid coded for by that codon.

Q **Why is this process called translation?.**

(Figure 8.10). Because mRNA is produced in the cytoplasm, the start codons of an mRNA being transcribed are available to ribosomes before the entire mRNA molecule is even made.

In eukaryotic cells, transcription takes place in the nucleus. The mRNA must be completely synthesized and moved through the nuclear membrane to the cytoplasm before translation can begin. In addition, the RNA undergoes processing before it leaves the nucleus. In eukaryotic cells the regions of genes that code for proteins are often interrupted by noncoding DNA. Thus, eukaryotic genes are composed of **exons,** the regions of DNA *expressed*, and **introns,** the *intervening* regions of DNA that do not encode protein (Figure 8.11).

1. In the nucleus of a eukaryotic cell, RNA polymerase synthesizes an RNA molecule containing exons and introns called an *RNA transcript*.
2. This long RNA is then processed by ribozymes, which remove the intron-derived RNA and splice together the exon-derived RNA, producing an mRNA.
3. The resulting mRNA leaves the nucleus to be used by rRNA and tRNA for protein synthesis.

* * *

To summarize, genes are the units of biological information encoded by the sequence of nucleotide bases in DNA. A

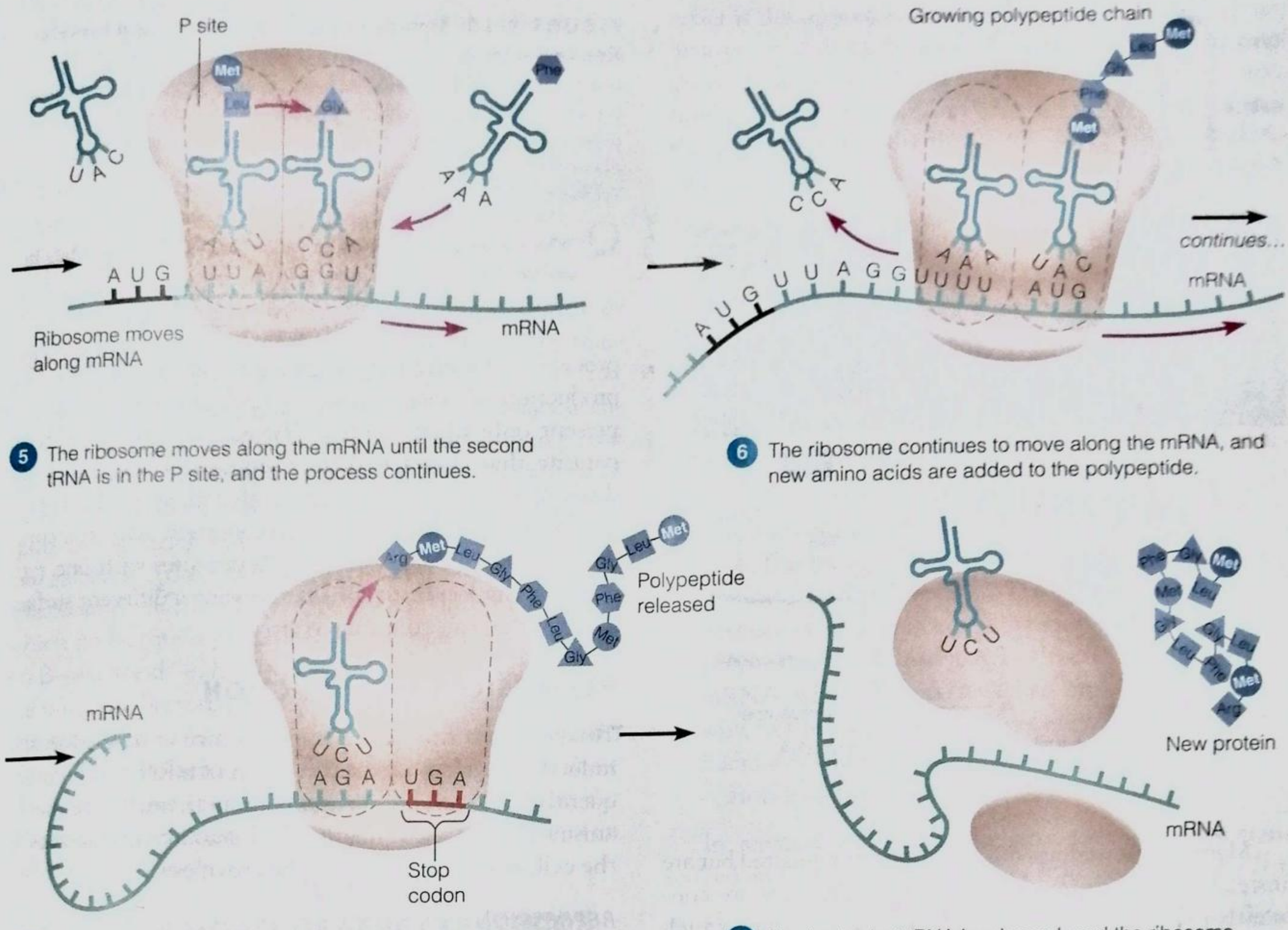


FIGURE 8.9 The process of translation. (continued)

gene is expressed, or turned into a product within the cell, through the processes of transcription and translation. The genetic information carried in DNA is transferred to a temporary mRNA molecule by transcription. Then, during translation, the mRNA directs the assembly of amino acids into a polypeptide chain: mRNA attaches to a ribosome, tRNAs deliver the amino acids to the ribosome as directed by the mRNA codon sequence, and the ribosome assembles the amino acids into the chain that will be the newly synthesized protein. ✳ **Animation: Go to The Microbiology Place website or CD-ROM and click "Animations" to view Translation.**

THE REGULATION OF BACTERIAL GENE EXPRESSION

LEARNING OBJECTIVE

- Explain the regulation of gene expression in bacteria by induction, repression, and catabolite repression.

A cell's genetic machinery and its metabolic machinery are integrated and interdependent. Recall from Chapter 5 that the bacterial cell carries out an enormous number of metabolic reactions. The common feature of all metabolic reactions is that they are catalyzed by enzymes. Also recall from Chapter 5 (page 122) that feedback inhibition stops a cell from performing unneeded chemical reactions. Feedback inhibition stops enzymes that have already been synthesized. We will now look at mechanisms to prevent synthesis of enzymes that are not needed.

We have seen that genes, through transcription and translation, direct the synthesis of proteins, many of which serve as enzymes—the very enzymes used for cellular metabolism. Because protein synthesis requires a tremendous expenditure of energy, the regulation of protein synthesis is important to the cell's energy economy. The cell conserves energy by making only those proteins needed at a particular time. We will next look at how chemical reactions are regulated by controlling the synthesis of the enzymes.

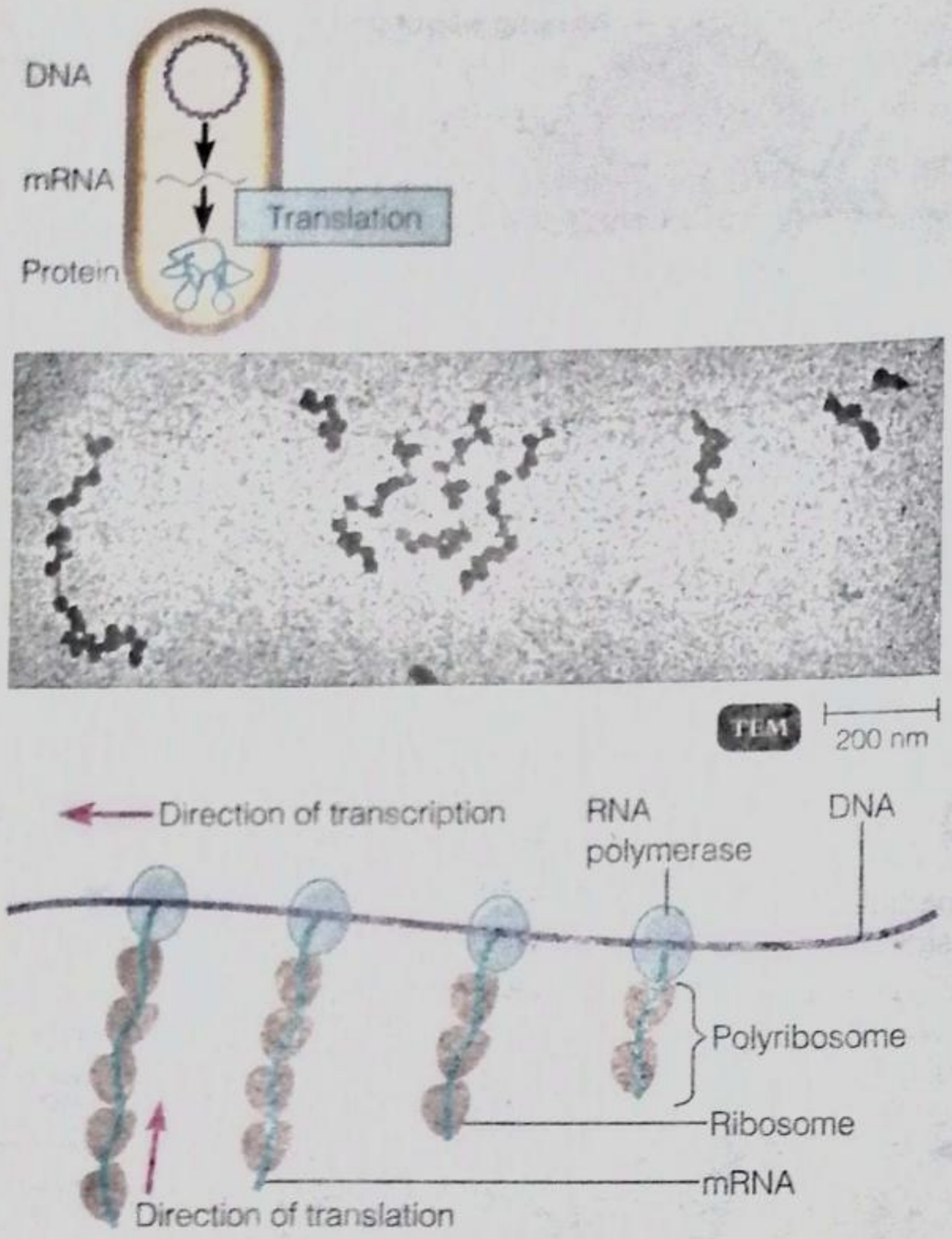


FIGURE 8.10 Simultaneous transcription and translation in bacteria. The micrograph and diagram show these processes in a single bacterial gene. Many molecules of mRNA are being synthesized simultaneously. The longest mRNA molecules were the first to be transcribed at the promoter. Notice the ribosomes attached to the newly forming mRNA. The newly synthesized polypeptides are not shown.

Q Why can translation begin before transcription is complete in prokaryotes but not in eukaryotes?

Many genes, perhaps 60–80%, are not regulated but are instead *constitutive*, meaning that their products are constantly produced at a fixed rate. Usually these genes, which are effectively turned on all the time, code for enzymes that the cell needs in fairly large amounts for its major life processes; the enzymes of glycolysis are examples. The production of other enzymes is regulated so that they are present only when needed. *Trypanosoma*, the protozoan parasite that causes African sleeping sickness, has hundreds of genes coding for surface glycoproteins. Each protozoan cell turns on only one glycoprotein gene at a time. As the host's immune system kills parasites with one type of surface molecule, parasites expressing a different surface glycoprotein can continue to grow.

REPRESSION AND INDUCTION

Two genetic control mechanisms known as repression and induction regulate the transcription of mRNA and consequently the synthesis of enzymes from them. These mechanisms control the formation and amounts of enzymes in the cell, not the activities of the enzymes.

REPRESSION

The regulatory mechanism that inhibits gene expression and decreases the synthesis of enzymes is called **repression.**

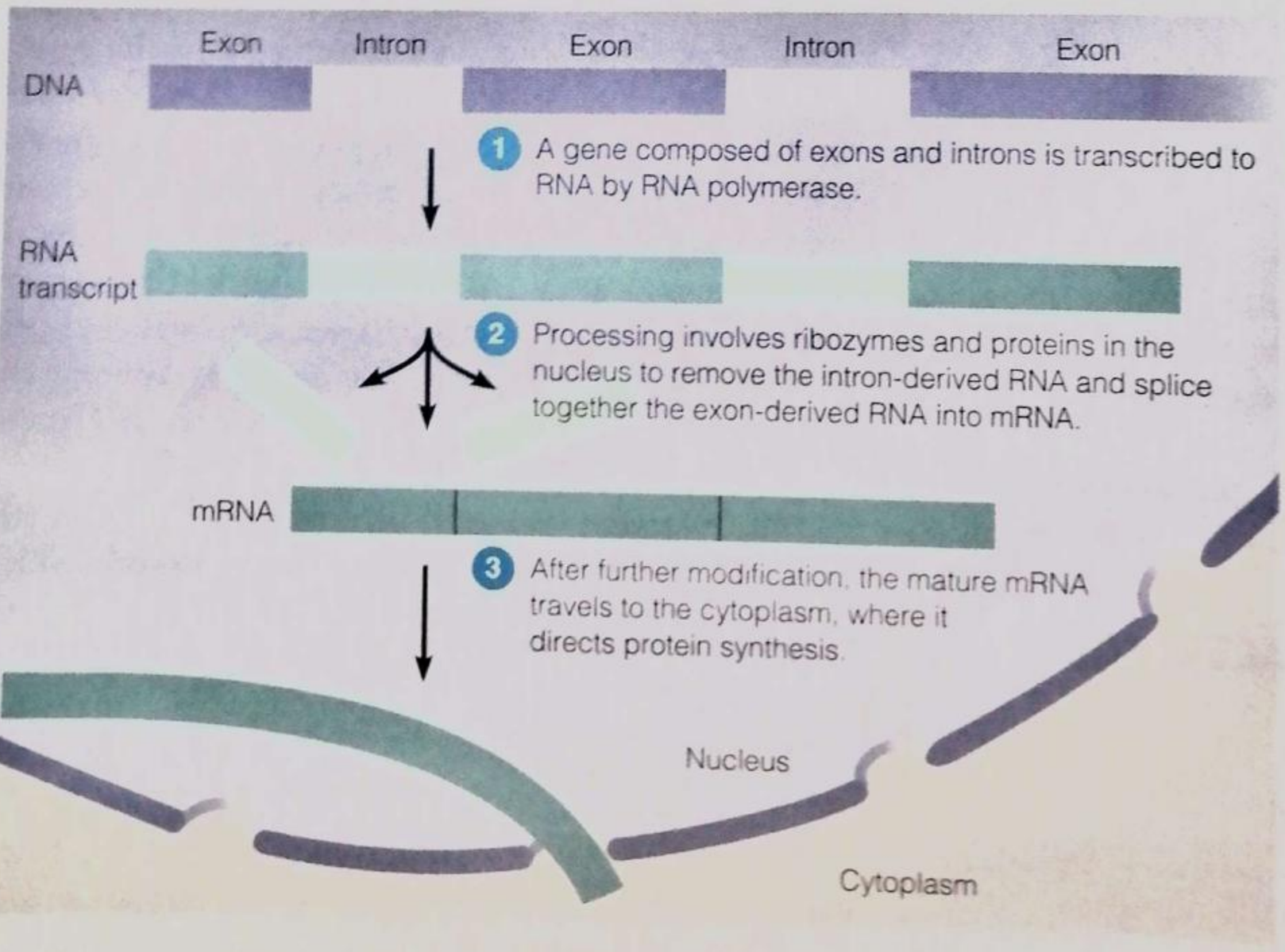


FIGURE 8.11 RNA processing in eukaryotic cells.

Q Why can't the RNA transcript be used for translation?

Repression is usually a response to the overabundance of an end-product of a metabolic pathway; it causes a decrease in the rate of synthesis of the enzymes leading to the formation of that product. Repression is mediated by regulatory proteins called **repressors,** which block the ability of RNA polymerase to initiate transcription from the repressed genes. The default position of a repressible gene is *on*.

INDUCTION

The process that turns on the transcription of a gene or genes is **induction.** A substance that acts to induce transcription of a gene is called an **inducer,** and enzymes that are synthesized in the presence of inducers are *inducible enzymes*. The genes required for lactose metabolism in *E. coli* are a well-known example of an inducible system. One of these genes codes for the enzyme β-galactosidase, which splits the substrate lactose into two simple sugars, glucose and galactose. (β refers to the type of linkage that joins the glucose and galactose.) If *E. coli* is placed into a medium in which no lactose is present, the organisms contain almost no β-galactosidase; however, when lactose is added to the medium, the bacterial cells produce a large quantity of the enzyme. Lactose is converted in the cell to the related compound allolactose, which is the inducer for these genes; the presence of lactose thus indirectly induces the cells to synthesize more enzyme. The default position of an inducible gene is *off*.

THE OPERON MODEL OF GENE EXPRESSION

Details of the control of gene expression by induction and repression are described by the operon model. François Jacob and Jacques Monod formulated this general model in 1961 to account for the regulation of protein synthesis. They based their model on studies of the induction of the enzymes of lactose catabolism in *E. coli*. In addition to β-galactosidase, these enzymes include lac permease, which is involved in the transport of lactose into the cell, and transacetylase, which metabolizes certain disaccharides other than lactose.

The genes for the three enzymes involved in lactose uptake and utilization are next to each other on the bacterial chromosome and are regulated together (Figure 8.12a). These genes, which determine the structures of proteins, are called **structural genes** to distinguish them from an adjoining control region on the DNA. When lactose is introduced into the culture medium, the *lac* structural genes are all transcribed and translated rapidly and simultaneously. We will now see how this regulation occurs.

In the control region of the *lac* operon are two relatively short segments of DNA. One, the *promoter*, is the region of DNA where RNA polymerase initiates transcription. The other is the **operator,** which is like a traffic light that acts as a go or stop signal for transcription of the structural genes. A set of operator and promoter sites, and the structural genes they control, are what define an **operon;** thus, the combination of the three *lac* structural genes and the adjoining control regions is called the *lac* operon.

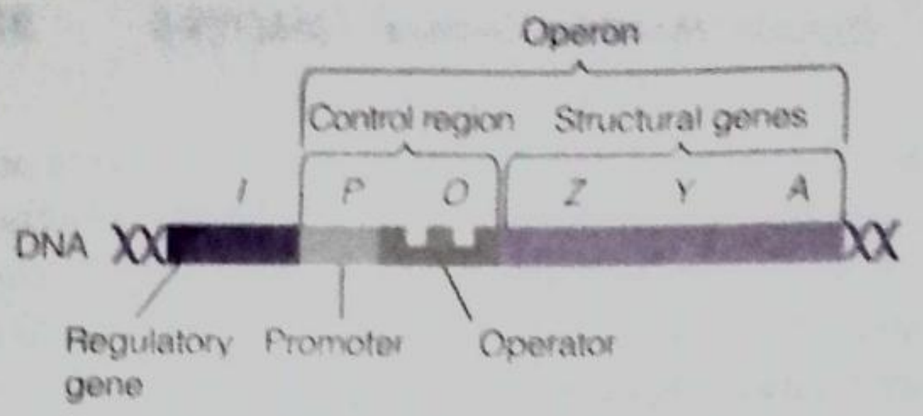
1. Near the *lac* operon on the bacterial DNA is a regulatory gene called the *I* gene, which codes for a repressor protein.
2. When lactose is absent, the repressor protein binds tightly to the operator site. This binding prevents RNA polymerase from transcribing the adjacent structural genes; consequently, no mRNA is made and no enzymes are synthesized.
3. But when lactose is present, some of it is transported into the cells and converted into the inducer allolactose. The inducer binds to the repressor protein and alters it so it cannot bind to the operator site. In the absence of an operator-bound repressor protein, RNA polymerase can transcribe the structural genes into mRNA, which is then translated into enzymes. This is why, in the presence of lactose, enzymes are produced. Lactose is said to induce enzyme synthesis, and the *lac* operon is called an inducible operon.

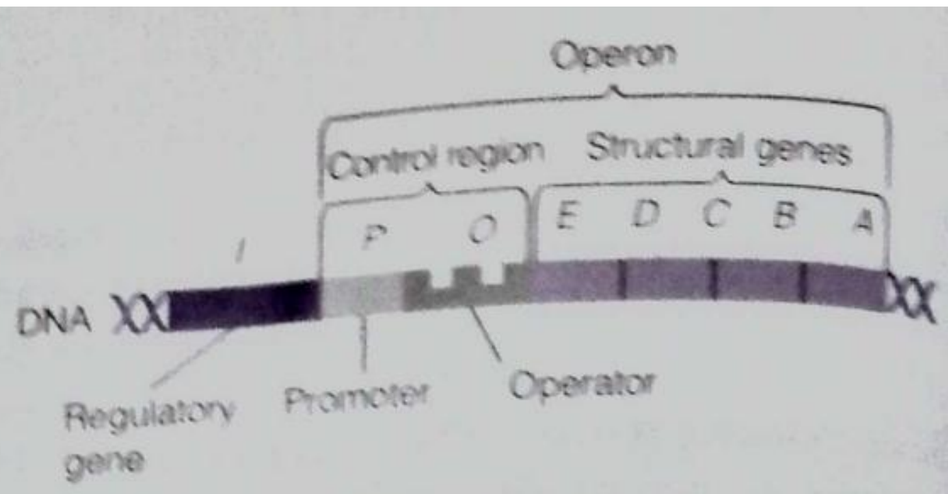
In repressible operons, the structural genes are transcribed until they are turned off, or *repressed* (Figure 8.12b).

1. The genes for the enzymes involved in the synthesis of tryptophan are regulated in this manner.
2. The structural genes are transcribed and translated, leading to tryptophan synthesis.
3. When excess tryptophan is present, the tryptophan acts as a **corepressor** binding to the repressor protein. The repressor protein can now bind to the operator, stopping further tryptophan synthesis.

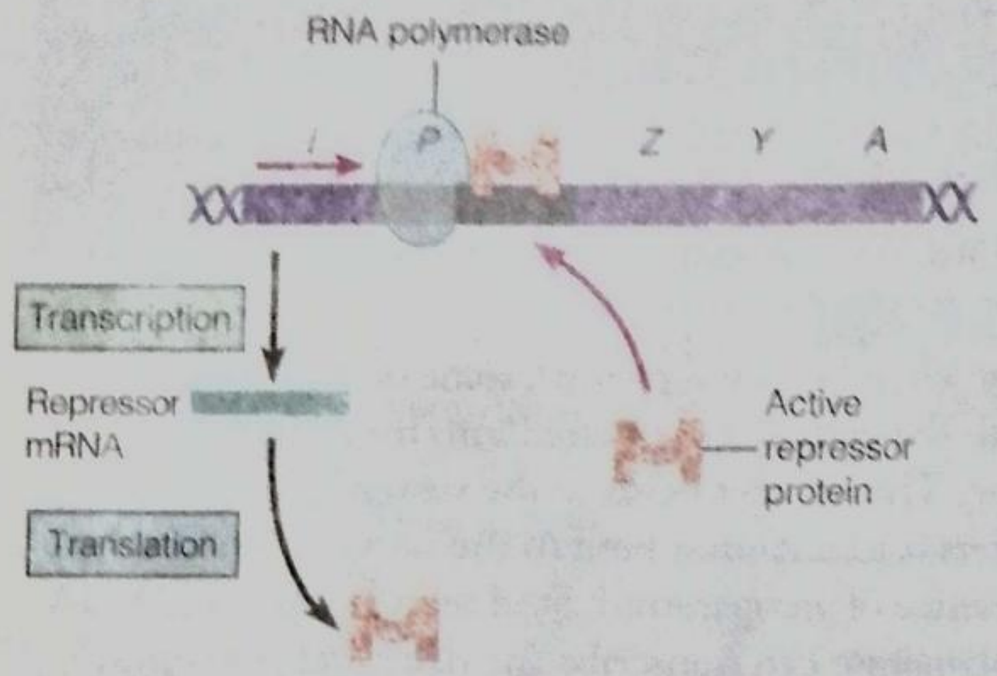
POSITIVE REGULATION

Regulation of the lactose operon also depends on the level of glucose in the medium, which in turn controls the intracellular level of the small molecule **cyclic AMP (cAMP),** a substance derived from ATP that serves as a cellular alarm signal. Enzymes that metabolize glucose are constitutive, and cells grow at their maximal rate with glucose as their carbon source because they can use it most efficiently (Figure 8.13). When glucose is no longer available, cAMP accumulates in the cell. The cAMP binds to the allosteric site of *catabolic activator protein* (CAP). CAP then binds to the *lac* promoter, which initiates transcription by making it easier for RNA polymerase to bind to the promoter. Thus transcription of the *lac*

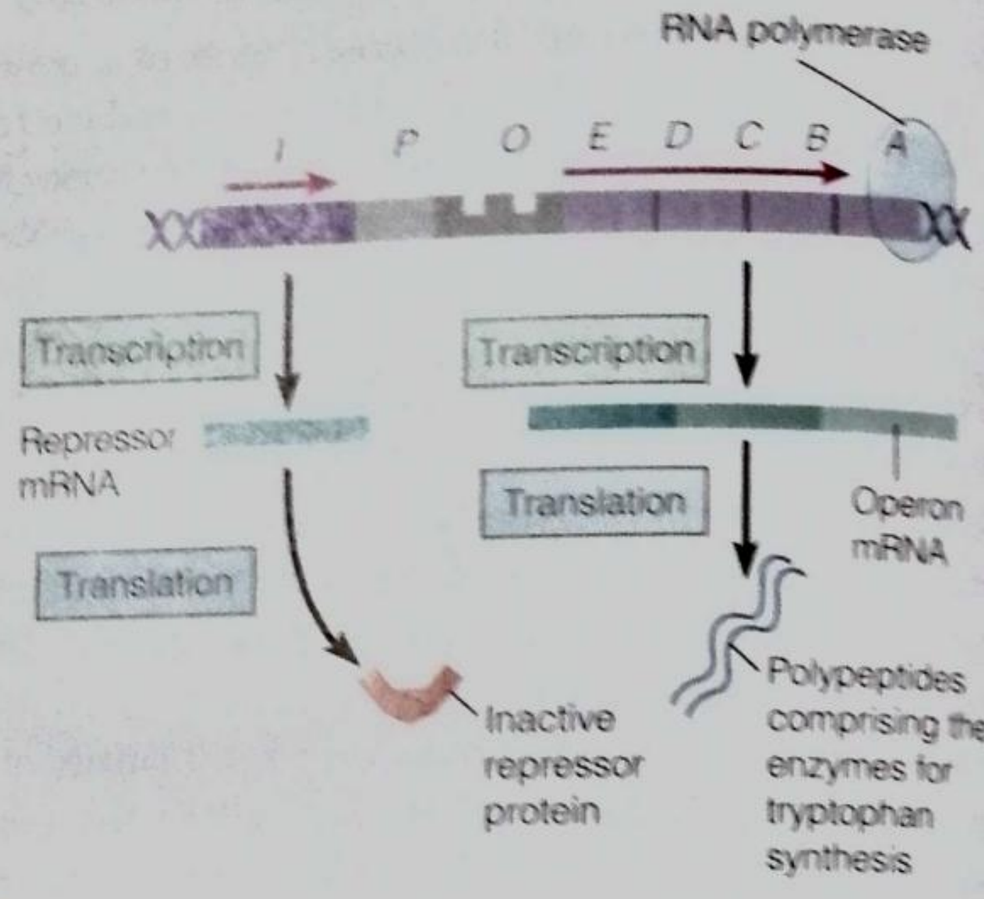




1 Structure of the operon. The operon consists of the promoter (*P*) and operator (*O*) sites and structural genes that code for the protein. The operon is regulated by the product of the regulatory gene (*I*).



2 Repressor active, operon off. The repressor protein binds with the operator, preventing transcription from the operon.



2 Repressor inactive, operon on. The repressor is inactive, and transcription and translation proceed, leading to the synthesis of tryptophan.

3 Repressor inactive, operon on. When the inducer allolactose binds to the repressor protein, the inactivated repressor can no longer block transcription. The structural genes are transcribed, ultimately resulting in the production of the enzymes needed for lactose catabolism.

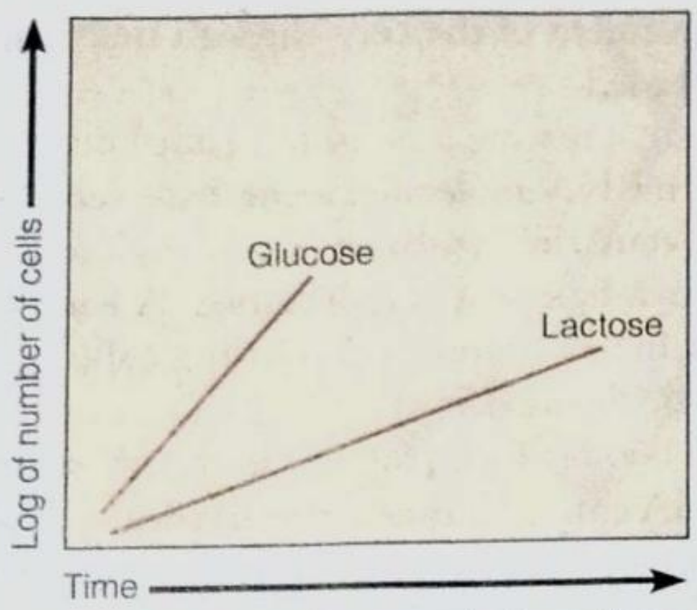
(a) An inducible operon

3 Repressor active, operon off. When the corepressor tryptophan binds to the repressor protein, the activated repressor binds with the operator, preventing transcription from the operon.

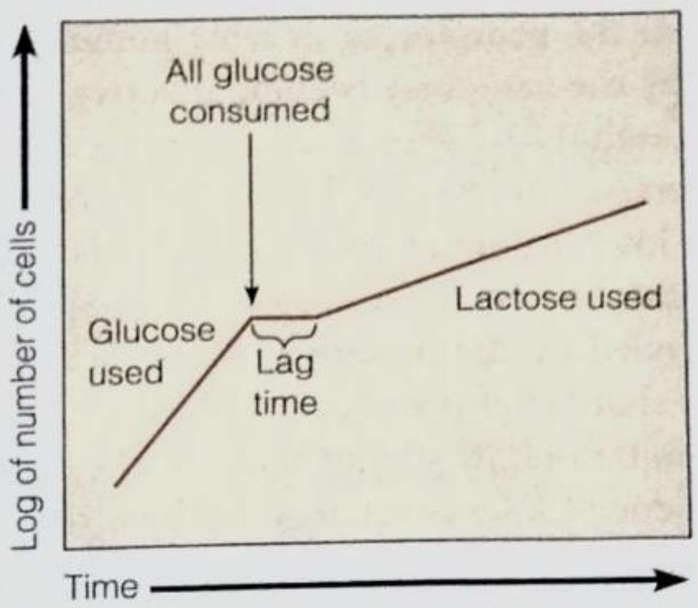
(b) A repressible operon

FIGURE 8.12 The operon: regulation of gene expression. (**a**) Lactose is digested by a catabolic pathway catalyzed by inducible enzymes. (**b**) Tryptophan is an amino acid produced by an anabolic pathway catalyzed by repressible enzymes.

How does a repressible enzyme differ from an inducible enzyme?



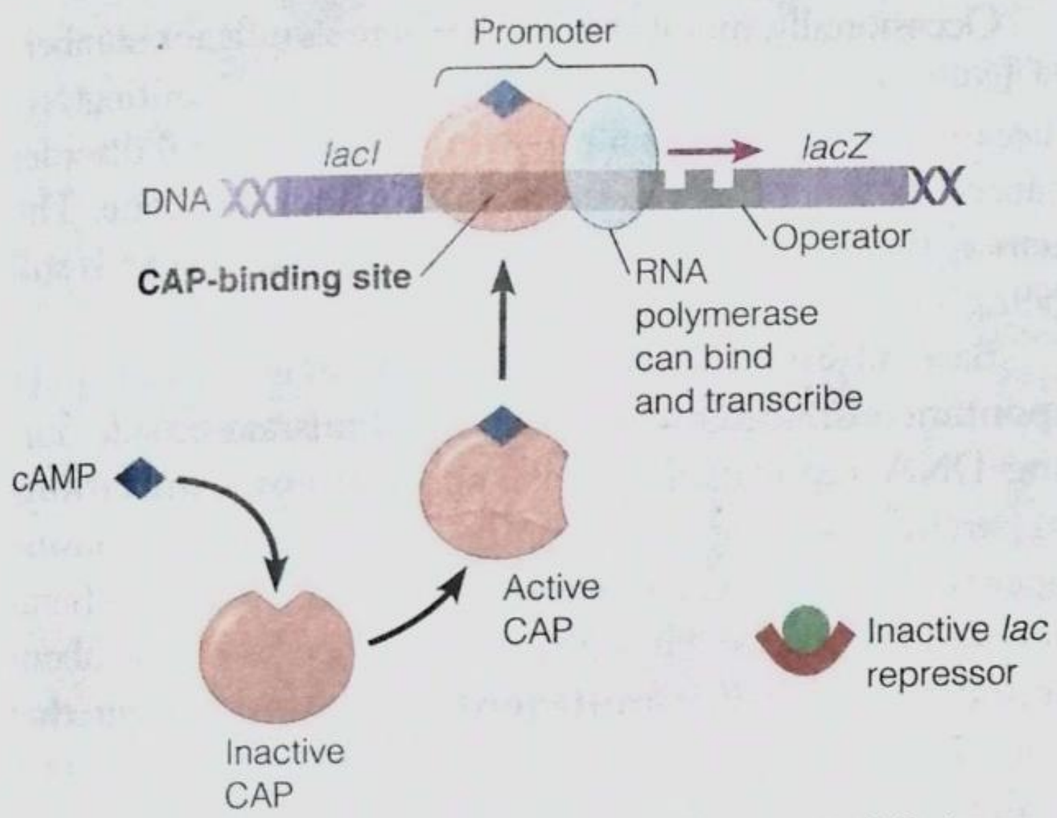
(a) Growth on glucose or lactose alone



(b) Growth on glucose and lactose combined

FIGURE 8.13 The growth rate of *E. coli* on glucose and lactose. The steeper the straight line, the faster the growth. (**a**) Bacteria growing on glucose as the sole carbon source grow faster than on lactose. (**b**) Bacteria growing in a medium containing glucose and lactose first consume the glucose, and then, after a short lag time, the lactose. During the lag time, intracellular cAMP increases, the *lac* operon is transcribed, more lactose is transported into the cell, and β-galactosidase is synthesized to break down lactose.

Q When both glucose and lactose are present, why will cells use glucose first?



(a) Lactose present, glucose scarce (cAMP level high) If glucose is scarce, the high level of cAMP activates CAP, and the *lac* operon produces large amounts of mRNA for lactose digestion.

(b) Lactose present, glucose present (cAMP level low) When glucose is present, cAMP is scarce, and CAP is unable to stimulate transcription.

FIGURE 8.14 Positive regulation of the lac operon. (**a**) The enzymes for lactose digestion are synthesized if lactose is present and the cell lacks an energy source. (**b**) The enzymes for lactose digestion are not produced if the cell has sufficient energy.

Q Will transcription of the *lac* operon occur in the presence of lactose and glucose? In the presence of lactose and the absence of glucose? In the presence of glucose and the absence of lactose?

operon requires both the presence of lactose and the absence of glucose (Figure 8.14).

Cyclic AMP is an example of an *alarmone*, a chemical alarm signal that promotes a cell's response to environmental or nutritional stress. (In this case, the stress is the lack of glucose.) The same mechanism involving cAMP allows the cell to grow on other sugars. Inhibition of the metabolism of alternative carbon sources by glucose is termed **catabolite repression** (or the *glucose effect*). When glucose is available, the level of cAMP in the cell is low, and consequently CAP is not bound. **Animation: Go to The Microbiology Place website or CD-ROM and click "Animations" to view Operons.**

MUTATION: CHANGE IN THE GENETIC MATERIAL

A **mutation** is a change in the base sequence of DNA. Such a change in the base sequence of a gene will sometimes cause a change in the product encoded by that gene.

For example, when the gene for an enzyme mutates, the enzyme encoded by the gene may become inactive or less active because its amino acid sequence has changed. Such a change in genotype may be disadvantageous, or even lethal, if the cell loses a phenotypic trait it needs. However, a mutation can be beneficial if, for instance, the altered enzyme encoded by the mutant gene has a new or enhanced activity that benefits the cell.

Many simple mutations are silent (neutral); the change in DNA base sequence causes no change in the activity of the product encoded by the gene. Silent mutations commonly occur when one nucleotide is substituted for another in the DNA, especially at a location corresponding to the third position of the mRNA codon. Because of the degeneracy of the genetic code, the resulting new codon might still code for the same amino acid. Even if the amino acid is changed, the function of the protein may not change if the amino acid is in a nonvital portion of the protein, or is chemically very similar to the original amino acid.

TYPES OF MUTATIONS

LEARNING OBJECTIVE

- Classify mutations by type, and describe how mutations are prevented or repaired.

The most common type of mutation involving single base pairs is **base substitution** (or *point mutation*), in which a single base at one point in the DNA sequence is replaced with a different base. When the DNA replicates, the result is a substituted base pair (Figure 8.15). For example, AT might be substituted for GC, or CG for GC. If a base substitution occurs within a gene that codes for a protein, the mRNA transcribed from the gene will carry an incorrect base at that position. When the mRNA is translated into protein, the incorrect base may cause the insertion of an incorrect amino acid in the protein. If the base substitution results in an amino acid substitution in the synthesized protein, this change in the DNA is known as a **missense mutation** (Figure 8.16a and b).

The effects of such mutations can be dramatic. For example, sickle cell disease is caused by a single change in the gene for globin, the protein component of hemoglobin. Hemoglobin is primarily responsible for transporting oxygen from the lungs to the tissues. A single missense mutation, a change from an A to a T at a specific site, results in the change from glutamic acid to valine in the protein. The effect of this change is that the shape of the hemoglobin molecule changes under conditions of low oxygen, altering the shape of the red blood cells such that movement of the cells through small capillaries is greatly impeded.

By creating a nonsense (stop) codon in the middle of an mRNA molecule, some base substitutions effectively prevent the synthesis of a complete functional protein; only a fragment is synthesized. A base substitution resulting in a nonsense codon is thus called a **nonsense mutation** (Figure 8.16c).

Besides base-pair mutations, there are also changes in DNA called **frameshift mutations,** in which one or a few nucleotide pairs are deleted or inserted in the DNA (Figure 8.16d). This mutation can shift the "translational reading frame"—that is, the three-by-three grouping of nucleotides recognized as codons by the tRNAs during translation. For example, deleting one nucleotide pair in the middle of a gene causes changes in many amino acids downstream from the site of the original mutation. Frameshift mutations almost always result in a long stretch of altered amino acids and the production of an inactive protein from the mutated gene. In most cases, a nonsense codon will eventually be encountered and thereby terminate translation.

Occasionally, mutations occur where significant numbers of bases are added to (inserted into) a gene. Huntington's disease, for example, is a progressive neurological disorder caused by extra bases inserted into a particular gene. The reason these insertions occur in this particular gene is still being studied.

Base substitutions and frameshift mutations may occur spontaneously because of occasional mistakes made during DNA replication. These **spontaneous mutations** apparently occur in the absence of any mutation-causing agents. Agents in the environment, such as certain chemicals and radiation, that directly or indirectly bring about mutations are called **mutagens.** Almost any agent that can chemically or physically react with DNA can potentially cause mutations. A wide variety of chemicals, many of which are common in nature or in households, are known to be mutagens. Many forms of radiation, including X rays and ultraviolet light, are also mutagenic, as discussed shortly.

In the microbial world, certain mutations result in resistance to antibiotics (see the box in Chapter 26, page 793) or altered pathogenicity. A mutation in a gene encoding the outer membrane may increase pathogenicity; for example, *Salmonella typhimurium* with an altered outer membrane can survive in phagocytes. A mutation in a capsule-encoding gene may result in decreased pathogenicity because phagocytes can destroy the bacteria, as in the cases of *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis*.

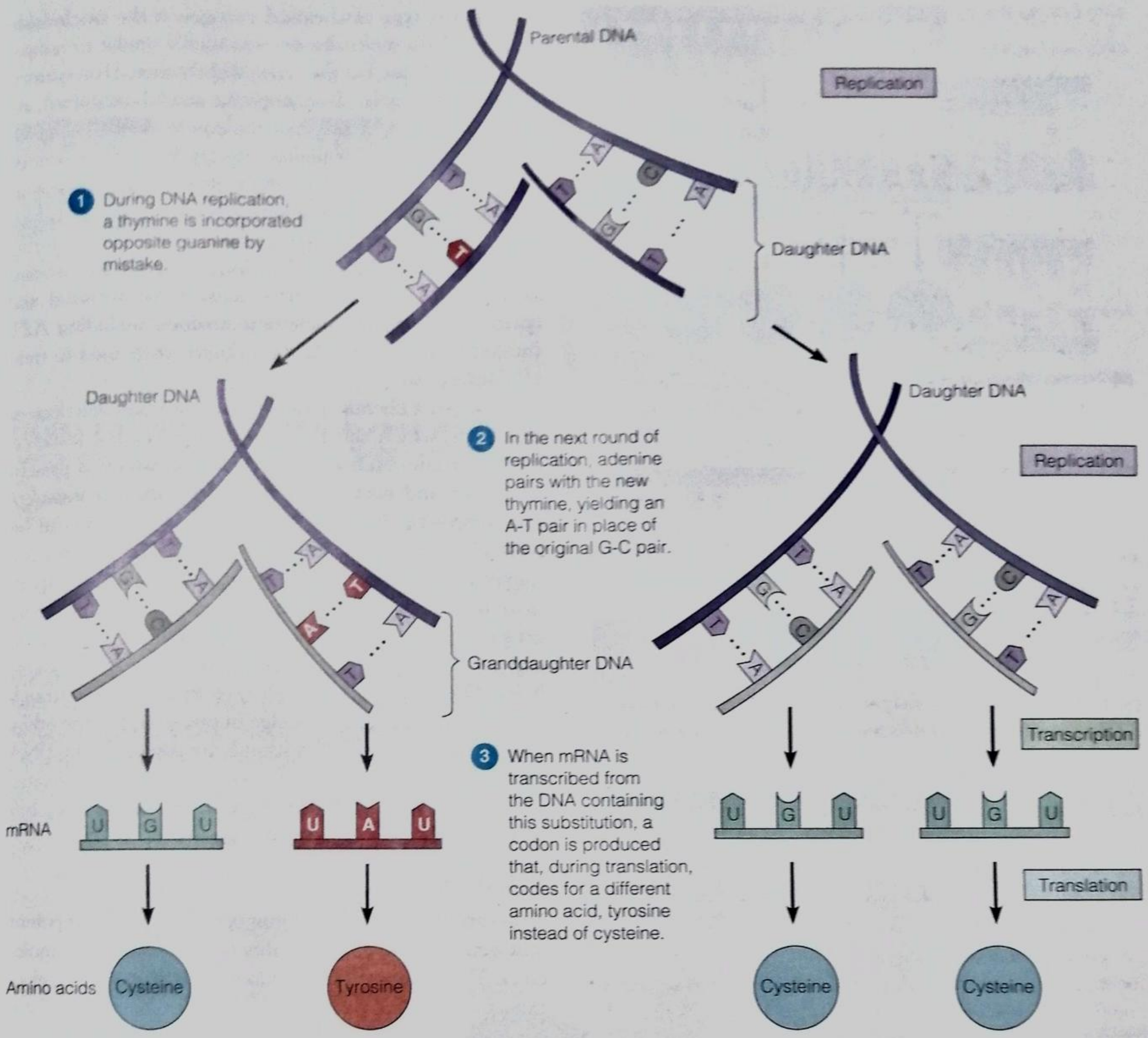


FIGURE 8.15 Base substitution. This mutation leads to an altered protein in a granddaughter cell.

Q **Does a base substitution always result in a different amino acid?**

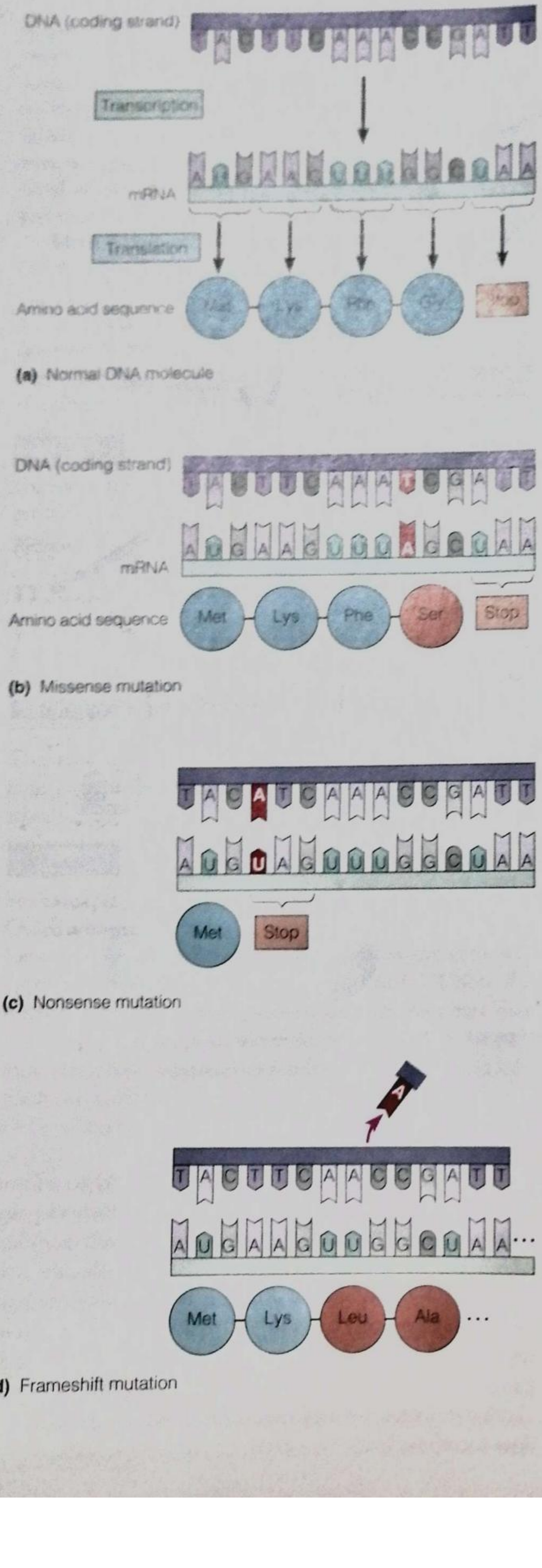
MUTAGENS

LEARNING OBJECTIVES

- Define *mutagen*.
- Describe two ways mutations can be repaired.

CHEMICAL MUTAGENS

One of the many chemicals known to be a mutagen is nitrous acid. Figure 8.17 shows how exposure of DNA to nitrous acid can convert the base adenine (A) to a form that no longer pairs with thymine (T) but instead pairs with cytosine (C). When DNA containing such modified adenines replicates, one daughter DNA molecule will have a base-pair sequence different from that of the parent DNA. Eventually, some AT base pairs of the parent will have been changed to GC base pairs in a granddaughter cell. Nitrous acid makes a specific base-pair change in DNA. Like all mutagens, it alters DNA at random locations.



Another type of chemical mutagen is the **nucleoside analog.** These molecules are structurally similar to normal nitrogenous bases, but they have slightly altered base-pairing properties. Examples, 2-aminopurine and 5-bromouracil, are shown in Figure 8.18. When nucleoside analogs are given to growing cells, the analogs are randomly incorporated into cellular DNA in place of the normal bases. Then, during DNA replication, the analogs cause mistakes in base pairing. The incorrectly paired bases will be copied during subsequent replication of the DNA, resulting in base-pair substitutions in the progeny cells. Some antiviral and antitumor drugs are nucleoside analogs, including AZT (azidothymidine), one of the primary drugs used to treat HIV infection.

Still other chemical mutagens cause small deletions or insertions, which can result in frameshifts. For instance, under certain conditions, benzopyrene, which is present in smoke and soot, is an effective *frameshift mutagen.* Aflatoxin—produced by *Aspergillus flavus* (a-spér-jil'lus flā'vus), a mold that grows on peanuts and grain—is a frameshift mutagen, as are the acridine dyes used experimentally against herpesvirus infections. Frameshift mutagens usually have the right size and chemical properties to slip between the stacked base pairs of the DNA double helix. They may work by slightly offsetting the two strands of DNA, leaving a gap or bulge in one strand or the other. When the staggered DNA strands are copied during DNA synthesis, one or more base pairs can be inserted or deleted in the new double-stranded DNA. Interestingly, frameshift mutagens are often potent carcinogens.

RADIATION

X rays and gamma rays are forms of radiation that are potent mutagens because of their ability to ionize atoms and molecules. The penetrating rays of ionizing radiation cause electrons to pop out of their usual shells (see Chapter 2). These electrons bombard other molecules and cause more damage, and many of the resulting ions and free radicals (molecular fragments with unpaired electrons) are very reactive. Some of these ions can combine with bases in DNA, resulting in errors in DNA replication and repair that produce mutations. An even more serious outcome is the breakage of covalent bonds in the sugar-phosphate backbone of DNA, which causes physical breaks in chromosomes.

FIGURE 8.16 Types of mutations and their effects on the amino acid sequences of proteins.

Q **On what basis are missense, nonsense, and frameshift mutations distinguished?**

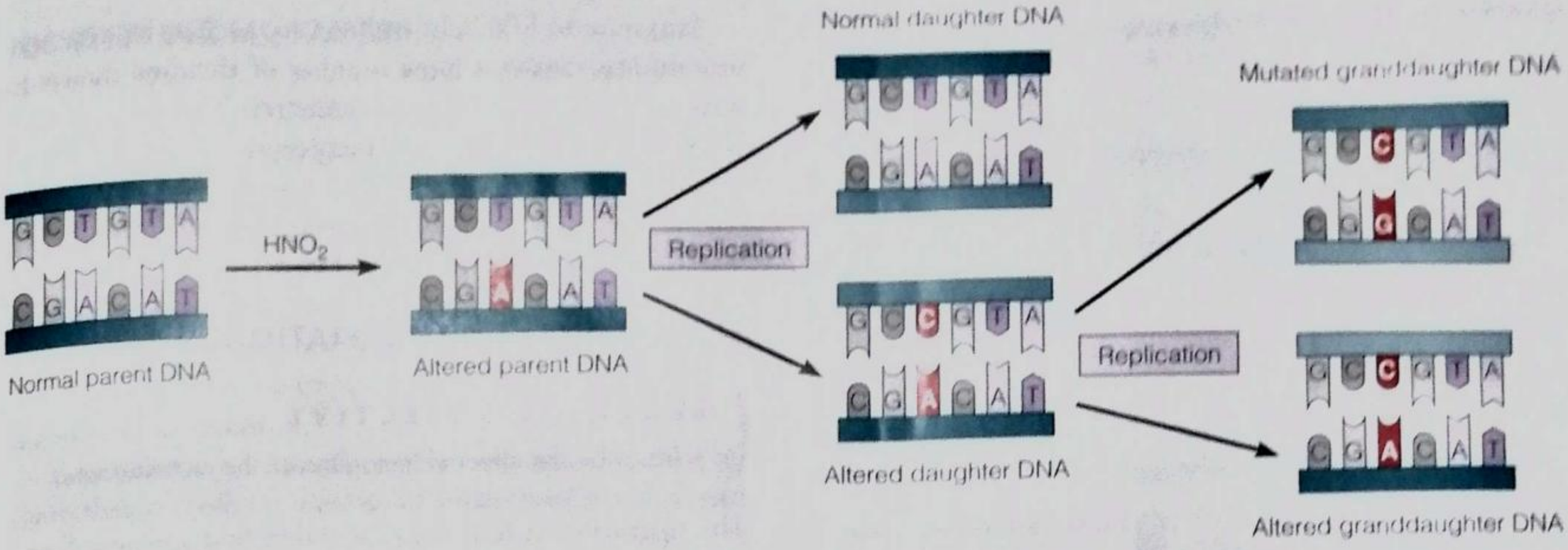


FIGURE 8.17 Nitrous acid (HNO_2) as a mutagen. The nitrous acid alters an adenine in such a way that it pairs with cytosine instead of thymine.

Q **What is a mutagen?**

Another form of mutagenic radiation is ultraviolet (UV) light, a nonionizing component of ordinary sunlight. However, the most mutagenic component of UV light (wavelength 260 nm) is screened out by the ozone layer of the atmosphere. The most important effect of direct UV light on DNA is the formation of harmful covalent bonds between certain bases. Adjacent thymines in a DNA strand can cross-link to form thymine dimers. Such dimers, unless repaired, may cause serious damage or death to the cell because it cannot properly transcribe or replicate such DNA.

Bacteria and other organisms have enzymes that can repair UV-induced damage. **Photolyases,** also known as *light-repair enzymes*, use visible light energy to separate the dimer back to the original two thymines. **Nucleotide excision repair,** shown in Figure 8.19, is not restricted to UV-induced damage; it can repair mutations from other causes as well. Enzymes cut out the incorrect base and fill in the gap with newly synthesized DNA that is complementary to the correct strand. For many years biologists questioned how the incorrect base could be distinguished from the correct base if it was not physically distorted like

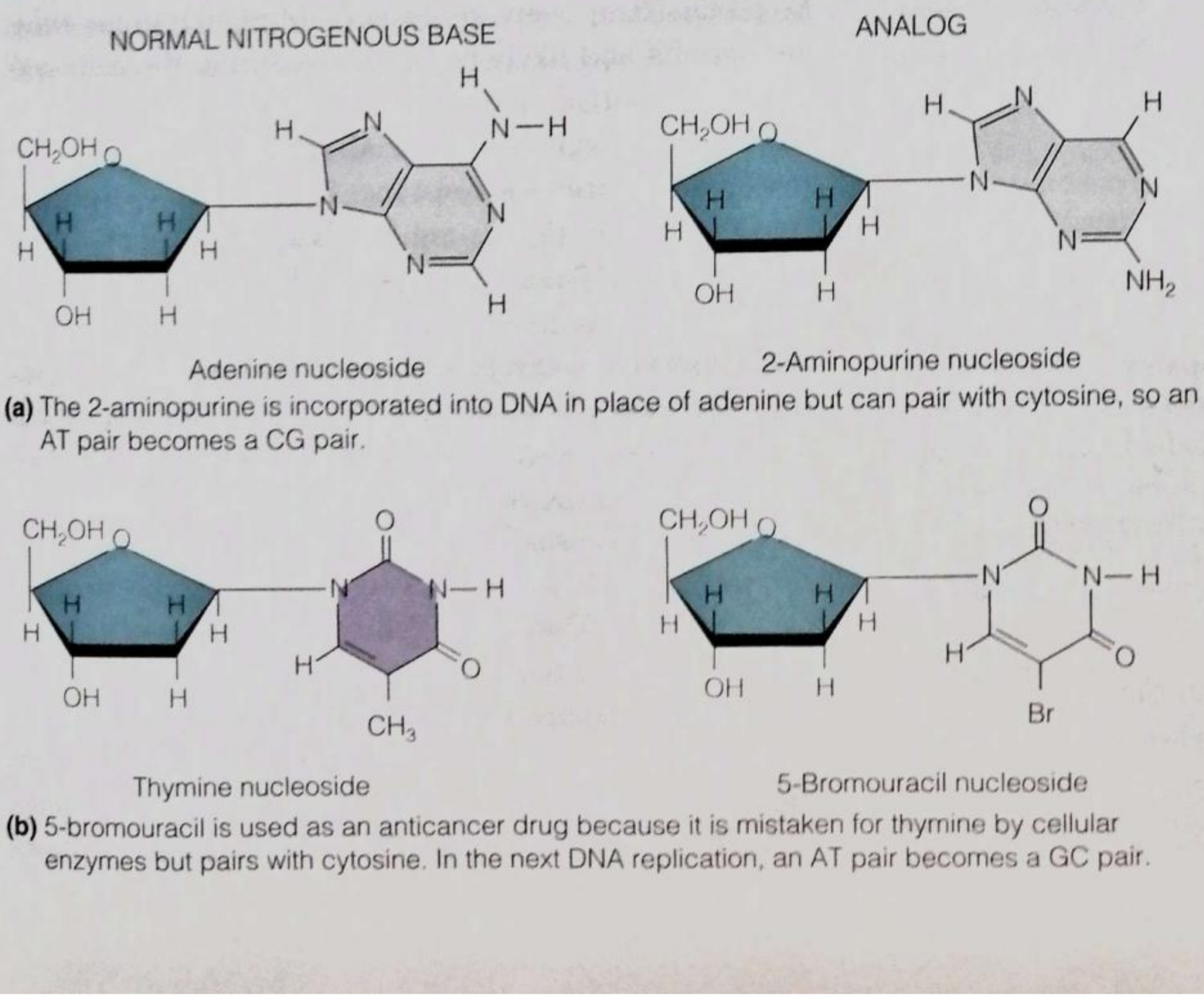


FIGURE 8.18 Nucleoside analogs and the nitrogenous bases they replace. A nucleoside is phosphorylated and the resulting nucleotide used to synthesize DNA.

Q **Why do these drugs kill cells?**

(a) The 2-aminopurine is incorporated into DNA in place of adenine but can pair with cytosine, so an AT pair becomes a CG pair.

(b) 5-bromouracil is used as an anticancer drug because it is mistaken for thymine by cellular enzymes but pairs with cytosine. In the next DNA replication, an AT pair becomes a GC pair.

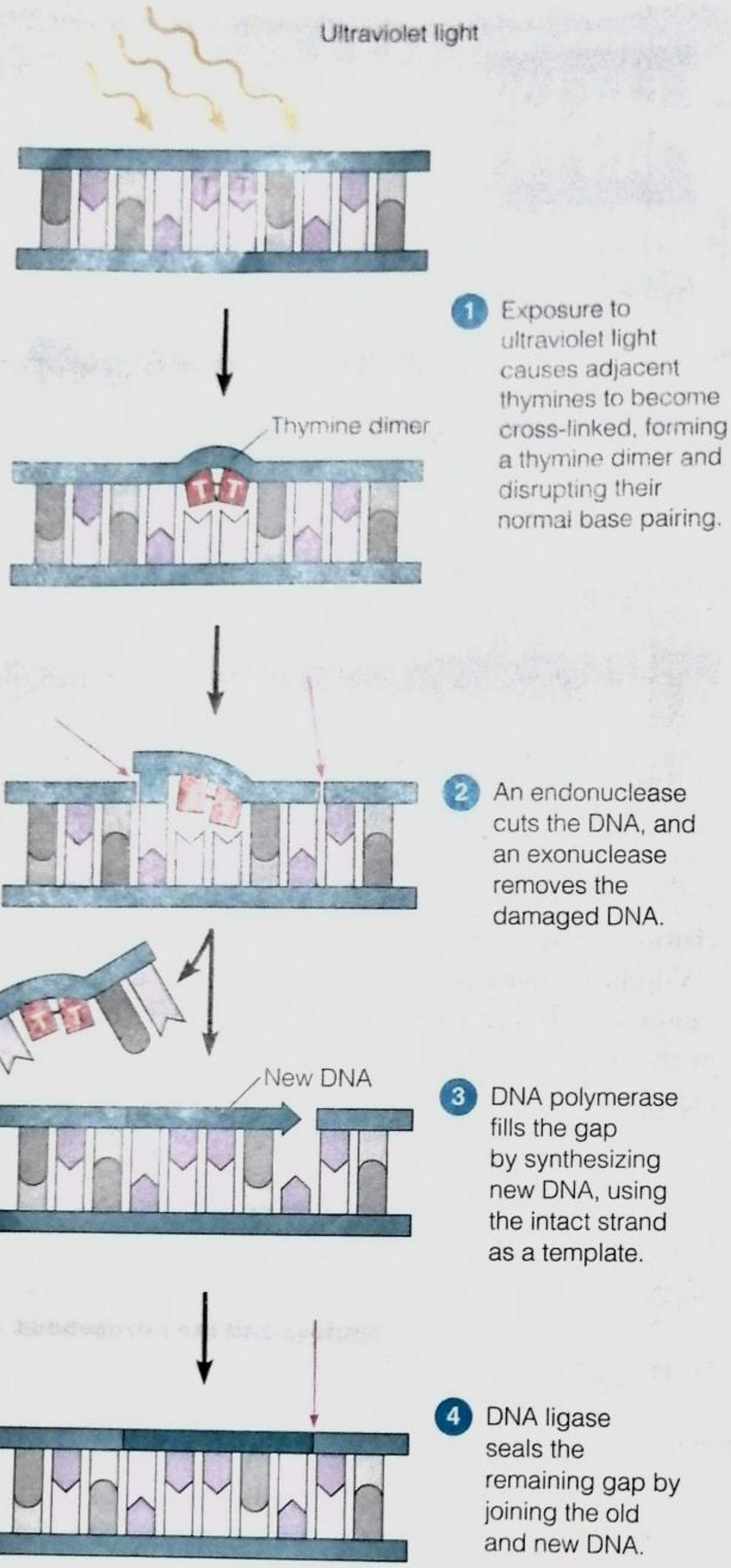


FIGURE 8.19 The creation and repair of a thymine dimer caused by ultraviolet light. After exposure to UV light, adjacent thymines can become cross-linked, forming a thymine dimer. In the absence of visible light, the nucleotide excision repair mechanism is used in a cell to repair the damage.

Q **How do excision repair enzymes "know" which strand is incorrect?**

a thymine dimer. In 1970, Hamilton Smith provided the answer with the discovery of **methylases.** These enzymes add a methyl group to selected bases soon after a DNA strand is made. A repair endonuclease then cuts the nonmethylated strand.

Exposure to UV light in humans, such as by excessive suntanning, causes a large number of thymine dimers in skin cells. Unrepaired dimers may result in skin cancers. Humans with xeroderma pigmentosum, an inherited condition that results in increased sensitivity to UV light, have a defect in nucleotide excision repair; consequently, they have an increased risk of skin cancer.

THE FREQUENCY OF MUTATION

LEARNING OBJECTIVE

- Describe the effect of mutagens on the mutation rate.

The **mutation rate** is the probability that a gene will mutate when a cell divides. The rate is usually stated as a power of 10, and because mutations are very rare, the exponent is always a negative number. For example, if there is one chance in 10,000 that a gene will mutate when the cell divides, the mutation rate is 1/10,000, which is expressed as 10^{-4}. Spontaneous mistakes in DNA replication occur at a very low rate, perhaps only once in 10^9 replicated base pairs (a mutation rate of 10^{-9}). Because the average gene has about 10^3 base pairs, the spontaneous rate of mutation is about one in 10^6 (a million) replicated genes.

Mutations usually occur more or less randomly along a chromosome. The occurrence of random mutations at low frequency is an essential aspect of the adaptation of species to their environment, for evolution requires that genetic diversity be generated randomly and at a low rate. For example, in a bacterial population of significant size—say, greater than 10^7 cells—a few new mutant cells will always be produced in every generation. Most mutations either are harmful and likely to be removed from the gene pool when the individual cell dies or are neutral. However, a few mutations may be beneficial. For example, a mutation that confers antibiotic resistance is beneficial to a population of bacteria that is regularly exposed to antibiotics. Once such a trait has appeared through mutation, cells carrying the mutated gene are more likely than other cells to survive and reproduce as long as the environment stays the same. Soon most of the cells in the population will have the gene; an evolutionary change will have occurred, although on a small scale.

A mutagen usually increases the spontaneous rate of mutation, which is about one in 10^6 replicated genes, by a factor of 10–1000 times. In other words, in the presence of a mutagen, the normal rate of 10^{-6} mutations per replicated gene becomes a rate of 10^{-5} to 10^{-3} per replicated gene. Mutagens are used experimentally to enhance the production of mutant cells for research on the genetic properties of microorganisms and for commercial purposes.