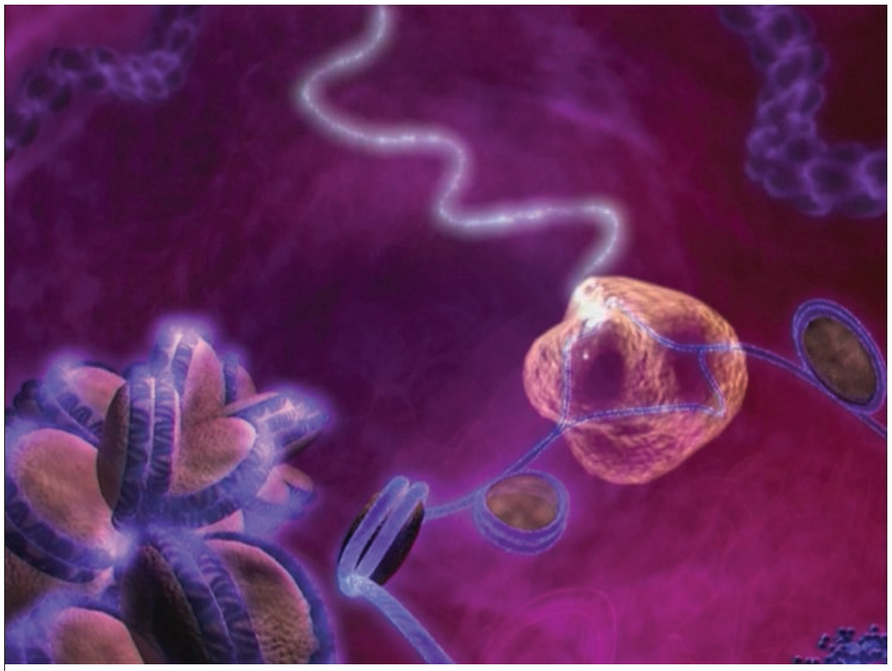


© LookatSciences/Phototake

Transcription of a eukaryotic gene to produce messenger RNA (mRNA), a type of RNA that acts as a template for protein synthesis. The DNA of the gene unwinds from the nucleosome (left side) and is copied by an RNA polymerase (center) into mRNA (exiting the top).

## Study Plan



# 15 From DNA to Protein

## Why It Matters

The marine mussel *Mytilus* **(Figure 15.1)** lives in one of the most demanding environments on Earth—it clings permanently to rocks pounded by surf day in and day out, constantly in danger of being dashed to pieces or torn loose by foraging predators. The mussel is remarkably resistant to disturbance, however; if you try to pry one loose you will find how difficult it is to tear the tough, elastic fibers that hold it fast, or even to cut them with a knife.

The fibers holding mussels to the rocks are a complex of proteins secreted by the muscular foot of the animal. The proteins, which include *keratin* (an intermediate filament protein; discussed in Section 5.3) form a tough, adhesive material called *byssus*.

Byssus is a premier underwater adhesives. It fascinates biochemists, adhesive manufacturers, dentists, and surgeons looking for better ways to hold repaired body parts together. Genetic engineers have inserted segments of mussel DNA into yeast cells, which reproduce in large numbers and serve as "factories" translating the mussel genes into byssus and other proteins. With byssus produced in this way, investigators are learning how to use or imitate the mussel glue for human needs. This exciting work, like the mussel's own byssus-building,

Dean Hallinan/FPG/Getty Images

Bob Evans/Peter Arnold, Inc.

**Figure 15.1**
The marine mussel *Mytilus* and its natural habitat.

starts with one of life's universal truths: *every protein is assembled on ribosomes according to instructions that are copied from DNA*.

In this chapter we trace the reactions by which proteins are made, beginning with the instructions encoded in DNA and leading through RNA to the sequence of amino acids in a protein. Many enzymes and other proteins are players as well as products in this story, as are several kinds of RNA and the cell's protein-making machines, the ribosomes. The same basic steps produce the proteins of all organisms. Our discussion begins with an overview of the entire process, starting with DNA and ending with a finished protein.

## 15.1 The Connection between DNA, RNA, and Protein

You have learned that genes encode proteins. In this section you will learn how that connection was discovered. This section also presents an overview of the molecular steps from gene to protein: transcription and translation.

### Proteins Are Specified by Genes

How do scientists know that genes encode—specify the amino acid sequence of—proteins? Two key pieces of research involving defects in metabolism proved this connection unequivocally. The first began in 1896 with Archibald Garrod, an English physician. He studied *alkaptonuria,* a human disease that does little harm but is easily detected: the patient's urine turns black in air. Garrod and an English geneticist, William Bateson, studied families of patients with the disease and concluded that it is an inherited trait. Garrod also found that people with alkaptonuria excrete a particular chemical in their urine. It is this chemical that turns black in air. Garrod deduced that normal people are able to metabolize the chemical, whereas people with alkaptonuria cannot. In 1908 Garrod concluded that the disease was an *inborn error of metabolism*. He did not know it at the time, but alkaptonuria results from a change in a gene that encodes an enzyme that metabolizes a key chemical. The altered gene causes a defect in the function of the enzyme, which leads to the phenotype of the disease. Garrod's work was the first evidence of a specific relationship between genes and metabolism.

In the second piece of research, George Beadle and Edward Tatum, working at Stanford University in the 1940s with the orange bread mold *Neurospora crassa,* obtained results showing a direct relationship between genes and enzymes. Beadle and Tatum chose *Neurospora* for their work because it is a haploid fungus, with simple nutritional needs. That is, wild-type *Neurospora*—the form of the mold found in nature—grows readily on a minimal medium (MM) consisting of a number of inorganic salts, sucrose, and a vitamin. The researchers reasoned that the fungus uses the simple chemicals in MM to synthesize all of the more complex molecules needed for growth and reproduction, including amino acids for proteins and nucleotides for DNA and RNA.

Beadle and Tatum exposed spores of wild-type *Neurospora* to X-rays. An X-ray is a *mutagen,* an agent that causes mutations. They found that some of the treated spores would not germinate and grow on MM unless they supplemented the medium with additional nutrients, such as amino acids or vitamins. Mutant strains that are unable to grow on MM are called *auxotrophs* (*auxo* = increased; *troph* = eater) or *nutritional mutants*. Beadle and Tatum hypothesized that each auxotrophic strain had a defect in a gene that codes for an enzyme needed to synthesize a particular nutrient. The wild-type strain could make the nutrient for itself from raw mate-

rials in the MM, but the mutant strain could grow only if the researchers supplied the nutrient. By testing each mutant strain on MM with a single added nutrient, they discovered what specific nutrient the strain needed in order to grow and, therefore, generally what gene defect it had. For example, a mutant that requires the addition of the amino acid arginine to grow has a defect in a gene for an enzyme involved in the synthesis of arginine. Such arginine auxotrophs are known as *arg* mutants. The assembly of arginine from raw materials is a multistep "assembly-line" process with a different enzyme catalyzing each step. Each *arg* mutant differs in the enzyme that is defective and therefore in the step of the assembly pathway that is blocked.

Beadle and Tatum studied four *arg* mutants—*argE, argF, argG,* and *argH*—to determine the metabolic defect each had; that is, where in the arginine synthesis pathway each was blocked. They took samples from each culture and tested whether the samples could grow on MM, or on MM supplemented either with ornithine, citrulline, argininosuccinate—three compounds known to be involved in the synthesis of arginine—or with arginine itself **(Figure 15.2).** None of the four mutants grew on MM, but all grew on MM + arginine. Each of the *arg* mutants showed a different pattern of growth on the supplemented MM (see Figure 15.2). Beadle and Tatum deduced that the biosynthesis of arginine occurs in a number of steps, with each step controlled by a gene that encodes the enzyme for the step (see Figure 15.2, Conclusion). For example, the *argH* mutant grows on MM + arginine, but not on MM + any of the other three compounds; this means that the mutant is blocked at the last step in the pathway that produces arginine. Similarly, the *argG* mutant grows on MM + arginine or argininosuccinate, but not on MM + any of the other supplements; this means that *argG* is blocked in the pathway before argininosuccinate is made (see Figure 15.2, Conclusion). With similar analysis, the researchers deduced the whole pathway from precursor to arginine and showed which gene encoded the enzyme that carried out each step. In sum, Beadle and Tatum had shown the direct relationship between genes and enzymes, which they put forward as the **one gene–one enzyme hypothesis.** Their experiment was a keystone in the development of molecular biology. As a result of their work, they were awarded a Nobel Prize in 1958.

As you learned in Chapter 3, enzymes are just one form of proteins, the amino acid-containing macromolecules that carry out many vital functions in living organisms. A functional protein consists of one or more subunits, called *polypeptides.* The protein hemoglobin, for instance, is made up of four polypeptides, two each of an α subunit and a β subunit. Hemoglobin's ability to transport oxygen is a functional property belonging only to the complete protein, and not to any of the polypeptides individually. Since a different gene encodes each distinct polypeptide, two different genes are needed to encode the hemoglobin protein: one for the α polypeptide and one for the β polypeptide. Due to the fact that some proteins consist of more than one polypeptide, and not all proteins are enzymes, Beadle and Tatum's hypothesis is now restated as the **one gene–one polypeptide hypothesis.** It is important to keep in mind the distinction between a protein, the functional molecule, and a polypeptide, the molecule encoded by a gene, as we discuss transcription and translation in the rest of this chapter.

## The Pathway from Gene to Polypeptide Involves Transcription and Translation

The pathway from gene to polypeptide has two major steps, *transcription* and *translation*. **Transcription** is the mechanism by which the information encoded in DNA is made into a complementary RNA copy. It is called transcription because the information in one nucleic acid type is transferred to another nucleic acid type. **Translation** is the use of the information encoded in the RNA to assemble amino acids into a polypeptide. It is called translation because the information in a nucleic acid, in the form of nucleotides, is converted into a different kind of molecule—amino acids. In 1956, Francis Crick gave the name *central dogma* to the flow of information from DNA to RNA to protein.

In transcription, the enzyme RNA polymerase copies the DNA sequence of a gene into an RNA sequence. The process is similar to DNA replication, except that only one of the two DNA strands—the **template strand**—is copied into an RNA strand, and only part of the DNA sequence of the genome is copied in any cell at any given time. A gene encoding a polypeptide is a *protein-coding gene,* and the RNA transcribed from it is called **messenger RNA (mRNA).**

In translation, an mRNA associates with a *ribosome,* a particle on which amino acids are linked into polypeptide chains. As the ribosome moves along the mRNA, the amino acids specified by the mRNA are joined one by one to form the polypeptide encoded by the gene.

The processes of transcription and translation are similar in prokaryotes and eukaryotes **(Figure 15.3).** One key difference is that in eukaryotes, RNA polymerase makes a precursor-mRNA (pre-mRNA) in the nucleus that is processed to produce the functional mRNA. That mRNA exits the nucleus and is translated in the cytoplasm.

## The Genetic Code Is Written in Three-Letter Words Using a Four-Letter Alphabet

Conceptually, the transcription of DNA into RNA is straightforward. The DNA "alphabet" consists of the four letters A, T, G, and C, representing the four DNA

## Figure 15.2 Experimental Research

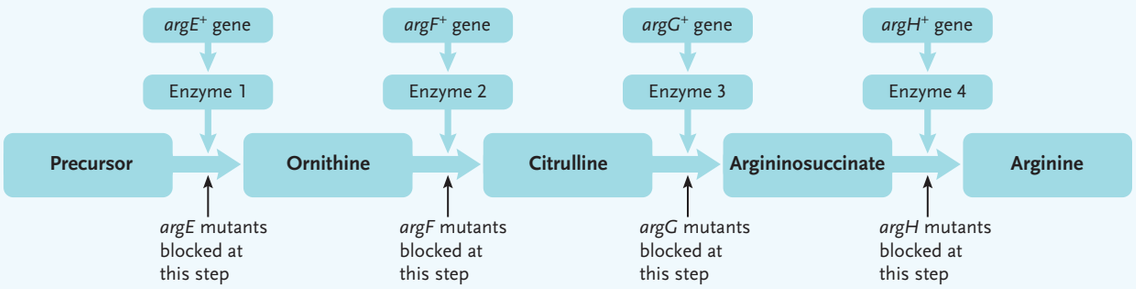
### *Relationship between Genes and Enzymes*

**QUESTION:** Do genes specify enzymes?

**EXPERIMENT:** Test *arg* mutants of the orange bread mold *Neurospora crassa* for growth on MM (minimal medium), MM + ornithine, MM + citrulline, MM + argininosuccinate, and MM + arginine. *Arg* mutants are unable to synthesize the amino acid arginine, which is essential for growth.

**RESULTS:**


**CONCLUSION:** Arginine is synthesized in a biochemical pathway. Each step of the pathway is catalyzed by an enzyme, and each enzyme is encoded by a gene:



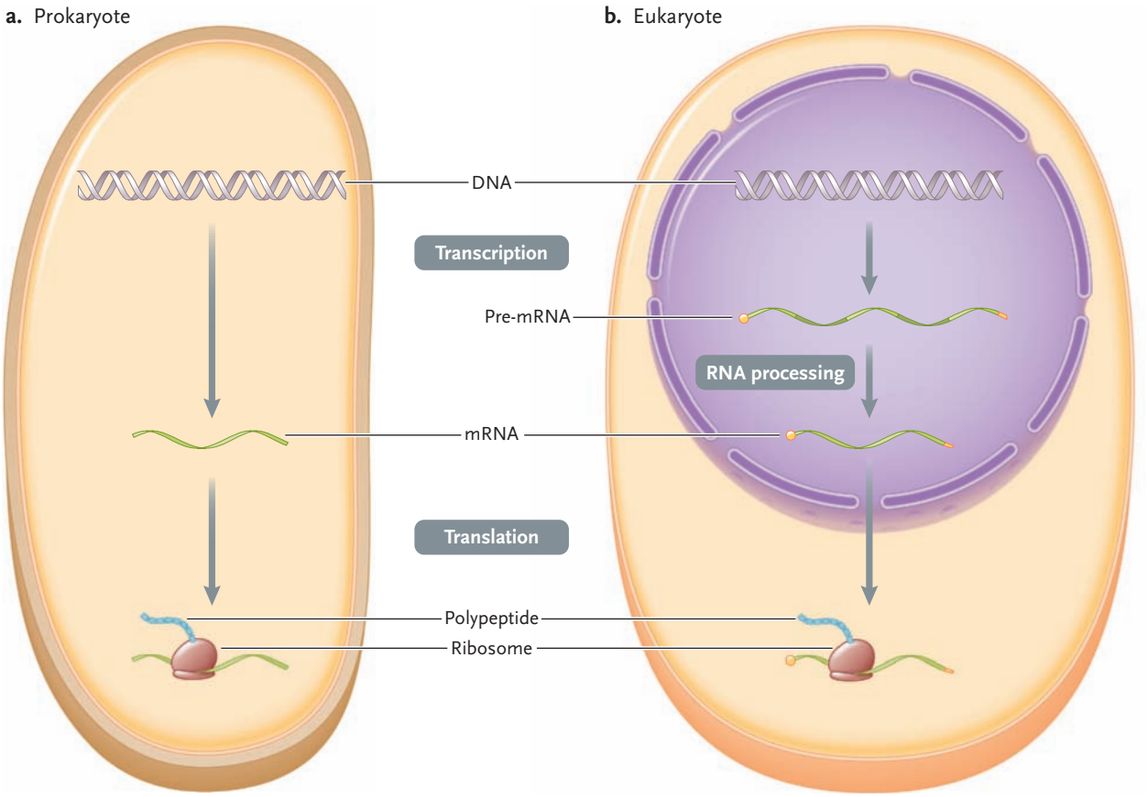



**Figure 15.3**
Transcription and translation in **(a)** prokaryotes and **(b)** eukaryotes. In prokaryotes RNA polymerase synthesizes an mRNA molecule that is ready for translation on ribosomes. In eukaryotes, RNA polymerase synthesizes a precursor-mRNA (pre-mRNA molecule) that has extra segments that are removed by RNA processing to produce a translatable mRNA. That mRNA exits the nucleus through a nuclear pore and is translated on ribosomes in the cytoplasm.

nucleotide bases: adenine, thymine, guanine, and cytosine. The RNA "alphabet" consists of the four letters A, U, G, and C, representing the four RNA bases: adenine, uracil, guanine, and cytosine. In other words, the nucleic acids share three of the four bases but differ in the other one; T in DNA is equivalent to U in RNA. But while there are four RNA bases, there are 20 amino acids. How is nucleotide information in an mRNA translated into the amino acid sequence of a polypeptide?

**Breaking the Genetic Code.** The nucleotide information that specifies the amino acid sequence of a polypeptide is called the **genetic code.** Scientists realized that the four bases in an mRNA (A, U, G, C) would have to be used in combinations of at least three to provide the capacity to code for 20 different amino acids. One- and two-letter words were eliminated because if the code used one-letter words, only four different amino acids could be specified (that is, $4^1$); if two-letter words were used, only 16 different amino acids could be specified (that is, $4^2$). But if the code used three-letter words, 64 different amino acids could be specified (that is, $4^3$), more than enough to specify 20 amino acids. We know now that the genetic code is a three-letter code; each three-letter word (triplet) of the code is called a **codon.** **Figure 15.4** illustrates the relationship between codons in a gene, codons in an mRNA, and the amino acid sequence of a polypeptide. The three-letter codons in DNA are first transcribed into complementary three-letter RNA codons. The process is similar to DNA replication except that in mRNA, the complement to adenine (A) in the template strand is uracil (U) instead of thymine (T) as in DNA replication.

How do the RNA codons correspond to the amino acids? The identity of most of the codons was established in 1964 by Marshall Nirenberg and Philip Leder of the National Institutes of Health (NIH). These researchers found that short, artificial mRNAs of codon length—three nucleotides—could bind to ribosomes in a test tube and cause a single transfer RNA (tRNA), with its linked amino acid, to bind to the ribosome. (As you will learn in Section 15.4, tRNAs are a special class of RNA molecules that bring amino acids to the ribosome for assembly into the polypeptide chain.) Nirenberg and Leder then made 64 of the short mRNAs, each consisting of a different, single codon. They added the mRNAs, one at a time, to a test tube containing ribosomes and all the different tRNAs, each linked to its own amino

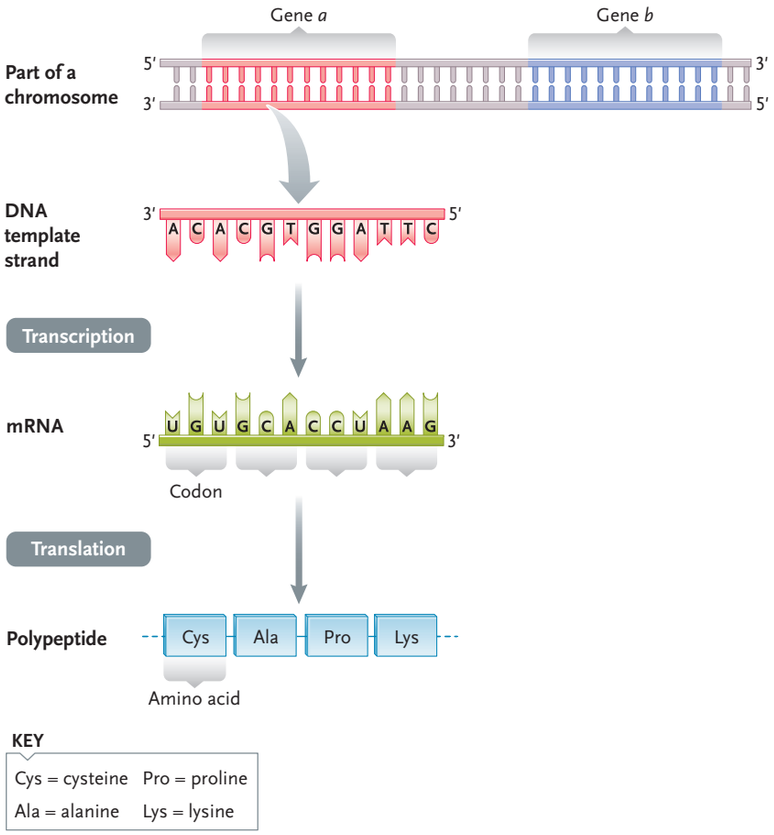


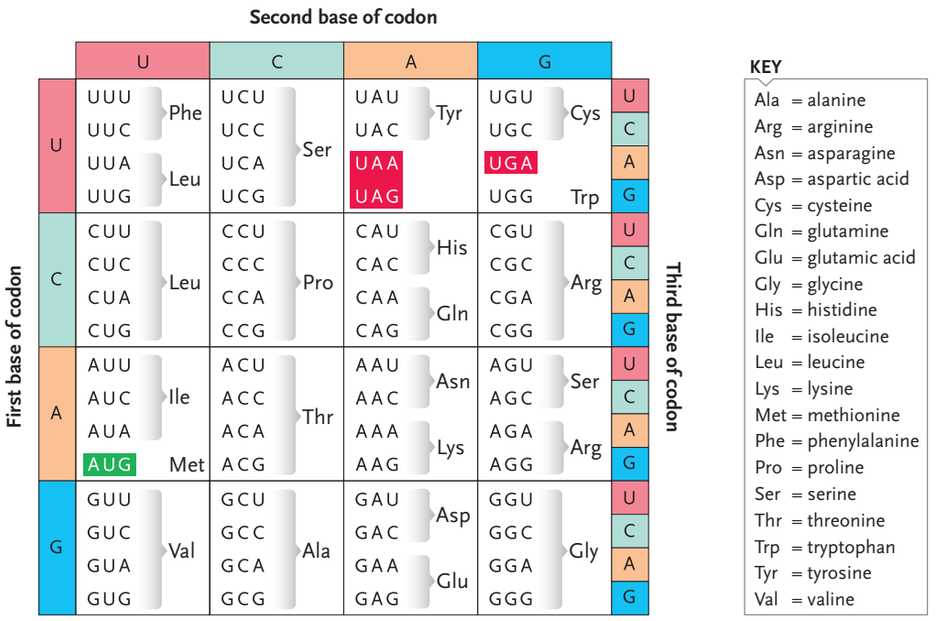

**Figure 15.4**
Relationship between a gene, codons in an mRNA, and the amino acid sequence of a polypeptide.

acid. The idea was that each single-codon mRNA would link to the tRNA in the mixture that carried the amino acid corresponding to the codon. The experiment worked for 50 of the 64 codons, allowing those codons to be assigned to amino acids definitively.

Another approach, carried out in 1966 by H. Gobind Khorana and his coworkers at Massachusetts Institute of Technology, used long, artificial mRNA molecules containing only one nucleotide repeated continuously, or different nucleotides in repeating patterns. Each artificial mRNA was added to ribosomes in a test tube, and the sequence of amino acids in the polypeptide chain made by the ribosomes was analyzed. For example, an artificial mRNA containing only uracil nucleotides in the sequence UUUUUU . . . resulted in a polypeptide containing only the amino acid phenylalanine: UUU must be the codon for phenylalanine. Khorana's approach, combined with the results of Nirenberg and Leder's experiments, identified the coding assignments of all the codons. Nirenberg and Khorana received a Nobel Prize in 1968 for their research in solving the nucleic acid code.

**Features of the Genetic Code.** **Figure 15.5** shows the genetic code of the 64 possible codons. By convention, scientists write the codons in the 5′→3′ direction, and as they appear in mRNAs, in which U substitutes for the T of DNA. Of the 64 codons, 61 specify amino acids. These are known as **sense codons.** One of these codons, AUG, specifying the amino acid methionine, is the first codon read in an mRNA in translation in both prokaryotes and eukaryotes. In that position, AUG is called a **start codon** or **initiator codon.** The three codons that do not specify amino acids—UAA, UAG, and UGA—are **stop codons** (also called **nonsense codons** and **termination codons**) that act as "periods" indicating the end of a polypeptide-encoding sentence. When a ribosome reaches one of the stop codons, poly-

**Figure 15.5**
The genetic code, written in the form in which the codons appear in mRNA. The AUG initiator codon, which codes for methionine, is shown in green; the three terminator codons are boxed in red.



peptide synthesis stops and the new polypeptide chain is released from the ribosome.

Only two amino acids, methionine and tryptophan, are specified by a single codon. All the rest are represented by more than one codon, some by as many as six. In other words, there are many *synonyms* in the nucleic acid code, a feature known as **degeneracy** (also called *redundancy*). For example, UGU and UGC both specify cysteine, and CCU, CCC, CCA, and CCG all specify proline.

Another feature of the genetic code is that it is **commaless;** that is, the words of the nucleic acid code are sequential, with no indicators such as commas or spaces to mark the end of one codon and the beginning of the next. The code can be read correctly only by starting at the right place—at the first base of the first three-letter codon at the beginning of a coded message—and reading three nucleotides at a time from this beginning codon. In other words, there is only one correct reading frame for each mRNA. For example, if you read the message SADMOMHASMOPCUTOFFBOYTOT three letters at a time, starting with the first letter of the first "codon," you would find that a mother reluctantly had her small child's hair cut. However, if you start incorrectly at the second letter of the first codon, you read the gibberish message ADM OMH ASM OPC UTO FFB OYT OT.

The code is also **universal.** With a few exceptions, the same codons specify the same amino acids in all living organisms, and also in viruses. The universality of the nucleic acid code indicates that it was established in its present form very early in the evolution of life and has remained virtually unchanged through billions of years of evolutionary history. (The evolution of life and the genetic code are discussed further in Chapter 24.) Minor exceptions to the universality of the genetic code have been found in a few organisms including a yeast, some protozoans, a prokaryote, and in the genetic systems of mitochondria and chloroplasts.

### Study Break

1. On the basis of their work with auxotrophic mutants of the fungus, *Neurospora crassa,* Beadle and Tatum proposed the one gene–one enzyme hypothesis. Why is it now known as the one gene–one polypeptide hypothesis?
2. If the codon were five bases long, how many different codons would exist in the genetic code?

## 15.2 Transcription: DNA-Directed RNA Synthesis

Transcription is the process by which information coded in DNA is transferred to a complementary RNA copy. The process of RNA transcription is similar to DNA replication **(Figure 15.6).** However, there are some important differences between transcription and replication. In transcription:

- Only one of the two DNA nucleotide strands acts as a template for synthesis of a complementary copy, instead of both as in replication.
- Only a relatively small part of a DNA molecule—the sequence encoding a single gene—serves as a template, rather than all of both strands as in DNA replication.
- **RNA polymerases** catalyze the assembly of nucleotides into an RNA strand, rather than the DNA polymerases that catalyze replication.
- The RNA molecules resulting from transcription are single polynucleotide chains, not double ones as in DNA replication.
- Where adenine appears in the DNA template chain, a uracil is matched to it in the RNA transcript instead of thymine as in DNA replication (see Figure 15.4 and Figure 15.6).

Transcription is similar in prokaryotes and eukaryotes. Throughout this section, we will point out the important differences between prokaryote and eukaryote processes.

### RNA Polymerases Work Like DNA Polymerases, but Require No Primer

Transcription begins as RNA polymerase binds to the DNA and unwinds it near the beginning of a gene (Figure 15.6, step 1). Unlike DNA polymerases, RNA polymerases can start the complementary copy with no need for a primer already in place (primers for DNA replication are discussed in Section 14.3). Like DNA, RNA is made in the 5′→3′ direction using the 3′→5′ DNA strand as template (step 2). Thus, we refer to the beginning of the RNA strand as the *5′ end,* and the other end as the *3′ end*. The RNA polymerase continues adding nucleotides one at a time until the gene is transcribed completely. At this point the newly synthesized RNA molecule and the enzyme are released from the DNA template (step 3).

### Specific Sequences of Nucleotides in the DNA Indicate Where Transcription of a Gene Begins and Ends

An organism's genome contains a large number of genes. For example, scientists analyzing the human genome sequence believe between 20,000 and 25,000 protein-coding genes are needed to make a human. Transcription is the process whereby particular genes are expressed in any given cell at a given time. Some of those genes are protein-coding genes which encode mRNAs that are translated; others are non–protein-coding genes which encode RNAs that are

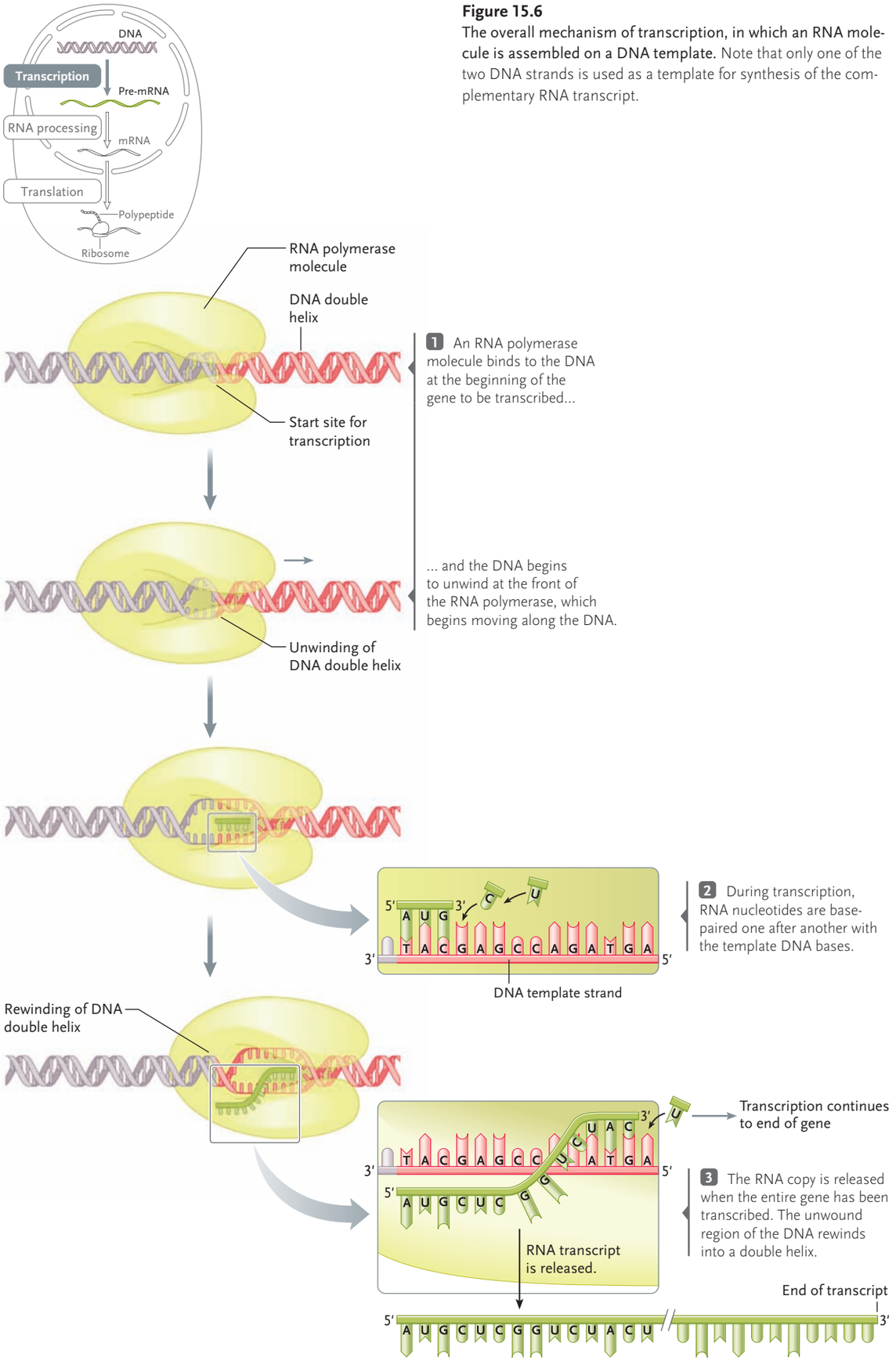



**Figure 15.6**
**The overall mechanism of transcription, in which an RNA molecule is assembled on a DNA template.** Note that only one of the two DNA strands is used as a template for synthesis of the complementary RNA transcript.

not translated, such as ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) The following sections describe the basic steps of transcription for a protein-coding gene.

**Organization of a Gene and the Steps of Transcription.** Let us first outline the structure of a gene and how it is transcribed into an RNA **(Figure 15.7).** At one end of a gene is a control sequence called a **promoter** (Figure 15.7, step 1). The part of the gene that is copied into RNA is called the **transcription unit.** To *initiate* transcription, RNA polymerase binds to the promoter, unwinds the DNA in that region, and starts synthesizing a new RNA molecule at the *transcription start point* (step 1). As RNA polymerase moves along the DNA, unwinding it at the forward end of the enzyme, the RNA molecule *elongates* as new nucleotides are added one by one (step 2). The newly synthesized portion of the RNA molecule winds temporarily with the template strand of the DNA into a hybrid RNA-DNA double helix. Beyond this short region the growing RNA strand unwinds from the DNA and extends from the RNA polymerase as a single nucleotide chain. At the back end of the RNA polymerase, the DNA double helix reforms. Elongation of the RNA chain continues until the end of the transcription unit, at which point RNA synthesis *terminates* and the completed RNA transcript and RNA polymerase are released from the DNA (step 3).

Once an RNA polymerase molecule has started transcription and progressed past the beginning of a gene, another molecule of RNA polymerase may start transcribing as soon as there is room at the promoter. In most genes this process continues until there are many RNA polymerase molecules spaced closely along a gene, each making an RNA transcript.

**The Promoter of Protein-Coding Genes and Transcription Initiation.** The promoter specifies where in the DNA transcription begins. In prokaryotes, the promoters are immediately upstream of the site where transcription initiates. RNA polymerase itself recognizes key DNA sequences in the promoter, binds to the promotor, and begins transcription of the mRNA. The same RNA polymerase also transcribes the tRNA and rRNA genes in prokaryotes; those genes have promoters highly similar in sequence to those of protein-coding genes.

In eukaryotes, RNA polymerase II transcribes protein-coding genes. RNA polymerases I and III transcribe genes for non–protein-coding RNAs, such as rRNAs in the ribosomes and tRNAs. Like prokaryotic promoters, the promoters of eukaryotic protein-coding genes are immediately upstream of the transcription start point, but they are typically more complex than those in prokaryotes. For example, a key element of the promoter of most eukaryotic protein-coding genes, the **TATA box**, plays an important role in transcription initiation. Other sequences further upstream of the gene are important for regulating the rate of transcription (discussed in Chapter 16).

**Transcription Termination.** In prokaryotes, specific DNA sequences called **terminators** signal the end of transcription of the gene. A protein binds to the terminator, triggering the termination of transcription and the release of the RNA and RNA polymerase from the template. Eukaryotic DNA has no equivalent sequences. Instead, the 3′ end of the mRNA is specified by a very different process, which will be discussed in the next section.

### Study Break

1. If the DNA template strand has the sequence 3′-CAAATTGGCTTATTACCGGATG-5′, what would be the sequence of an RNA transcribed from it?
2. What is the role of the promoter in transcription?

## 15.3 Production of mRNAs in Eukaryotes

Both prokaryotic and eukaryotic mRNAs contain regions that code for proteins, along with noncoding regions that play key roles in the process of protein synthesis. In prokaryotic mRNAs, the coding region is flanked by untranslated ends, the 5′ untranslated region (5′ UTR) and a 3′ untranslated region (3′ UTR). The same elements are present in eukaryotic mRNAs along with additional noncoding elements. The synthesis of mRNA in eukaryotes is the focus of this section. Roger Kornberg of Stanford University received a Nobel Prize in 2006 for describing the molecular structure of the eukaryotic transcription apparatus and how it acts in transcription.

### Eukaryotic Protein-Coding Genes Are Transcribed into Precursor-mRNAs That Are Modified in the Nucleus

A eukaryotic protein-coding gene is typically transcribed into a **precursor-mRNA (pre-mRNA)** that must be processed in the nucleus to produce the translatable mRNA (**Figure 15.8;** and see Figure 15.3). The mRNA exits the nucleus and is translated in the cytoplasm.

**Modifications of Pre-mRNA and mRNA Ends.** At the 5′ end of the pre-mRNA is the **5′ cap,** consisting of a guanine-containing nucleotide that is reversed so that its 3′-OH group faces the beginning rather than the end of the molecule. A *capping enzyme* adds the 5′ cap to the pre-mRNA soon after RNA polymerase II begins

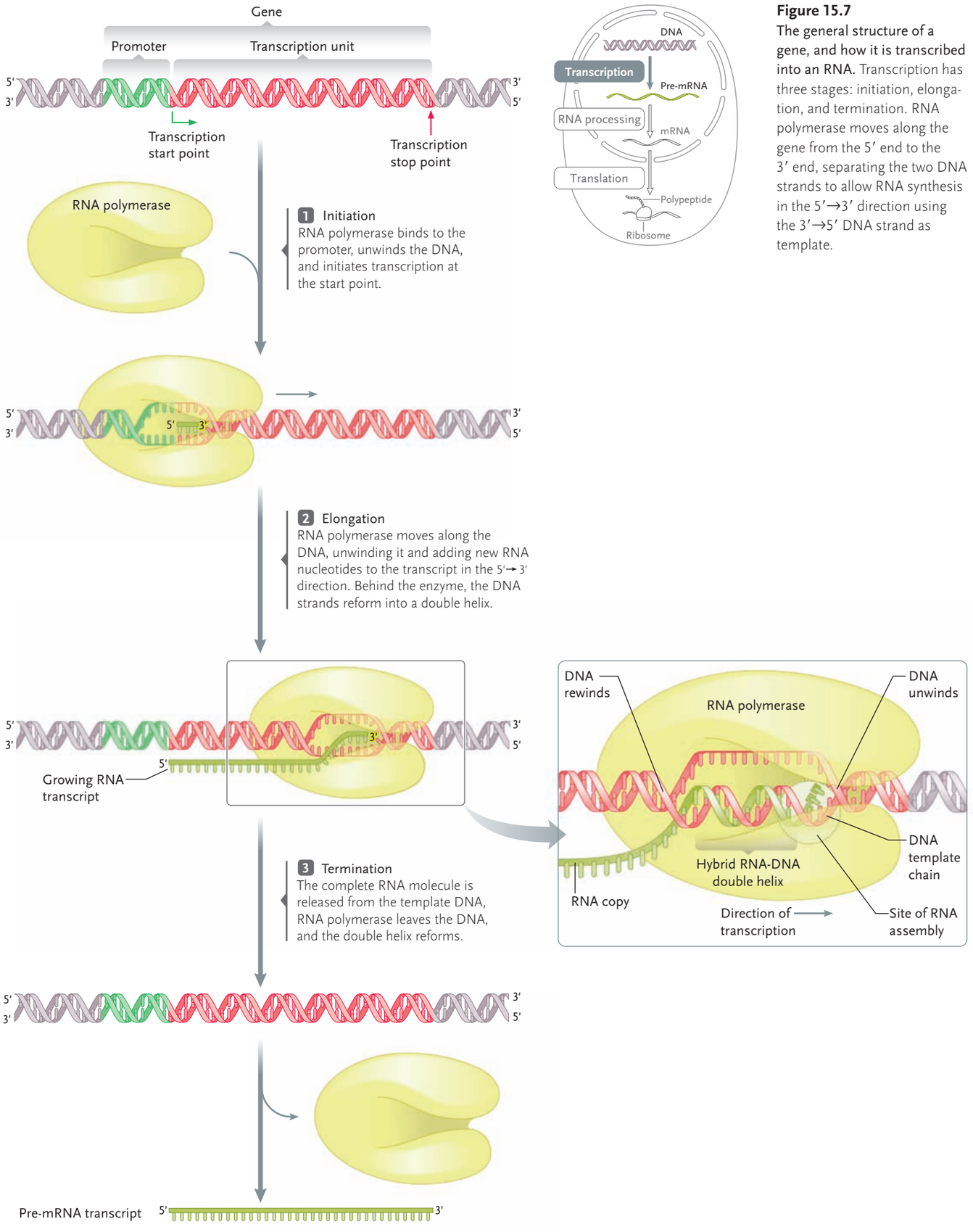



**Figure 15.7**
**The general structure of a gene, and how it is transcribed into an RNA.** Transcription has three stages: initiation, elongation, and termination. RNA polymerase moves along the gene from the 5′ end to the 3′ end, separating the two DNA strands to allow RNA synthesis in the 5′→3′ direction using the 3′→5′ DNA strand as template.

transcription. The cap, which is connected to the rest of the chain by three phosphate groups, remains when pre-mRNA is processed to mRNA. The cap is the site where ribosomes attach to mRNAs at the start of translation.

The termination of transcription of a eukaryotic protein-coding gene is different from that of a prokaryotic gene in that there is no terminator sequence at the end of the gene in the DNA that signals RNA polymerase to stop transcription. Instead, at the 3′ end of the gene is a sequence that is transcribed into the pre-mRNA. Proteins bind to this *polyadenylation signal,* and cleave the pre-mRNA at that point. Then, the enzyme *poly(A) polymerase* adds a chain of 50 to 250 adenine nucleotides, one nucleotide at a time, to that 3′ end of the pre-mRNA. This string of A nucleotides, called the **poly(A) tail,** enables the mRNA produced from the pre-mRNA to be translated efficiently, and protects it from attack by RNA-digesting enzymes in the cytoplasm. If it is removed experimentally, the mRNAs are quickly degraded inside cells.

**Sequences Interrupting the Protein-Coding Sequence.** The transcription unit of a eukaryotic protein-coding gene—the RNA-coding sequence—also contains non–protein-coding sequences called **introns** that interrupt the protein-coding sequence (shown in Figure 15.8). The introns are transcribed into pre-mRNAs, but removed from pre-mRNAs during processing in the nucleus, so that the coded messages in the finished mRNAs are read continuously, without interruptions. The amino acid–coding sequences that are retained in finished mRNAs are called **exons.** The mechanisms by which introns originated in genes is a mystery.

Introns were discovered by several methods, including direct comparisons between the nucleotide sequences of mature mRNAs and either pre-mRNAs or the genes encoding them. The majority of known eukaryotic genes contain at least one intron; some contain more than 60. The original discoverers of introns, Richard Roberts of New England Biolabs and Phillip Sharp of Massachusetts Institute of Technology, received a Nobel Prize in 1993 for their findings.

## Introns Are Removed During Pre-mRNA Processing to Produce the Translatable mRNA

A process called **mRNA splicing**, which occurs in the nucleus, removes introns from pre-mRNAs and joins exons together **(Figure 15.9).** mRNA splicing takes place in a **spliceosome,** a complex formed between the pre-mRNA and a handful of **small ribonucleoprotein particles.** A ribonucleoprotein particle is a complex of RNA and proteins. The small ribonucleoprotein particles involved in mRNA splicing are located in the nucleus; each consists of a relatively short *small nuclear RNA* (snRNA) bound to a number of proteins. The particles are therefore known as snRNPs, pronounced "snurps."

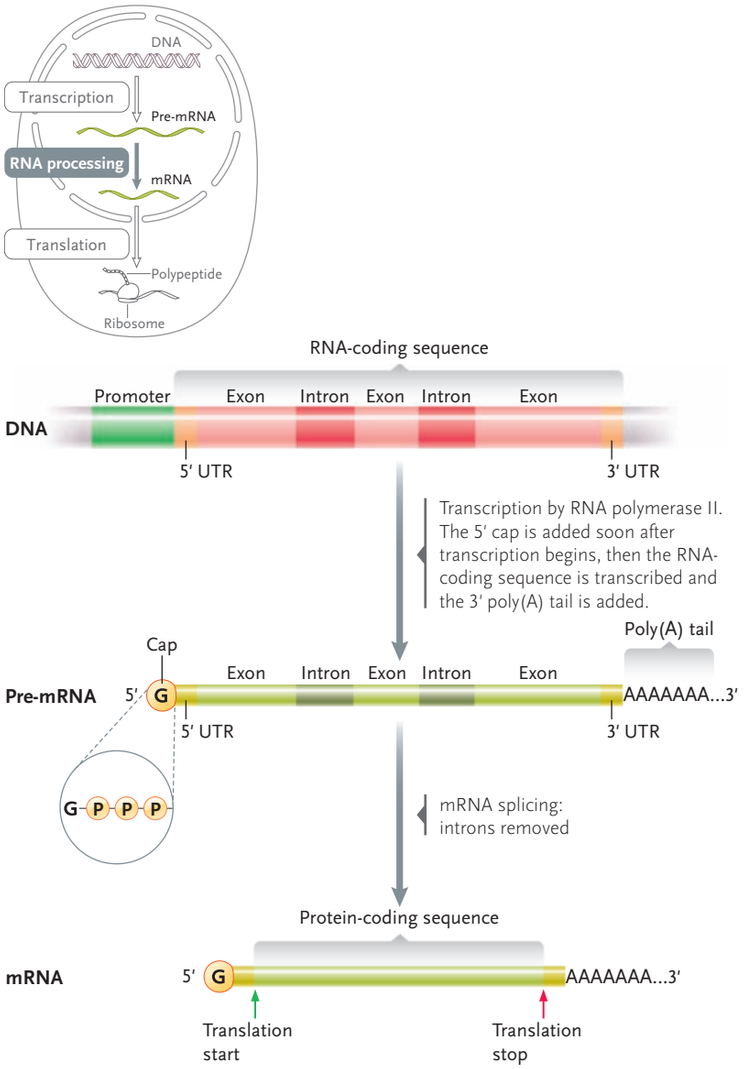



**Figure 15.8** Relationship between a eukaryotic protein-coding gene, the pre-mRNA transcribed from it, and the mRNA processed from the pre-mRNA.

The snRNPs bind in a particular order to an intron in the pre-mRNA. The first snRNPs are those with snRNAs that recognize and pair with sequences at the junctions of the intron with the adjacent exons. The complex then recruits the other snRNPs, producing a larger complex that loops out the intron and brings the two exon ends close together. At this point the active spliceosome has been formed. The spliceosome cleaves the pre-mRNA at the junction between the 3′ end of the exon (exon 1 in Figure 15.9) and the 5′ end of the intron, and the intron loops back to bond with itself near its 3′ end. The spliceosome then cleaves the pre-mRNA at the junction between the 3′ end of the intron and exon 2, releasing the intron

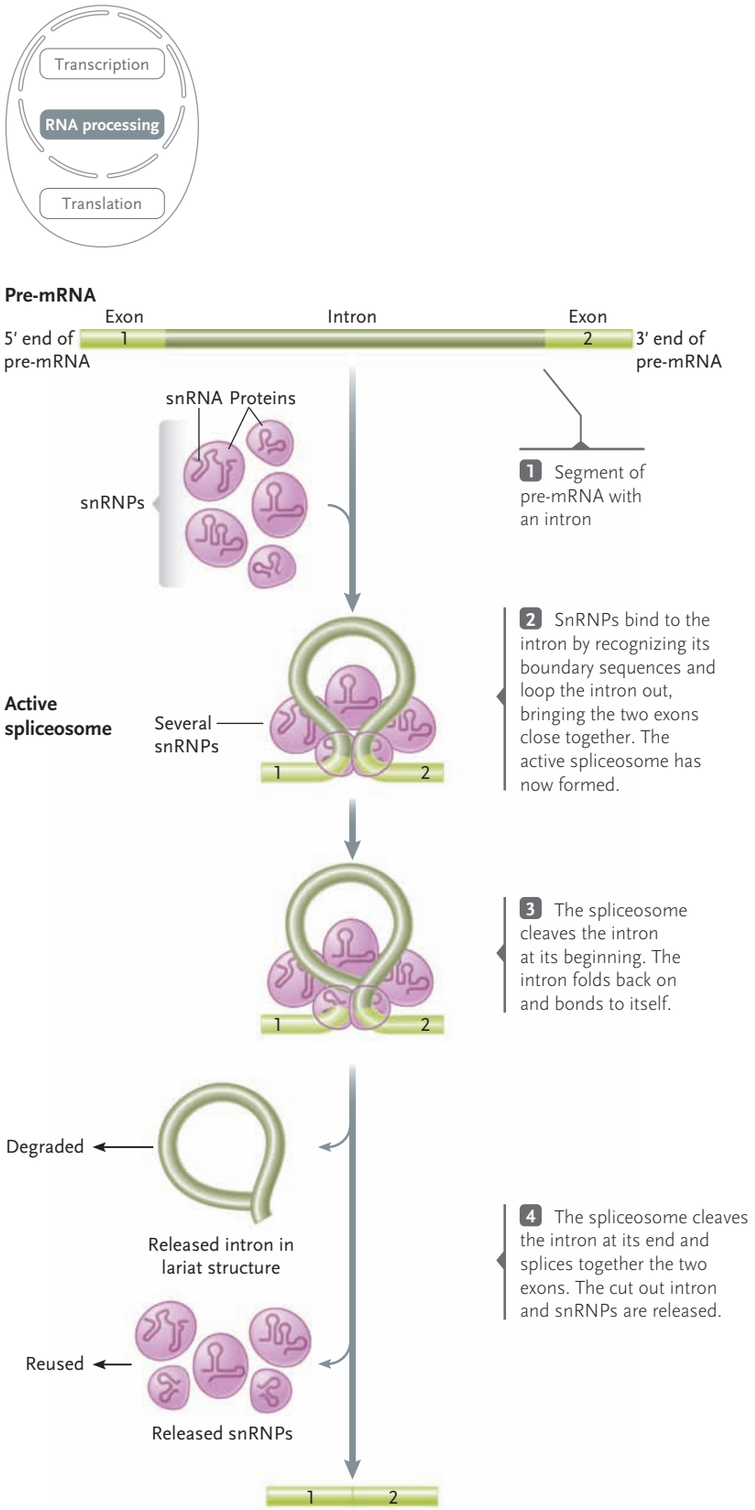



**Figure 15.9**
mRNA splicing—the removal from pre-mRNA of introns and joining of exons in the spliceosome.

and joining together the two exons (exon 1 and exon 2 in Figure 15.9). Because of the shape of the released intron, it is called a *lariat structure*. Enzymes degrade the intron and the snRNPs are released and used in other mRNA splicing reactions.

The cutting and splicing are so exact that not a single base of an intron is retained in the finished mRNA, nor is a single base removed from the exons. Without this precision, removing introns would change the reading frame of the coding portion of the mRNA, producing gibberish from the point of a mistake onward.

## Introns Contribute to Protein Variability

Introns seem wasteful in terms of the energy and raw materials required to replicate and transcribe them, and the elaborate cellular machinery required to remove them during pre-mRNA processing. Why are they present in mRNA-encoding genes? Among a number of possibilities, introns may provide advantages by increasing the coding capacity of existing genes through a process called *alternative splicing* and by generating new proteins through a process called *exon shuffling*.

**Alternative Splicing.** Many pre-mRNAs are processed by reactions that join exons in different combinations to produce different mRNAs from a single gene. The mechanism, called **alternative splicing**, greatly increases the number and variety of proteins encoded in the cell nucleus without increasing the size of the genome. For example, geneticists estimate that three-quarters of all human pre-mRNAs are subjected to alternative splicing. In each case the different mRNAs produced from the "parent" pre-mRNA are translated to produce a family of related proteins with various combinations of amino acid sequences derived from the exons. Each protein in the family, then, will vary to a degree in its function. Alternative splicing helps us understand why humans have only about 25,000 genes. As a result of the alternative splicing process, the number of proteins produced far exceeds the number of genes, and it is proteins that direct an organism's functions.

**Figure 15.10** shows an example of alternative splicing that occurs in mammals, including humans. The pre-mRNA transcript of the α-tropomyosin gene is alternatively spliced in various ways in different tissues—smooth muscle (for example, muscles of the intestine and bladder), skeletal muscle (for example, biceps, glutes), fibroblast (connective tissue cell that makes collagen), liver, and brain—to produce different forms of the α-tropomyosin protein that are functionally optimized for each tissue type. **Figure 15.10** shows the alternative splicing of the α-tropomyosin pre-mRNA to the

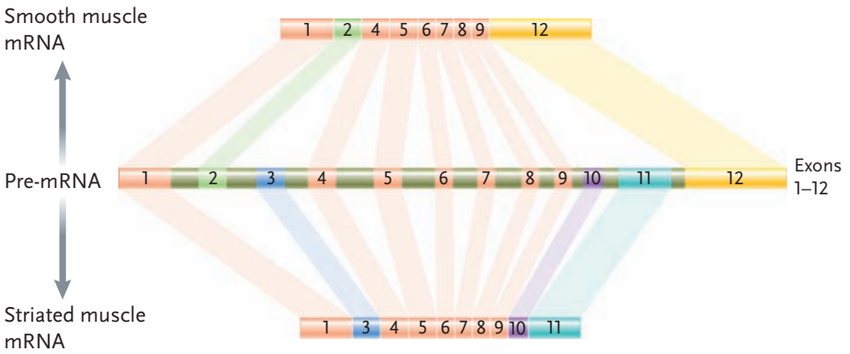



**Figure 15.10**
Alternative splicing of the α-tropomyosin pre-mRNA to distinct mRNA forms found in smooth muscle and striated muscle. All of the introns are removed in both mRNA splicing pathways, but exons 3, 10, and 11 are also removed to produce the smooth muscle mRNA, and exons 2 and 12 are also removed to produce the striated muscle mRNA.

mRNAs found in smooth muscle and striated muscle. Exons 2 and 12 are exclusive to the smooth muscle mRNA, while exons 3, 10, and 11 are exclusive to the striated muscle mRNA.

**Exon Shuffling.** Another advantage provided by introns may come from the fact that intron-exon junctions often fall at points dividing major functional regions in encoded proteins, as they do in the genes for antibody proteins, hemoglobin blood proteins, and the peptide hormone insulin. The functional divisions may have allowed new proteins to evolve by **exon shuffling**, a process by which existing protein regions or domains, already selected for their functions by the evolutionary process, are mixed into novel combinations to create new proteins. Evolution of new proteins by this mechanism would produce changes much more quickly and efficiently than by alterations in individual amino acids at random points. The process resembles automobile design, in which new models are produced by combining proven parts and substructures of previous models, rather than starting with an entirely new design each year.

### Study Break

1. What are the similarities and differences between pre-mRNAs and mRNAs?
2. What is the role of snRNPs in mRNA splicing?

## 15.4 Translation: mRNA-Directed Polypeptide Synthesis

Translation is the assembly of amino acids into polypeptides. In prokaryotes, translation takes place throughout the cell, while in eukaryotes it takes place mostly in the cytoplasm, although, as you will see, a few specialized genes are transcribed and translated in mitochondria and chloroplasts.

**Figure 15.11** summarizes the translation process. For prokaryotes, the mRNA produced by transcription is immediately available for translation. For eukaryotes, the mRNA produced by splicing of the pre-mRNA first exits the nucleus, and then is translated in the cytoplasm. In translation, the mRNA associates with a ribosome, and tRNAs, another type of RNA, bring amino acids to the complex to be joined one by one into the polypeptide chain. The sequence of amino acids in the polypeptide chain is determined by the sequence of codons in the mRNA, while the ribosome is simply a facilitator of the translation process.

In this section we start by discussing the key players in the process, the tRNAs and ribosomes, and then walk through the translation process from a start codon to a stop codon.

### tRNAs Are Small, Highly Specialized RNAs That Bring Amino Acids to the Ribosome

**Transfer RNAs (tRNAs)** bring amino acids to the ribosome for addition to the polypeptide chain.

**tRNA Structure.** tRNAs are small RNAs, about 75 to 90 nucleotides long (mRNAs are typically hundreds of nucleotides long), with a highly distinctive structure that accomplishes their role in translation **(Figure 15.12).** All tRNAs can wind into four double-helical segments, forming in two dimensions what is known as the *cloverleaf* pattern. At the tip of one of the double-helical segments is the **anticodon**, the three-nucleotide segment that pairs with a codon in mRNAs. Opposite the anticodon, at the other end of the cloverleaf, is a double-helical segment that links to the amino acid corresponding to the anticodon. For example, a tRNA that is linked to serine (Ser) pairs with the codon 5′-AGU-3′ (see Figure 15.11). The anticodon of the tRNA that pairs with this codon is 3′-UCA-5′. (The an-

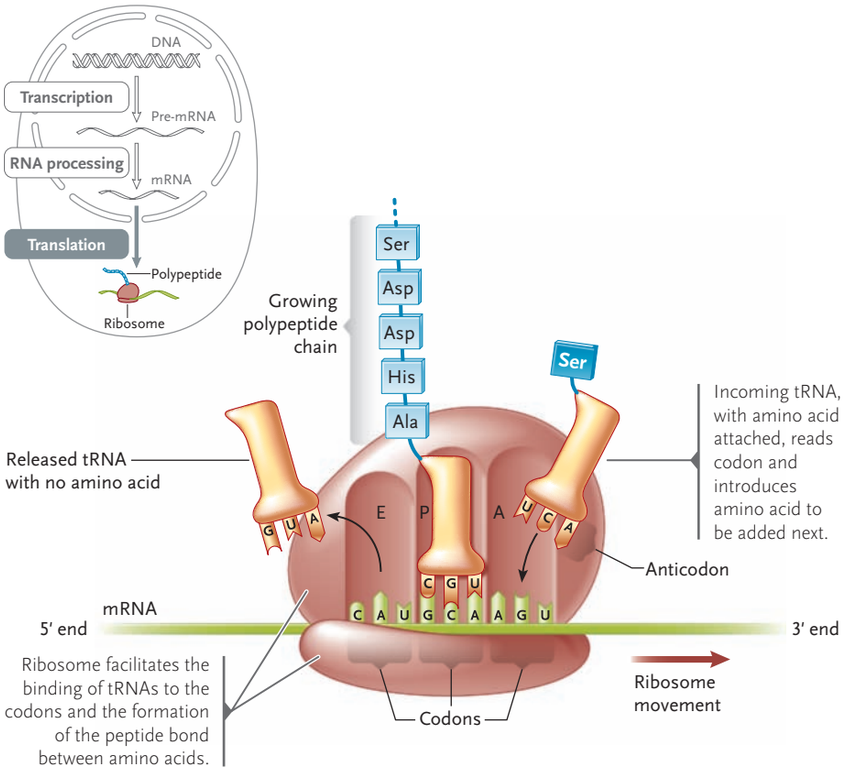
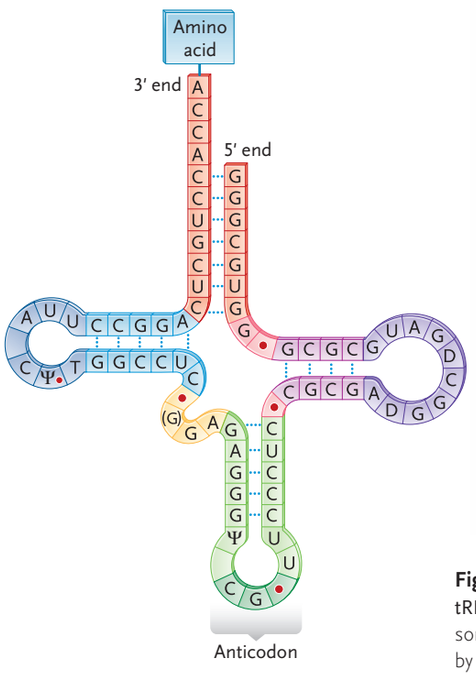
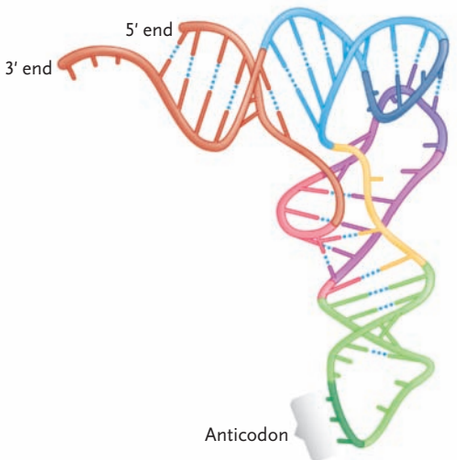



**Figure 15.11**
**An overview of translation, in which ribosomes assemble amino acids into a polypeptide chain.** The figure shows a ribosome in the process of translation. A tRNA molecule with an amino acid bound to it is entering the ribosome on the right. The anticodon on the tRNA will pair with the codon in the mRNA. Its amino acid will then be added to the growing polypeptide that is currently attached to the tRNA in the middle of the ribosome.

**a.** A tRNA molecule in two dimensions (yeast alanine tRNA)




**b.** A tRNA molecule in three dimensions




**c.** How an amino acid–tRNA complex is shown in this book

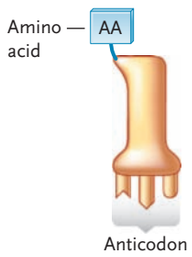



**Figure 15.12**
**tRNA structure.** The red dots show sites where bases are chemically modified into other forms. Note that some of the base pairs are unusual combinations such as G-A and G-U; these unusual base pairs, allowed by the greater flexibility of short RNA chains, are common in tRNAs.

ticodon and codon pair in an antiparallel manner, as do the strands in DNA. We will write anticodons in the 3′→5′ direction to make it easy to see how they pair with codons.)

The tRNA cloverleaf folds in three dimensions into the L-shaped structure shown in Figure 15.12b. The anticodon and the segment binding the amino acid are located at the opposite tips of the L.

We learned earlier that 61 of the 64 codons of the genetic code specify an amino acid. Does this mean that 61 different tRNAs read the sense codons? The answer is no. Francis Crick's **wobble hypothesis** states that the complete set of 61 sense codons can be read by fewer than 61 distinct tRNAs because of particular pairing properties of the bases in the anticodons. That is, the pairing of the anticodon with the first two nucleotides of the codon is always precise, but the anticodon has more flexibility in pairing with the third nucleotide of the codon. In many cases the same tRNA anticodon can read codons that have either U or C in the third position; for example, a tRNA carrying phenylalanine can read both codons UUU and UUC. Similarly the same tRNA anticodon can read two codons that have A or G in the third position; for example, a tRNA carrying glutamine can read both codons CAA and CAG.

**Addition of Amino Acids to Their Corresponding tRNAs.** The correct amino acid must be present on a tRNA if translation is to be accurate. The process of adding an amino acid to a tRNA is called **aminoacylation** (literally, the addition of an amino acid) or **charging** (because the process adds free energy as the amino acid-tRNA combinations are formed).

The finished product of charging, a tRNA linked to its "correct" amino acid, is called an **aminoacyl-tRNA.** Twenty different enzymes called **aminoacyl-tRNA synthetases**—one synthetase for each of the 20 amino acids—catalyze aminoacylation **(Figure 15.13).** First, a molecule of ATP and the amino acid (AA) bind to the enzyme, and the enzyme links the two, with the release of two phosphate groups (Figure 15.13, step 1):

$$AA + ATP \rightarrow AA\text{-}AMP + 2\ P_i$$

Much of the energy released by the breakdown is retained in the aminoacyl-AMP molecule. Next, the correct tRNA binds to the enzyme (step 2). Third, the enzyme transfers the amino acid from the AA-AMP to the tRNA to form the aminoacyl-tRNA (step 3):

$$AA\text{-}AMP + tRNA \rightarrow AA\text{-}tRNA + AMP$$

Finally, the charged tRNA is released from the enzyme; the enzyme can then perform other activation reactions (step 4). The aminoacyl-tRNA also retains much of the energy released by ATP breakdown. This energy eventually drives the formation of the peptide bond linking amino acids during translation.

With the tRNAs attached to their corresponding amino acids, our attention moves to the ribosome, where the amino acids are removed from their tRNAs and linked into polypeptide chains.

## Ribosomes Are rRNA-Protein Complexes That Work as Automated Protein Assembly Machines

**Ribosomes** are ribonucleoprotein particles that carry out protein synthesis by translating mRNA into chains of amino acids. Like some automated machines, such as those forming complicated metal parts by a series of machining steps, ribosomes use an information tape—an mRNA molecule—as the directions required to accomplish a task. For ribosomes, the task is joining amino acids in ordered sequences to make a polypeptide chain.

In prokaryotes, ribosomes carry out their assembly functions throughout the cell. In eukaryotes, ribosomes function in the cytoplasm, either suspended freely in the cytoplasmic solution, or attached to the membranes of the endoplasmic reticulum (ER), the system of tubular or flattened sacs in the cytoplasm (discussed in Section 5.5).

A finished ribosome is made up of two parts of dissimilar size, called the *large* and *small ribosomal subunits* **(Figure 15.14).** Each subunit is a combination of **ribosomal RNA (rRNA)** and ribosomal proteins.

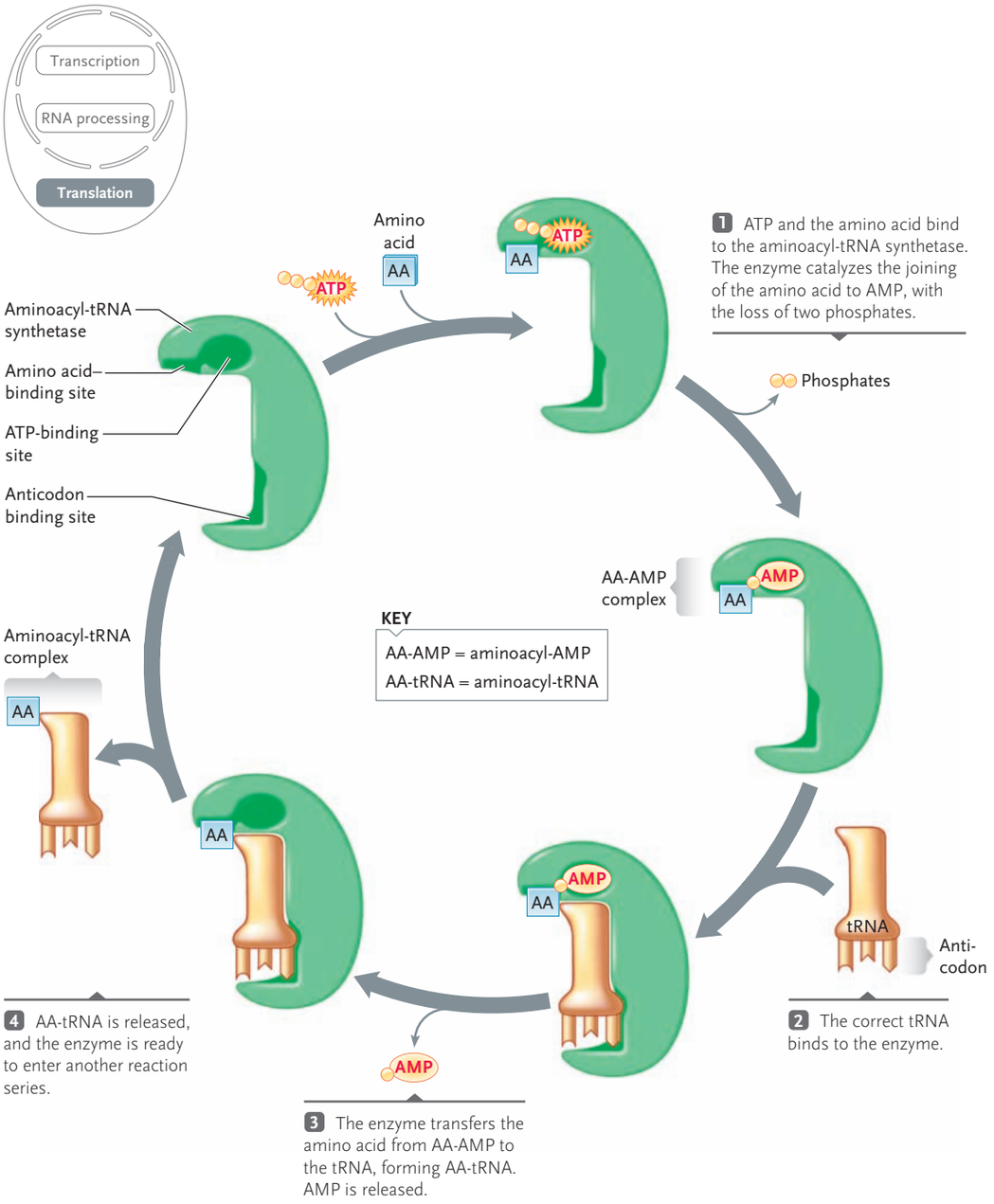
Prokaryotic and eukaryotic ribosomes are similar in structure and function. However, the differences in their molecular structure, particularly in the ribosomal proteins, give them distinguishable properties. For example, the antibiotics streptomycin and erythromycin are effective antibacterial agents because they inhibit the function of the bacterial ribosome, but not the eukaryotic ribosome. Streptomycin and erythromycin affect translation activities in the small and large ribosomal subunit, respectively.

In translation, the mRNA moves through a groove in the ribosome. The ribosome also has binding sites where tRNAs interact with the mRNA (see Figure 15.14 and refer also to Figure 15.11). The **A site** (aminoacyl site) is where the incoming aminoacyl-tRNA carrying the next amino acid to be added to the polypeptide chain binds to the mRNA. The **P site** (peptidyl site) is where the tRNA carrying the growing polypeptide chain is bound. The **E site** (exit site) is where an exiting tRNA binds prior to release from the ribosome. You will learn more about these functional sites as we discuss the stages of translation.

## Translation Initiation Brings the Ribosomal Subunits, an mRNA, and the First Aminoacyl-tRNA Together

Translation is similar in prokaryotes and eukaryotes. We will present translation from a eukaryotic perspective, and indicate how it differs in prokaryotes.

**Figure 15.13**
Aminoacylation (aka charging): the addition of an amino acid to a tRNA.




There are three major stages of translation: initiation, elongation, and termination. During initiation the translation components assemble on the start codon of the mRNA. In elongation the assembled complex reads the string of codons in the mRNA one at a time while joining the specified amino acids into the polypeptide. Termination completes the translation process when the complex disassembles after the last amino acid of the polypeptide specified by the mRNA has been added to the polypeptide.

In translation initiation, a large and a small ribosomal subunit associates with an mRNA molecule and the first aminoacyl-tRNA of the new protein chain binds to the AUG start codon **(Figure 15.15).** The aminoacyl-tRNA used for initiation is a specialized **initiator tRNA** with an anticodon to the methionine-specifying AUG start codon. Each step in translation initiation is aided by proteins called *initiation factors*.

In the first step of the initiation process, the initiator methionine-tRNA (Met-tRNA—anticodon 3′-UAC-5′) with a molecule of GTP bound to it forms a complex with the small ribosomal subunit (Figure 15.15, step 1). The complex binds to the mRNA at the 5′ cap and then moves along the mRNA—a process called *scanning*—until it

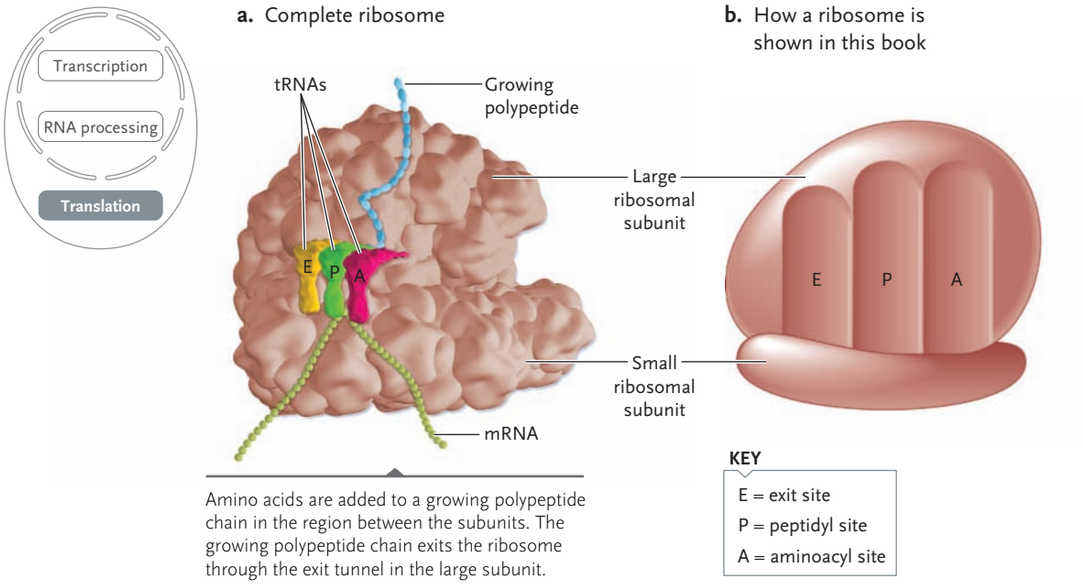



**Figure 15.14**
Ribosome structure. **(a)** Computer model of a ribosome in the process of translation. **(b)** The ribosome as we will show it during translation. (a: Michael W. Davidson/ Molecular Expressions, Florida State Research Foundation.)

reaches the first AUG codon (step 2). This is the start codon and it is recognized by the anticodon of the Met-tRNA. The large ribosomal subunit then binds, completing the ribosome (step 3). The initiation factors are released when GTP is hydrolyzed to GDP + phosphate. At the end of initiation, the initiator Met-tRNA is in the P site.

In prokaryotes, translation initiation is different: Rather than scanning from the 5′ end of the mRNA, the small ribosomal subunit, the initiator Met-tRNA, GTP, and protein initiator factors bind directly to the region of the mRNA with the AUG start codon. A **ribosome binding site** just upstream of the start codon directs the small ribosomal subunit in this initiation step. The large ribosomal subunit then binds to the small subunit to complete the ribosome.

After the initiator tRNA pairs with the AUG initiator codon, the subsequent stages of translation simply read the nucleotide bases three at a time on the mRNA. The initiator tRNA-AUG pairing thus establishes the correct *reading frame*—the series of codons for the polypeptide encoded by the mRNA.

## Polypeptide Chains Grow during the Elongation Stage of Translation

The central reactions of translation take place in the elongation stage, which adds amino acids one at a time to a growing polypeptide chain. The individual steps of elongation depend on the binding properties of the P, A, and E sites of the ribosome. Protein *elongation factors* aid the elongation events.

The P site, with one exception, can bind only to a **peptidyl-tRNA**—a tRNA linked to a growing polypeptide chain containing two or more amino acids. The exception is the initiator tRNA, which is recognized by the P site as a peptidyl-tRNA even though it carries only a single amino acid, methionine. The A site can bind only to an aminoacyl-tRNA. The tRNA that previously was in the P site binds to the E site and then leaves the ribosome.

How the P, A, and E sites are used through the elongation cycle is shown in **Figure 15.16.** The cycle begins at the point when an initiator tRNA with its attached methionine is bound to the P site, and the A site is empty. First, an aminoacyl-tRNA with an appropriate anticodon binds to the codon in the A site of the ribosome; GTP is hydrolyzed in this step (Figure 15.16, step 1).

Next, the amino acid (here, the initiator methionine) is cleaved from the tRNA in the P site and forms a peptide bond with the amino acid on the tRNA in the A site (step 2). **Peptidyl transferase** catalyzes this reaction. Researchers were surprised to discover that this enzyme is not a protein but a part of an rRNA of the large ribosomal subunit. An RNA molecule that catalyzes a reaction like a protein enzyme does is called a *catalytic RNA* or a **ribozyme** (*ribo*nucleic acid en*zyme*).

At the close of the reaction, the (now) polypeptide chain is attached to the tRNA in the A site and an "empty" tRNA remains at the P site. Next, the ribosome moves—translocates—along the mRNA to the next codon, using energy from GTP hydrolysis (step 3). The two tRNAs remain bound to their respective codons, so this step positions the just-formed peptidyl-tRNA in the P site, and generates a new vacant A site. The empty tRNA that was in the P site is now in the E site, from where it is released from the ribosome (step 4). With the A site empty and a peptidyl tRNA in the P site, the ribosome repeats the elongation cycle. In subsequent

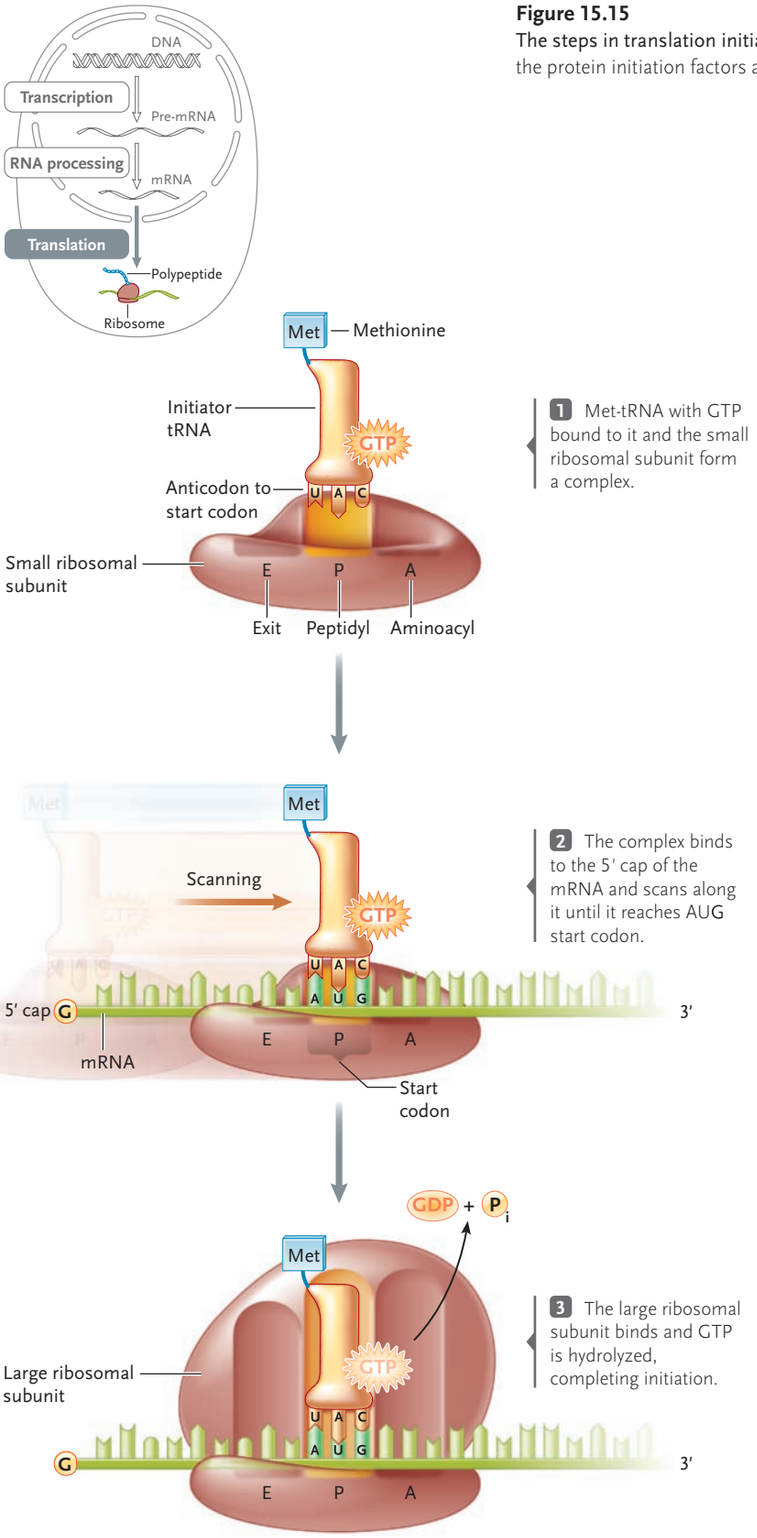



**Figure 15.15**
**The steps in translation initiation in eukaryotes.** (For simplicity, the protein initiation factors are not shown in the figure.)

turns of the cycle, the growing polypeptide on the tRNA in the P site is transferred to the amino acid on the A site tRNA. (*Insights from the Molecular Revolution* describes a technique for investigating the structure of the P and A sites.)

Elongation is highly similar in prokaryotes and eukaryotes, with no significant conceptual differences. During each elongation cycle, which turns at the rate of about one to three times per second in eukaryotes and 15 to 20 times per second in prokaryotes, the ribo-

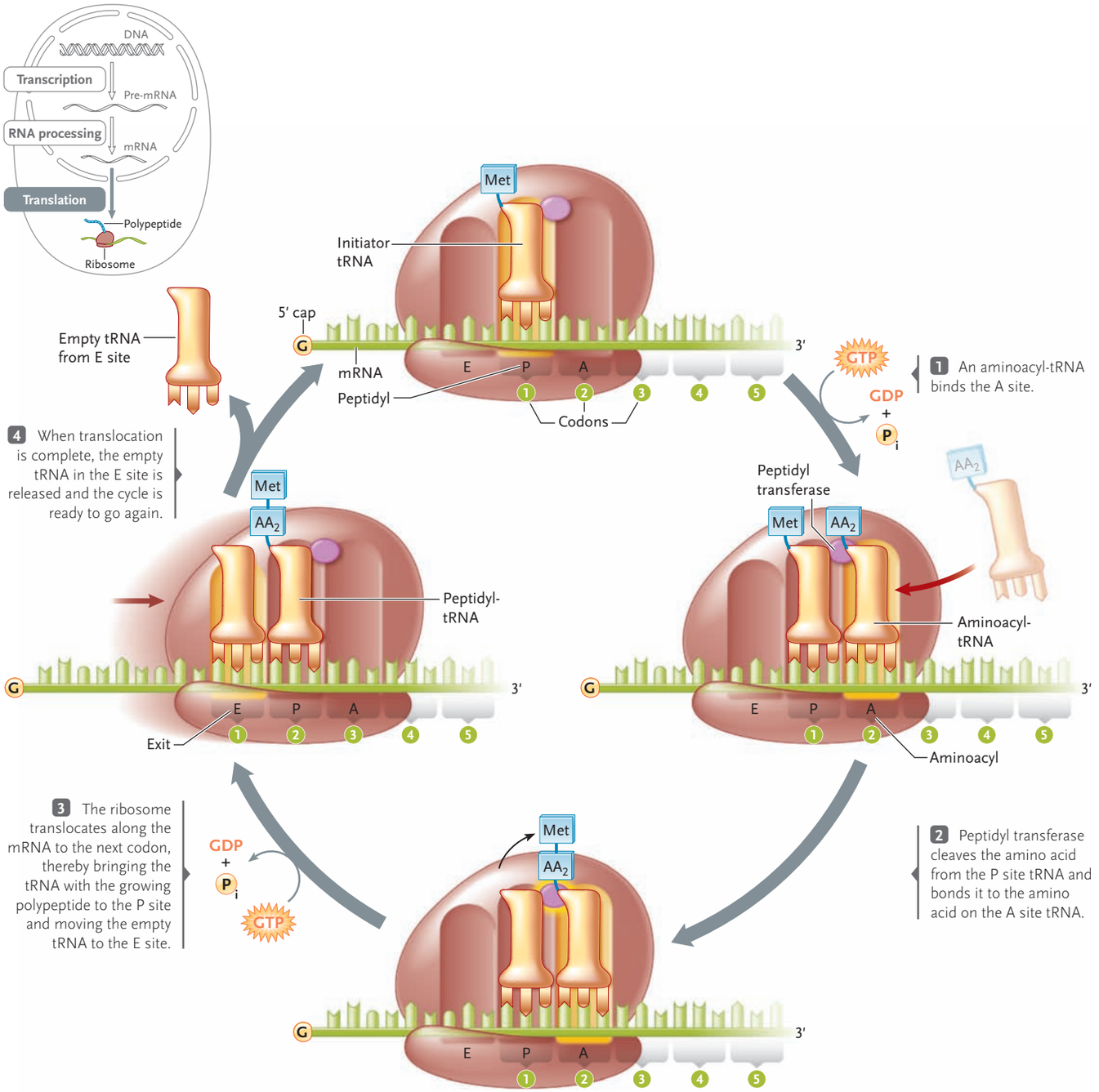



**Figure 15.16**
The steps in the elongation stage of translation. (For simplicity, protein elongation factors and GTP are not shown in the figure.)

some advances by one codon along the mRNA and adds one amino acid to the growing polypeptide chain. The growing polypeptide chain extends from the ribosome through the exit tunnel (see Figure 15.14) as elongation continues.

## Termination Releases a Completed Polypeptide from the Ribosome

Translation switches from the elongation to the termination stage when the A site of a ribosome arrives at one of the UAA, UAG, or UGA stop codons on the mRNA (**Figure 15.17,** step 1). When a stop codon appears at the A site, a **release factor** (RF; also called a **termination factor**) binds at this site instead of an aminoacyl-tRNA (step 2). In response, the polypeptide chain is released from the tRNA at the P site as usual (step 3). However, because no amino acid is present at the A site, the freed polypeptide chain is released from the ribosome (step 4). At the same time, the ribosomal subunits separate and detach from the mRNA. The empty tRNA and the release factor are also released. Termination is highly similar in prokaryotes and eukaryotes.

## Insights from the Molecular Revolution

### Measuring Ribosomes with a Molecular Ruler

One of the major problems in the research on the three-dimensional structure of ribosomes has been unraveling the structure and arrangement of the rRNA molecules of ribosomes. It is important to know this because the rRNAs play important roles in translation; the peptidyl transferase, for example, is an enzyme activity of the large rRNA of the large ribosomal subunit. The problem of determining rRNA positions is difficult because the rRNAs twist and fold throughout the large or small ribosomal subunits. Which part of each rRNA type is located in active sites of the ribosome, such as the E, P, and A sites and the peptidyl transferase site, and what are the functions of the rRNA segments that wind through these locations? An ingenious molecular approach, developed by Harry F. Noller and his coworkers at the University of California at Santa Cruz, provides a method for answering some of these questions. The technique provides a *molecular scissors* that leaves an identifying snip at points in the rRNA molecules of the ribosome.

Noller and his colleagues prepared artificial tRNAs that extend in a straight line, with an anticodon at one end and a sequence at the other end that mimics the part of a tRNA that binds to an amino acid. Instead of an amino acid, Noller's team attached an $Fe^{2+}$ atom to the acceptor end of the structure (see figure). The iron atom can break the sugar-phosphate backbone of an RNA molecule. Thus the artificial tRNAs can work as a molecular scissors, with one end able to bind to an mRNA in either the A, P, or E site, and the other end able to make an identifying snip in the part of an rRNA molecule that is located in its vicinity. And, the scissors were made in different, precisely calculated lengths, running in regular increments from 4 to 33 base pairs. The artificial tRNAs would thus act as "molecular rulers" that could measure distances between the anticodon-binding region and parts of the rRNA molecules in ribosomes.

The molecular scissors were reacted with ribosomes, one length at a time, and then the rRNAs were extracted to see where the snips occurred. To locate the sites of the snips, the researchers sorted the extracted rRNAs by gel electrophoresis, a method that is capable of separating RNA molecules that differ in

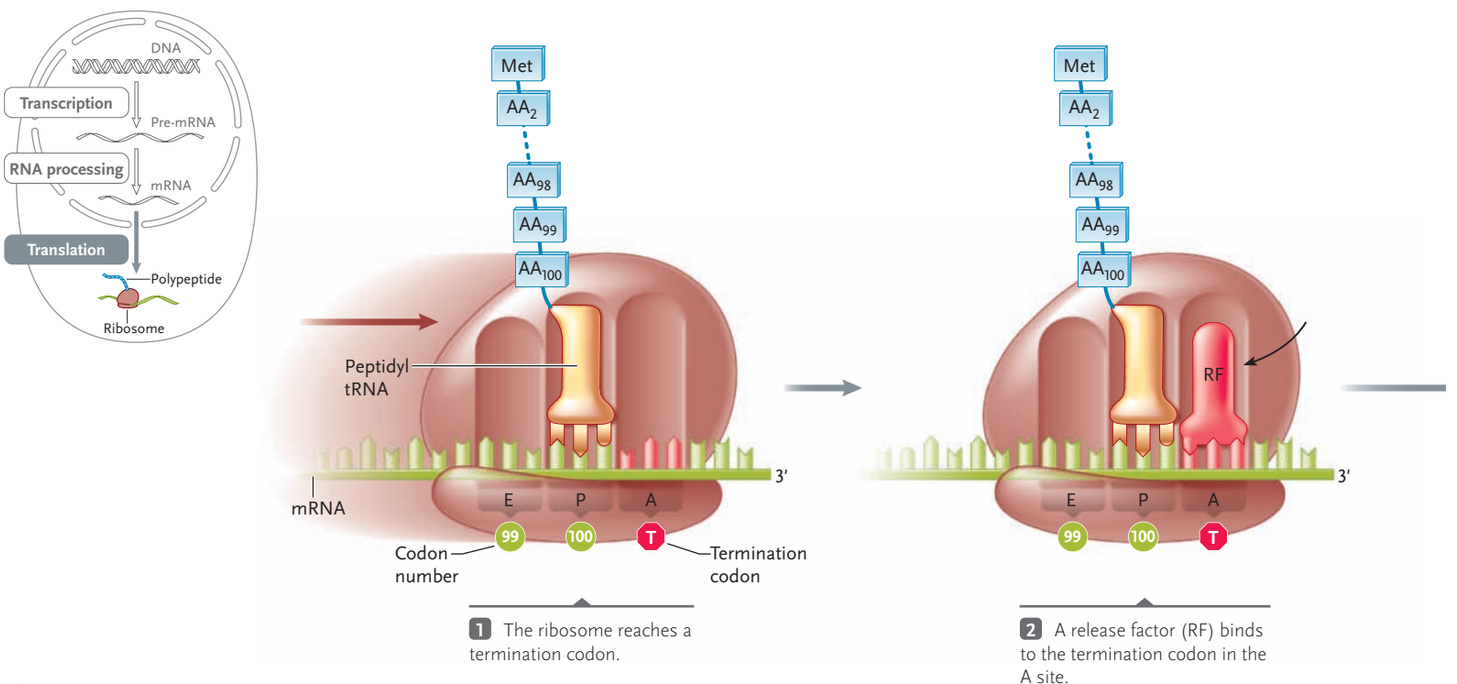
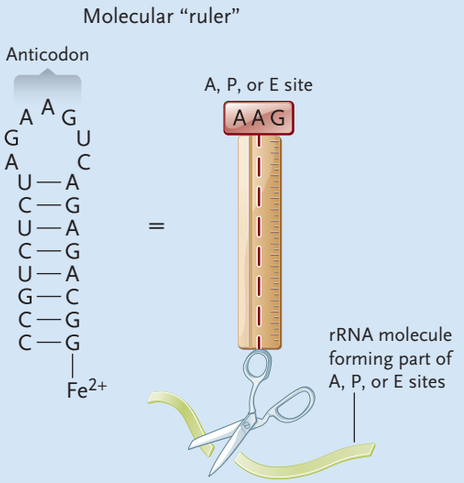



**Figure 15.17**
The steps in the termination stage of translation.

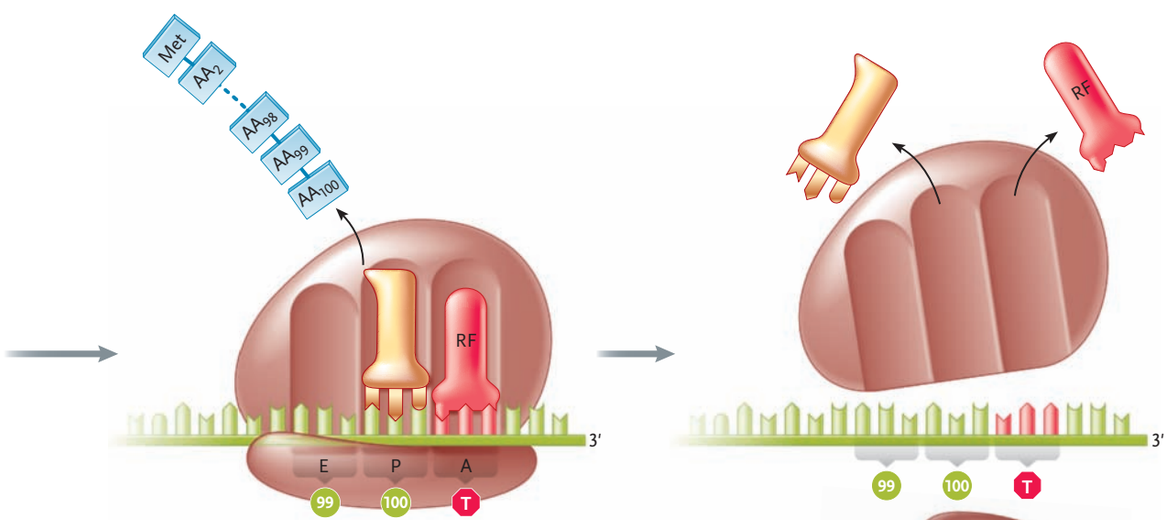
length by only a single base. (Gel electrophoresis is described in Figure 18.7.) By comparing the lengths of the snipped rRNA molecules with those of intact molecules, the investigators could locate the exact points in the sequences at which the RNAs were cut. This information allowed them to map the segments of the rRNAs forming parts of the E, P, and A sites.

The rulers were bound to bacterial ribosomes, which contain three types of rRNAs: 5S; 23S, which forms part of the large subunit; and 16S, which forms part of the small subunit. (The S values are a measure of molecular size, with the larger numbers indicating larger size.) When the rulers attached to the A site, the snipping end of the artificial tRNAs made cuts in both 23S rRNA and 16S rRNA, showing that both the large and small ribosomal subunits cooperate to form the A site.

When bound to the P site, the molecular rulers made cuts in all three rRNA types. These results show that, as in the A site, both the large and small subunits cooperate to form the P site and that all three rRNA types evidently form parts of the P site. The locations of the cuts also gave clues about the three-dimensional structure of the 16S and other rRNAs in the A and P sites. This method may allow the three-dimensional structure of the E, P, and A sites to be reconstructed as well as identify the rRNA types forming the sites. The next step is to try to work out the functions of these segments in polypeptide assembly.




**A molecular ruler used to measure the arrangement of tRNAs and rRNAs in ribosomes.** By varying the length of the ruler, rRNA molecules located at different points within the A, P, and E sites can be marked by cuts.




**3** The polypeptide chain is released from the peptidyl-tRNA in the P site.

**4** The empty tRNA and release factor are released, and the ribosomal subunits separate.

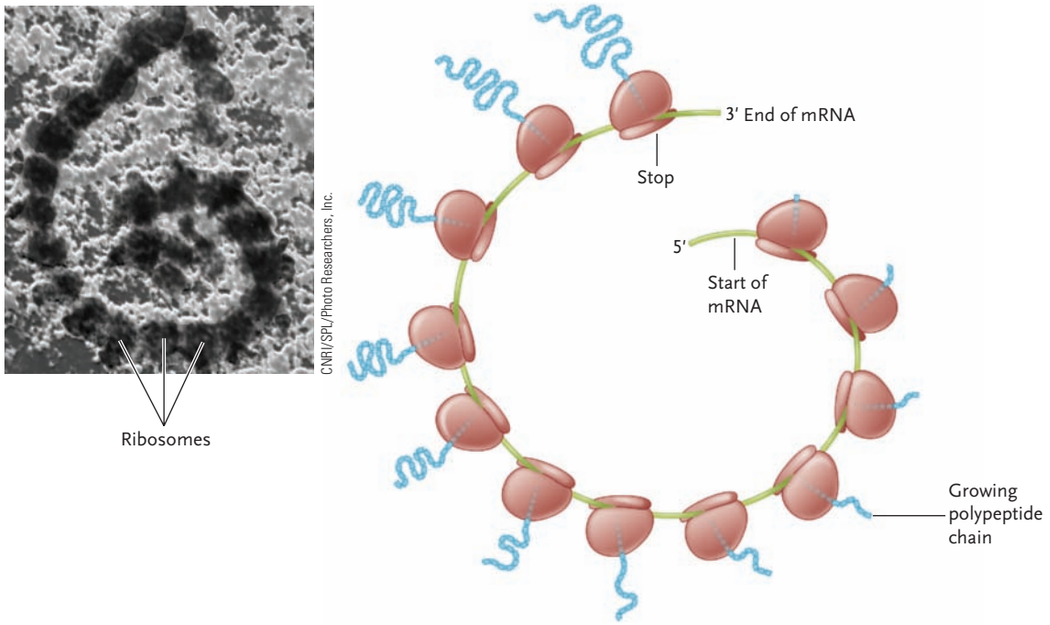



**Figure 15.18**
Polysomes, consisting of a series of ribosomes "reading" the same mRNA.

## Multiple Ribosomes Simultaneously Translate a Single mRNA

Once the first ribosome has begun translation, another one can assemble with an initiator tRNA as soon as there is room on the mRNA. Ribosomes continue to attach as translation continues and become spaced along the mRNA like beads on a string. The entire structure of an mRNA molecule and the multiple ribosomes attached to it is known as a **polysome** (a contraction of *polyribosome;* **Figure 15.18**). The multiple ribosomes greatly increase the overall rate of polypeptide synthesis from a single mRNA. The total number of ribosomes in a polysome depends on the length of the coding region of its mRNA molecule, ranging from a minimum of one or two ribosomes on the smallest mRNAs to as many as 100 on the longest mRNAs.

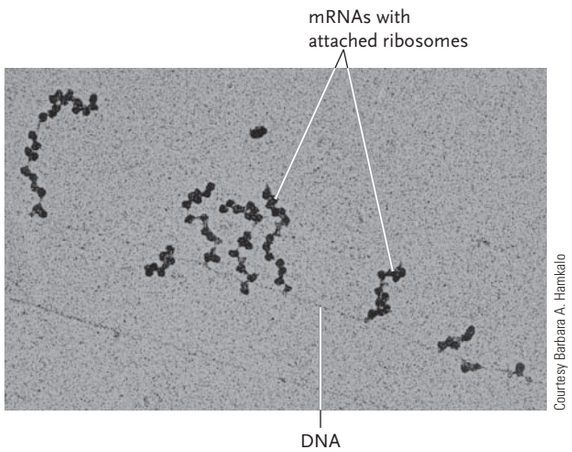



**Figure 15.19**
Simultaneous transcription and translation in progress in an electron microscope preparation extracted from *E. coli.* ×57,000.

In prokaryotes, because of the absence of a nuclear envelope, transcription and translation typically are coupled. That is, as soon as the 5′ end of a new mRNA emerges from the RNA polymerase, ribosomal subunits attach and initiate translation **(Figure 15.19).** In essence the polysome forms while the mRNA is still being made. By the time the mRNA is completely transcribed, it is covered with ribosomes from end to end, each assembling a copy of the encoded polypeptide.

## Newly Synthesized Polypeptides Are Processed and Folded into Finished Form

Most eukaryotic proteins are in an inactive, unfinished form when ribosomes release them. Processing reactions that convert the new proteins into finished form include the removal of amino acids from the ends or interior of the polypeptide chain and the addition of larger organic groups, including carbohydrate or lipid structures.

Proteins fold into their final three-dimensional shapes as the processing reactions take place. For many proteins, helper proteins called *chaperones* or *chaperonins* assist the folding process by combining with the folding protein, promoting "correct" three-dimensional structures, and inhibiting incorrect ones (see Section 3.5 and Figure 3.24).

In some cases the same initial polypeptide may be processed by alternative pathways that produce different mature polypeptides, usually by removing different, long stretches of amino acids from the interior of the polypeptide chain. Alternative processing is another mechanism that increases the number of polypeptides encoded by a single gene.

Other proteins are processed into an initial, inactive form that is later activated at a particular time or location by removal of a covering segment of the amino acid chain. The digestive enzyme pepsin, for example, is made by processing reactions within cells lining the stomach into an inactive form called *pepsinogen*. When the cells secrete pepsinogen into the stomach, the high acidity of that organ triggers removal of a segment of amino acids from one end of the protein's amino acid chain; the removal converts the enzyme into the active form in which it rapidly breaks proteins in food particles into shorter pieces. The initial production of the protein as inactive pepsinogen protects the cells that make it from having their proteins degraded by the enzyme.

## Finished Proteins Contain Sorting Signals That Direct Them to Cellular Locations

Proteins are found in all parts of the eukaryotic cell, including the cytosol, the nucleus, the plasma membrane, and the membranes or interior of various organelles; they are also transported to the cell exterior. How are newly synthesized proteins directed to these locations? Proteins that remain in the cytosol, such as microtubule proteins or the enzymes of glycolysis, have no signals. They are made on ribosomes called *free ribosomes,* which are suspended in the cytosol, and they enter the cytosol as they are made.

For all other proteins, a system of "zip codes," in the form of amino acid sequences that form *sorting signals* in the proteins, direct the proteins to their cellular locations, or out of the cell. The signals are coded in the DNA, transcribed into mRNAs, and "printed" in proteins as they are made. The signals, first discovered by Günter Blobel, Peter Walter, and their coworkers at Rockefeller University, are recognized and bound by receptors in the locations to which the proteins are addressed. Blobel received a Nobel Prize in 1999 for his work with the mechanism sorting proteins in cells.

One major signal pathway sorts proteins to the endoplasmic reticulum **(Figure 15.20).** In these proteins, a short segment of amino acids called the **signal peptide** (also called a **signal sequence**) is in the very first part of the polypeptide chain to be made. When the signal peptide emerges from the ribosome, a protein-RNA complex called the **signal recognition particle (SRP)** binds to it and temporarily blocks further translation (Figure 15.20, step 1). Next, the SRP binds a protein in the ER membrane called the **SRP receptor;** this step "docks" the ribosome on the ER membrane (step 2). (The docked ribosomes give the dotted appearance to the rough ER; see Section 5.3.) The ribosome now can continue protein synthesis, and the growing polypeptide is pushed through the ER membrane into the rough ER lumen (see step 2) where an enzyme, *signal peptidase,* removes the signal sequence (step 3). Synthesis of the polypeptide continues until it is complete (step 4). Depending on other built-in signals, the polypeptide may move to any part of the ER-based system: the ER itself, the Golgi complex, the plasma membrane,

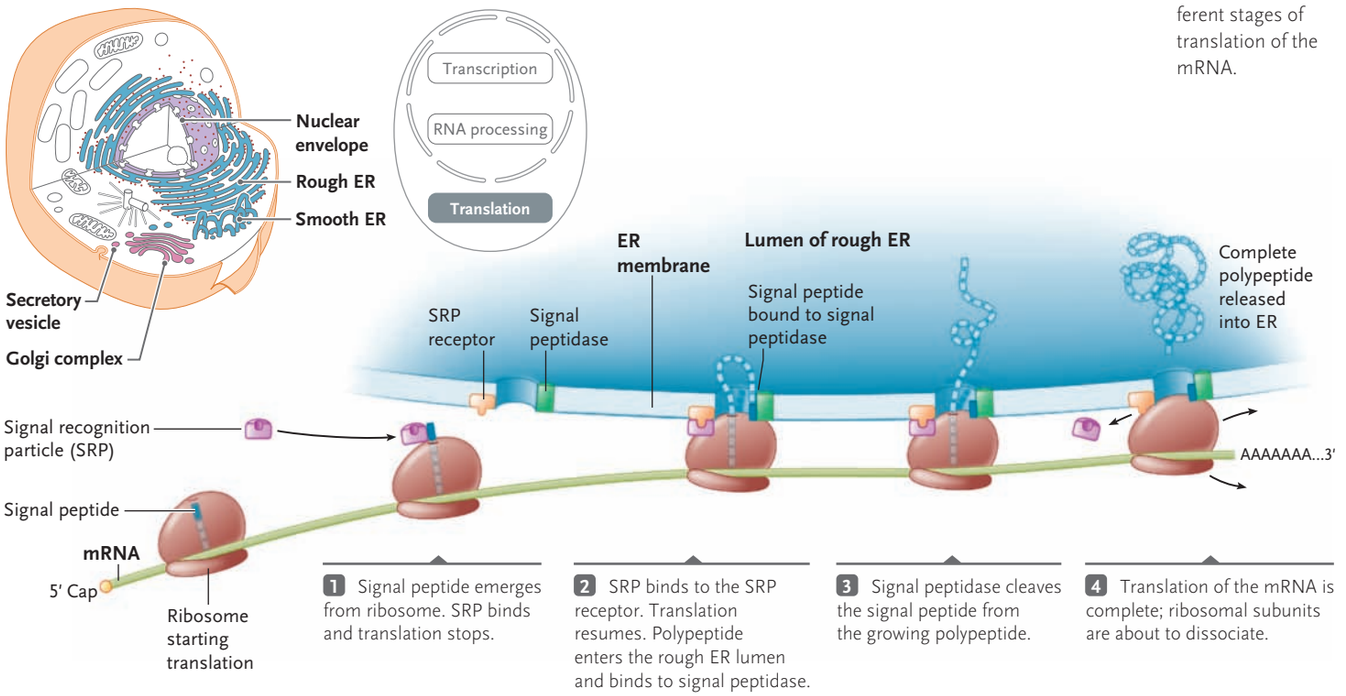



**Figure 15.20**
**The signal mechanism directing proteins to the ER.** The figure shows several ribosomes at different stages of translation of the mRNA.

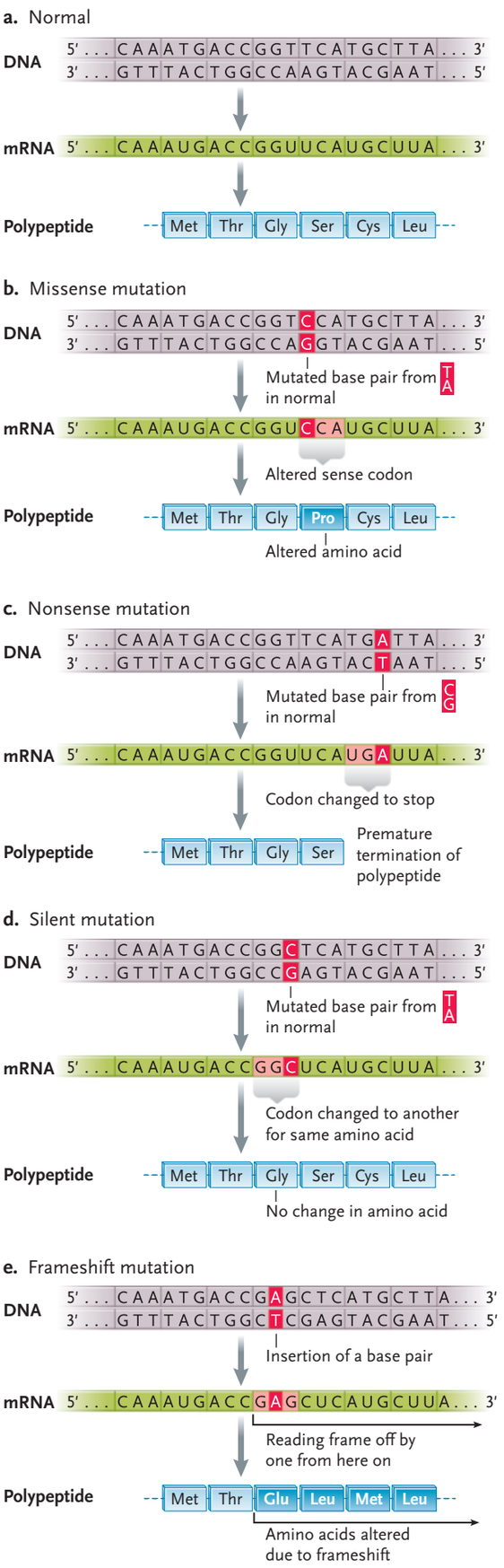



**Figure 15.21**
Effects of mutations in protein-coding genes on the amino acid sequence of the encoded polypeptide.

the nuclear envelope, secretory vesicles, or via secretory vesicles to the cell exterior (these destinations are shown in the inset to Figure 15.20).

Nuclear proteins include a signal that is bound by receptors in the pore complexes of the nuclear envelope (shown in Figure 5.11). Once bound, they are pushed through the pore complex into the nuclear interior, in a process that requires ATP energy. For these proteins, the signal remains because the proteins need to enter the nucleus each time the nuclear envelope breaks down and reforms during the cell division cycle.

Many proteins that are to become part of organelles such as mitochondria, chloroplasts, or microbodies are also made on free ribosomes. However, these proteins have signals that are bound by receptors in the organelle membranes, targeting them for entry into the organelles. Further signals on the proteins direct them to the different membranes or compartments inside the organelles.

The sorting system, in all its remarkable complexity, routes newly synthesized proteins and gets them to their final destinations in the eukaryotic cell. Without it, cells would wind up as a jumble of proteins floating about in the cytoplasm, with none of the spatial organization that makes cellular life possible.

The same basic system of sorting signals distributes proteins throughout prokaryotic cells, indicating that this mechanism probably evolved with the first cells. In prokaryotes, signals similar to the ER-directing signals of eukaryotes direct newly synthesized bacterial proteins to the plasma membrane (bacteria do not have ER membranes); further information built into the proteins keeps them in the plasma membrane or allows them to enter the cell wall or to be secreted outside the cell. Proteins without sorting signals remain in the cytoplasmic solution.

Interestingly, the bacterial and eukaryotic routing signals are interchangeable. That is, a bacterial signal peptide grafted to a polypeptide made in a eukaryotic cell routes the molecule to the ER membrane, and a eukaryotic ER-directing signal peptide grafted to a polypeptide made in a bacterial cell directs the molecule to the plasma membrane. The interchangeability of the bacterial and eukaryotic signal peptides indicates that the sorting mechanism appeared early in the evolution of cellular life.

## Base-Pair Mutations Can Affect Protein Structure and Function

**Mutations** are changes to the genetic material. How do mutations affect protein structure and function? **Base-pair substitution mutations** are particular mutations involving changes to individual base pairs in the genetic material. If a base-pair substitution mutation occurs in the protein-coding portion of a gene, it can affect the structure and function of the protein.

That is, a base-pair change will change a base in a codon.

Let us consider a theoretical stretch of DNA encoding a string of amino acids in a polypeptide to see the possible consequences of base-pair substitution mutations in the coding region of a gene. The normal (unmutated) DNA and amino acid sequences are shown in **Figure 15.21a.** If the codon is altered to specify a different amino acid, then the protein will have a different amino acid sequence. A mutation such as this is called a **missense mutation**, because the amino acid that is placed in the polypeptide is not the normal one **(Figure 15.21b).** Whether the function of a polypeptide is altered significantly depends on the amino acid change that occurs. Individuals homozygous for a missense mutation in the gene for one of the two polypeptide types found in the oxygen-carrying protein hemoglobin **(Figure 15.22)** have the genetic disease sickle-cell disease, described in Chapter 12 (pp. 235–236). Many other human genetic diseases are caused by homozygous missense mutations.

A second type of base-pair substitution mutation is a **nonsense mutation (Figure 15.21c).** In this case the base-pair change in the DNA results in a change from a sense (amino acid–coding) codon to a nonsense (termination) codon in the mRNA. Translation of an mRNA containing a nonsense mutation results in a shorter-than-normal polypeptide and, in many cases, this polypeptide will be only partially functional at best.

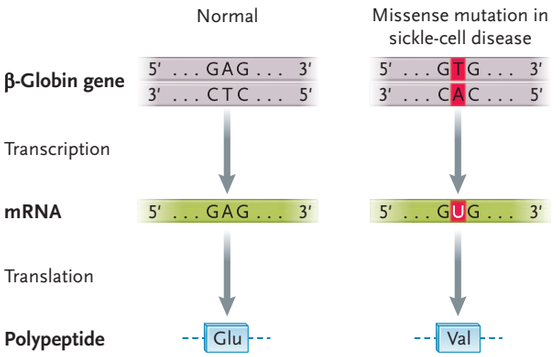



**Figure 15.22**
Missense mutation in a gene for one of the two polypeptides of hemoglobin that is the cause of sickle-cell disease.

Because of the degeneracy of the genetic code, some base-pair substitution mutations do not alter the amino acid specified by the gene because the changed codon specifies the same amino acid as in the normal polypeptide. Such mutations are known as **silent mutations (Figure 15.21d).**

If a single base pair is deleted or inserted in the coding region of a gene, the reading frame of the resulting mRNA is altered. That is, after that point, the ribosome reads codons that are not the same as for the normal mRNA, producing a different amino acid sequence in the polypeptide from then on. This type of

## Unanswered Questions

### What are the structures and specific functions of the proteins involved in transcription?

Even though molecular geneticists have a good understanding of the initiation, elongation, and termination stages of transcription, they have a lot to learn about the precise molecular events involved. Researchers are actively exploring the molecular structures of RNA polymerase, the transcription factors, and other proteins, and are developing models for how they interact to initiate transcription and how that interaction is controlled. For example, in yeast, the transcription machinery assembled at a promoter includes nearly 100 polypeptides! What does each of these components do? How are they assembled and positioned correctly on the template? Among the researchers studying this topic is Richard Ebright, a Howard Hughes Medical Institute investigator at Rutgers University.

### How is alternative splicing regulated?

The discovery that a majority of human genes show alternative splicing was very surprising. How is alternative splicing regulated? This is not a peculiarly human question, because alternative splicing is prevalent in all eukaryotes. Not only will this research provide basic information about the mechanism of mRNA splicing, but it is likely also to have significant implications for human medicine. For example, the same disease gene often produces variable symptoms in different individuals. Is this because of effects on the regulation of alternative splicing resulting in different defective proteins with different effects? If so, how does this altered regulation occur, and what regulates it? The answers to these questions would open the door to the development of drugs to treat the disease symptoms.

### How does the ribosome work?

The ribosome is a large complex of rRNAs and proteins. Research in recent years resulted in the unexpected conclusion that the rRNAs are functionally important in protein synthesis, rather than simply being a scaffold on which the ribosomal proteins hang. Peptidyl transferase activity is the property of rRNA, for instance, as we discussed in this chapter. Current research is directed at understanding how the three-dimensional structure of the ribosome gives it its functional properties. For example, how does the ribosome translocate along the mRNA? This process must involve changes in the three-dimensional shape of the ribosome. What are the ribosome's moving parts? A surprising result about this from Harry Noller's lab at the University of California, Santa Cruz (see *Insights from the Molecular Revolution* for other research from this lab), indicates that translocation is a property of the ribosome itself and not of the protein factors and GTP. The molecular basis of translocation is being intensely researched in light of this new knowledge.

Peter J. Russell

mutation is called a **frameshift mutation** (**Figure 15.21e,** insertion mutation shown); the resulting polypeptide often is nonfunctional because of the significantly altered amino acid sequence.

Both transcription and translation are steps in the process of gene expression, the realization of the gene's coded information in the makeup and activities of a cell. However, the flow of information is not one way; organisms and cells also exert control over how their genes are expressed, as you will see in the next chapter.

### Study Break

1. How does translation initiation occur in eukaryotes versus prokaryotes?
2. Distinguish between the P, A, and E sites of the ribosome.
3. How are proteins directed to different parts of a eukaryotic cell?

## Review

Go to ThomsonNOW™ at www.thomsonedu.com/login to access quizzing, animations, exercises, articles, and personalized homework help.

### 15.1 The Connection between DNA, RNA, and Protein

- In their genetic experiments with *Neurospora crassa,* Beadle and Tatum found a direct correspondence between gene mutations and alterations of enzymes. Their one gene–one enzyme hypothesis is now restated as the one gene–one polypeptide hypothesis (Figure 15.2).
- The pathway from genes to proteins involves transcription then translation. In transcription, a sequence of nucleotides in DNA is copied into a complementary sequence in an RNA molecule. In translation, the sequence of nucleotides in an mRNA molecule specifies an amino acid sequence in a polypeptide (Figure 15.3).
- The genetic code is a triplet code. AUG at the beginning of a coded message establishes a reading frame for reading the codons three nucleotides at a time. The code is redundant: most of the amino acids are specified by more than one codon (Figures 15.4 and 15.5).
- The genetic code is essentially universal.

**Animation: Uracil–thymine comparison**

**Animation: Protein synthesis summary**

**Practice: The major differences between prokaryotic and eukaryotic protein synthesis**

### 15.2 Transcription: DNA-Directed RNA Synthesis

- Transcription is the process by which information coded in DNA is transferred to a complementary RNA copy (Figure 15.6).
- Transcription begins when an RNA polymerase binds to a promoter sequence in the DNA and starts synthesizing an RNA molecule. The enzyme then adds RNA nucleotides in sequence according to the DNA template. At the end of the transcribed sequence, the enzyme and the completed RNA transcript release from the DNA template. The mechanism of termination is different in eukaryotes and prokaryotes (Figure 15.7).

**Animation: Gene transcription details**

### 15.3 Production of mRNAs in Eukaryotes

- A gene encoding an mRNA molecule includes the promoter, which is recognized by the regulatory proteins and transcription factors that promote DNA unwinding and the initiation of transcription by an RNA polymerase. Transcription in eukaryotes produces a pre-mRNA molecule that consists of a 5′ cap, the 5′ untranslated region, interspersed exons (amino acid-coding segments) and introns, the 3′ untranslated region, and the 3′ poly(A) tail. All are copied from DNA except the 5′ cap and poly(A) tail, which are added during transcription (Figure 15.8).
- Introns in pre-mRNAs are removed to produce functional mRNAs by splicing. snRNPs bind to the introns, loop them out of the pre-mRNA, clip the intron at each exon boundary, and join the adjacent exons together (Figure 15.9).
- Many pre-mRNAs are subjected to alternative splicing, a process that joins exons in different combinations to produce different mRNAs encoded by the same gene. Translation of each mRNA produced in this way generates a protein with different function (Figure 15.10).

**Animation: Pre-mRNA transcript processing**

### 15.4 Translation: mRNA-Directed Polypeptide Synthesis

- Translation is the assembly of amino acids into polypeptides. Translation occurs on ribosomes. The P, A, and E sites of the ribosome are used for the stepwise addition of amino acids to the polypeptide as directed by the mRNA (Figures 15.11 and 15.14).
- Amino acids are brought to the ribosome attached to specific tRNAs. Amino acids are linked to their corresponding tRNAs by aminoacyl-tRNA synthetases. By matching amino acids with tRNAs, the reactions also provide the ultimate basis for the accuracy of translation (Figures 15.12 and 15.13).
- Translation proceeds through the stages of initiation, elongation, and termination. In initiation, a ribosome assembles with an mRNA molecule and an initiator methionine-tRNA. In elongation, amino acids linked to tRNAs add one at a time to the growing polypeptide chain. In termination, the new polypeptide is released from the ribosome and the ribosomal subunits separate from the mRNA (Figures 15.15–15.17).
- After they are synthesized on ribosomes, polypeptides are converted into finished form by processing reactions, which include removal of one or more amino acids from the protein chains, addition of organic groups, and folding guided by chaperones.
- Proteins are distributed in cells by means of signals spelled out by amino acid sequences (Figure 15.20).
- Base-pair substitution mutations alter the mRNA and can lead to changes in the amino acid sequence of the encoded polypeptide. A missense mutation changes one sense codon to one that specifies a different amino acid, a nonsense mutation changes a sense

codon to a stop codon, and a silent mutation changes one sense codon to another sense codon that specifies the same amino acid. A base-pair insertion or deletion is a frameshift mutation that alters the reading frame beyond the point of the mutation, leading to a different amino acid sequence from then on in the polypeptide (Figures 15.21 and 15.22).

**Animation: Structure of a ribosome**

**Animation: Translation**

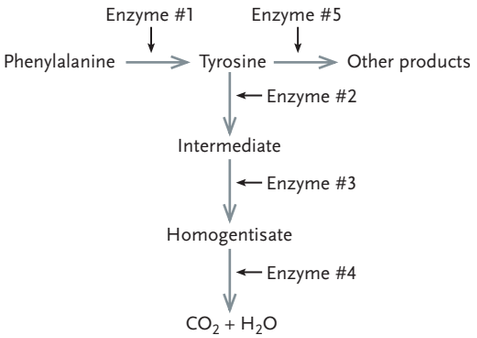
**Animation: Base-pair substitution**

**Animation: Frameshift mutation**

# Questions

## Self-Test Questions

1. Which statement about the following pathway is false?




   a. A mutation for enzyme #1 causes phenylalanine to build up.
   b. A mutation for enzyme #2 prevents tyrosine from being synthesized.
   c. A mutation at enzyme #3 prevents homogentistate from being synthesized.
   d. A mutation for enzyme #2 could hide a mutation in enzyme #4.
   e. Each step in a pathway such as this is catalyzed by an enzyme, which is coded by a gene.

2. Eukaryotic mRNA:
   a. uses snRNPs to cut out introns and seal together translatable exons.
   b. uses a spliceosome mechanism made of DNA to recognize consensus regions to cut and splice.
   c. has a guanine cap on its 3′ end and a poly(A) tail on its 5′ end.
   d. is composed of adenine, thymine, guanine, and cytosine.
   e. codes the guanine cap and poly(A) tail from the DNA template.

3. A segment strand of DNA has a base sequence of 5′-GCATTAGAC-3′. What would be the sequence of an RNA molecule complementary to that sequence?
   a. 5′-GUCTAATGC-3′
   b. 5′-GCAUUAGAC-3′
   c. 5′-CGTAATCTG-3′
   d. 5′-GUCUAAUGC-3′
   e. 5′-CGUAAUCUG-3′

4. Which of the following statements about the initiation phase of translation is false?
   a. An initiation factor allows 5′ mRNA to attach to the small ribosomal subunit.
   b. Initiation factors complex with GTP to help Met-tRNA and AUG pair.
   c. mRNA attaches first to the small ribosomal subunit.
   d. GTP is synthesized.
   e. 3′-UAC-5′ on the tRNA binds 5′-AUG-3′ on mRNA.

5. Which of the following statements about aminoacylation is false?
   a. It precedes translation.
   b. It occurs in the ribosome.
   c. It requires ATP to bind an aminoacyl-tRNA synthetase.
   d. It joins the correct amino acid to a specific tRNA based on the tRNA's anticodon.
   e. It uses three binding sites on aminoacyl-tRNA synthetase.

6. Translation is in progress, with methionine bound to a tRNA in the P site, and a phenylalanine bound to a tRNA in the A site. The order of the next steps in the elongation cycle is:
   a. the ribosome translocates → a new aminoacyl-tRNA enters the A site → peptidyl transferase catalyzes a peptide bond between the two amino acids → empty tRNA is released from the ribosome.
   b. peptidyl transferase catalyzes a peptide bond between the two amino acids → a new aminoacyl-tRNA enters the A site → empty tRNA is released from the ribosome → the ribosome translocates.
   c. peptidyl transferase catalyzes a peptide bond between the two amino acids → empty tRNA is released from the ribosome → a new aminoacyl-tRNA enters the A site → the ribosome translocates.
   d. peptidyl transferase catalyzes a peptide bond between the two amino acids → the ribosome translocates → empty tRNA is released from the ribosome → a new aminoacyl-tRNA enters the A site.
   e. the ribosome translocates → peptidyl transferase catalyzes a peptide bond between the two amino acids → empty tRNA is released from the ribosome → a new aminoacyl-tRNA enters the A site.

7. Which of the following statements is false?
   a. GTP is an energy source during various stages of translation.
   b. In the ribosome, peptidyl transferase catalyses peptide bond formation between amino acids.
   c. When the mRNA code UAA reaches the ribosome, there is no tRNA to bind to it.
   d. A long polypeptide is cut off the tRNA in the A site so its Met amino acid links to the amino acid in the P site.
   e. Forty-two amino acids of a protein are encoded by 126 nucleotides of the mRNA.

8. Which item binds to SRP receptor and to the signal sequence to guide a newly synthesized protein to be secreted to its proper "channel"?
   a. ribosome
   b. signal recognition particle
   c. endoplasmic reticulum
   d. signal peptidase
   e. receptor protein

9. A part of an mRNA molecule with the sequence 5′-UGC GCA-3′ is being translated by a ribosome. The following activated tRNA molecules are available. Two of them can correctly bind the mRNA so that a dipeptide can form.

| tRNA Anticodon | Amino Acid |
|---|---|
| 3′-GGC-5′ | Proline |
| 3′-CGU-5′ | Alanine |
| 3′-UGC-5′ | Threonine |
| 3′-CCG-5′ | Glycine |
| 3′-ACG-5′ | Cysteine |
| 3′-CGG-5′ | Alanine |

a. cysteine-alanine
b. proline-cysteine
c. glycine-cysteine
d. alanine-alanine
e. threonine-glycine

10. A missense mutation cannot be:
    a. the code for the sickle-cell gene.
    b. caused by a frameshift.
    c. the deletion of a base in a coding sequence.
    d. the addition of two bases in a coding sequence.
    e. the same as a silent mutation.

## Questions for Discussion

1. Which do you think are more important to the accuracy by which amino acids are linked into proteins: nucleic acids or enzymatic proteins? Why?
2. A mutation appears that alters an anticodon in a tRNA from AAU to AUU. What effect will this change have on protein synthesis in cells carrying this mutation?
3. The normal form of a gene contains the nucleotide sequence:

   5′-ATGCCCGCCTTTGCTACTTGGTAG-3′
   3′-TACGGGCGGAAACGATGAACCATC-5′

   When this gene is transcribed, the result is the following mRNA molecule:

   5′-AUGCCCGCCUUUGCUACUUGGUAG-3′

   In a mutated form of the gene, two extra base pairs (underlined) are inserted:

   5′-ATGCCCGCCTAATTGCTACTTGGTAG-3′
   3′-TACGGGCGGATTAACGATGAACCATC-5′

   What effect will this particular mutation have on the structure of the protein encoded in the gene?
4. A geneticist is attempting to isolate mutations in the genes for four enzymes acting in a metabolic pathway in the bacterium *Escherichia coli*. The end product *E* of the pathway is absolutely essential for life:


   The geneticist has been able to isolate mutations in the genes for enzymes 1 and 2, but not for enzymes 3 and 4. Develop a hypothesis to explain why.

## Experimental Analysis

How could you show experimentally that the genetic code is universal; namely, that it is the same in bacteria as it is in eukaryotes such as fungi, plants, and animals?

## Evolution Link Question

How might the process of alternative splicing and exon shuffling affect the rate at which new proteins evolve?

## How Would You Vote?

Ricin, a molecule that inactivates ribosomes, is difficult to disperse through the air and is unlikely to be used in a large-scale terrorist attack. However, ricin powder did turn up in a Senate office building. Scientists are working to develop a vaccine against ricin. If mass immunizations were to be offered, would you sign up to be vaccinated? Go to www.thomsonedu.com/login to investigate both sides of the issue and then vote.