

From Gene to Protein



▲ **Figure 17.1** How does a single faulty gene result in the dramatic appearance of an albino deer?

KEY CONCEPTS

- 17.1** Genes specify proteins via transcription and translation
- 17.2** Transcription is the DNA-directed synthesis of RNA: *a closer look*
- 17.3** Eukaryotic cells modify RNA after transcription
- 17.4** Translation is the RNA-directed synthesis of a polypeptide: *a closer look*
- 17.5** Mutations of one or a few nucleotides can affect protein structure and function
- 17.6** While gene expression differs among the domains of life, the concept of a gene is universal

OVERVIEW

The Flow of Genetic Information

In 2006, a young albino deer seen frolicking with several brown deer in the mountains of eastern Germany elicited a public outcry (**Figure 17.1**). A local hunting organization announced that the albino deer suffered from a “genetic disorder” and should be shot. Some argued that the deer should merely be prevented from mating with other deer to safeguard the population’s gene pool. Others favored relocating the albino deer to a nature reserve because they worried that it might be more noticeable to predators if left in the wild. A German rock star even held a benefit concert to raise funds for the relocation. What led to the striking phenotype of this deer, the cause of this lively debate?

You learned in Chapter 14 that inherited traits are determined by genes and that the trait of albinism is caused by a recessive allele of a pigmentation gene. The information content of genes is in the form of specific sequences of nucleotides along strands of DNA, the genetic material. But how does this information determine an organism’s traits? Put another way, what does a gene actually say? And how is its message translated by cells into a specific trait, such as brown hair, type A blood, or, in the case of an albino deer, a total lack of pigment? The albino deer has a faulty version of a key protein, an enzyme required for pigment synthesis, and this protein is faulty because the gene that codes for it contains incorrect information.

This example illustrates the main point of this chapter: The DNA inherited by an organism leads to specific traits by dictating the synthesis of proteins and of RNA molecules involved in protein synthesis. In other words, proteins are the link between genotype and phenotype. **Gene expression** is the process by which DNA directs the synthesis of proteins (or, in some cases, just RNAs). The expression of genes that code for proteins includes two stages: transcription and translation. This chapter describes the flow of information from gene to protein in detail and explains how genetic mutations affect organisms through their proteins. Understanding the processes of gene expression, which are similar in all three domains of life, will allow us to revisit the concept of the gene in more detail at the end of the chapter.

CONCEPT 17.1

Genes specify proteins via transcription and translation

Before going into the details of how genes direct protein synthesis, let’s step back and examine how the fundamental relationship between genes and proteins was discovered.

Evidence from the Study of Metabolic Defects

In 1902, British physician Archibald Garrod was the first to suggest that genes dictate phenotypes through enzymes that catalyze specific chemical reactions in the cell. Garrod postulated that the symptoms of an inherited disease reflect a person's inability to make a particular enzyme. He later referred to such diseases as "inborn errors of metabolism." Garrod gave as one example the hereditary condition called alkaptonuria. In this disorder, the urine is black because it contains the chemical alkapton, which darkens upon exposure to air. Garrod reasoned that most people have an enzyme that metabolizes alkapton, whereas people with alkaptonuria have inherited an inability to make that enzyme.

Garrod may have been the first to recognize that Mendel's principles of heredity apply to humans as well as peas. Garrod's realization was ahead of its time, but research several decades later supported his hypothesis that a gene dictates the production of a specific enzyme. Biochemists accumulated much evidence that cells synthesize and degrade most organic molecules via metabolic pathways, in which each chemical reaction in a sequence is catalyzed by a specific enzyme (see p. 188). Such metabolic pathways lead, for instance, to the synthesis of the pigments that give the brown deer in Figure 17.1 their fur color or fruit flies (*Drosophila*) their eye color (see Figure 15.3). In the 1930s, the American biochemist and geneticist George Beadle and his French colleague Boris Ephrussi speculated that in *Drosophila*, each of the various mutations affecting eye color blocks pigment synthesis at a specific step by preventing production of the enzyme that catalyzes that step. But neither the chemical reactions nor the enzymes that catalyze them were known at the time.

Nutritional Mutants in *Neurospora*: Scientific Inquiry

A breakthrough in demonstrating the relationship between genes and enzymes came a few years later at Stanford University, where Beadle and Edward Tatum began working with a bread mold, *Neurospora crassa*. They bombarded *Neurospora* with X-rays, shown in the 1920s to cause genetic changes, and then looked among the survivors for mutants that differed in their nutritional needs from the wild-type bread mold. Wild-type *Neurospora* has modest food requirements. It can grow in the laboratory on a simple solution of inorganic salts, glucose, and the vitamin biotin, incorporated into agar, a support medium. From this *minimal medium*, the mold cells use their metabolic pathways to produce all the other molecules they need. Beadle and Tatum identified mutants that could not survive on minimal medium, apparently because they were unable to synthesize certain essential molecules from the minimal ingredients. To ensure survival of these nutritional mutants, Beadle and Tatum allowed them to grow on a *complete growth medium*, which consisted of minimal medium supplemented with all 20 amino acids and a few other nutrients. The complete growth medium could support any mutant that couldn't synthesize one of the supplements.

To characterize the metabolic defect in each nutritional mutant, Beadle and Tatum took samples from the mutant growing on complete medium and distributed them to a number of different vials. Each vial contained minimal medium plus a single additional nutrient. The particular supplement that allowed growth indicated the metabolic defect. For example, if the only supplemented vial that supported growth of the mutant was the one fortified with the amino acid arginine, the researchers could conclude that the mutant was defective in the biochemical pathway that wild-type cells use to synthesize arginine.

In fact, such arginine-requiring mutants were obtained and studied by two colleagues of Beadle and Tatum, Adrian Srb and Norman Horowitz, who wanted to investigate the biochemical pathway for arginine synthesis in *Neurospora* (Figure 17.2). Srb and Horowitz pinned down each mutant's defect more specifically, using additional tests to distinguish among three classes of arginine-requiring mutants. Mutants in each class required a different set of compounds along the arginine-synthesizing pathway, which has three steps. These results, and those of many similar experiments done by Beadle and Tatum, suggested that each class was blocked at a different step in this pathway because mutants in that class lacked the enzyme that catalyzes the blocked step.

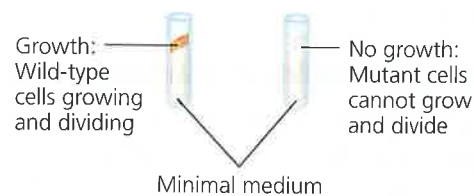
Because each mutant was defective in a single gene, Beadle and Tatum saw that, taken together, the collected results provided strong support for a working hypothesis they had proposed earlier. The *one gene–one enzyme hypothesis*, as they dubbed it, states that the function of a gene is to dictate the production of a specific enzyme. Further support for this hypothesis came from experiments that identified the specific enzymes lacking in the mutants. Beadle and Tatum shared a Nobel Prize in 1958 for "their discovery that genes act by regulating definite chemical events" (in the words of the Nobel committee).

The Products of Gene Expression: A Developing Story

As researchers learned more about proteins, they made revisions to the one gene–one enzyme hypothesis. First of all, not all proteins are enzymes. Keratin, the structural protein of animal hair, and the hormone insulin are two examples of nonenzyme proteins. Because proteins that are not enzymes are nevertheless gene products, molecular biologists began to think in terms of one gene–one protein. However, many proteins are constructed from two or more different polypeptide chains, and each polypeptide is specified by its own gene. For example, hemoglobin, the oxygen-transporting protein of vertebrate red blood cells, contains two kinds of polypeptides, and thus two genes code for this protein (see Figure 5.20). Beadle and Tatum's idea was therefore restated as the *one gene–one polypeptide hypothesis*. Even this description is not entirely accurate, though. First, many eukaryotic genes can each code for a set of closely related polypeptides via a process called alternative splicing, which you will learn about later in this chapter. Second, quite a few genes code for RNA molecules that have important functions in cells

Do individual genes specify the enzymes that function in a biochemical pathway?

EXPERIMENT Working with the mold *Neurospora crassa*, Adrian Srb and Norman Horowitz, then at Stanford University, used Beadle and Tatum's experimental approach to isolate mutants that required arginine in their growth medium. The researchers showed that these mutants fell into three classes, each defective in a different gene. From other considerations, they suspected that the metabolic pathway of arginine biosynthesis involved a precursor nutrient and the intermediate molecules ornithine and citrulline. Their most famous experiment, shown here, tested both the one gene-one enzyme hypothesis and their postulated arginine-synthesizing pathway. In this experiment, they grew their three classes of mutants under the four different conditions shown in the Results section below. They included minimal medium (MM) as a control because they knew that wild-type cells could grow on MM but mutant cells could not. (See test tubes on the right.)



RESULTS The wild-type strain was capable of growth under all experimental conditions, requiring only the minimal medium. The three classes of mutants each had a specific set of growth requirements. For example, class II mutants could not grow when ornithine alone was added but could grow when either citrulline or arginine was added.

Classes of <i>Neurospora crassa</i>					
Condition		Wild type	Class I mutants	Class II mutants	Class III mutants
	Minimal medium (MM) (control)				
	MM + ornithine				
	MM + citrulline				
	MM + arginine (control)				
	Summary of results	Can grow with or without any supplements	Can grow on ornithine, citrulline, or arginine	Can grow only on citrulline or arginine	Require arginine to grow

CONCLUSION From the growth requirements of the mutants, Srb and Horowitz deduced that each class of mutant was unable to carry out one step in the pathway for synthesizing arginine, presumably because it lacked the necessary enzyme. Because each of their mutants was mutated in a single gene, they concluded that each mutated gene must normally dictate the production of one enzyme. Their results supported the one gene-one enzyme hypothesis proposed by Beadle and Tatum and also confirmed that the arginine pathway described in the mammalian liver also operates in *Neurospora*. (Notice in the Results that a mutant can grow only if supplied with a compound made *after* the defective step because this bypasses the defect.)

Gene (codes for enzyme)	Wild type	Class I mutants (mutation in gene A)	Class II mutants (mutation in gene B)	Class III mutants (mutation in gene C)
Gene A → Precursor → Enzyme A	Precursor → Enzyme A	Precursor 	Precursor → Enzyme A	Precursor → Enzyme A
Gene B → Ornithine → Enzyme B	Ornithine → Enzyme B	Ornithine → Enzyme B	Ornithine 	Ornithine → Enzyme B
Gene C → Citrulline → Enzyme C	Citrulline → Enzyme C	Citrulline → Enzyme C	Citrulline → Enzyme C	Citrulline
	Arginine	Arginine	Arginine	Arginine

SOURCE A. M. Srb and N. H. Horowitz, The ornithine cycle in *Neurospora* and its genetic control, *Journal of Biological Chemistry* 154:129–139 (1944).

WHAT IF? Suppose the experiment had shown that class I mutants could grow only in MM supplemented by ornithine or arginine and that class II mutants could grow in MM supplemented by citrulline, ornithine, or arginine. What conclusions would the researchers have drawn from those results regarding the biochemical pathway and the defect in class I and class II mutants?

even though they are never translated into protein. For now, we will focus on genes that do code for polypeptides. (Note that it is common to refer to these gene products as proteins—a practice you will encounter in this book—rather than more precisely as polypeptides.)

Basic Principles of Transcription and Translation

Genes provide the instructions for making specific proteins. But a gene does not build a protein directly. The bridge between DNA and protein synthesis is the nucleic acid RNA. You learned in Chapter 5 that RNA is chemically similar to DNA except that it contains ribose instead of deoxyribose as its sugar and has the nitrogenous base uracil rather than thymine (see Figure 5.26). Thus, each nucleotide along a DNA strand has A, G, C, or T as its base, and each nucleotide along an RNA strand has A, G, C, or U as its base. An RNA molecule usually consists of a single strand.

It is customary to describe the flow of information from gene to protein in linguistic terms because both nucleic acids and proteins are polymers with specific sequences of monomers that convey information, much as specific sequences of letters communicate information in a language like English. In DNA or RNA, the monomers are the four types of nucleotides, which differ in their nitrogenous bases. Genes are typically hundreds or thousands of nucleotides long, each gene having a specific sequence of nucleotides. Each polypeptide of a protein also has monomers arranged in a particular linear order (the protein's primary structure), but its monomers are amino acids. Thus, nucleic acids and proteins contain information written in two different chemical languages. Getting from DNA to protein requires two major stages: transcription and translation.

Transcription is the synthesis of RNA using information in the DNA. The two nucleic acids are written in different forms of the same language, and the information is simply transcribed, or “rewritten,” from DNA to RNA. Just as a DNA strand provides a template for making a new complementary strand during DNA replication, it also can serve as a template for assembling a complementary sequence of RNA nucleotides. For a protein-coding gene, the resulting RNA molecule is a faithful transcript of the gene's protein-building instructions. This type of RNA molecule is called **messenger RNA (mRNA)** because it carries a genetic message from the DNA to the protein-synthesizing machinery of the cell. (Transcription is the general term for the synthesis of *any* kind of RNA on a DNA template. Later, you will learn about some other types of RNA produced by transcription.)

Translation is the synthesis of a polypeptide using the information in the mRNA. During this stage, there is a change in language: The cell must translate the nucleotide sequence of an mRNA molecule into the amino acid sequence of a polypeptide. The sites of translation are **ribosomes**, complex particles that facilitate the orderly linking of amino acids into polypeptide chains.

Transcription and translation occur in all organisms, both those that lack a membrane-bounded nucleus (bacteria and archaea) and those that have one (eukaryotes). Because most studies of transcription and translation have used bacteria and eukaryotic cells, these are our main focus in this chapter. Our understanding of transcription and translation in archaea lags behind, but in the last section of the chapter we will discuss a few aspects of archaeal gene expression.

The basic mechanics of transcription and translation are similar for bacteria and eukaryotes, but there is an important difference in the flow of genetic information within the cells. Because bacteria do not have nuclei, their DNA is not separated by nuclear membranes from ribosomes and the other protein-synthesizing equipment (**Figure 17.3a**). As you will see later, this lack of compartmentalization allows translation of an mRNA to begin while its transcription is still in progress. In a eukaryotic cell, by contrast, the nuclear envelope separates transcription from translation in space and time (**Figure 17.3b**). Transcription occurs in the nucleus, and mRNA is then transported to the cytoplasm, where translation occurs. But before eukaryotic RNA transcripts from protein-coding genes can leave the nucleus, they are modified in various ways to produce the final, functional mRNA. The transcription of a protein-coding eukaryotic gene results in *pre-mRNA*, and further processing yields the finished mRNA. The initial RNA transcript from any gene, including those specifying RNA that is not translated into protein, is more generally called a **primary transcript**.

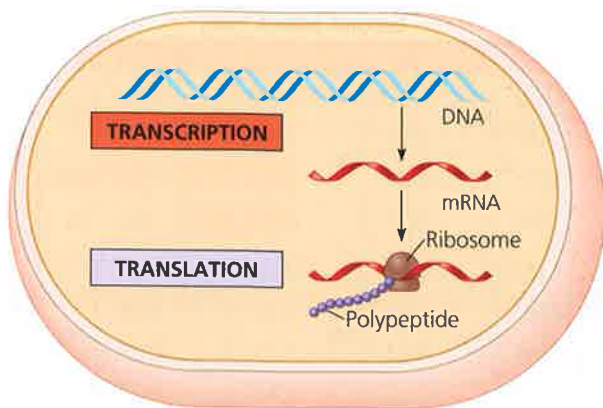
To summarize: Genes program protein synthesis via genetic messages in the form of messenger RNA. Put another way, cells are governed by a molecular chain of command with a directional flow of genetic information, shown here by arrows:



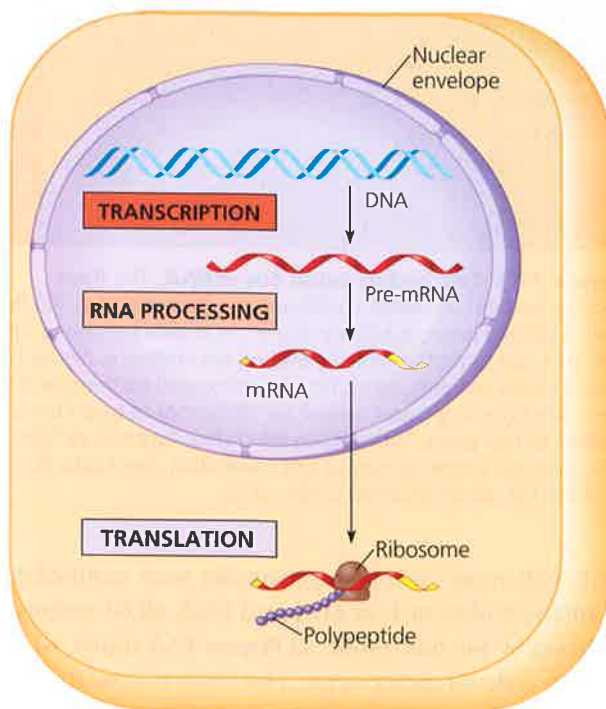
This concept was dubbed the *central dogma* by Francis Crick in 1956. How has the concept held up over time? In the 1970s, scientists were surprised to discover that some RNA molecules can act as templates for DNA synthesis, a process you'll read about in Chapter 19. However, these exceptions do not invalidate the idea that, in general, genetic information flows from DNA to RNA to protein. In the next section, we discuss how the instructions for assembling amino acids into a specific order are encoded in nucleic acids.

The Genetic Code

When biologists began to suspect that the instructions for protein synthesis were encoded in DNA, they recognized a problem: There are only four nucleotide bases to specify 20 amino acids. Thus, the genetic code cannot be a language like Chinese, where each written symbol corresponds to a word. How many nucleotides, then, correspond to an amino acid?



(a) **Bacterial cell.** In a bacterial cell, which lacks a nucleus, mRNA produced by transcription is immediately translated without additional processing.



(b) **Eukaryotic cell.** The nucleus provides a separate compartment for transcription. The original RNA transcript, called pre-mRNA, is processed in various ways before leaving the nucleus as mRNA.

▲ Figure 17.3 Overview: the roles of transcription and translation in the flow of genetic information. In a cell, inherited information flows from DNA to RNA to protein. The two main stages of information flow are transcription and translation. A miniature version of part (a) or (b) accompanies several figures later in the chapter as an orientation diagram to help you see where a particular figure fits into the overall scheme.

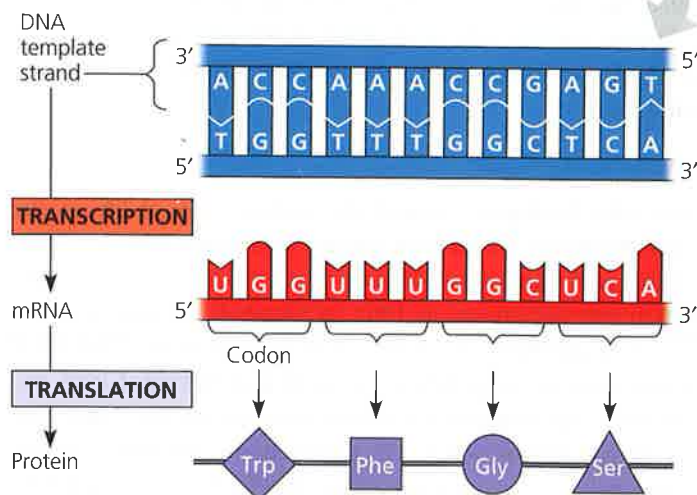
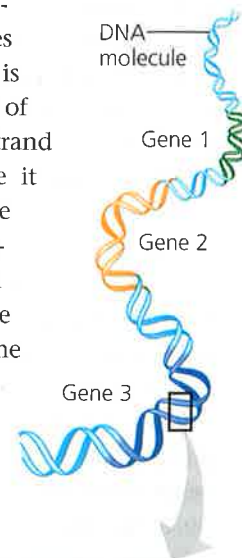
Codons: Triplets of Nucleotides

If each kind of nucleotide base were translated into an amino acid, only 4 of the 20 amino acids could be specified. Would a language of two-letter code words suffice? The two-nucleotide sequence AG, for example, could specify one amino acid, and GT could specify another. Since there are four possible

nucleotide bases in each position, this would give us 16 (that is, 4^2) possible arrangements—still not enough to code for all 20 amino acids.

Triplets of nucleotide bases are the smallest units of uniform length that can code for all the amino acids. If each arrangement of three consecutive nucleotide bases specifies an amino acid, there can be 64 (that is, 4^3) possible code words—more than enough to specify all the amino acids. Experiments have verified that the flow of information from gene to protein is based on a **triplet code**: The genetic instructions for a polypeptide chain are written in the DNA as a series of nonoverlapping, three-nucleotide words. The series of words in a gene is transcribed into a complementary series of nonoverlapping, three-nucleotide words in mRNA, which is then translated into a chain of amino acids (**Figure 17.4**).

During transcription, the gene determines the sequence of nucleotide bases along the length of the RNA molecule that is being synthesized. For each gene, only one of the two DNA strands is transcribed. This strand is called the **template strand** because it provides the pattern, or template, for the sequence of nucleotides in an RNA transcript. For any given gene, the same strand is used as the template every time the gene is transcribed. For other genes on the same DNA molecule, however, the opposite strand may be the one that always functions as the template.



▲ Figure 17.4 The triplet code. For each gene, one DNA strand functions as a template for transcription of RNAs, such as mRNA. The base-pairing rules for DNA synthesis also guide transcription, except that uracil (U) takes the place of thymine (T) in RNA. During translation, the mRNA is read as a sequence of nucleotide triplets, called codons. Each codon specifies an amino acid to be added to the growing polypeptide chain. The mRNA is read in the 5' → 3' direction.

? Compare the sequence of the mRNA to that of the nontemplate DNA strand, in both cases reading from 5' → 3'.

An mRNA molecule is complementary rather than identical to its DNA template because RNA nucleotides are assembled on the template according to base-pairing rules (see Figure 17.4). The pairs are similar to those that form during DNA replication, except that U, the RNA substitute for T, pairs with A and the mRNA nucleotides contain ribose instead of deoxyribose. Like a new strand of DNA, the RNA molecule is synthesized in an antiparallel direction to the template strand of DNA. (To review what is meant by “antiparallel” and the 5′ and 3′ ends of a nucleic acid chain, see Figure 16.7.) In the example in Figure 17.4, the nucleotide triplet ACC along the DNA (written as 3′-ACC-5′) provides a template for 5′-UGG-3′ in the mRNA molecule. The mRNA nucleotide triplets are called **codons**, and they are customarily written in the 5′ → 3′ direction. In our example, UGG is the codon for the amino acid tryptophan (abbreviated Trp). The term *codon* is also used for the DNA nucleotide triplets along the *nontemplate* strand. These codons are complementary to the template strand and thus identical in sequence to the mRNA, except that they have T instead of U. (For this reason, the nontemplate DNA strand is sometimes called the “coding strand.”)

During translation, the sequence of codons along an mRNA molecule is decoded, or translated, into a sequence of amino acids making up a polypeptide chain. The codons are read by the translation machinery in the 5′ → 3′ direction along the mRNA. Each codon specifies which one of the 20 amino acids will be incorporated at the corresponding position along a polypeptide. Because codons are nucleotide triplets, the number of nucleotides making up a genetic message must be three times the number of amino acids in the protein product. For example, it takes 300 nucleotides along an mRNA strand to code for the amino acids in a polypeptide that is 100 amino acids long.

Cracking the Code

Molecular biologists cracked the genetic code of life in the early 1960s when a series of elegant experiments disclosed the amino acid translations of each of the RNA codons. The first codon was deciphered in 1961 by Marshall Nirenberg, of the National Institutes of Health, and his colleagues. Nirenberg synthesized an artificial mRNA by linking identical RNA nucleotides containing uracil as their base. No matter where this message started or stopped, it could contain only one codon in repetition: UUU. Nirenberg added this “poly-U” to a test-tube mixture containing amino acids, ribosomes, and the other components required for protein synthesis. His artificial system translated the poly-U into a polypeptide containing many units of the amino acid phenylalanine (Phe), strung together as a long polyphenylalanine chain. Thus, Nirenberg determined that the mRNA codon UUU specifies the amino acid phenylalanine. Soon, the amino acids specified by the codons AAA, GGG, and CCC were also determined.

		Second mRNA base				
		U	C	A	G	
First mRNA base (5' end of codon)	U	UUU } Phe	UCU } Ser	UAU } Tyr	UGU } Cys	U
		UUC }	UCC }	UAC }	UGC }	C
		UUA } Leu	UCA }	UAA Stop	UGA Stop	A
		UUG }	UCG }	UAG Stop	UGG Trp	G
C		CUU } Leu	CCU }	CAU } His	CGU } Arg	U
		CUC }	CCC }	CAC }	CGC }	C
		CUA }	CCA }	CAA } Gln	CGA }	A
		CUG }	CCG }	CAG }	CGG }	G
A		AUU } Ile	ACU } Thr	AAU } Asn	AGU } Ser	U
		AUC }	ACC }	AAC }	AGC }	C
		AUA }	ACA }	AAA } Lys	AGA } Arg	A
		AUG Met or start	ACG }	AAG }	AGG }	G
G		GUU } Val	GCU } Ala	GAU } Asp	GGU } Gly	U
		GUC }	GCC }	GAC }	GGC }	C
		GUA }	GCA }	GAA } Glu	GGA }	A
		GUG }	GCG }	GAG }	GGG }	G
		Third mRNA base (3' end of codon)				

▲ **Figure 17.5 The codon table for mRNA.** The three nucleotide bases of an mRNA codon are designated here as the first, second, and third bases, reading in the 5′ → 3′ direction along the mRNA. (Practice using this table by finding the codons in Figure 17.4.) The codon AUG not only stands for the amino acid methionine (Met) but also functions as a “start” signal for ribosomes to begin translating the mRNA at that point. Three of the 64 codons function as “stop” signals, marking where ribosomes end translation. See Figure 5.16 for a list of the full names of all the amino acids.

Although more elaborate techniques were required to decode mixed triplets such as AUA and CGA, all 64 codons were deciphered by the mid-1960s. As Figure 17.5 shows, 61 of the 64 triplets code for amino acids. The three codons that do not designate amino acids are “stop” signals, or termination codons, marking the end of translation. Notice that the codon AUG has a dual function: It codes for the amino acid methionine (Met) and also functions as a “start” signal, or initiation codon. Genetic messages usually begin with the mRNA codon AUG, which signals the protein-synthesizing machinery to begin translating the mRNA at that location. (Because AUG also stands for methionine, polypeptide chains begin with methionine when they are synthesized. However, an enzyme may subsequently remove this starter amino acid from the chain.)

Notice in Figure 17.5 that there is redundancy in the genetic code, but no ambiguity. For example, although codons GAA and GAG both specify glutamic acid (redundancy), neither of them ever specifies any other amino acid (no ambiguity). The redundancy in the code is not altogether random. In many cases, codons that are synonyms for a particular amino acid differ only in the third nucleotide base of the triplet. We will consider a possible benefit of this redundancy later in the chapter.

Our ability to extract the intended message from a written language depends on reading the symbols in the correct groupings—that is, in the correct **reading frame**. Consider this statement: “The red dog ate the bug.” Group the letters incorrectly by starting at the wrong point, and the result will probably be gibberish: for example, “her edd oga tet heb ug.” The reading frame is also important in the molecular language of cells. The short stretch of polypeptide shown in Figure 17.4, for instance, will be made correctly only if the mRNA nucleotides are read from left to right (5′ → 3′) in the groups of three shown in the figure: UGG UUU GGC UCA. Although a genetic message is written with no spaces between the codons, the cell’s protein-synthesizing machinery reads the message as a series of nonoverlapping three-letter words. The message is *not* read as a series of overlapping words—UGGUUU, and so on—which would convey a very different message.

Evolution of the Genetic Code

EVOLUTION The genetic code is nearly universal, shared by organisms from the simplest bacteria to the most complex plants and animals. The RNA codon CCG, for instance, is translated as the amino acid proline in all organisms whose genetic code has been examined. In laboratory experiments, genes can be transcribed and translated after being transplanted from one species to another, sometimes with quite striking results, as shown in **Figure 17.6**! Bacteria can be pro-



(a) Tobacco plant expressing a firefly gene. The yellow glow is produced by a chemical reaction catalyzed by the protein product of the firefly gene.



(b) Pig expressing a jellyfish gene. Researchers injected the gene for a fluorescent protein into fertilized pig eggs. One of the eggs developed into this fluorescent pig.

▲ **Figure 17.6 Expression of genes from different species.** Because diverse forms of life share a common genetic code, one species can be programmed to produce proteins characteristic of a second species by introducing DNA from the second species into the first.

grammed by the insertion of human genes to synthesize certain human proteins for medical use, such as insulin. Such applications have produced many exciting developments in the area of biotechnology (see Chapter 20).

Exceptions to the universality of the genetic code include translation systems in which a few codons differ from the standard ones. Slight variations in the genetic code exist in certain unicellular eukaryotes and in the organelle genes of some species. Despite these exceptions, the evolutionary significance of the code’s *near* universality is clear. A language shared by all living things must have been operating very early in the history of life—early enough to be present in the common ancestor of all present-day organisms. A shared genetic vocabulary is a reminder of the kinship that bonds all life on Earth.

CONCEPT CHECK 17.1

- MAKE CONNECTIONS** In a research article about alkaptonuria published in 1902, Garrod suggested that humans inherit two “characters” (alleles) for a particular enzyme and that both parents must contribute a faulty version for the offspring to have the disorder. Today, would this disorder be called dominant or recessive? See Concept 14.4, pages 322–324.
- What polypeptide product would you expect from a poly-G mRNA that is 30 nucleotides long?
- DRAW IT** The template strand of a gene contains the sequence 3′-TTCAGTCGT-5′. Draw the nontemplate sequence and the mRNA sequence, indicating 5′ and 3′ ends of each. Compare the two sequences.
- WHAT IF? DRAW IT** Imagine that the nontemplate sequence in question 3 was transcribed instead of the template sequence. Draw the mRNA sequence and translate it using Figure 17.5. (Be sure to pay attention to the 5′ and 3′ ends.) Predict how well the protein synthesized from the nontemplate strand would function, if at all.

For suggested answers, see Appendix A.

CONCEPT 17.2

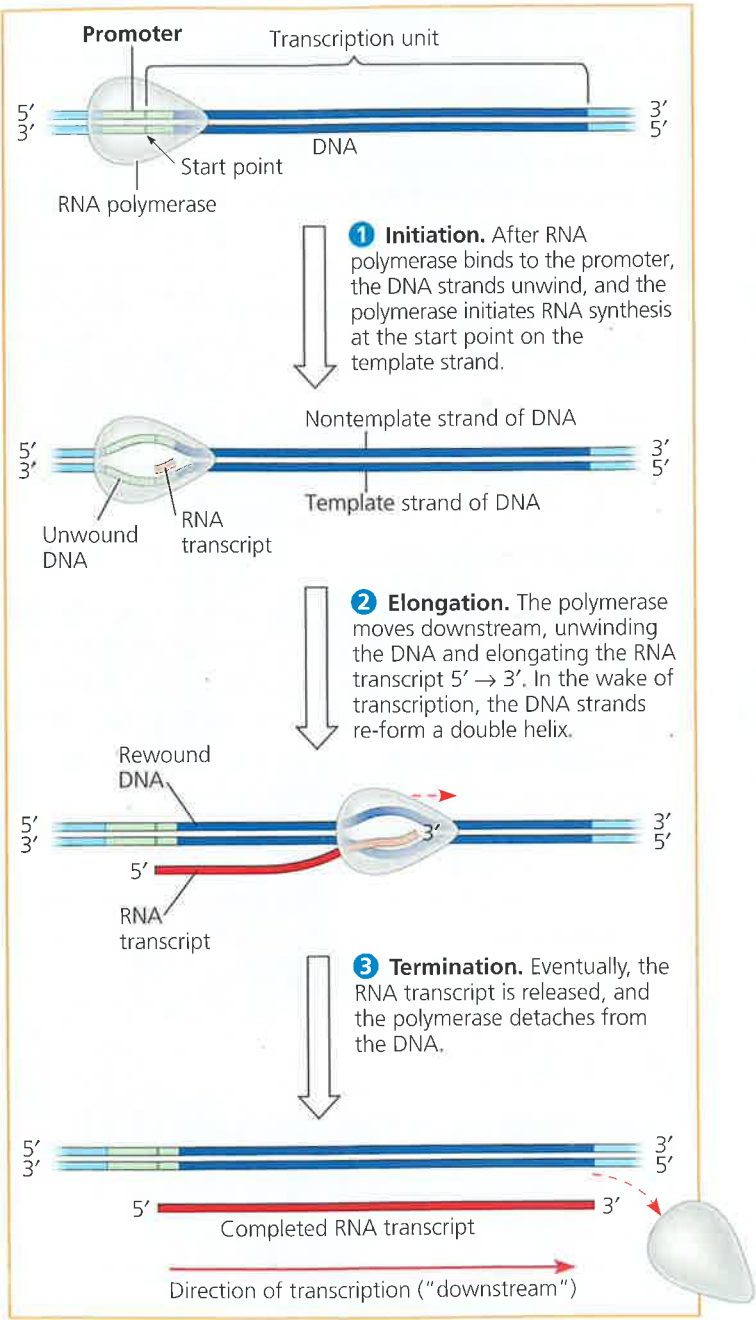
Transcription is the DNA-directed synthesis of RNA: a closer look

Now that we have considered the linguistic logic and evolutionary significance of the genetic code, we are ready to re-examine transcription, the first stage of gene expression, in more detail.

Molecular Components of Transcription

Messenger RNA, the carrier of information from DNA to the cell’s protein-synthesizing machinery, is transcribed from the template strand of a gene. An enzyme called an **RNA polymerase** pries the two strands of DNA apart and joins

together RNA nucleotides complementary to the DNA template strand, thus elongating the RNA polynucleotide (Figure 17.7). Like the DNA polymerases that function in DNA replication, RNA polymerases can assemble a polynucleotide only in its 5' → 3' direction. Unlike DNA polymerases, however, RNA polymerases are able to start a chain from scratch; they don't need a primer.



▲ Figure 17.7 The stages of transcription: initiation, elongation, and termination. This general depiction of transcription applies to both bacteria and eukaryotes, but the details of termination differ, as described in the text. Also, in a bacterium, the RNA transcript is immediately usable as mRNA; in a eukaryote, the RNA transcript must first undergo processing.

MAKE CONNECTIONS Compare the use of a template strand during transcription and replication. See Figure 16.17, page 363.

Specific sequences of nucleotides along the DNA mark where transcription of a gene begins and ends. The DNA sequence where RNA polymerase attaches and initiates transcription is known as the **promoter**; in bacteria, the sequence that signals the end of transcription is called the **terminator**. (The termination mechanism is different in eukaryotes; we'll describe it later.) Molecular biologists refer to the direction of transcription as "downstream" and the other direction as "upstream." These terms are also used to describe the positions of nucleotide sequences within the DNA or RNA. Thus, the promoter sequence in DNA is said to be upstream from the terminator. The stretch of DNA that is transcribed into an RNA molecule is called a **transcription unit**.

Bacteria have a single type of RNA polymerase that synthesizes not only mRNA but also other types of RNA that function in protein synthesis, such as ribosomal RNA. In contrast, eukaryotes have at least three types of RNA polymerase in their nuclei. The one used for mRNA synthesis is called RNA polymerase II. The other RNA polymerases transcribe RNA molecules that are not translated into protein. In the discussion of transcription that follows, we start with the features of mRNA synthesis common to both bacteria and eukaryotes and then describe some key differences.

Synthesis of an RNA Transcript

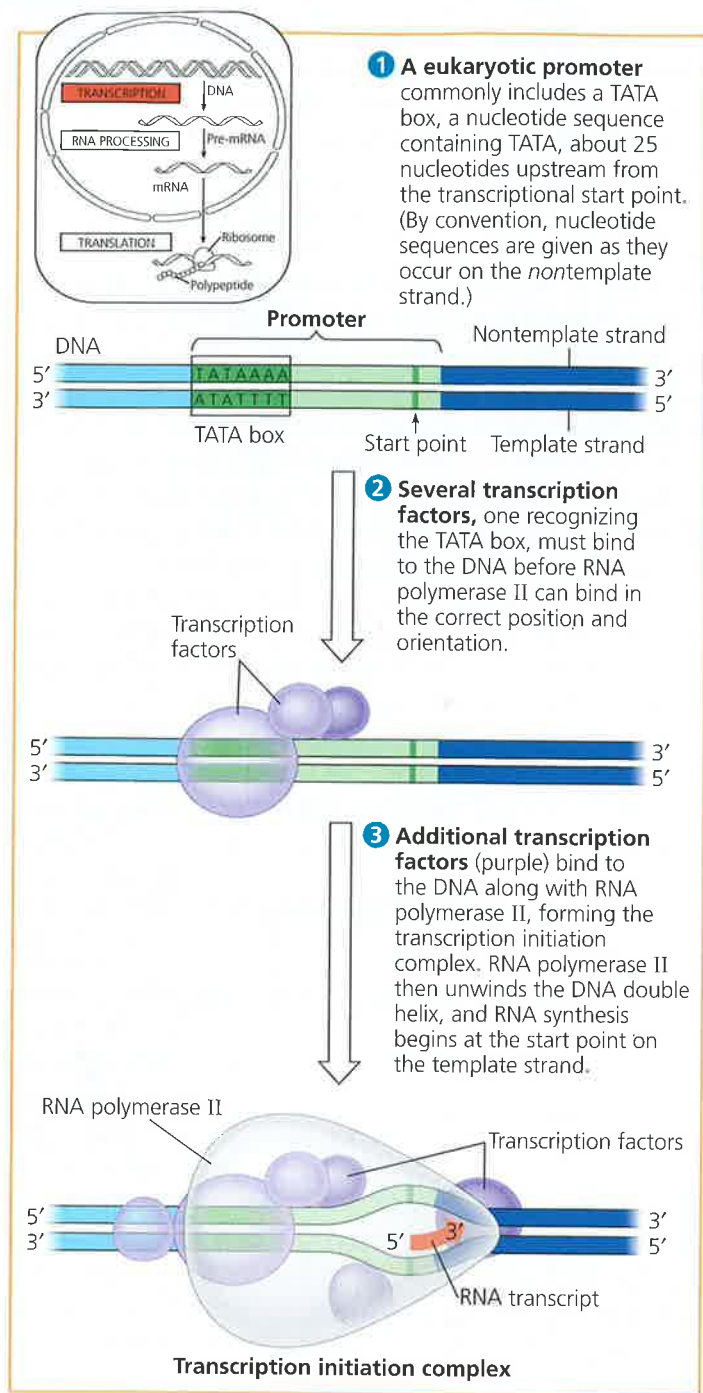
The three stages of transcription, as shown in Figure 17.7 and described next, are initiation, elongation, and termination of the RNA chain. Study Figure 17.7 to familiarize yourself with the stages and the terms used to describe them.

RNA Polymerase Binding and Initiation of Transcription

The promoter of a gene includes within it the transcription **start point** (the nucleotide where RNA synthesis actually begins) and typically extends several dozen or more nucleotide pairs upstream from the start point. RNA polymerase binds in a precise location and orientation on the promoter, therefore determining where transcription starts and which of the two strands of the DNA helix is used as the template.

Certain sections of a promoter are especially important for binding RNA polymerase. In bacteria, the RNA polymerase itself specifically recognizes and binds to the promoter. In eukaryotes, a collection of proteins called **transcription factors** mediate the binding of RNA polymerase and the initiation of transcription. Only after transcription factors are attached to the promoter does RNA polymerase II bind to it. The whole complex of transcription factors and RNA polymerase II bound to the promoter is called a **transcription initiation complex**. Figure 17.8 shows the role of transcription factors and a crucial promoter DNA sequence called a **TATA box** in forming the initiation complex at a eukaryotic promoter.

The interaction between eukaryotic RNA polymerase II and transcription factors is an example of the importance of protein-protein interactions in controlling eukaryotic



▲ **Figure 17.8 The initiation of transcription at a eukaryotic promoter.** In eukaryotic cells, proteins called transcription factors mediate the initiation of transcription by RNA polymerase II.

? Explain how the interaction of RNA polymerase with the promoter would differ if the figure showed transcription initiation for bacteria.

transcription. (And as you learned in Figure 16.22, the DNA of a eukaryotic chromosome is complexed with histones and other proteins in the form of chromatin. The roles of these proteins in making the DNA accessible to transcription factors will be discussed in Chapter 18). Once the appropriate transcription factors are firmly attached to the promoter DNA and the polymerase is bound in the correct orientation,

the enzyme unwinds the two DNA strands and starts transcribing the template strand.

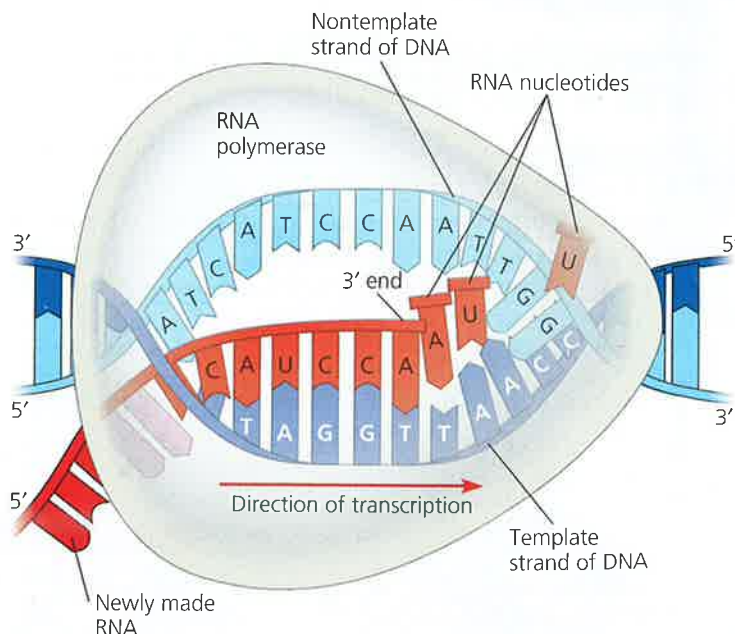
Elongation of the RNA Strand

As RNA polymerase moves along the DNA, it continues to untwist the double helix, exposing about 10–20 DNA nucleotides at a time for pairing with RNA nucleotides (Figure 17.9). The enzyme adds nucleotides to the 3' end of the growing RNA molecule as it continues along the double helix. In the wake of this advancing wave of RNA synthesis, the new RNA molecule peels away from its DNA template, and the DNA double helix re-forms. Transcription progresses at a rate of about 40 nucleotides per second in eukaryotes.

A single gene can be transcribed simultaneously by several molecules of RNA polymerase following each other like trucks in a convoy. A growing strand of RNA trails off from each polymerase, with the length of each new strand reflecting how far along the template the enzyme has traveled from the start point (see the mRNA molecules in Figure 17.25). The congregation of many polymerase molecules simultaneously transcribing a single gene increases the amount of mRNA transcribed from it, which helps the cell make the encoded protein in large amounts.

Termination of Transcription

The mechanism of termination differs between bacteria and eukaryotes. In bacteria, transcription proceeds through a terminator sequence in the DNA. The transcribed terminator (an RNA sequence) functions as the termination signal,



▲ **Figure 17.9 Transcription elongation.** RNA polymerase moves along the DNA template strand, joining complementary RNA nucleotides to the 3' end of the growing RNA transcript. Behind the polymerase, the new RNA peels away from the template strand, which re-forms a double helix with the nontemplate strand.

causing the polymerase to detach from the DNA and release the transcript, which requires no further modification before translation. In eukaryotes, RNA polymerase II transcribes a sequence on the DNA called the polyadenylation signal sequence, which codes for a polyadenylation signal (AAUAAA) in the pre-mRNA. Then, at a point about 10–35 nucleotides downstream from the AAUAAA signal, proteins associated with the growing RNA transcript cut it free from the polymerase, releasing the pre-mRNA. The pre-mRNA then undergoes processing, the topic of the next section.

CONCEPT CHECK 17.2

- MAKE CONNECTIONS** Compare DNA polymerase and RNA polymerase in terms of how they function, the requirement for a template and primer, the direction of synthesis, and the type of nucleotides used. See Figure 16.17, page 353.
- What is a promoter, and is it located at the upstream or downstream end of a transcription unit?
- What enables RNA polymerase to start transcribing a gene at the right place on the DNA in a bacterial cell? In a eukaryotic cell?
- WHAT IF?** Suppose X-rays caused a sequence change in the TATA box of a particular gene's promoter. How would that affect transcription of the gene? (See Figure 17.8.)

For suggested answers, see Appendix A.

CONCEPT 17.3

Eukaryotic cells modify RNA after transcription

Enzymes in the eukaryotic nucleus modify pre-mRNA in specific ways before the genetic messages are dispatched to the cytoplasm. During this **RNA processing**, both ends of the primary transcript are altered. Also, in most cases, certain interior sections of the RNA molecule are cut out and the

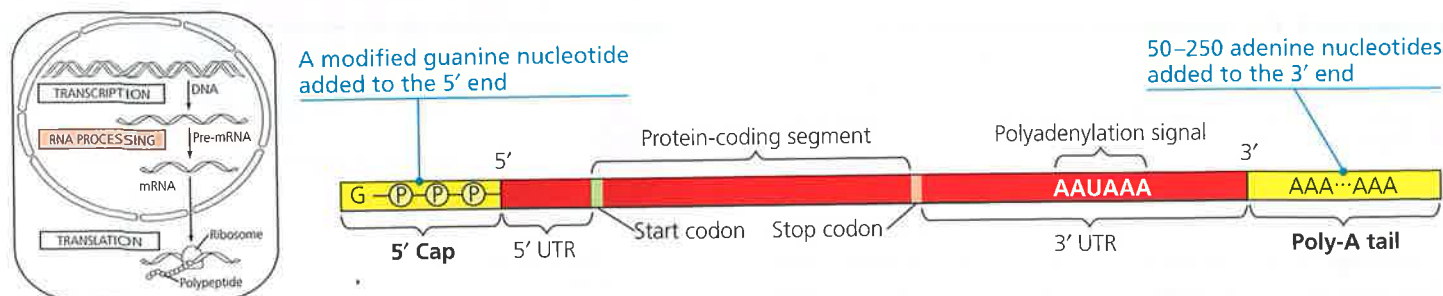
remaining parts spliced together. These modifications produce an mRNA molecule ready for translation.

Alteration of mRNA Ends

Each end of a pre-mRNA molecule is modified in a particular way (**Figure 17.10**). The 5' end is synthesized first; it receives a **5' cap**, a modified form of a guanine (G) nucleotide added onto the 5' end after transcription of the first 20–40 nucleotides. The 3' end of the pre-mRNA molecule is also modified before the mRNA exits the nucleus. Recall that the pre-mRNA is released soon after the polyadenylation signal, AAUAAA, is transcribed. At the 3' end, an enzyme adds 50–250 more adenine (A) nucleotides, forming a **poly-A tail**. The 5' cap and poly-A tail share several important functions. First, they seem to facilitate the export of the mature mRNA from the nucleus. Second, they help protect the mRNA from degradation by hydrolytic enzymes. And third, they help ribosomes attach to the 5' end of the mRNA once the mRNA reaches the cytoplasm. Figure 17.10 shows a diagram of a eukaryotic mRNA molecule with cap and tail. The figure also shows the untranslated regions (UTRs) at the 5' and 3' ends of the mRNA (referred to as the 5' UTR and 3' UTR). The UTRs are parts of the mRNA that will not be translated into protein, but they have other functions, such as ribosome binding.

Split Genes and RNA Splicing

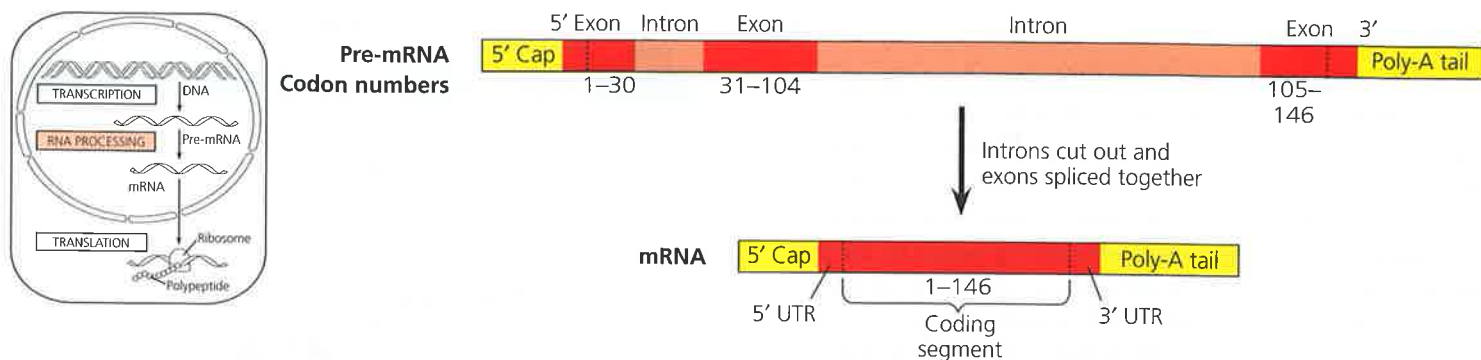
A remarkable stage of RNA processing in the eukaryotic nucleus is the removal of large portions of the RNA molecule that is initially synthesized—a cut-and-paste job called **RNA splicing**, similar to editing a video (**Figure 17.11**). The average length of a transcription unit along a human DNA molecule is about 27,000 nucleotide pairs, so the primary RNA transcript is also that long. However, it takes only 1,200 nucleotides in RNA to code for the average-sized protein of 400 amino acids. (Remember, each amino acid is encoded by a *triplet* of nucleotides.) This means that most eukaryotic genes and their RNA transcripts have long noncoding stretches of nucleotides, regions that are not translated. Even more surprising



▲ Figure 17.10 RNA processing: Addition of the 5' cap and poly-A tail. Enzymes modify the two ends of a eukaryotic pre-mRNA molecule. The modified ends may promote the export of mRNA from the nucleus,

and they help protect the mRNA from degradation. When the mRNA reaches the cytoplasm, the modified ends, in conjunction with certain cytoplasmic proteins, facilitate

ribosome attachment. The 5' cap and poly-A tail are not translated into protein, nor are the regions called the 5' untranslated region (5' UTR) and 3' untranslated region (3' UTR).



▲ **Figure 17.11 RNA processing: RNA splicing.** The RNA molecule shown here codes for β -globin, one of the polypeptides of hemoglobin. The numbers under the RNA refer to codons; β -globin is 146 amino acids long.

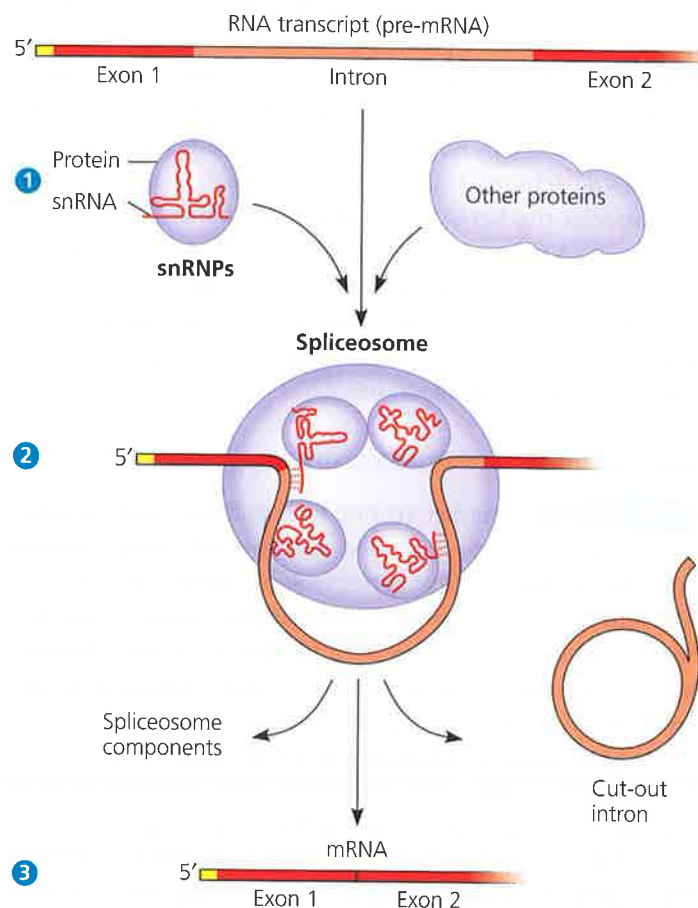
The β -globin gene and its pre-mRNA transcript have three exons, corresponding to sequences that will leave the nucleus as mRNA. (The 5' UTR and 3' UTR are parts of exons because they are included in the mRNA; however, they

do not code for protein.) During RNA processing, the introns are cut out and the exons spliced together. In many genes, the introns are much larger than the exons.

is that most of these noncoding sequences are interspersed between coding segments of the gene and thus between coding segments of the pre-mRNA. In other words, the sequence of DNA nucleotides that codes for a eukaryotic polypeptide is usually not continuous; it is split into segments. The noncoding segments of nucleic acid that lie between coding regions are called *intervening sequences*, or **introns**. The other regions are called **exons**, because they are eventually expressed, usually by being translated into amino acid sequences. (Exceptions include the UTRs of the exons at the ends of the RNA, which make up part of the mRNA but are not translated into protein. Because of these exceptions, you may find it helpful to think of exons as sequences of RNA that *exit* the nucleus.) The terms *intron* and *exon* are used for both RNA sequences and the DNA sequences that encode them.

In making a primary transcript from a gene, RNA polymerase II transcribes both introns and exons from the DNA, but the mRNA molecule that enters the cytoplasm is an abridged version. The introns are cut out from the molecule and the exons joined together, forming an mRNA molecule with a continuous coding sequence. This is the process of RNA splicing.

How is pre-mRNA splicing carried out? Researchers have learned that the signal for RNA splicing is a short nucleotide sequence at each end of an intron. Joan Steitz, our interviewee for this unit (see pp. 292–293), discovered in 1979 that particles called *small nuclear ribonucleoproteins*, abbreviated *snRNPs* (pronounced “snurps”), recognize these splice sites. As the full name implies, snRNPs are located in the cell nucleus and are composed of RNA and protein molecules. The RNA in a snRNP particle is called a *small nuclear RNA* (*snRNA*); each snRNA molecule is about 150 nucleotides long. Several different snRNPs join with additional proteins to form an even larger assembly called a **spliceosome**, which is almost as big as a ribosome. The spliceosome interacts with certain sites along an intron, releasing the intron, which is rapidly degraded, and joining together the two exons that flanked the intron (Figure 17.12). It turns out that snRNAs catalyze these processes, as well as participating in spliceosome assembly and splice site recognition.



▲ **Figure 17.12 The roles of snRNPs and spliceosomes in pre-mRNA splicing.** The diagram shows only a portion of the pre-mRNA transcript; additional introns and exons lie downstream from the ones pictured here. **1** Small nuclear ribonucleoproteins (snRNPs) and other proteins form a molecular complex called a spliceosome on a pre-mRNA molecule containing exons and introns. **2** Within the spliceosome, snRNA base-pairs with nucleotides at specific sites along the intron. **3** The spliceosome cuts the pre-mRNA, releasing the intron for rapid degradation, and at the same time splices the exons together. The spliceosome then comes apart, releasing mRNA, which now contains only exons.

Ribozymes

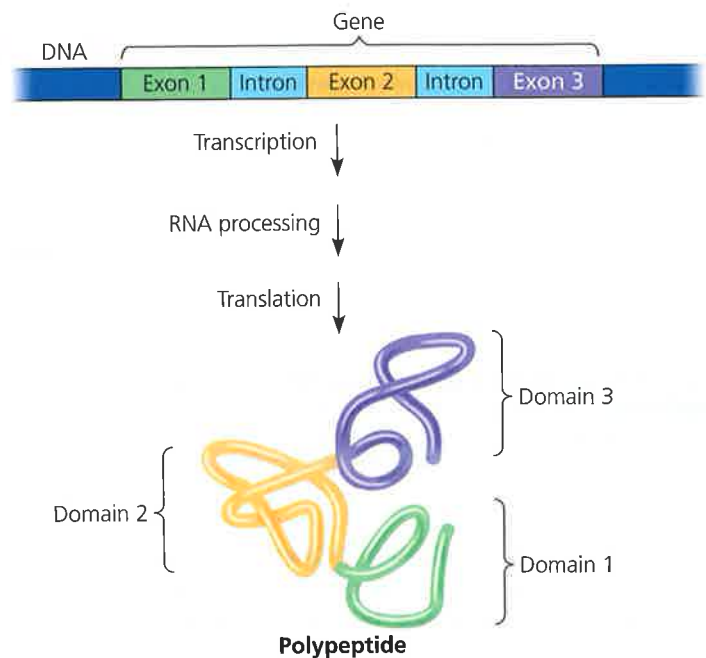
The idea of a catalytic role for snRNA arose from the discovery of **ribozymes**, RNA molecules that function as enzymes. In some organisms, RNA splicing can occur without proteins or even additional RNA molecules: The intron RNA functions as a ribozyme and catalyzes its own excision! For example, in the ciliate protist *Tetrahymena*, self-splicing occurs in the production of ribosomal RNA (rRNA), a component of the organism's ribosomes. The pre-rRNA actually removes its own introns. The discovery of ribozymes rendered obsolete the idea that all biological catalysts are proteins.

Three properties of RNA enable some RNA molecules to function as enzymes. First, because RNA is single-stranded, a region of an RNA molecule may base-pair with a complementary region elsewhere in the same molecule, which gives the molecule a particular three-dimensional structure. A specific structure is essential to the catalytic function of ribozymes, just as it is for enzymatic proteins. Second, like certain amino acids in an enzymatic protein, some of the bases in RNA contain functional groups that may participate in catalysis. Third, the ability of RNA to hydrogen-bond with other nucleic acid molecules (either RNA or DNA) adds specificity to its catalytic activity. For example, complementary base pairing between the RNA of the spliceosome and the RNA of a primary RNA transcript precisely locates the region where the ribozyme catalyzes splicing. Later in this chapter, you will see how these properties of RNA also allow it to perform important noncatalytic roles in the cell, such as recognition of the three-nucleotide codons on mRNA.

The Functional and Evolutionary Importance of Introns

EVOLUTION Whether or not RNA splicing and the presence of introns have provided selective advantages during evolutionary history is a matter of some debate. In any case, it is informative to consider their possible adaptive benefits. Specific functions have not been identified for most introns, but at least some contain sequences that regulate gene expression, and many affect gene products.

One important consequence of the presence of introns in genes is that a single gene can encode more than one kind of polypeptide. Many genes are known to give rise to two or more different polypeptides, depending on which segments are treated as exons during RNA processing; this is called **alternative RNA splicing** (see Figure 18.13). For example, sex differences in fruit flies are largely due to differences in how males and females splice the RNA transcribed from certain genes. Results from the Human Genome Project (discussed in Chapter 21) suggest that alternative RNA splicing is one reason humans can get along with about the same number of genes as a nematode (roundworm). Because of alternative splicing, the number of different protein products an organism produces can be much greater than its number of genes.



▲ **Figure 17.13** Correspondence between exons and protein domains.

Proteins often have a modular architecture consisting of discrete structural and functional regions called **domains**. One domain of an enzyme, for example, might include the active site, while another might allow the enzyme to bind to a cellular membrane. In quite a few cases, different exons code for the different domains of a protein (**Figure 17.13**).

The presence of introns in a gene may facilitate the evolution of new and potentially beneficial proteins as a result of a process known as *exon shuffling*. Introns increase the probability of crossing over between the exons of alleles of a gene—simply by providing more terrain for crossovers without interrupting coding sequences. This might result in new combinations of exons and proteins with altered structure and function. We can also imagine the occasional mixing and matching of exons between completely different (non-allelic) genes. Exon shuffling of either sort could lead to new proteins with novel combinations of functions. While most of the shuffling would result in nonbeneficial changes, occasionally a beneficial variant might arise.

CONCEPT CHECK 17.3

1. How can human cells make 75,000–100,000 different proteins, given that there are about 20,000 human genes?
2. How is RNA splicing similar to editing a video? What would introns correspond to in this analogy?
3. **WHAT IF?** What would be the effect of treating cells with an agent that removed the cap from mRNAs?

For suggested answers, see Appendix A.

Translation is the RNA-directed synthesis of a polypeptide: a closer look

We will now examine in greater detail how genetic information flows from mRNA to protein—the process of translation. As we did for transcription, we'll concentrate on the basic steps of translation that occur in both bacteria and eukaryotes, while pointing out key differences.

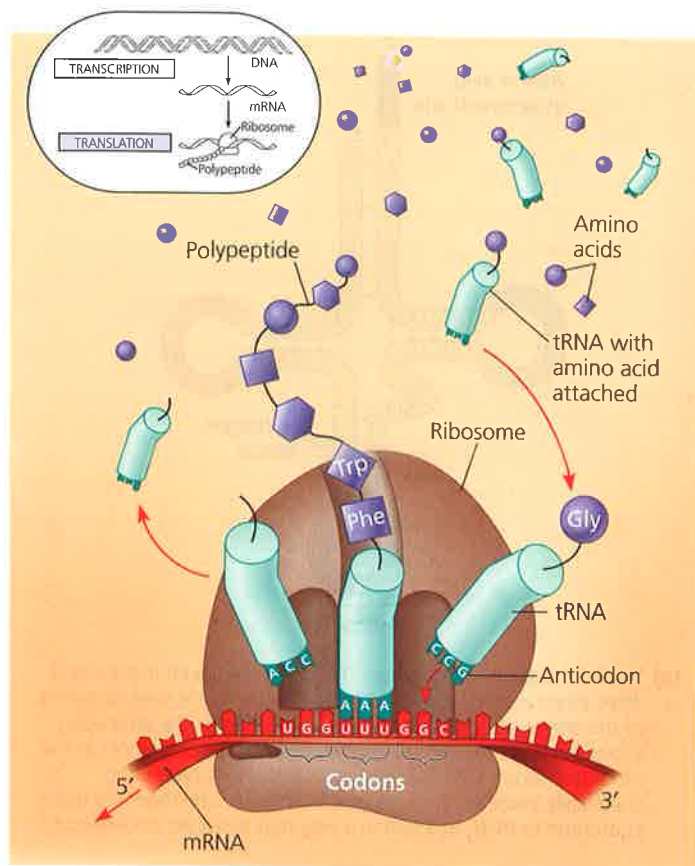
Molecular Components of Translation

In the process of translation, a cell “reads” a genetic message and builds a polypeptide accordingly. The message is a series of codons along an mRNA molecule, and the translator is called **transfer RNA (tRNA)**. The function of tRNA is to transfer amino acids from the cytoplasmic pool of amino acids to a growing polypeptide in a ribosome. A cell keeps its cytoplasm stocked with all 20 amino acids, either by synthesizing them from other compounds or by taking them up from the surrounding solution. The ribosome, a structure made of proteins and RNAs, adds each amino acid brought to it by tRNA to the growing end of a polypeptide chain (**Figure 17.14**).

Translation is simple in principle but complex in its biochemistry and mechanics, especially in the eukaryotic cell. In dissecting translation, we'll concentrate on the slightly less complicated version of the process that occurs in bacteria. We'll begin by looking at the major players in this cellular process and then see how they act together in making a polypeptide.

The Structure and Function of Transfer RNA

The key to translating a genetic message into a specific amino acid sequence is the fact that molecules of tRNA are not all identical, and each type of tRNA molecule translates a particular mRNA codon into a particular amino acid. A tRNA molecule arrives at a ribosome bearing a specific amino acid at one end. At the other end of the tRNA is a nucleotide triplet called an **anticodon**, which base-pairs with a complementary codon on mRNA. For example, consider the mRNA codon GGC, which is translated as the amino acid glycine. The tRNA that base-pairs with this codon by hydrogen bonding has CCG as its anticodon and carries glycine at its other end (see the incoming tRNA approaching the ribosome in Figure 17.14). As an mRNA molecule is moved through a ribosome, glycine will be added to the polypeptide chain whenever the codon GGC is presented for translation. Codon by codon, the genetic message is translated as tRNAs deposit amino acids in the order prescribed, and the ribosome joins the amino acids into a chain. The tRNA molecule is a translator in the sense that it



▲ Figure 17.14 Translation: the basic concept. As a molecule of mRNA is moved through a ribosome, codons are translated into amino acids, one by one. The interpreters are tRNA molecules, each type with a specific anticodon at one end and a corresponding amino acid at the other end. A tRNA adds its amino acid cargo to a growing polypeptide chain when the anticodon hydrogen-bonds to a complementary codon on the mRNA. The figures that follow show some of the details of translation in a bacterial cell.

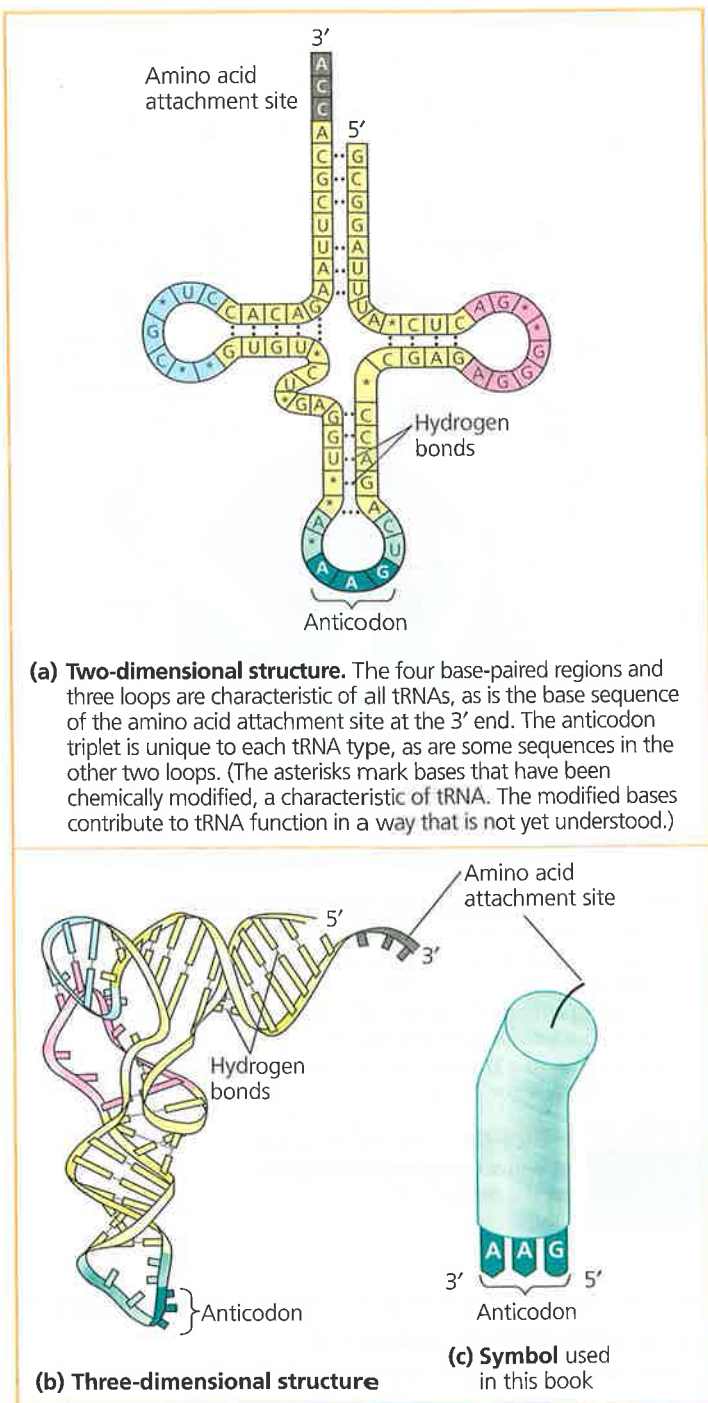


BioFlix Visit the Study Area at www.masteringbiology.com for the BioFlix® 3-D Animation on Protein Synthesis.

can read a nucleic acid word (the mRNA codon) and interpret it as a protein word (the amino acid).

Like mRNA and other types of cellular RNA, transfer RNA molecules are transcribed from DNA templates. In a eukaryotic cell, tRNA, like mRNA, is made in the nucleus and then travels from the nucleus to the cytoplasm, where translation occurs. In both bacterial and eukaryotic cells, each tRNA molecule is used repeatedly, picking up its designated amino acid in the cytosol, depositing this cargo onto a polypeptide chain at the ribosome, and then leaving the ribosome, ready to pick up another amino acid.

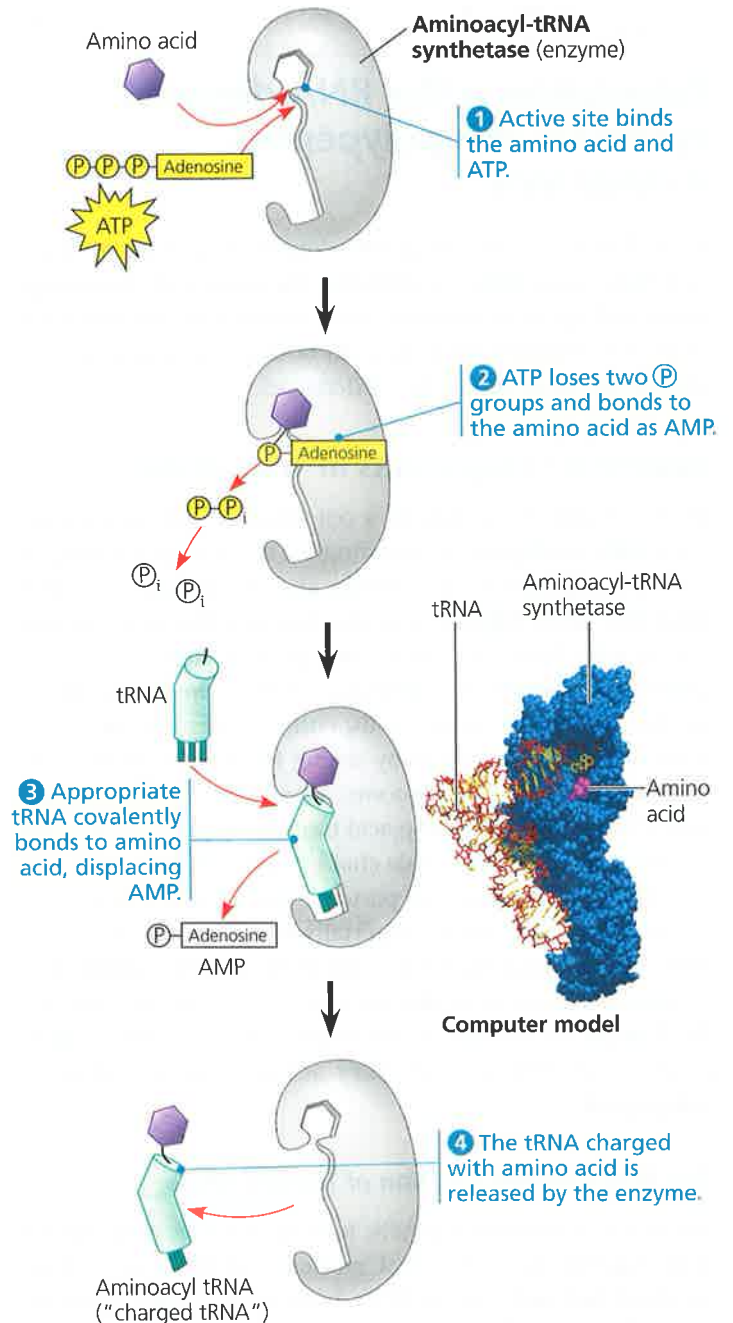
A tRNA molecule consists of a single RNA strand that is only about 80 nucleotides long (compared to hundreds of nucleotides for most mRNA molecules). Because of the presence of complementary stretches of nucleotide bases that can hydrogen-bond to each other, this single strand can fold back upon itself



▲ **Figure 17.15 The structure of transfer RNA (tRNA).**

Anticodons are conventionally written 3' → 5' to align properly with codons written 5' → 3' (see Figure 17.14). For base pairing, RNA strands must be antiparallel, like DNA. For example, anticodon 3'-AAG-5' pairs with mRNA codon 5'-UUC-3'.

and form a molecule with a three-dimensional structure. Flattened into one plane to clarify this base pairing, a tRNA molecule looks like a cloverleaf (Figure 17.15a). The tRNA actually twists and folds into a compact three-dimensional structure that is roughly L-shaped (Figure 17.15b). The loop extending from one end of the L includes the anticodon, the particular nucleotide triplet that base-pairs to a specific mRNA codon.



▲ **Figure 17.16 An aminoacyl-tRNA synthetase joining a specific amino acid to a tRNA.**

Linkage of the tRNA and amino acid is an endergonic process that occurs at the expense of ATP. The ATP loses two phosphate groups, becoming AMP (adenosine monophosphate).

From the other end of the L-shaped tRNA molecule protrudes its 3' end, which is the attachment site for an amino acid. Thus, the structure of a tRNA molecule fits its function.

The accurate translation of a genetic message requires two instances of molecular recognition. First, a tRNA that binds to an mRNA codon specifying a particular amino acid must carry that amino acid, and no other, to the ribosome. The correct matching up of tRNA and amino acid is carried out by a family of related enzymes called **aminoacyl-tRNA synthetases** (Figure 17.16). The active site of each type of aminoacyl-tRNA

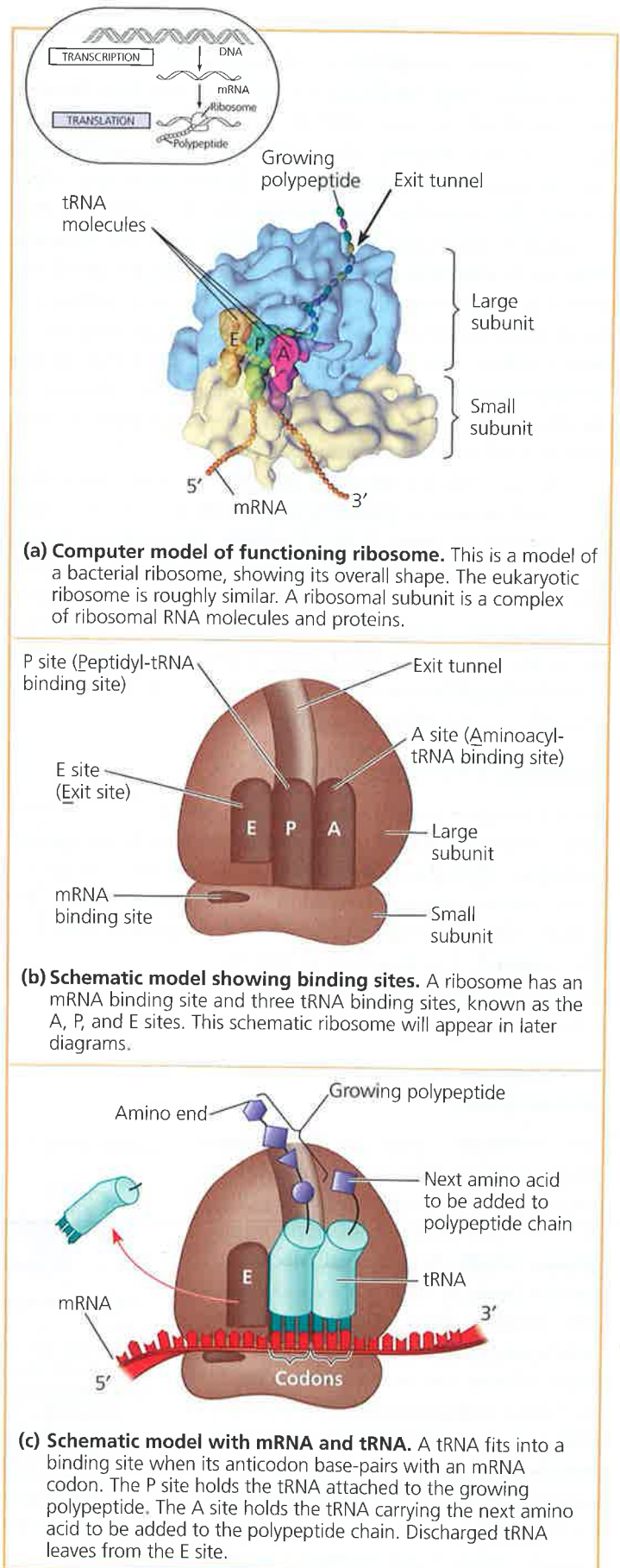
synthetase fits only a specific combination of amino acid and tRNA. (Regions of both the amino acid attachment end and the anticodon end of the tRNA are instrumental in ensuring the specific fit.) There are 20 different synthetases, one for each amino acid; each synthetase is able to bind all the different tRNAs that code for its particular amino acid. The synthetase catalyzes the covalent attachment of the amino acid to its tRNA in a process driven by the hydrolysis of ATP. The resulting aminoacyl tRNA, also called a charged tRNA, is released from the enzyme and is then available to deliver its amino acid to a growing polypeptide chain on a ribosome.

The second instance of molecular recognition is the pairing of the tRNA anticodon with the appropriate mRNA codon. If one tRNA variety existed for each mRNA codon specifying an amino acid, there would be 61 tRNAs (see Figure 17.5). In fact, there are only about 45, signifying that some tRNAs must be able to bind to more than one codon. Such versatility is possible because the rules for base pairing between the third nucleotide base of a codon and the corresponding base of a tRNA anticodon are relaxed compared to those at other codon positions. For example, the nucleotide base U at the 5' end of a tRNA anticodon can pair with either A or G in the third position (at the 3' end) of an mRNA codon. The flexible base pairing at this codon position is called **wobble**. Wobble explains why the synonymous codons for a given amino acid most often differ in their third nucleotide base, but not in the other bases. For example, a tRNA with the anticodon 3'-UCU-5' can base-pair with either the mRNA codon 5'-AGA-3' or 5'-AGG-3', both of which code for arginine (see Figure 17.5).

Ribosomes

Ribosomes facilitate the specific coupling of tRNA anticodons with mRNA codons during protein synthesis. A ribosome consists of a large subunit and a small subunit, each made up of proteins and one or more **ribosomal RNAs (rRNAs)** (Figure 17.17). In eukaryotes, the subunits are made in the nucleolus. Ribosomal RNA genes are transcribed, and the RNA is processed and assembled with proteins imported from the cytoplasm. The resulting ribosomal subunits are then exported via nuclear pores to the cytoplasm. In both bacteria and eukaryotes, large and small subunits join to form a functional ribosome only when they attach to an mRNA molecule. About one-third of the mass of a ribosome is made up of proteins; the rest consists of rRNAs, either three molecules (in bacteria) or four (in eukaryotes). Because most cells contain thousands of ribosomes, rRNA is the most abundant type of cellular RNA.

Although the ribosomes of bacteria and eukaryotes are very similar in structure and function, those of eukaryotes are slightly larger and differ somewhat from bacterial ribosomes in their molecular composition. The differences are medically significant. Certain antibiotic drugs can inactivate bacterial ribosomes without inhibiting the ability of eukaryotic ribosomes to make proteins. These drugs, including tetracycline and streptomycin, are used to combat bacterial infections.



▲ **Figure 17.17** The anatomy of a functioning ribosome.

CONCEPT CHECK 17.4

1. What two processes ensure that the correct amino acid is added to a growing polypeptide chain?
2. Discuss the ways in which rRNA structure likely contributes to ribosomal function.
3. Describe how a polypeptide to be secreted is transported to the endomembrane system.
4. **WHAT IF? DRAW IT** Draw a tRNA with the anticodon 3'-CGU-5'. What two different codons could it bind to? Draw each codon on an mRNA, labeling all 5' and 3' ends. Add the amino acid carried by this tRNA.

For suggested answers, see Appendix A.

CONCEPT 17.5

Mutations of one or a few nucleotides can affect protein structure and function

Now that you have explored the process of gene expression, you are ready to understand the effects of changes to the genetic information of a cell (or virus). These changes, called **mutations**, are responsible for the huge diversity of genes found among organisms because mutations are the ultimate source of new genes. In Figure 15.14, we considered chromosomal rearrangements that affect long segments of DNA, which can be considered large-scale mutations. Here we examine small-scale mutations of one or a few nucleotide pairs, including **point mutations**, changes in a single nucleotide pair of a gene.

If a point mutation occurs in a gamete or in a cell that gives rise to gametes, it may be transmitted to offspring and to a succession of future generations. If the mutation has an adverse effect on the phenotype of an organism, the mutant condition is referred to as a genetic disorder or hereditary disease. For example, we can trace the genetic basis of sickle-cell disease to the mutation of a single nucleotide pair in the gene that encodes the β -globin polypeptide of hemoglobin. The change of a single nucleotide in the DNA's template strand leads to the production of an abnormal protein (**Figure 17.23**; also see Figure 5.21). In individuals who are homozygous for the mutant allele, the sickling of red blood cells caused by the altered hemoglobin produces the multiple symptoms associated with sickle-cell disease (see Chapter 14). Another disorder caused by a

point mutation is a heart condition, familial cardiomyopathy, that is responsible for some incidents of sudden death in young athletes. Point mutations in several genes have been identified, any of which can lead to this disorder.

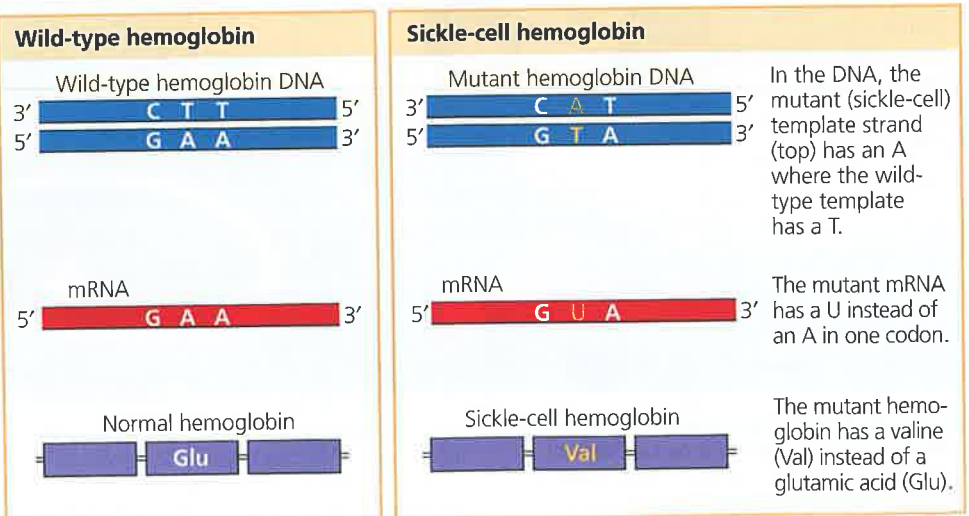
Types of Small-Scale Mutations

Let's now consider how small-scale mutations affect proteins. Point mutations within a gene can be divided into two general categories: (1) single nucleotide-pair substitutions and (2) nucleotide-pair insertions or deletions. Insertions and deletions can involve one or more nucleotide pairs.

Substitutions

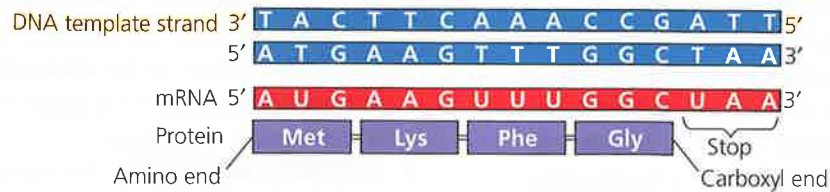
A **nucleotide-pair substitution** is the replacement of one nucleotide and its partner with another pair of nucleotides (**Figure 17.24a**). Some substitutions have no effect on the encoded protein, owing to the redundancy of the genetic code. For example, if 3'-CCG-5' on the template strand mutated to 3'-CCA-5', the mRNA codon that used to be GGC would become GGU, but a glycine would still be inserted at the proper location in the protein (see Figure 17.5). In other words, a change in a nucleotide pair may transform one codon into another that is translated into the same amino acid. Such a change is an example of a **silent mutation**, which has no observable effect on the phenotype. (Silent mutations can occur outside genes as well.) Substitutions that change one amino acid to another one are called **missense mutations**. Such a mutation may have little effect on the protein: The new amino acid may have properties similar to those of the amino acid it replaces, or it may be in a region of the protein where the exact sequence of amino acids is not essential to the protein's function.

▼ Figure 17.23 The molecular basis of sickle-cell disease: a point mutation. The allele that causes sickle-cell disease differs from the wild-type (normal) allele by a single DNA nucleotide pair.

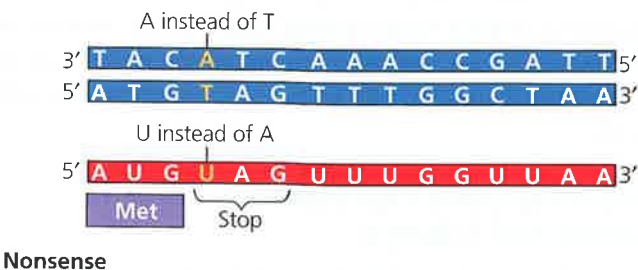
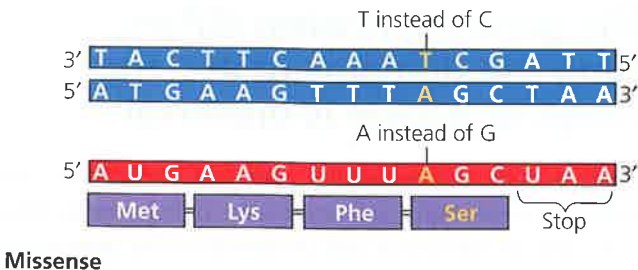
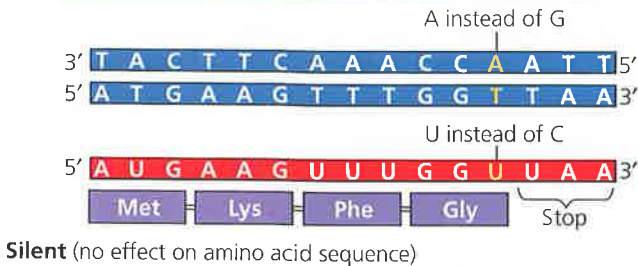


▼ **Figure 17.24** Types of small-scale mutations that affect mRNA sequence. All but one of the types shown here also affect the amino acid sequence of the encoded polypeptide.

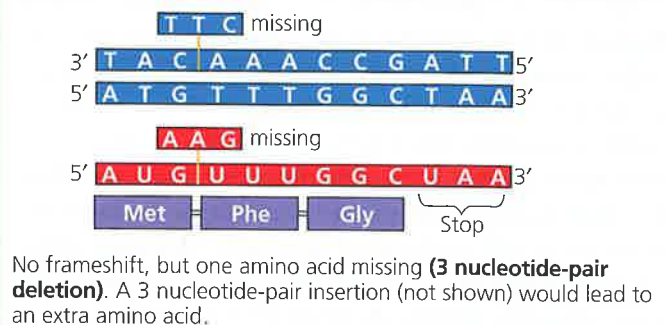
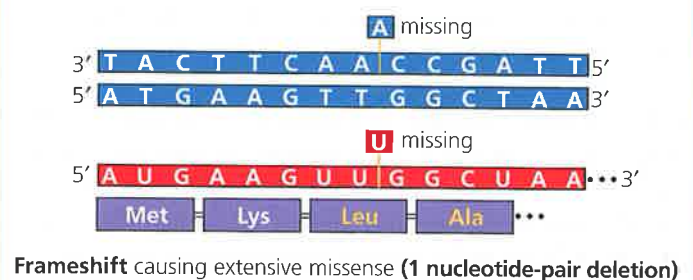
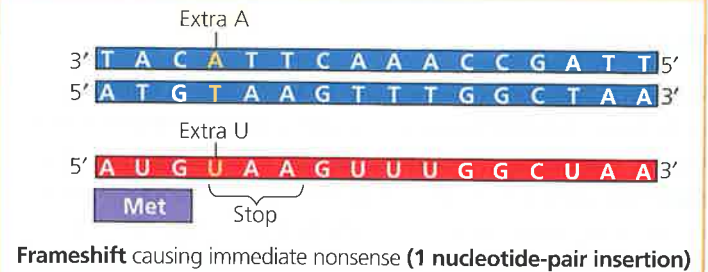
Wild type



(a) Nucleotide-pair substitution



(b) Nucleotide-pair insertion or deletion



However, the nucleotide-pair substitutions of greatest interest are those that cause a major change in a protein. The alteration of a single amino acid in a crucial area of a protein—such as in the part of hemoglobin shown in Figure 17.23 or in the active site of an enzyme as shown in Figure 8.18—will significantly alter protein activity. Occasionally, such a mutation leads to an improved protein or one with novel capabilities, but much more often such mutations are detrimental, leading to a useless or less active protein that impairs cellular function.

Substitution mutations are usually missense mutations; that is, the altered codon still codes for an amino acid and

thus makes sense, although not necessarily the *right* sense. But a point mutation can also change a codon for an amino acid into a stop codon. This is called a **nonsense mutation**, and it causes translation to be terminated prematurely; the resulting polypeptide will be shorter than the polypeptide encoded by the normal gene. Nearly all nonsense mutations lead to nonfunctional proteins.

Insertions and Deletions

Insertions and **deletions** are additions or losses of nucleotide pairs in a gene (Figure 17.24b). These mutations have

a disastrous effect on the resulting protein more often than substitutions do. Insertion or deletion of nucleotides may alter the reading frame of the genetic message, the triplet grouping of nucleotides on the mRNA that is read during translation. Such a mutation, called a **frameshift mutation**, will occur whenever the number of nucleotides inserted or deleted is not a multiple of three. All the nucleotides that are downstream of the deletion or insertion will be improperly grouped into codons, and the result will be extensive missense, usually ending sooner or later in nonsense and premature termination. Unless the frameshift is very near the end of the gene, the protein is almost certain to be nonfunctional.

Mutagens

Mutations can arise in a number of ways. Errors during DNA replication or recombination can lead to nucleotide-pair substitutions, insertions, or deletions, as well as to mutations affecting longer stretches of DNA. If an incorrect nucleotide is added to a growing chain during replication, for example, the base on that nucleotide will then be mismatched with the nucleotide base on the other strand. In many cases, the error will be corrected by systems you learned about in Chapter 16. Otherwise, the incorrect base will be used as a template in the next round of replication, resulting in a mutation. Such mutations are called *spontaneous mutations*. It is difficult to calculate the rate at which such mutations occur. Rough estimates have been made of the rate of mutation during DNA replication for both *E. coli* and eukaryotes, and the numbers are similar: About one nucleotide in every 10^{10} is altered, and the change is passed on to the next generation of cells.

A number of physical and chemical agents, called **mutagens**, interact with DNA in ways that cause mutations. In the 1920s, Hermann Muller discovered that X-rays caused genetic changes in fruit flies, and he used X-rays to make *Drosophila* mutants for his genetic studies. But he also recognized an alarming implication of his discovery: X-rays and other forms of high-energy radiation pose hazards to the genetic material of people as well as laboratory organisms. Mutagenic radiation, a physical mutagen, includes ultraviolet (UV) light, which can cause disruptive thymine dimers in DNA (see Figure 16.19).

Chemical mutagens fall into several categories. Nucleotide analogs are chemicals that are similar to normal DNA nucleotides but that pair incorrectly during DNA replication. Some other chemical mutagens interfere with correct DNA replication by inserting themselves into the DNA and distorting the double helix. Still other mutagens cause chemical changes in bases that change their pairing properties.

Researchers have developed a variety of methods to test the mutagenic activity of chemicals. A major application of these tests is the preliminary screening of chemicals to identify those that may cause cancer. This approach makes sense because most carcinogens (cancer-causing chemicals) are mutagenic, and conversely, most mutagens are carcinogenic.

CONCEPT CHECK 17.5

1. What happens when one nucleotide pair is lost from the middle of the coding sequence of a gene?
2. **MAKE CONNECTIONS** Individuals heterozygous for the sickle-cell allele are generally healthy but show phenotypic effects of the allele under some circumstances; see Concept 14.4, pages 323–324. Explain in terms of gene expression.
3. **WHAT IF? DRAW IT** The template strand of a gene includes this sequence:
3'-TACTTGTCCGATATC-5'. It is mutated to
3'-TACTTGTCCAATATC-5'. For both normal and mutant sequences, draw the double-stranded DNA, the resulting mRNA, and the amino acid sequence each encodes. What is the effect of the mutation on the amino acid sequence?

For suggested answers, see Appendix A.

CONCEPT 17.6

While gene expression differs among the domains of life, the concept of a gene is universal

Although bacteria and eukaryotes carry out transcription and translation in very similar ways, we have noted certain differences in cellular machinery and in details of the processes in these two domains. The division of organisms into three domains was established about 40 years ago, when archaea were recognized as distinct from bacteria. Like bacteria, archaea are prokaryotes. However, archaea share many aspects of the mechanisms of gene expression with eukaryotes, as well as a few with bacteria.

Comparing Gene Expression in Bacteria, Archaea, and Eukarya

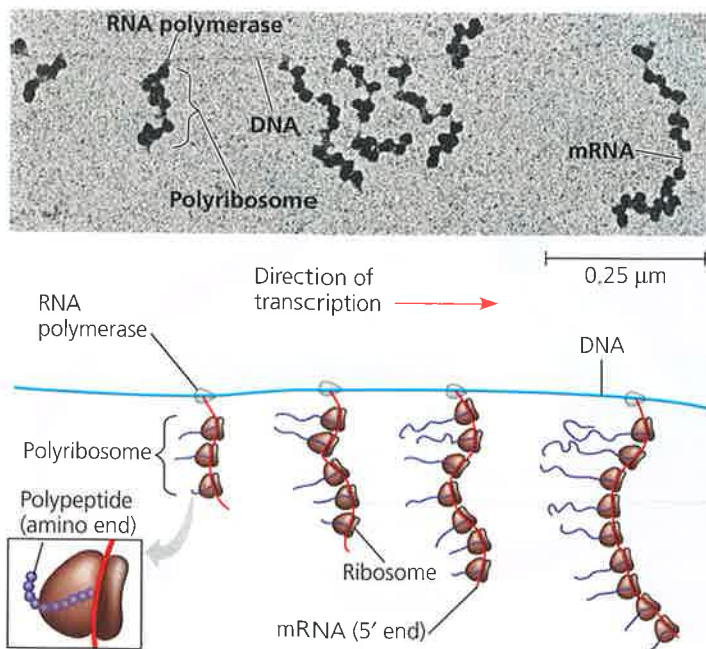
Recent advances in molecular biology have enabled researchers to determine the complete nucleotide sequences of hundreds of genomes, including many genomes from each domain. This wealth of data allows us to compare gene and protein sequences across domains. Foremost among genes of interest are those that encode components of such fundamental biological processes as transcription and translation.

Bacterial and eukaryotic RNA polymerases differ significantly from each other, while the single archaeal RNA polymerase resembles the three eukaryotic ones. Archaea and eukaryotes use a complex set of transcription factors, unlike the smaller set of accessory proteins in bacteria. Transcription is terminated differently in bacteria and eukaryotes. The little that is known about archaeal transcription termination suggests that it is similar to the eukaryotic process.

As far as translation is concerned, archaeal ribosomes are the same size as bacterial ribosomes, but their sensitivities to chemical inhibitors more closely match those of eukaryotic ribosomes. We mentioned earlier that initiation of translation is slightly different in bacteria and eukaryotes. In this respect, the archaeal process is more like that of bacteria.

The most important differences between bacteria and eukaryotes with regard to gene expression arise from the bacterial cell's lack of compartmental organization. Like a one-room workshop, a bacterial cell ensures a streamlined operation. In the absence of a nucleus, it can simultaneously transcribe and translate the same gene (**Figure 17.25**), and the newly made protein can quickly diffuse to its site of function. Most researchers suspect that transcription and translation are coupled like this in archaeal cells as well, since archaea lack a nuclear envelope. In contrast, the eukaryotic cell's nuclear envelope segregates transcription from translation and provides a compartment for extensive RNA processing. This processing stage includes additional steps whose regulation can help coordinate the eukaryotic cell's elaborate activities (see Chapter 18).

Learning more about the proteins and RNAs involved in archaeal transcription and translation will tell us much about the evolution of these processes in all three domains. In spite of the differences in gene expression cataloged here, however, the idea of the gene itself is a unifying concept among all forms of life.



▲ Figure 17.25 Coupled transcription and translation in bacteria. In bacterial cells, the translation of mRNA can begin as soon as the leading (5') end of the mRNA molecule peels away from the DNA template. The micrograph (TEM) shows a strand of *E. coli* DNA being transcribed by RNA polymerase molecules. Attached to each RNA polymerase molecule is a growing strand of mRNA, which is already being translated by ribosomes. The newly synthesized polypeptides are not visible in the micrograph but are shown in the diagram.

? Which one of the mRNA molecules started transcription first? On that mRNA, which ribosome started translating first?

What Is a Gene? Revisiting the Question

Our definition of a gene has evolved over the past few chapters, as it has through the history of genetics. We began with the Mendelian concept of a gene as a discrete unit of inheritance that affects a phenotypic character (Chapter 14). We saw that Morgan and his colleagues assigned such genes to specific loci on chromosomes (Chapter 15). We went on to view a gene as a region of specific nucleotide sequence along the length of the DNA molecule of a chromosome (Chapter 16). Finally, in this chapter, we have considered a functional definition of a gene as a DNA sequence that codes for a specific polypeptide chain. (**Figure 17.26**, on the next page, summarizes the path from gene to polypeptide in a eukaryotic cell.) All these definitions are useful, depending on the context in which genes are being studied.

Clearly, the statement that a gene codes for a polypeptide is too simple. Most eukaryotic genes contain noncoding segments (such as introns), so large portions of these genes have no corresponding segments in polypeptides. Molecular biologists also often include promoters and certain other regulatory regions of DNA within the boundaries of a gene. These DNA sequences are not transcribed, but they can be considered part of the functional gene because they must be present for transcription to occur. Our definition of a gene must also be broad enough to include the DNA that is transcribed into rRNA, tRNA, and other RNAs that are not translated. These genes have no polypeptide products but play crucial roles in the cell. Thus, we arrive at the following definition: *A gene is a region of DNA that can be expressed to produce a final functional product that is either a polypeptide or an RNA molecule.*

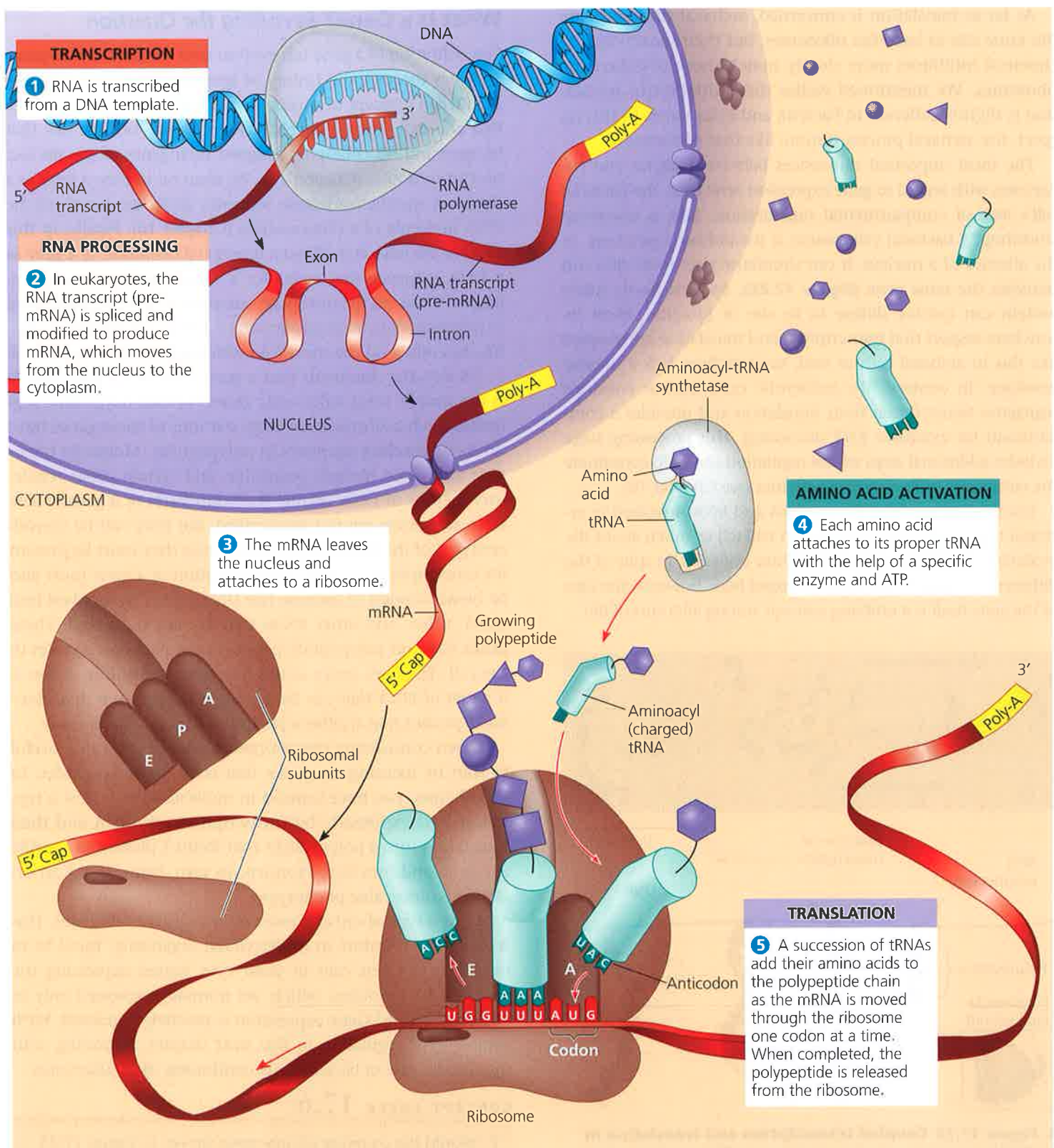
When considering phenotypes, however, it is often useful to start by focusing on genes that code for polypeptides. In this chapter, you have learned in molecular terms how a typical gene is expressed—by transcription into RNA and then translation into a polypeptide that forms a protein of specific structure and function. Proteins, in turn, bring about an organism's observable phenotype.

A given type of cell expresses only a subset of its genes. This is an essential feature in multicellular organisms: You'd be in trouble if the lens cells in your eyes started expressing the genes for hair proteins, which are normally expressed only in hair follicle cells! Gene expression is precisely regulated. We'll explore gene regulation in the next chapter, beginning with the simpler case of bacteria and continuing with eukaryotes.

CONCEPT CHECK 17.6

1. Would the coupling of processes shown in Figure 17.25 be found in a eukaryotic cell? Explain.
2. **WHAT IF?** In eukaryotic cells, mRNAs have been found to have a circular arrangement in which proteins hold the poly-A tail near the 5' cap. How might this increase translation efficiency?

For suggested answers, see Appendix A.



▲ **Figure 17.26 A summary of transcription and translation in a eukaryotic cell.** This diagram shows the path from one gene to one polypeptide. Keep in mind that each gene in the DNA can be transcribed repeatedly into many identical RNA molecules and that each mRNA can be

translated repeatedly to yield many identical polypeptide molecules. (Also, remember that the final products of some genes are not polypeptides but RNA molecules, including tRNA and rRNA.) In general, the steps of transcription and translation are similar in bacterial, archaeal, and eukaryotic cells. The

major difference is the occurrence of RNA processing in the eukaryotic nucleus. Other significant differences are found in the initiation stages of both transcription and translation and in the termination of transcription.

17 CHAPTER REVIEW

SUMMARY OF KEY CONCEPTS

CONCEPT 17.1

Genes specify proteins via transcription and translation (pp. 371–377)

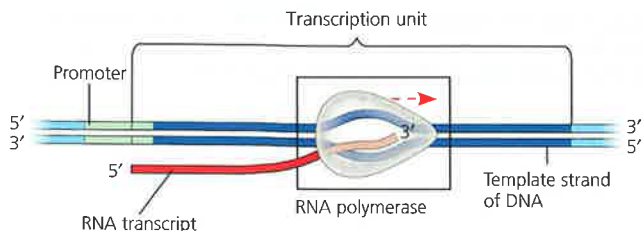
- DNA controls metabolism by directing cells to make specific enzymes and other proteins, via the process of **gene expression**. Beadle and Tatum's studies of mutant strains of *Neurospora* led to the one gene–one polypeptide hypothesis. Genes code for polypeptide chains or specify RNA molecules.
- Transcription** is the synthesis of RNA complementary to a **template strand** of DNA, providing a nucleotide-to-nucleotide transfer of information. **Translation** is the synthesis of a polypeptide whose amino acid sequence is specified by the nucleotide sequence in **mRNA**; this informational transfer thus involves a change of language, from that of nucleotides to that of amino acids.
- Genetic information is encoded as a sequence of nonoverlapping nucleotide triplets, or **codons**. A codon in messenger RNA (mRNA) either is translated into an amino acid (61 of the 64 codons) or serves as a stop signal (3 codons). Codons must be read in the correct **reading frame**.

? Describe the process of gene expression, by which a gene affects the phenotype of an organism.

CONCEPT 17.2

Transcription is the DNA-directed synthesis of RNA: a closer look (pp. 377–380)

- RNA synthesis is catalyzed by **RNA polymerase**, which links together RNA nucleotides complementary to a DNA template strand. This process follows the same base-pairing rules as DNA replication, except that in RNA, uracil substitutes for thymine.



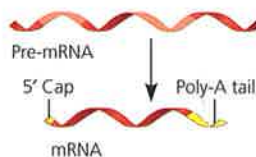
- The three stages of transcription are initiation, elongation, and termination. A **promoter**, often including a **TATA box** in eukaryotes, establishes where RNA synthesis is initiated. **Transcription factors** help eukaryotic RNA polymerase recognize promoter sequences, forming a **transcription initiation complex**. The mechanisms of termination are different in bacteria and eukaryotes.

? What are the similarities and differences in the initiation of gene transcription in bacteria and eukaryotes?

CONCEPT 17.3

Eukaryotic cells modify RNA after transcription (pp. 380–382)

- Before leaving the nucleus, eukaryotic mRNA molecules undergo **RNA processing**, which includes RNA splicing, the addition of a modified nucleotide **5' cap** to the 5' end, and the addition of a **poly-A tail** to the 3' end.



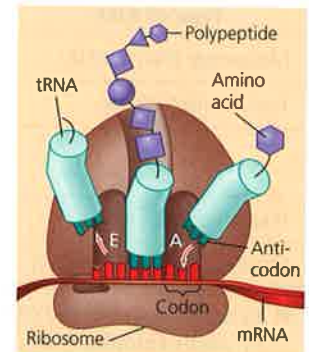
- Most eukaryotic genes are split into segments: They have **introns** interspersed among the **exons** (the regions included in the mRNA). In **RNA splicing**, introns are removed and exons joined. RNA splicing is typically carried out by **spliceosomes**, but in some cases, RNA alone catalyzes its own splicing. The catalytic ability of some RNA molecules, called **ribozymes**, derives from the inherent properties of RNA. The presence of introns allows for **alternative RNA splicing**.

? What function do the 5' cap and the poly-A tail serve on a eukaryotic mRNA?

CONCEPT 17.4

Translation is the RNA-directed synthesis of a polypeptide: a closer look (pp. 383–390)

- A cell translates an mRNA message into protein using **transfer RNAs (tRNAs)**. After being bound to a specific amino acid by an **aminoacyl-tRNA synthetase**, a tRNA lines up via its **anticodon** at the complementary codon on mRNA. A **ribosome**, made up of **ribosomal RNAs (rRNAs)** and proteins, facilitates this coupling with binding sites for mRNA and tRNA.
- Ribosomes coordinate the three stages of translation: initiation, elongation, and termination. The formation of peptide bonds between amino acids is catalyzed by rRNA as tRNAs move through the **A** and **P** sites and exit through the **E** site.
- A single mRNA molecule can be translated simultaneously by a number of ribosomes, forming a **polyribosome**.
- After translation, modifications to proteins can affect their three-dimensional shape. Free ribosomes in the cytosol initiate synthesis of all proteins, but proteins destined for the endomembrane system or for secretion are transported into the ER. Such proteins have a **signal peptide** to which a **signal-recognition particle (SRP)** binds, enabling the translating ribosome to bind to the ER.



? What function do tRNAs serve in the process of translation?

CONCEPT 17.5

Mutations of one or a few nucleotides can affect protein structure and function (pp. 390–392)

- Small-scale **mutations** include **point mutations**, changes in one DNA nucleotide pair, which may lead to production of nonfunctional proteins. **Nucleotide-pair substitutions** can cause **missense** or **nonsense mutations**. Nucleotide-pair **insertions** or **deletions** may produce **frameshift mutations**.
- Spontaneous mutations can occur during DNA replication, recombination, or repair. Chemical and physical **mutagens** cause DNA damage that can alter genes.

? What will be the results of chemically modifying one nucleotide base of a gene? What role is played by DNA repair systems in the cell?

CONCEPT 17.6

While gene expression differs among the domains of life, the concept of a gene is universal (pp. 392–394)

- There are some differences in gene expression among bacteria, archaea, and eukaryotes. Because bacterial cells lack a nuclear envelope, translation can begin while transcription is still in progress. Archaeal cells show similarities to both eukaryotic and bacterial cells in their processes of gene expression. In a eukaryotic cell, the nuclear envelope separates transcription from translation, and extensive RNA processing occurs in the nucleus.
- A gene is a region of DNA whose final functional product is either a polypeptide or an RNA molecule.

? How does the presence of a nuclear envelope affect gene expression in eukaryotes?

TEST YOUR UNDERSTANDING



Multiple-choice Self-Quiz questions #1–7 can be found in the Study Area at www.masteringbiology.com.

8. **DRAW IT** Fill in the following table:

Type of RNA	Functions
Messenger RNA (mRNA)	
Transfer RNA (tRNA)	
	Plays catalytic (ribozyme) roles and structural roles in ribosomes
Primary transcript	
Small nuclear RNA (snRNA)	

9. **EVOLUTION CONNECTION** Most amino acids are coded for by a set of similar codons (see Figure 17.5). What evolutionary explanations can you give for this pattern? (*Hint:* There is one explanation relating to ancestry, and some less obvious ones of a “form-fits-function” type.)

10. SCIENTIFIC INQUIRY

Knowing that the genetic code is almost universal, a scientist uses molecular biological methods to insert the human β -globin gene (shown in Figure 17.11) into bacterial cells, hoping the cells will express it and synthesize functional β -globin protein. Instead, the protein produced is nonfunctional and is found to contain many fewer amino acids than does β -globin made by a eukaryotic cell. Explain why.

11. WRITE ABOUT A THEME

Evolution and The Genetic Basis of Life Evolution accounts for the unity and diversity of life, and the continuity of life is based on heritable information in the form of DNA. In a short essay (100–150 words), discuss how the fidelity with which DNA is inherited is related to the processes of evolution. (Review the discussion of proofreading and DNA repair in Concept 16.2, pp. 362–364.)

For selected answers, see Appendix A.



www.masteringbiology.com

1. MasteringBiology® Assignments

Make Connections Tutorial Point Mutations (Chapter 17) and Protein Structure (Chapter 5)

BioFlix Tutorials Protein Synthesis: Overview • Transcription and RNA Processing • Translation and Protein Targeting Pathways
Tutorials The Genetic Code • Following the Instructions in DNA • Types of RNA • Point Mutations

Activities Overview of Protein Synthesis • RNA Synthesis • Transcription • RNA Processing • Synthesizing Proteins • Translation • The Triplet Nature of the Genetic Code

Questions Student Misconceptions • Reading Quiz • Multiple Choice • End-of-Chapter

2. eText

Read your book online, search, take notes, highlight text, and more.

3. The Study Area

Practice Tests • Cumulative Test • **BioFlix** 3-D Animations • MP3 Tutor Sessions • Videos • Activities • Investigations • Lab Media • Audio Glossary • Word Study Tools • Art