CHAPTER **SEVEN** 7

From DNA to Protein: How Cells Read the Genome

Once the double-helical structure of DNA (deoxyribonucleic acid) had been determined in the early 1950s, it became clear that the hereditary information in cells is encoded in the linear order—or *sequence*—of the four different nucleotide subunits that make up the DNA. We saw in Chapter 6 how this information can be passed on unchanged from a cell to its descendants through the process of DNA replication. But how does the cell decode and use the information? How do genetic instructions written in an alphabet of just four "letters" direct the formation of a bacterium, a fruit fly, or a human? We still have a lot to learn about how the information stored in an organism's genes produces even the simplest unicellular bacterium, let alone how it directs the development of complex multicellular organisms like ourselves. But the DNA code itself has been deciphered, and we have come a long way in understanding how cells read it.

Even before the code was broken, it was known that the information contained in genes somehow directed the synthesis of proteins. Proteins are the principal constituents of cells and determine not only cell structure but also cell function. In previous chapters, we encountered some of the thousands of different kinds of proteins that cells can make. We saw in Chapter 4 that the properties and function of a protein molecule are determined by the sequence of the 20 different amino acid subunits in its polypeptide chain: each type of protein has its own unique amino acid sequence, which dictates how the chain will fold to form a molecule with a distinctive shape and chemistry. The genetic instructions carried by DNA must therefore specify the amino acid sequences of proteins. We will see in this chapter exactly how this happens.

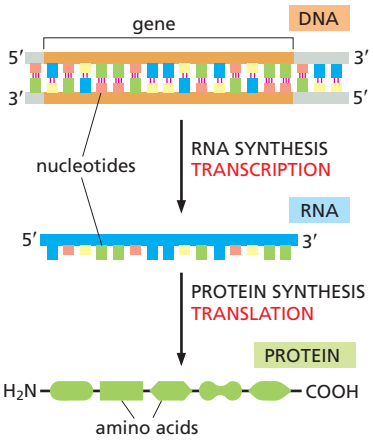


Figure 7–1 Genetic information directs the synthesis of proteins. The flow of genetic information from DNA to RNA (transcription) and from RNA to protein (translation) occurs in all living cells. DNA can also be copied—or replicated—to produce new DNA molecules, as we saw in Chapter 6. The segments of DNA that are transcribed into RNA are called genes (*orange*).

DNA does not synthesize proteins on its own: it acts more like a manager, delegating the various tasks to a team of workers. When a particular protein is needed by the cell, the nucleotide sequence of the appropriate segment of a DNA molecule is first copied into another type of nucleic acid—RNA (*ribonucleic acid*). That segment of DNA is called a **gene**, and the resulting RNA copies are then used to direct the synthesis of the protein. Many thousands of these conversions from DNA to protein occur every second in each cell in our body. The flow of genetic information in cells is therefore from DNA to RNA to protein (**Figure 7–1**). All cells, from bacteria to those in humans, express their genetic information in this way—a principle so fundamental that it has been termed the *central dogma* of molecular biology.

In this chapter, we explain the mechanisms by which cells copy DNA into RNA (a process called *transcription*) and then use the information in RNA to make protein (a process called *translation*). We also discuss a few of the key variations on this basic scheme. Principal among these is *RNA splicing*, a process in eukaryotic cells in which segments of an *RNA transcript* are removed—and the remaining segments stitched back together—before the RNA is translated into protein. We will also learn that, for some genes, it is the RNA, not a protein, that is the final product. In the final section, we consider how the present scheme of information storage, transcription, and translation might have arisen from much simpler systems in the earliest stages of cell evolution.

QUESTION 7–1

Consider the expression "central dogma," which refers to the flow of genetic information from DNA to RNA to protein. Is the word "dogma" appropriate in this context?

FROM DNA TO RNA

The first step in *gene expression*, the process by which cells read out the instructions in their *genes*, is transcription. Many identical RNA copies can be made from the same gene. For most genes, RNA serves solely as an intermediary on the pathway to making a protein. For these genes, each RNA molecule can direct the synthesis, or translation, of many identical protein molecules. This successive amplification enables cells to rapidly synthesize large amounts of protein whenever necessary. At the same time, each gene can be transcribed, and its RNA translated, at different rates, providing the cell with a way to make vast quantities of some proteins and tiny quantities of others (**Figure 7–2**). Moreover, as we discuss in Chapter 8, a cell can change (or regulate) the expression of each of its genes according to the needs of the moment. In this section, we focus on the production of RNA. We describe how the transcriptional machinery recognizes genes and copies the instructions they contain into molecules

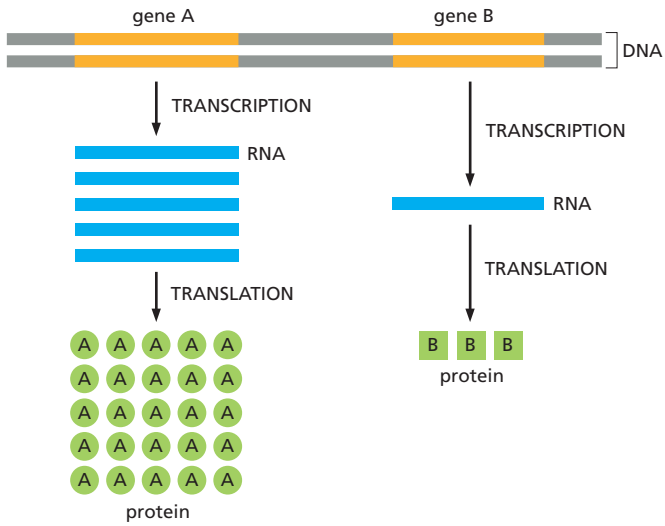


Figure 7–2 A cell can express different genes at different rates. In this and later figures, the portions of the DNA that are not transcribed are shown in *gray*.

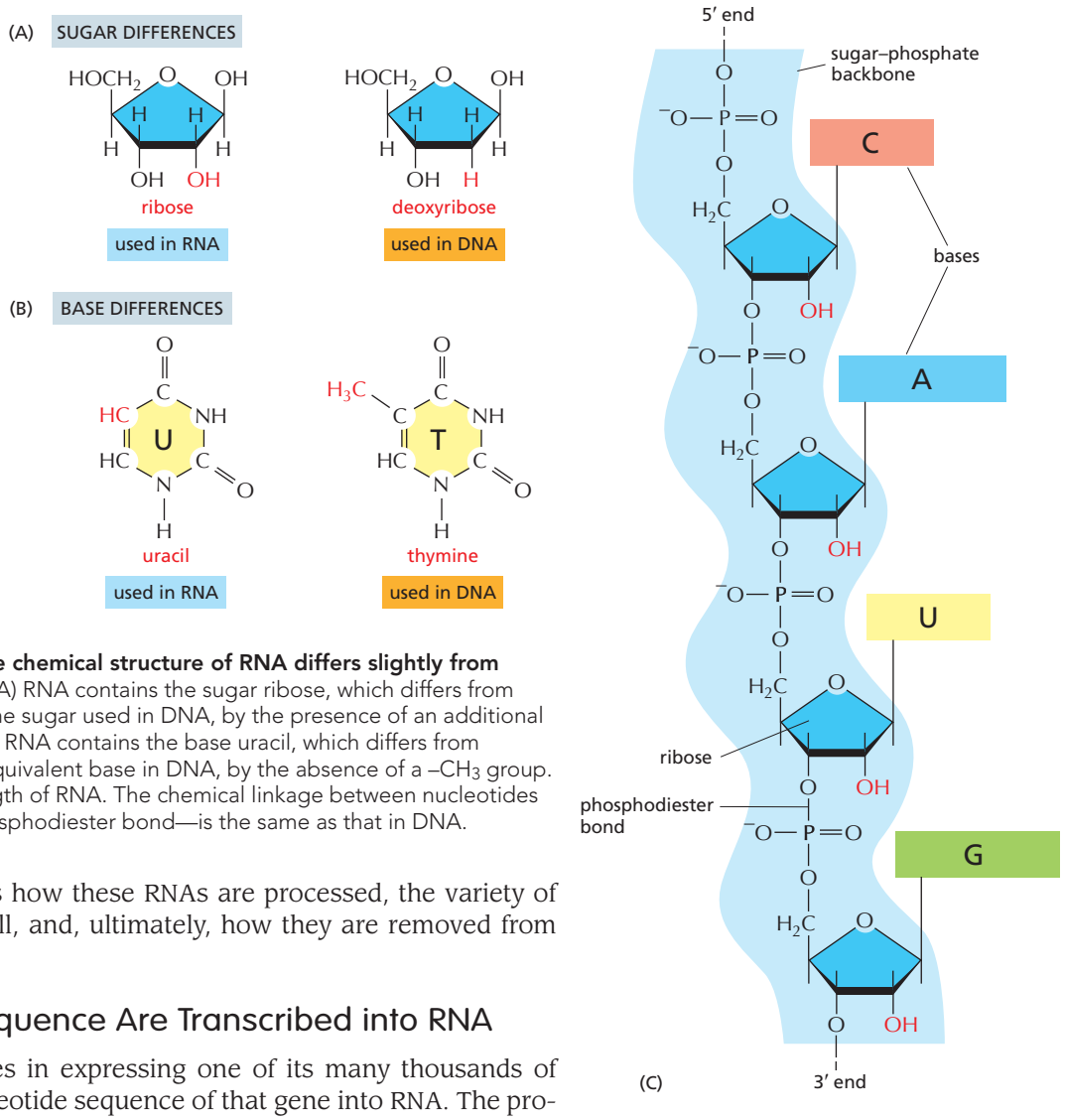


Figure 7–3 The chemical structure of RNA differs slightly from that of DNA. (A) RNA contains the sugar ribose, which differs from deoxyribose, the sugar used in DNA, by the presence of an additional –OH group. (B) RNA contains the base uracil, which differs from thymine, the equivalent base in DNA, by the absence of a $–CH_3$ group. (C) A short length of RNA. The chemical linkage between nucleotides in RNA—a phosphodiester bond—is the same as that in DNA.

of RNA. We then discuss how these RNAs are processed, the variety of roles they play in the cell, and, ultimately, how they are removed from circulation.

Portions of DNA Sequence Are Transcribed into RNA

The first step a cell takes in expressing one of its many thousands of genes is to copy the nucleotide sequence of that gene into RNA. The process is called **transcription** because the information, though copied into another chemical form, is still written in essentially the same language—the language of nucleotides. Like DNA, **RNA** is a linear polymer made of four different nucleotide subunits, linked together by phosphodiester bonds. It differs from DNA chemically in two respects: (1) the nucleotides in RNA are *ribonucleotides*—that is, they contain the sugar ribose (hence the name *ribo*nucleic acid) rather than the deoxyribose found in DNA; and (2) although, like DNA, RNA contains the bases adenine (A), guanine (G), and cytosine (C), it contains uracil (U) instead of the thymine (T) found in DNA (**Figure 7–3**). Because U, like T, can base-pair by hydrogen-bonding with A (**Figure 7–4**), the complementary base-pairing properties described for DNA in Chapter 5 apply also to RNA.

Although their chemical differences are small, DNA and RNA differ quite dramatically in overall structure. Whereas DNA always occurs in cells as a double-stranded helix, RNA is largely single-stranded. This difference has important functional consequences. Because an RNA chain is single-stranded, it can fold up into a variety of shapes, just as a polypeptide chain folds up to form the final shape of a protein (**Figure 7–5**);

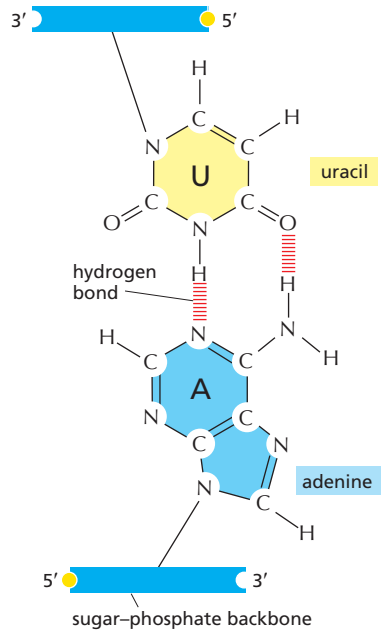


Figure 7–4 Uracil forms a base pair with adenine. The hydrogen bonds that hold the base pair together are shown in *red*. Uracil has the same base-pairing properties as thymine. Thus U-A base pairs in RNA closely resemble T-A base pairs in DNA (see Figure 5–4A).

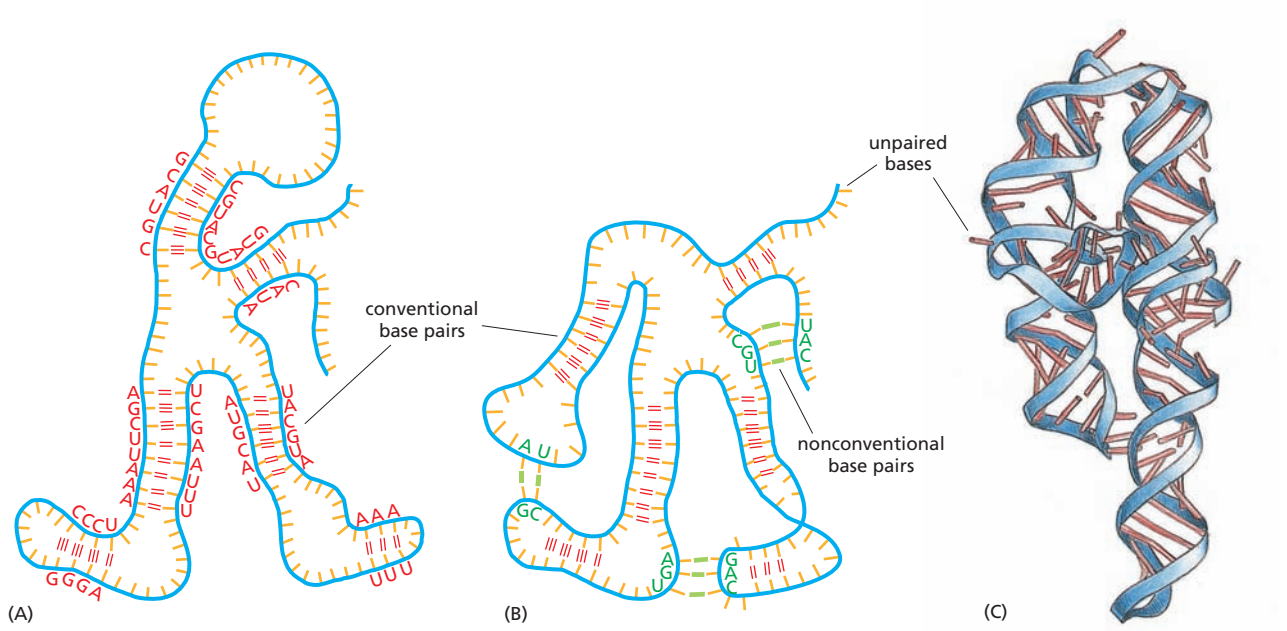


Figure 7–5 RNA molecules can fold into specific structures that are held together by hydrogen bonds between different base pairs. RNA is largely single-stranded, but it often contains short stretches of nucleotides that can base-pair with complementary sequences found elsewhere on the same molecule. These interactions—along with some "nonconventional" base-pair interactions (e.g., A-G)—allow an RNA molecule to fold into a three-dimensional structure that is determined by its sequence of nucleotides. (A) A diagram of a hypothetical, folded RNA structure showing only conventional (G-C and A-U) base-pair interactions (*red*). (B) Formation of nonconventional base-pair interactions (*green*) folds the structure of the hypothetical RNA shown in (A) even further. (C) Structure of an actual RNA molecule that is involved in RNA splicing. The considerable amount of double-helical structure displayed by this RNA is produced by conventional base pairing. For an additional view of RNA structure, see **Movie 7.1**.

double-stranded DNA cannot fold in this fashion. As we discuss later in the chapter, the ability to fold into a complex three-dimensional shape allows RNA to carry out various functions in cells, in addition to conveying information between DNA and protein. Whereas DNA functions solely as an information store, some RNAs have structural, regulatory, or catalytic roles.

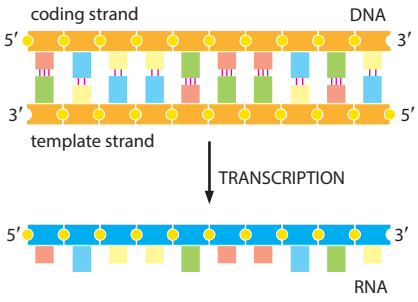


Figure 7–6 Transcription of a gene produces an RNA complementary to one strand of DNA. The *bottom* strand of DNA in this example is called the template strand because it is used to guide the synthesis of the RNA molecule. The nontemplate strand of the gene (here, shown at the *top*) is sometimes called the *coding strand* because its sequence is equivalent to the RNA product, as shown. Which DNA strand serves as the template varies, depending on the gene, as we discuss later. By convention, an RNA molecule is usually depicted with its 5′ end—the first part to be synthesized—to the left.

Transcription Produces RNA That Is Complementary to One Strand of DNA

All the RNA in a cell is made by transcription, a process that has certain similarities to DNA replication (discussed in Chapter 6). Transcription begins with the opening of a small portion of the DNA double helix to expose the bases on each DNA strand. One of the two strands of the DNA double helix then serves as a template for the synthesis of RNA. Ribonucleotides are added, one by one, to the growing RNA chain; as in DNA replication, the nucleotide sequence of the RNA chain is determined by complementary base-pairing with the DNA template strand. When a good match is made, the incoming ribonucleoside triphosphate is covalently linked to the growing RNA chain by the enzyme *RNA polymerase*. The RNA chain produced by transcription—the **RNA transcript**—therefore has a nucleotide sequence exactly complementary to the strand of DNA used as the template (**Figure 7–6**).

Transcription differs from DNA replication, however, in several crucial respects. Unlike a newly formed DNA strand, the RNA strand does not remain hydrogen-bonded to the DNA template strand. Instead, just behind the region where the ribonucleotides are being added, the RNA chain is displaced and the DNA helix re-forms. For this reason—and because only one strand of the DNA molecule is transcribed—RNA molecules are

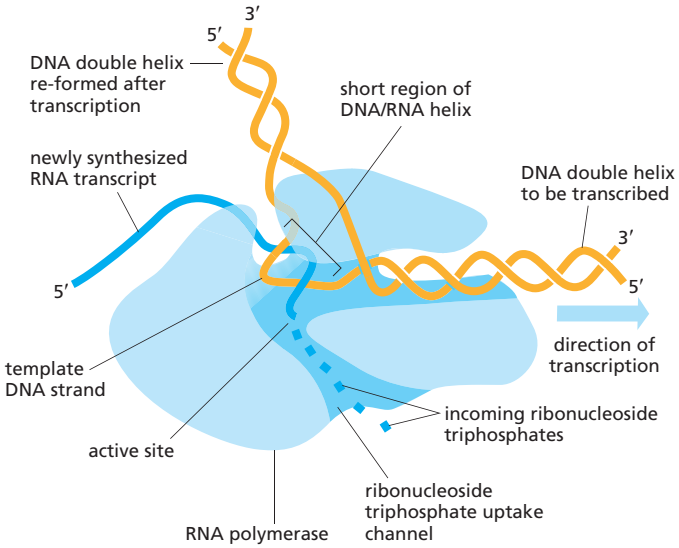


Figure 7–7 DNA is transcribed into RNA by the enzyme RNA polymerase. (A) RNA polymerase (*pale blue*) moves stepwise along the DNA, unwinding the DNA helix in front of it. As it progresses, the polymerase adds ribonucleotides one-by-one to the RNA chain, using an exposed DNA strand as a template. The resulting RNA transcript is thus single-stranded and complementary to the template strand (see Figure 7–6). As the polymerase moves along the DNA template, it displaces the newly formed RNA, allowing the two strands of DNA behind the polymerase to rewind. A short region of hybrid DNA/RNA helix (approximately nine nucleotides in length) therefore forms only transiently, causing a "window" of DNA/RNA helix to move along the DNA with the polymerase. Note that although the primase discussed in Chapter 6 and RNA polymerase both synthesize RNA using a DNA template, they are different enzymes, encoded by different genes.

single-stranded. Furthermore, a given RNA molecule is copied from only a limited region of DNA, making it much shorter than the DNA molecule from which it is made. A DNA molecule in a human chromosome can be up to 250 million nucleotide pairs long, whereas most mature RNAs are no more than a few thousand nucleotides long, and many are much shorter than that.

Like the DNA polymerase that carries out DNA replication (discussed in Chapter 6), **RNA polymerases** catalyze the formation of the phosphodiester bonds that link the nucleotides together and form the sugar–phosphate backbone of the RNA chain (see Figure 7–3). The RNA polymerase moves stepwise along the DNA, unwinding the DNA helix just ahead to expose a new region of the template strand for complementary base-pairing. In this way, the growing RNA chain is elongated by one nucleotide at a time in the 5′-to-3′ direction (**Figure 7–7**). The incoming ribonucleoside triphosphates (ATP, CTP, UTP, and GTP) provide the energy needed to drive the reaction forward, analogous to the process of DNA synthesis (see Figure 6–11).

The almost immediate release of the RNA strand from the DNA as it is synthesized means that many RNA copies can be made from the same gene in a relatively short time; the synthesis of the next RNA is usually started before the first RNA has been completed (**Figure 7–8**). A medium-sized gene—say, 1500 nucleotide pairs—requires approximately 50 seconds for a molecule of RNA polymerase to transcribe it (**Movie 7.2**). At any given time, there could be dozens of polymerases speeding along this single stretch of DNA, hard on one another's heels, allowing more than 1000 transcripts to be synthesized in an hour. For most genes, however, the amount of transcription is much less than this.

QUESTION 7–2

In the electron micrograph in Figure 7–8, are the RNA polymerase molecules moving from right to left or from left to right? Why are the RNA transcripts so much shorter than the DNA segments (genes) that encode them?

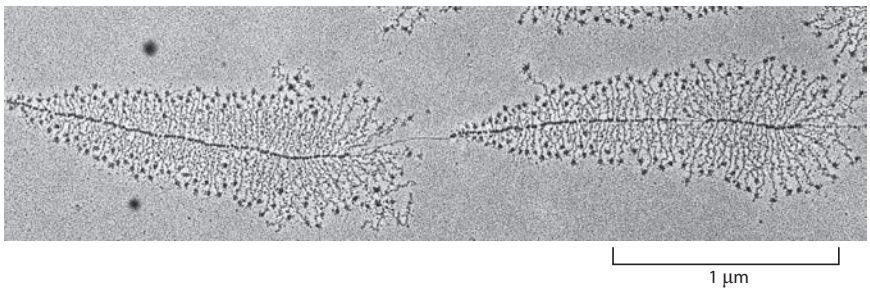


Figure 7–8 Many molecules of RNA polymerase can simultaneously transcribe the same gene. Shown in this electron micrograph are two adjacent ribosomal genes on a single DNA molecule. Molecules of RNA polymerase are barely visible as a series of tiny dots along the spine of the DNA molecule; each polymerase has an RNA transcript (a short, fine thread) radiating from it. The RNA molecules being transcribed from the two ribosomal genes—ribosomal RNAs (rRNAs)—are not translated into protein, but are instead used directly as components of ribosomes, macromolecular machines made of RNA and protein. The large particles that can be seen at the free, 5′ end of each rRNA transcript are ribosomal proteins that have assembled on the ends of the growing transcripts. These proteins will be discussed later in the chapter. (Courtesy of Ulrich Scheer.)

Although RNA polymerase catalyzes essentially the same chemical reaction as DNA polymerase, there are some important differences between the two enzymes. First, and most obviously, RNA polymerase uses ribonucleoside triphosphates as substrates, so it catalyzes the linkage of ribonucleotides, not deoxyribonucleotides. Second, unlike the DNA polymerase involved in DNA replication, RNA polymerases can start an RNA chain without a primer and do not accurately proofread their work. This sloppiness is tolerated because RNA, unlike DNA, is not used as the permanent storage form of genetic information in cells, so mistakes in RNA transcripts have relatively minor consequences for a cell. RNA polymerases make about one mistake for every 10^4 nucleotides copied into RNA, whereas DNA polymerase makes only one mistake for every 10^7 nucleotides copied.

Cells Produce Various Types of RNA

The majority of genes carried in a cell's DNA specify the amino acid sequences of proteins. The RNA molecules encoded by these genes—which ultimately direct the synthesis of proteins—are called **messenger RNAs** (**mRNAs**). In eukaryotes, each mRNA typically carries information transcribed from just one gene, which codes for a single protein; in bacteria, a set of adjacent genes is often transcribed as a single mRNA, which therefore carries the information for several different proteins.

The final product of other genes, however, is the RNA itself. As we see later, these *noncoding RNAs*, like proteins, have various roles, serving as regulatory, structural, and catalytic components of cells. They play key parts, for example, in translating the genetic message into protein: *ribosomal RNAs* (*rRNAs*) form the structural and catalytic core of the ribosomes, which translate mRNAs into protein, and *transfer RNAs* (*tRNAs*) act as adaptors that select specific amino acids and hold them in place on a ribosome for their incorporation into protein. Other small RNAs, called *microRNAs* (*miRNAs*), serve as key regulators of eukaryotic gene expression, as we discuss in Chapter 8. The most common types of RNA are summarized in Table 7–1.

In the broadest sense, the term **gene expression** refers to the process by which the information encoded in a DNA sequence is converted into a product, whether RNA or protein, that has some effect on a cell or organism. In cases where the final product of the gene is a protein, gene expression includes both transcription and translation. When an RNA molecule is the gene's final product, however, gene expression does not require translation.

TABLE 7–1 TYPES OF RNA PRODUCED IN CELLS

Type of RNA	Function
messenger RNAs (mRNAs)	code for proteins
ribosomal RNAs (rRNAs)	form the core of the ribosome's structure and catalyze protein synthesis
microRNAs (miRNAs)	regulate gene expression
transfer RNAs (tRNAs)	serve as adaptors between mRNA and amino acids during protein synthesis
Other noncoding RNAs	used in RNA splicing, gene regulation, telomere maintenance, and many other processes

Signals in the DNA Tell RNA Polymerase Where to Start and Stop Transcription

The initiation of transcription is an especially critical process because it is the main point at which the cell selects which RNAs are to be produced. To begin transcription, RNA polymerase must be able to recognize the start of a gene and bind firmly to the DNA at this site. The way in which RNA polymerases recognize the *transcription start site* of a gene differs somewhat between bacteria and eukaryotes. Because the situation in bacteria is simpler, we describe it first.

When an RNA polymerase collides randomly with a DNA molecule, the enzyme sticks weakly to the double helix and then slides rapidly along its length. RNA polymerase latches on tightly only after it has encountered a gene region called a **promoter**, which contains a specific sequence of nucleotides that lies immediately upstream of the starting point for RNA synthesis. As it binds tightly to this sequence, the RNA polymerase opens up the double helix immediately in front of the promoter to expose the nucleotides on each strand of a short stretch of DNA. One of the two exposed DNA strands then acts as a template for complementary base-pairing with incoming ribonucleoside triphosphates, two of which are joined together by the polymerase to begin synthesis of the RNA strand. Elongation then continues until the enzyme encounters a second signal in the DNA, the *terminator* (or stop site), where the polymerase halts and releases both the DNA template and the newly made RNA transcript (**Figure 7–9**). The terminator sequence itself is also transcribed, and it is the interaction of this 3′ segment of RNA with the polymerase that causes the enzyme to let go of the template DNA.

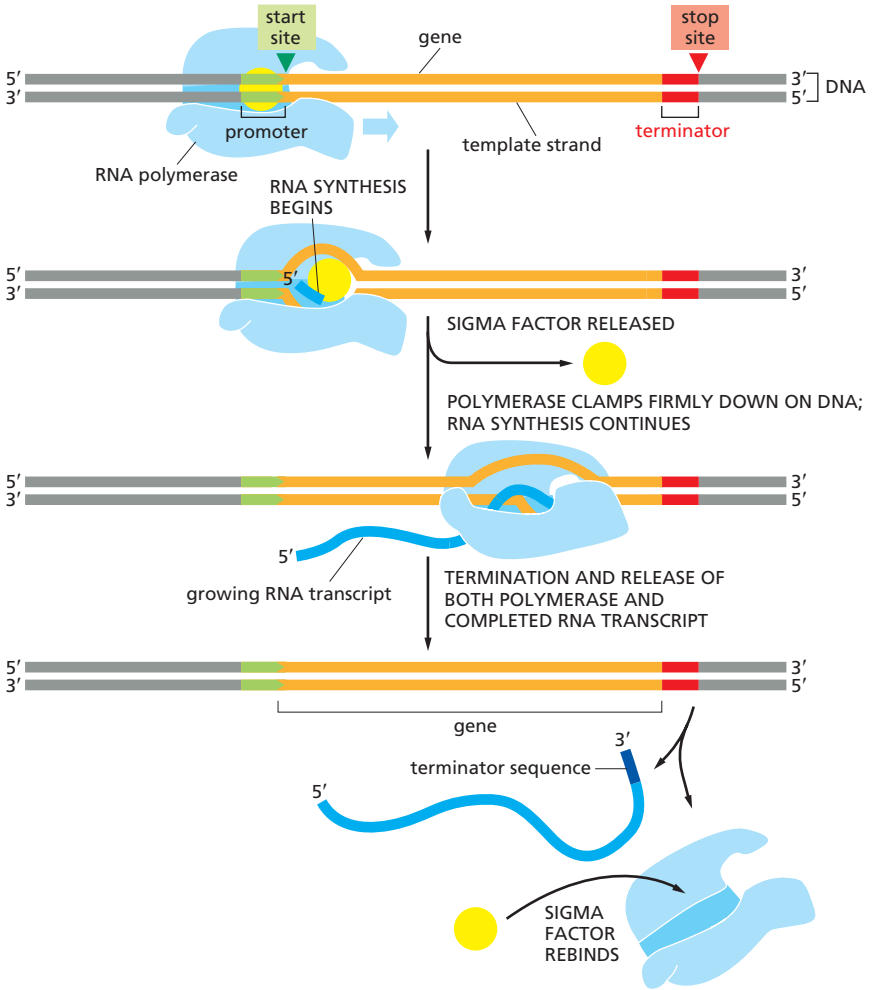


Figure 7–9 Signals in the nucleotide sequence of a gene tell bacterial RNA polymerase where to start and stop transcription. Bacterial RNA polymerase (*light blue*) contains a subunit called sigma factor (*yellow*) that recognizes the promoter of a gene (*green*). Once transcription has begun, sigma factor is released, and the polymerase moves forward and continues synthesizing the RNA. Elongation continues until the polymerase encounters a sequence in the gene called the terminator (*red*). After transcribing this sequence into RNA (*dark blue*), the enzyme halts and releases both the DNA template and the newly made RNA transcript. The polymerase then reassociates with a free sigma factor and searches for another promoter to begin the process again.

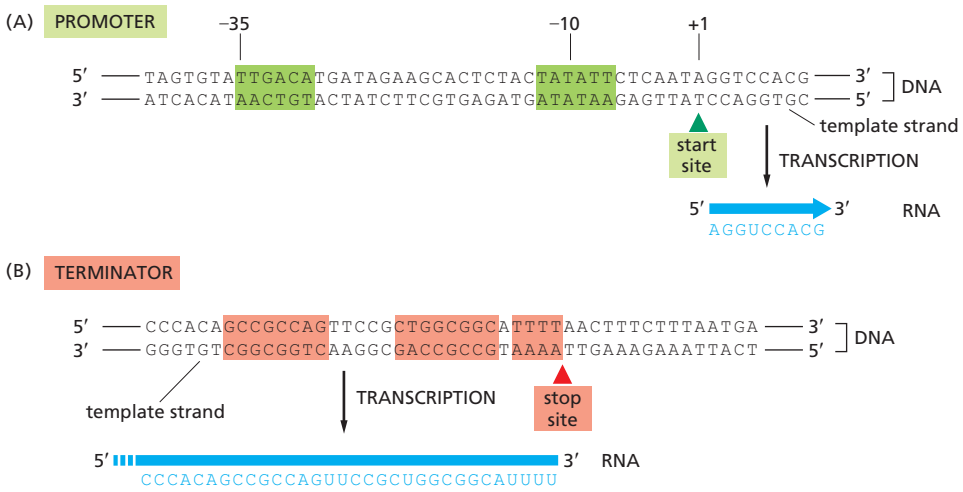


Figure 7–10 Bacterial promoters and terminators have specific nucleotide sequences that are recognized by RNA polymerase. (A) The *green*-shaded regions represent the nucleotide sequences that specify a promoter. The numbers above the DNA indicate the positions of nucleotides counting from the first nucleotide transcribed, which is designated +1. The polarity of the promoter orients the polymerase and determines which DNA strand is transcribed. All bacterial promoters contain DNA sequences at –10 and –35 that closely resemble those shown here. (B) The *red*-shaded regions represent sequences in the gene that signal the RNA polymerase to terminate transcription. Note that the regions transcribed into RNA contain the terminator but not the promoter nucleotide sequences.

Because the polymerase must bind tightly to DNA before transcription can begin, a segment of DNA will be transcribed only if it is preceded by a promoter. This ensures that only those portions of a DNA molecule that contain a gene will be transcribed into RNA. The nucleotide sequences of a typical promoter—and a typical terminator—are presented in Figure 7–10.

In bacteria, it is a subunit of RNA polymerase, the *sigma* (σ) factor (see Figure 7–9), that is primarily responsible for recognizing the promoter sequence on the DNA. But how can this factor "see" the promoter, given that the base pairs in question are situated in the interior of the DNA double helix? It turns out that each base presents unique features to the outside of the double helix, allowing the sigma factor to initially identify the promoter sequence without having to separate the entwined DNA strands. As it begins to open the DNA double helix, the sigma factor then binds to the exposed base pairs, keeping the double helix open.

The next problem an RNA polymerase faces is determining which of the two DNA strands to use as a template for transcription: each strand has a different nucleotide sequence and would produce a different RNA transcript. The secret lies in the structure of the promoter itself. Every promoter has a certain polarity: it contains two different nucleotide sequences, laid out in a specific 5′-to-3′ order, upstream of the transcriptional start site. These asymmetric sequences position the RNA polymerase such that it binds to the promoter in the correct orientation (see Figure 7–10A). Because the polymerase can only synthesize RNA in the 5′-to-3′ direction, once the enzyme is bound it must use the DNA strand that is oriented in the 3′-to-5′ direction as its template.

This selection of a template strand does not mean that on a given chromosome, all transcription proceeds in the same direction. With respect to the chromosome as a whole, the direction of transcription can vary from one gene to the next. But because each gene typically has only one promoter, the orientation of its promoter determines in which direction that gene is transcribed and therefore which strand is the template strand (Figure 7–11).

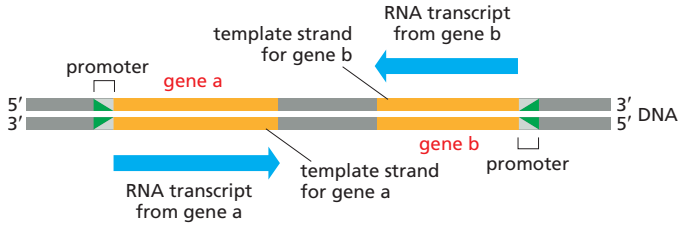


Figure 7–11 On an individual chromosome, some genes are transcribed using one DNA strand as a template, and others are transcribed from the other DNA strand. RNA polymerase always moves in the 3′-to-5′ direction with respect to the template DNA strand. Which strand will serve as the template is determined by the polarity of the promoter sequences (*green* arrowheads) at the beginning of each gene. In this drawing, gene a, which is transcribed from left to right, uses the bottom DNA strand as its template (see Figure 7–10); gene b, which is transcribed from right to left, uses the top strand as its template.

Initiation of Eukaryotic Gene Transcription Is a Complex Process

Many of the principles we just outlined for bacterial transcription also apply to eukaryotes. However, the initiation of transcription in eukaryotes differs in several important ways from the process in bacteria:

1. While bacteria use a single type of RNA polymerase for transcription, eukaryotic cells employ three: *RNA polymerase I*, *RNA polymerase II*, and *RNA polymerase III*. These polymerases are responsible for transcribing different types of genes. RNA polymerases I and III transcribe the genes encoding transfer RNA, ribosomal RNA, and various other RNAs that play structural and catalytic roles in the cell (**Table 7–2**). RNA polymerase II transcribes the rest, including all those that encode proteins—which constitutes the majority of genes in eukaryotes (**Movie 7.3**). Our subsequent discussion will therefore focus on RNA polymerase II.
2. Whereas the bacterial RNA polymerase (along with its sigma subunit) is able to initiate transcription on its own, eukaryotic RNA polymerases require the assistance of a large set of accessory proteins. Principal among these are the *general transcription factors*, which must assemble at each promoter, along with the polymerase, before transcription can begin.
3. The mechanisms that control the initiation of transcription in eukaryotes are much more elaborate than those that operate in prokaryotes—a point we discuss in detail in Chapter 8. In bacteria, genes tend to lie very close to one another, with only very short lengths of nontranscribed DNA between them. But in plants and animals, including humans, individual genes are spread out along the DNA, with stretches of up to 100,000 nucleotide pairs between one gene and the next. This architecture allows a single gene to be controlled by a large variety of *regulatory DNA sequences* scattered along the DNA, and it enables eukaryotes to engage in more complex forms of transcriptional regulation than do bacteria.
4. Eukaryotic transcription initiation must deal with the packing of DNA into *nucleosomes* and higher-order forms of chromatin structure, as we describe in Chapter 8.

To begin our discussion of eukaryotic transcription, we take a look at the general transcription factors and see how they help RNA polymerase II initiate the process.

Eukaryotic RNA Polymerase Requires General Transcription Factors

The initial finding that, unlike bacterial RNA polymerase, purified eukaryotic RNA polymerase II cannot initiate transcription on its own in a test tube led to the discovery and purification of the **general transcription**

QUESTION 7–3

Could the RNA polymerase used for transcription also be used to make the RNA primers required for DNA replication (discussed in Chapter 6)?

TABLE 7–2 THE THREE RNA POLYMERASES IN EUKARYOTIC CELLS

Type of Polymerase	Genes Transcribed
RNA polymerase I	most rRNA genes
RNA polymerase II	all protein-coding genes, miRNA genes, plus genes for other noncoding RNAs (e.g., those of the spliceosome)
RNA polymerase III	tRNA genes, 5S rRNA gene, genes for many other small RNAs

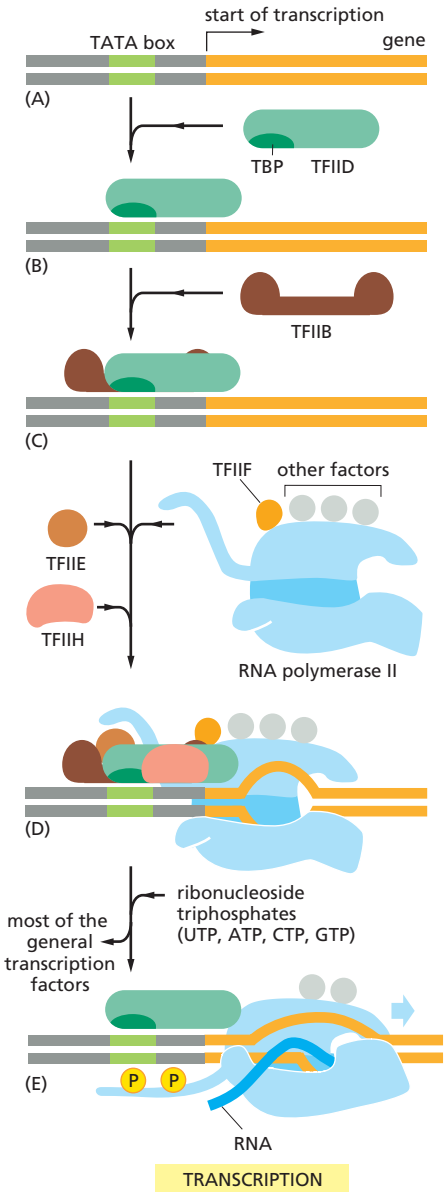


Figure 7–12 To begin transcription, eukaryotic RNA polymerase II requires a set of general transcription factors. These factors are designated TFIIB, TFIID, and so on. (A) Most eukaryotic promoters contain a DNA sequence called the TATA box. (B) The TATA box is recognized by a subunit of the general transcription factor TFIID, called the TATA-binding protein (TBP). For simplicity, the DNA distortion produced by the binding of the TBP (see Figure 7–13) is not shown. (C) The binding of TFIID enables the adjacent binding of TFIIB. (D) The rest of the general transcription factors, as well as the RNA polymerase itself, then assemble at the promoter. (E) TFIIH pries apart the double helix at the transcription start point, using the energy of ATP hydrolysis, which exposes the template strand of the gene. TFIIH also phosphorylates RNA polymerase II, releasing the polymerase from most of the general transcription factors, so it can begin transcription. The site of phosphorylation is a long polypeptide "tail" that extends from the polymerase. Once the polymerase moves away from the promoter, most of the general transcription factors are released from the DNA; the exception is TFIID, which remains bound through multiple rounds of transcription initiation.

factors. These accessory proteins assemble on the promoter, where they position the RNA polymerase and pull apart the DNA double helix to expose the template strand, allowing the polymerase to begin transcription. Thus, the general transcription factors have a similar role in eukaryotic transcription as sigma factor has in bacterial transcription.

Figure 7–12 shows the assembly of the general transcription factors at a promoter used by RNA polymerase II. The process begins with the binding of the general transcription factor TFIID to a short segment of DNA double helix composed primarily of T and A nucleotides; because of its composition, this part of the promoter is known as the *TATA box*. Upon binding to DNA, TFIID causes a dramatic local distortion in the DNA double helix (**Figure 7–13**); this structure helps to serve as a landmark for the subsequent assembly of other proteins at the promoter. The TATA box is a key component of many promoters used by RNA polymerase II, and it is typically located about 30 nucleotides upstream from the transcription start site. Once TFIID has bound to the TATA box, the other factors assemble, along with RNA polymerase II, to form a complete *transcription initiation complex*. Although Figure 7–12 shows the general transcription factors loading onto the promoter in a certain sequence, the actual order of assembly probably differs somewhat from one promoter to the next. Like bacterial promoters, eukaryotic promoters are composed of several distinct DNA sequences; these direct the general transcription factors where to assemble, and they orient the RNA polymerase so that it will begin transcription in the correct direction and on the correct DNA template strand (**Figure 7–14**).

Once RNA polymerase II has been positioned on the promoter, it must be released from the complex of general transcription factors to begin its task of making an RNA molecule. A key step in liberating the RNA polymerase is the addition of phosphate groups to its "tail" (see Figure

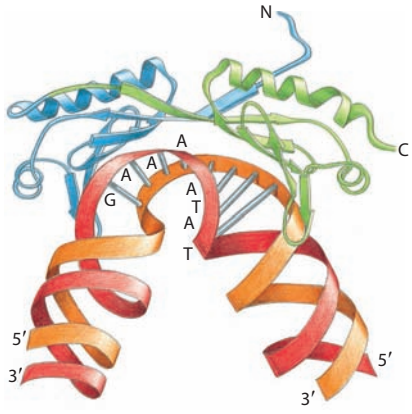
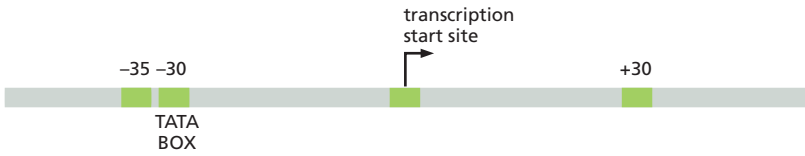


Figure 7–13 TATA-binding protein (TBP) binds to the TATA box (indicated by letters) and bends the DNA double helix. TBP, a subunit of TFIID (see Figure 7–12), distorts the DNA when it binds. TBP is a single polypeptide chain that is folded into two very similar domains (*blue* and *green*). The protein sits atop the DNA double helix like a saddle on a bucking horse (**Movie 7.4**).



location	DNA sequence	general transcription factor
–35	G/C G/C G/A C G C C	TFIIB
–30	T A T A A/T A A/T	TBP subunit of TFIID
transcription start site	C/T C/T A N T/A C/T C/T	TFIID
+30	A/G G A/T C G T G	TFIID

Figure 7–14 Eukaryotic promoters contain sequences that promote the binding of the general transcription factors. The location of each sequence and the general transcription factor that recognizes it are indicated. N stands for any nucleotide, and a *slash* (/) indicates that either nucleotide can be found at the indicated position. For most start sites transcribed by RNA polymerase II, only two or three of the four sequences are needed. Although most of these DNA sequences are located upstream of the transcription start site, one, at +30, is located within the transcribed region of the gene.

7–12E). This action is initiated by the general transcription factor TFIIH, which contains a protein kinase as one of its subunits. Once transcription has begun, most of the general transcription factors dissociate from the DNA and then are available to initiate another round of transcription with a new RNA polymerase molecule. When RNA polymerase II finishes transcribing a gene, it too is released from the DNA; the phosphates on its tail are stripped off by protein phosphatases, and the polymerase is then ready to find a new promoter. Only the dephosphorylated form of RNA polymerase II can re-initiate RNA synthesis.

Eukaryotic mRNAs Are Processed in the Nucleus

The principle of templating, by which DNA is transcribed into RNA, is the same in all organisms; however, the way in which the resulting RNA transcripts are handled before they are translated into protein differs between bacteria and eukaryotes. Because bacteria lack a nucleus, their DNA is directly exposed to the cytosol, which contains the *ribosomes* on which protein synthesis takes place. As an mRNA molecule in a bacterium starts to be synthesized, ribosomes immediately attach to the free 5′ end of the RNA transcript and begin translating it into protein.

In eukaryotic cells, by contrast, DNA is enclosed within the *nucleus*, which is where transcription takes place. Translation, however, occurs on ribosomes that are located in the cytosol. So, before a eukaryotic mRNA can be translated into protein, it must be transported out of the nucleus through small pores in the nuclear envelope (**Figure 7–15**). And before it can be exported to the cytosol, a eukaryotic RNA must go through several **RNA processing** steps, which include *capping*, *splicing*, and *polyadenylation*, as we discuss shortly. These steps take place as the RNA is being synthesized. The enzymes responsible for RNA processing ride on the phosphorylated tail of eukaryotic RNA polymerase II as it synthesizes an RNA molecule (see Figure 7–12), and they process the transcript as it emerges from the polymerase (**Figure 7–16**).

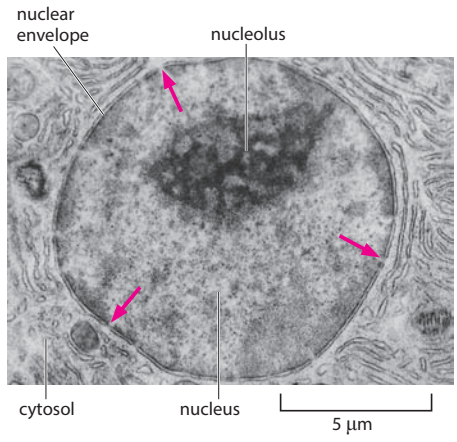


Figure 7–15 Before they can be translated, mRNA molecules made in the nucleus must be exported to the cytosol via pores in the nuclear envelope (*red* arrows). Shown here is a section of a liver cell nucleus. The nucleolus is where ribosomal RNAs are synthesized and combined with proteins to form ribosomes, which are then exported to the cytosol. (From D.W. Fawcett, *A Textbook of Histology*, 12th ed. 1994. With permission from Taylor & Francis Books UK.)

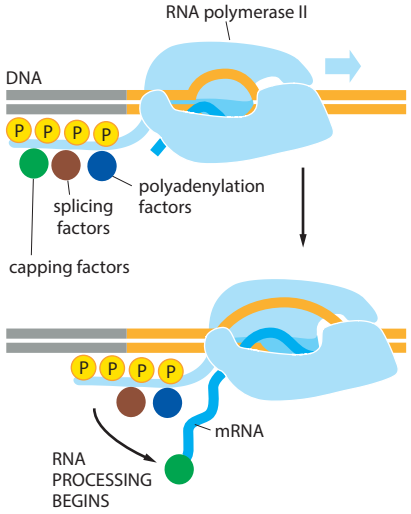


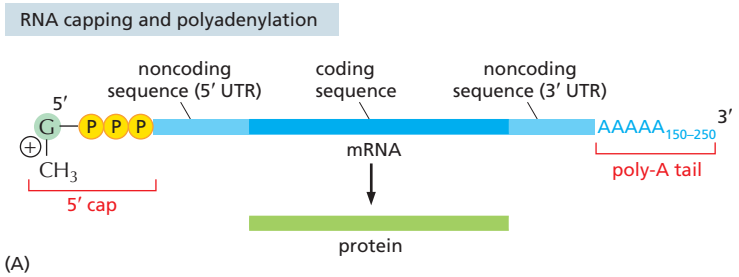
Figure 7–16 Phosphorylation of the tail of RNA polymerase II allows RNA-processing proteins to assemble there. Capping, polyadenylation, and splicing are all modifications that occur as the RNA is being synthesized. Note that the phosphates shown here are in addition to the ones required for transcription initiation (see Figure 7–12).

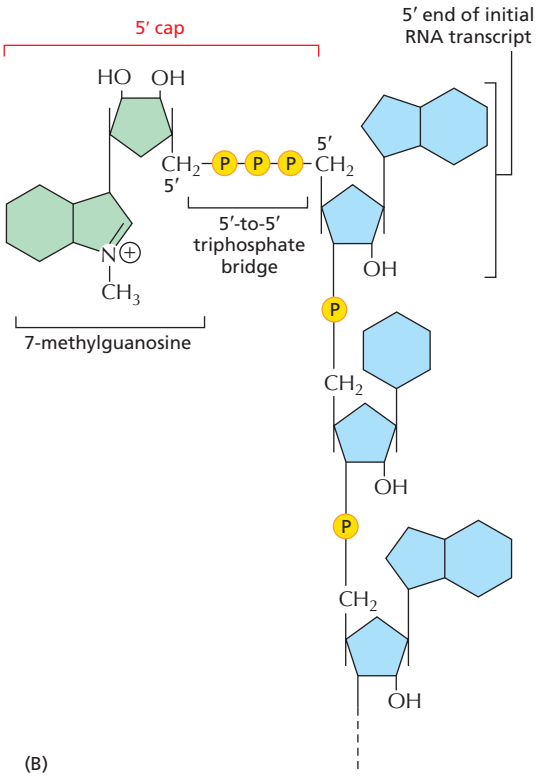
Two of these processing steps, capping and polyadenylation, occur on all RNA transcripts destined to become mRNA molecules.

1. **RNA capping** modifies the 5′ end of the RNA transcript, the part of the RNA that is synthesized first. The RNA cap includes an atypical nucleotide: a guanine (G) nucleotide bearing a methyl group is attached to the 5′ end of the RNA in an unusual way (**Figure 7–17**). In bacteria, by contrast, the 5′ end of an mRNA molecule is simply the first nucleotide of the transcript. In eukaryotic cells, capping takes place after RNA polymerase II has produced about 25 nucleotides of RNA, long before it has completed transcribing the whole gene.
2. **Polyadenylation** provides a newly transcribed mRNA with a special structure at its 3′ end. In contrast with bacteria, where the 3′ end of an mRNA is simply the end of the chain synthesized by the RNA polymerase, the 3′ end of a eukaryotic mRNA is first trimmed by an enzyme that cuts the RNA chain at a particular sequence of nucleotides. The transcript is then finished off by a second enzyme that adds a series of repeated adenine (A) nucleotides to the trimmed end. This *poly-A tail* is generally a few hundred nucleotides long (see Figure 7–17A).

These two modifications—capping and polyadenylation—increase the stability of a eukaryotic mRNA molecule, facilitate its export from the nucleus to the cytosol, and generally mark the RNA molecule as an mRNA. They are also used by the protein-synthesis machinery to make sure that both ends of the mRNA are present and that the message is therefore complete before protein synthesis begins.

Figure 7–17 Eukaryotic mRNA molecules are modified by capping and polyadenylation. (A) A eukaryotic mRNA has a cap at the 5′ end and a poly-A tail at the 3′ end. In addition to the nucleotide sequences that code for protein, most mRNAs also contain extra, noncoding sequences, as shown. The noncoding portion at the 5′ end is called the 5′ untranslated region, or 5′ UTR, and that at the 3′ end is called the 3′ UTR. (B) The structure of the 5′ cap. Many eukaryotic mRNA caps carry an additional modification: the 2′-hydroxyl group on the second ribose sugar in the mRNA is methylated (not shown).





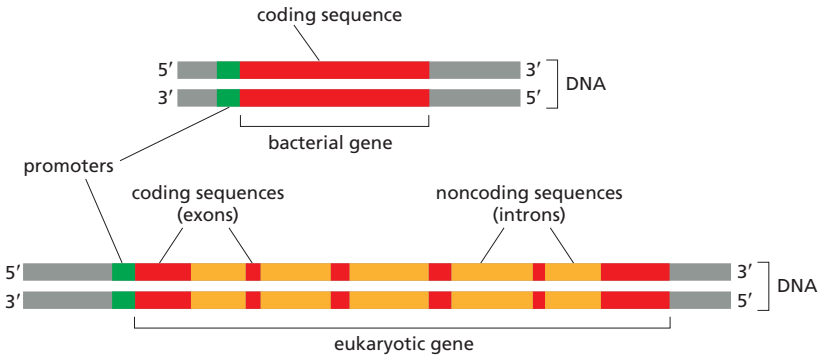


Figure 7–18 Eukaryotic and bacterial genes are organized differently. A bacterial gene consists of a single stretch of uninterrupted nucleotide sequence that encodes the amino acid sequence of a protein. In contrast, the protein-coding sequences of most eukaryotic genes (*exons*) are interrupted by noncoding sequences (*introns*). Promoter sequences are indicated in *green*.

In Eukaryotes, Protein-Coding Genes Are Interrupted by Noncoding Sequences Called Introns

Most eukaryotic mRNAs have to undergo an additional processing step before they become functional. This step involves a far more radical modification of the RNA transcript than capping or polyadenylation, and it is the consequence of a surprising feature of most eukaryotic genes. In bacteria, most proteins are encoded by an uninterrupted stretch of DNA sequence that is transcribed into an mRNA that, without any further processing, can be translated into protein. Most protein-coding eukaryotic genes, in contrast, have their coding sequences interrupted by long, noncoding, *intervening sequences* called **introns**. The scattered pieces of coding sequence—called *expressed sequences* or **exons**—are usually shorter than the introns, and they often represent only a small fraction of the total length of the gene (**Figure 7–18**). Introns range in length from a single nucleotide to more than 10,000 nucleotides. Some protein-coding eukaryotic genes lack introns altogether, some have only a few, but most have many (**Figure 7–19**). Note that the terms "exon" and "intron" apply to both the DNA and the corresponding RNA sequences.

Introns Are Removed from Pre-mRNAs by RNA Splicing

To produce an mRNA in a eukaryotic cell, the entire length of the gene, introns as well as exons, is transcribed into RNA. After capping, and as RNA polymerase II continues to transcribe the gene, **RNA splicing** begins. In this process, the introns are removed from the newly synthesized RNA and the exons are stitched together. Each transcript ultimately receives a poly-A tail; in many cases, this happens after splicing, whereas in other cases, it occurs before the final splicing reactions have been completed. Once a transcript has been spliced and its 5′ and 3′ ends have been modified, the RNA is now a functional mRNA molecule that can leave the nucleus and be translated into protein. Before these steps are completed, the RNA transcript is known as a *precursor-mRNA* or *pre-mRNA* for short.

How does the cell determine which parts of the RNA transcript to remove during splicing? Unlike the coding sequence of an exon, most of the nucleotide sequence of an intron is unimportant. Although there is little overall resemblance between the nucleotide sequences of different

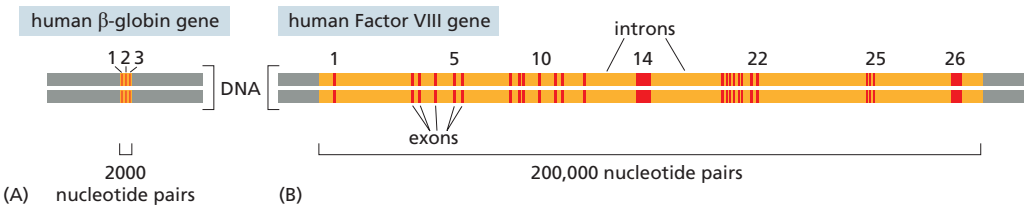


Figure 7–19 Most protein-coding human genes are broken into multiple exons and introns. (A) The β-globin gene, which encodes one of the subunits of the oxygen-carrying protein hemoglobin, contains 3 exons. (B) The gene that encodes Factor VIII, a protein that functions in the blood-clotting pathway, contains 26 exons. Mutations in this large gene are responsible for the most prevalent form of the blood disorder hemophilia.

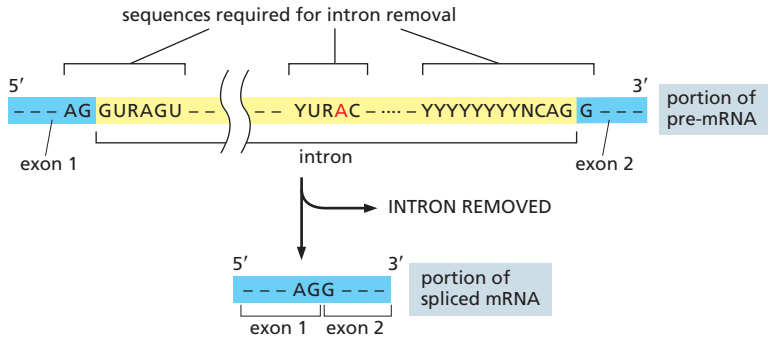


Figure 7–20 Special nucleotide sequences in a pre-mRNA transcript signal the beginning and the end of an intron. Only the nucleotide sequences shown are required to remove an intron; the other positions in an intron can be occupied by any nucleotide. The special sequences are recognized primarily by small nuclear ribonucleoproteins (snRNPs), which direct the cleavage of the RNA at the intron–exon borders and catalyze the covalent linkage of the exon sequences. Here, in addition to the standard symbols for nucleotides (A, C, G, U), R stands for either A or G; Y stands for either C or U; and N stands for any nucleotide. The A shown in *red* forms the branch point of the lariat produced in the splicing reaction shown in Figure 7–21. The distances along the RNA between the three splicing sequences are highly variable; however, the distance between the branch point and the 5′ splice junction is typically much longer than that between the 3′ splice junction and the branch point (see Figure 7–21). The splicing sequences shown are from humans; similar sequences direct RNA splicing in other eukaryotes.

introns, each intron contains a few short nucleotide sequences that act as cues for its removal from the pre-mRNA. These special sequences are found at or near each end of the intron and are the same or very similar in all introns (**Figure 7–20**). Guided by these sequences, an elaborate splicing machine cuts out the intron in the form of a "lariat" structure (**Figure 7–21**), formed by the reaction of an adenine nucleotide, highlighted in *red* in both Figures 7–20 and 7–21, with the beginning of the intron.

Although we will not describe the splicing process in detail, it is worthwhile to note that, unlike the other steps of mRNA production, RNA splicing is carried out largely by RNA molecules rather than proteins. These RNA molecules, called **small nuclear RNAs** (**snRNAs**), are packaged with additional proteins to form *small nuclear ribonucleoproteins* (*snRNPs*, pronounced "snurps"). The snRNPs recognize splice-site sequences through complementary base-pairing between their RNA components and the sequences in the pre-mRNA, and they carry out the chemistry of splicing (**Figure 7–22**). RNA molecules that catalyze reactions in this way are known as *ribozymes*, and we discuss them in more detail later in the chapter. Together, these snRNPs form the core of the **spliceosome**, the large assembly of RNA and protein molecules that carries out RNA splicing in the nucleus. To watch the spliceosome in action, see **Movie 7.5**.

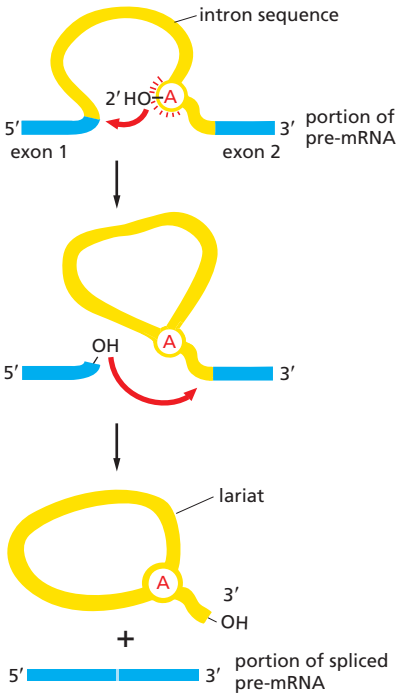


Figure 7–21 An intron in a pre-mRNA molecule forms a branched structure during RNA splicing. In the first step, the branch-point adenine (*red* A) in the intron sequence attacks the 5′ splice site and cuts the sugar–phosphate backbone of the RNA at this point (this is the same A highlighted in *red* in Figure 7–20). In this process, the released 5′ end of the intron becomes covalently linked to the 2′-OH group of the ribose of the adenine nucleotide to form a branched structure. In the second step of splicing, the free 3′-OH end of the exon sequence reacts with the start of the next exon sequence, joining the two exons together into a continuous coding sequence. The intron is released as a lariat structure, which is eventually degraded in the nucleus.

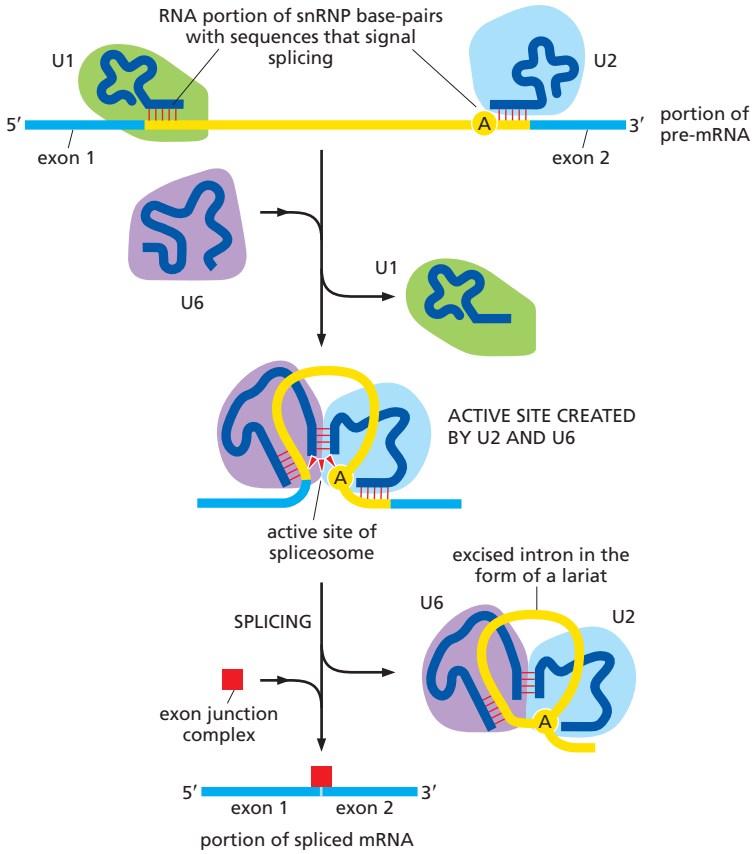


Figure 7–22 Splicing is carried out by a collection of RNA–protein complexes called snRNPs. Although there are five snRNPs and about 200 additional proteins required for splicing, only the three most important snRNPs—called U1, U2, and U6—are shown here. In the first steps of splicing, U1 recognizes the 5′ splice site and U2 recognizes the lariat branch-point site through complementary base-pairing. U6 then "re-checks" the 5′ splice site by displacing U1 and base-pairing with this intron sequence itself. This "re-reading" step improves the accuracy of splicing by double-checking the 5′ splice site before carrying out the splicing reaction. In the next steps, conformational changes in U2 and U6—triggered by the hydrolysis of ATP by spliceosomal proteins (not shown)—drive the formation of the spliceosome active site. Once the splicing reactions have occurred (see Figure 7–21), the spliceosome deposits a group of RNA-binding proteins, known as the exon junction complex (*red*), on the mRNA to mark the splice site as successfully completed.

The intron–exon type of gene arrangement in eukaryotes might seem wasteful, but it does provide some important benefits. First, the transcripts of many eukaryotic genes can be spliced in different ways, each of which can produce a distinct protein. Such **alternative splicing** thereby allows many different proteins to be produced from the same gene (**Figure 7–23**). About 95% of human genes are thought to undergo alternative splicing. Thus RNA splicing enables eukaryotes to increase the already enormous coding potential of their genomes. In Chapter 9, we will encounter another advantage of splicing—the production of novel proteins—when we discuss how proteins evolve.

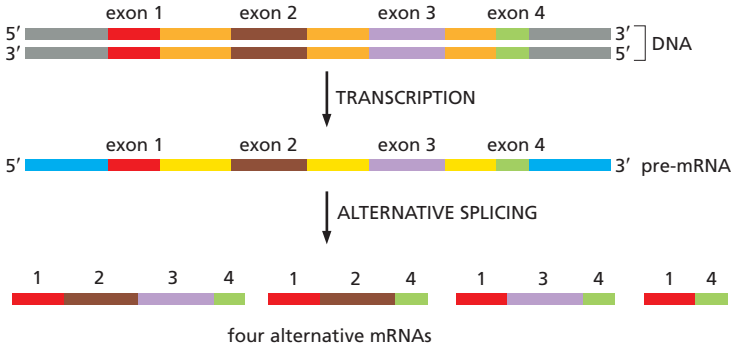


Figure 7–23 Some pre-mRNAs undergo alternative RNA splicing to produce different mRNAs and proteins from the same gene. Whereas all exons are transcribed, they can be skipped over by the spliceosome to produce alternatively spliced mRNAs, as shown. Such skipping occurs when the splicing signals at the 5′ end of one intron are paired up with the branch-point and 3′ end of a different intron. An important feature of alternative splicing is that exons can be skipped or included; however, their order—which is specified in the DNA sequence—cannot be rearranged.

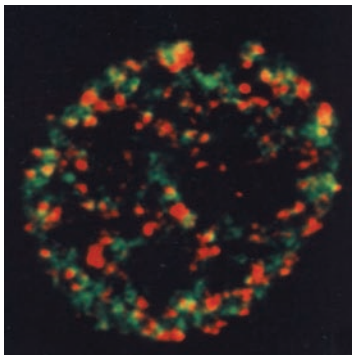


Figure 7–24 RNAs are produced by factories within the nucleus. RNAs are synthesized and processed (*red*) and DNA is replicated (*green*) in intracellular condensates that form discrete compartments within a mammalian nucleus. In this micrograph, these loose aggregates of protein and nucleic acid were visualized by detecting newly synthesized DNA and RNA. In some instances, both replication and transcription are taking place at the same site (*yellow*). (From D.G. Wansink et al., *J. Cell Sci.* 107:1449–1456, 1994. With permission from The Company of Biologists.)

RNA Synthesis and Processing Takes Place in "Factories" Within the Nucleus

RNA synthesis and processing in eukaryotes requires the coordinated action of a large number of proteins, from the RNA polymerases and accessory proteins that carry out transcription to the enzymes responsible for capping, polyadenylation, and splicing. With so many components required to produce and process every one of the RNA molecules that are being transcribed, how do all these factors manage to find one another?

We have already seen that the enzymes responsible for RNA processing ride on the phosphorylated tail of eukaryotic RNA polymerase II as it synthesizes an RNA molecule, so that the RNA transcript can be processed as it is being synthesized (see Figure 7–16). In addition to this association, RNA polymerases and RNA-processing proteins also form loose molecular aggregates—generally termed *intracellular condensates*—that act as "factories" for the production of RNA. These factories, which bring together the numerous RNA polymerases, RNA-processing components, and the genes being expressed, are large enough to be seen microscopically (**Figure 7–24**).

The aggregation of components needed to perform a specific task is not unique to RNA transcription. Proteins involved in DNA replication and repair also converge to form functional factories dedicated to their specific tasks. And genes encoding ribosomal RNAs cluster together in the nucleolus (see Figure 5–17), where their RNA products are combined with proteins to form ribosomes. These ribosomes, along with the mature mRNAs they will decode, must then be exported to the cytosol, where translation will take place.

Mature Eukaryotic mRNAs Are Exported from the Nucleus

Of all the pre-mRNA that is synthesized by a cell, only a small fraction—the sequences contained within mature mRNAs—will be useful. The remaining RNA fragments—excised introns, broken RNAs, and aberrantly spliced transcripts—are not only useless, but they could be dangerous to the cell if allowed to leave the nucleus. How, then, does the cell distinguish between the relatively rare mature mRNA molecules it needs to export to the cytosol and the overwhelming amount of debris generated by RNA processing?

The answer is that the transport of mRNA from the nucleus to the cytosol is highly selective: only correctly processed mRNAs are exported and therefore available to be translated. This selective transport is mediated by *nuclear pore complexes*, which connect the nucleoplasm with the cytosol and act as gates that control which macromolecules can enter or leave the nucleus (discussed in Chapter 15). To be "export ready," an mRNA molecule must be bound to an appropriate set of proteins, each of which recognizes different parts of a mature mRNA molecule. These proteins include poly-A-binding proteins, a cap-binding complex, and proteins that bind to mRNAs that have been appropriately spliced (**Figure 7–25**). The entire set of bound proteins, rather than any single protein, ultimately determines whether an mRNA molecule will leave the nucleus. The "waste RNAs" that remain behind in the nucleus are degraded there, and their nucleotide building blocks are reused for transcription.

mRNA Molecules Are Eventually Degraded in the Cytosol

Because a single mRNA molecule can be translated into protein many times (see Figure 7–2), the length of time that a mature mRNA molecule persists in the cell greatly influences the amount of protein it produces.

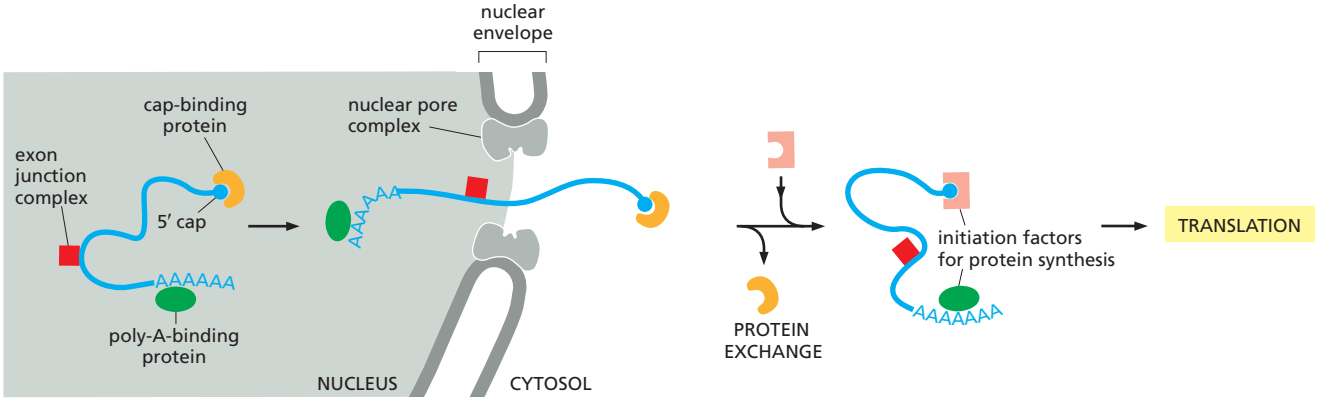


Figure 7–25 A specialized set of RNA-binding proteins signals that a completed mRNA is ready for export to the cytosol. As indicated on the *left*, the 5′ cap and poly-A tail of a mature mRNA molecule are "marked" by proteins that recognize these modifications. Successful splices are marked by exon junction complexes (see Figure 7–22). Once an mRNA is deemed "export ready," a nuclear transport receptor (discussed in Chapter 15) associates with the mRNA and guides it through the nuclear pore. In the cytosol, the mRNA can shed some of these proteins and bind new ones, which, along with poly-A-binding protein, act as initiation factors for protein synthesis, as we discuss in the next section of the chapter.

Each mRNA molecule is eventually degraded into nucleotides by ribonucleases (RNAses) present in the cytosol, but the lifespans of mRNA molecules differ considerably—depending on the nucleotide sequence of the mRNA and the type of cell. In bacteria, most mRNAs are degraded rapidly, having a typical lifespan of about 3 minutes. The mRNAs in eukaryotic cells usually persist longer: some, such as those encoding β-globin, have lifespans of more than 10 hours, whereas others stick around for less than 30 minutes.

These different lifespans are in part controlled by nucleotide sequences in the mRNA itself, most often in the portion of RNA called the 3′ *untranslated region*, which lies between the 3′ end of the coding sequence and the poly-A tail (see Figure 7–17). The lifespans of different mRNAs help the cell control how much protein will be produced. In general, proteins made in large amounts, such as β-globin, are translated from mRNAs that have long lifespans, whereas proteins made in smaller amounts, or whose levels must change rapidly in response to signals, are typically synthesized from short-lived mRNAs.

The synthesis, processing, and degradation of RNA in eukaryotes and prokaryotes is summarized and compared in **Figure 7–26**.

FROM RNA TO PROTEIN

By the end of the 1950s, biologists had demonstrated that the information encoded in DNA is copied first into RNA and then into protein. The debate then shifted to the "coding problem": How is the information in a linear sequence of nucleotides in an RNA molecule translated into the linear sequence of a chemically quite different set of subunits—the amino acids in a protein? This fascinating question intrigued scientists from many different disciplines, including physics, mathematics, and chemistry. Here was a cryptogram set up by nature that, after more than 3 billion years of evolution, could finally be solved by one of the products of evolution—the human brain! Indeed, scientists have not only cracked the code but have revealed, in atomic detail, the precise workings of the machinery by which cells read this code.

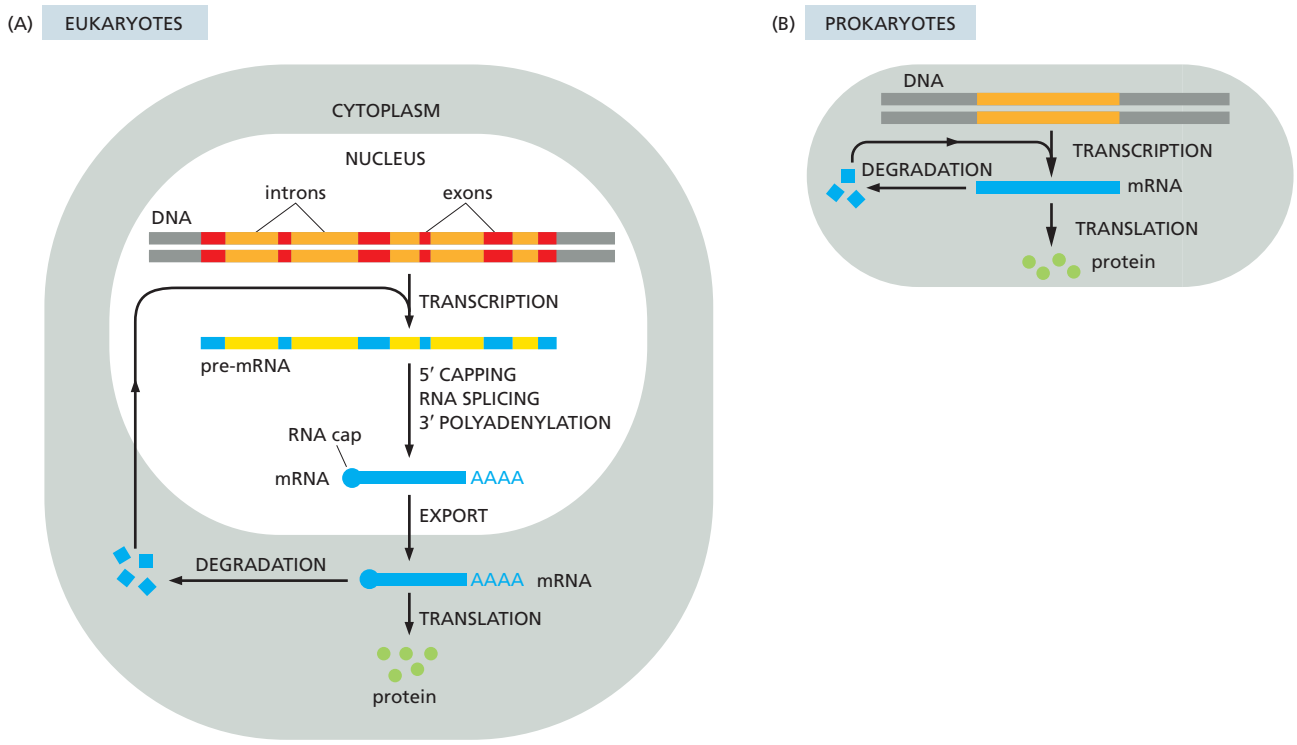


Figure 7–26 Producing mRNA molecules is more complex in eukaryotes than it is in prokaryotes. (A) In eukaryotic cells, the pre-mRNA molecule produced by transcription contains both intron and exon sequences. Its two ends are modified by capping and polyadenylation, and the introns are removed by RNA splicing. The completed mRNA is then transported from the nucleus to the cytosol, where it is translated into protein. Although these steps are depicted as occurring one after the other, in reality they occur simultaneously. For example, the RNA cap is usually added and splicing usually begins before transcription has been completed. Because of this overlap, transcripts of the entire gene (including all introns and exons) do not typically exist in the cell. Ultimately, mRNAs are degraded by RNAses in the cytosol and their nucleotide building blocks are reused for transcription. (B) In prokaryotes, the production of mRNA molecules is simpler. The 5′ end of an mRNA molecule is produced by the initiation of transcription by RNA polymerase, and the 3′ end is produced by the termination of transcription. Because prokaryotic cells lack a nucleus, transcription and translation—as well as degradation—take place in a common compartment. Translation of a prokaryotic mRNA can therefore begin before its synthesis has been completed. In both eukaryotes and prokaryotes, the amount of a protein in a cell depends on the rates of each of these steps, as well as on the rates of degradation of the mRNA and protein molecules.

An mRNA Sequence Is Decoded in Sets of Three Nucleotides

Transcription as a means of information transfer is simple to understand: DNA and RNA are chemically and structurally similar, and DNA can act as a direct template for the synthesis of RNA through complementary base-pairing. As the term transcription signifies, it is as if a message written out by hand were being converted, say, into a typewritten text. The language itself and the form of the message do not change, and the symbols used are closely related.

In contrast, the conversion of the information from RNA into protein represents a **translation** of the information into another language that uses different symbols. Because there are only 4 different nucleotides in mRNA but 20 different types of amino acids in a protein, this translation cannot be accounted for by a direct one-to-one correspondence between a nucleotide in RNA and an amino acid in protein. The set of rules by which the nucleotide sequence of a gene, through an intermediary mRNA molecule, is translated into the amino acid sequence of a protein is known as the **genetic code**.

In 1961, it was discovered that the sequence of nucleotides in an mRNA molecule is read consecutively in groups of three. And because RNA is made of 4 different nucleotides, there are $4 \times 4 \times 4 = 64$ possible combinations of three nucleotides: AAA, AUA, AUG, and so on. However, only 20 different amino acids are commonly found in proteins. Either some nucleotide triplets are never used, or the code is redundant, with some amino acids being specified by more than one triplet. The second possibility turned out to be correct, as shown by the completely deciphered genetic code shown in **Figure 7–27**. Each group of three consecutive nucleotides in RNA is called a **codon**, and each codon specifies one amino acid. The strategy by which this code was cracked is described in **How We Know**, pp. 246–247.

The same basic genetic code is used in all present-day organisms. Although a few slight differences have been found, these occur chiefly in

	Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
codons	GCA GCC GCG GCU	AGA AGG CGA CGC CGG CGU	GAC GAU	AAC AAU	UGC UGU	GAA GAG	CAA CAG	GGA GGC GGG GGU	CAC CAU	AUA AUC AUU	UUA UUG CUA CUC CUG CUU	AAA AAG	AUG	UUC UUU	CCA CCC CCG CCU	AGC AGU UCA UCC UCG UCU	ACA ACC ACG ACU	UGG	UAC UAU	GUA GUC GUG GUU	UAA UAG UGA
amino acids	A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

Figure 7–27 The nucleotide sequence of an mRNA is translated into the amino acid sequence of a protein via the genetic code. All of the three-nucleotide codons in mRNAs that specify a given amino acid are listed above that amino acid, which is given in both its three-letter and one-letter abbreviations (see Panel 2–6, pp. 76–77, for the full name of each amino acid and its structure). Like RNA molecules, codons are usually written with the 5′-terminal nucleotide to the left. Note that most amino acids are represented by more than one codon, and there are some regularities in the set of codons that specify each amino acid. For example, codons for the same amino acid tend to contain the same nucleotides at the first and second positions and vary at the third position. There are three codons that do not specify any amino acid but act as termination sites (*stop codons*), signaling the end of the protein-coding sequence in an mRNA. One codon—AUG—acts both as an initiation codon, signaling the start of a protein-coding message, and as the codon that specifies the amino acid methionine.

the mRNA of mitochondria and of some fungi and protozoa. Mitochondria have their own DNA replication, transcription, and protein-synthesis machinery, which operates independently of the corresponding machinery in the rest of the cell (discussed in Chapter 14), and they have been able to accommodate minor changes to the otherwise universal genetic code. Even in fungi and protozoa, the similarities in the code far outweigh the differences.

In principle, an mRNA sequence can be translated in any one of three different **reading frames**, depending on where the decoding process begins (**Figure 7–28**). However, only one of the three possible reading frames in an mRNA specifies the correct protein. We discuss later how a special signal at the beginning of each mRNA molecule sets the correct reading frame.

tRNA Molecules Match Amino Acids to Codons in mRNA

The codons in an mRNA molecule do not directly recognize the amino acids they specify: the set of three nucleotides does not, for example, bind directly to the amino acid. Rather, the translation of mRNA into protein depends on adaptor molecules that bind to a codon with one part of the adaptor and to an amino acid with another. These adaptors consist of a set of small RNA molecules known as **transfer RNAs** (**tRNAs**), each about 80 nucleotides in length.

We saw earlier that an RNA molecule generally folds into a three-dimensional structure by forming internal base pairs between different regions of the molecule. If the base-paired regions are sufficiently extensive, they will fold back on themselves to form a double-helical structure, like that of double-stranded DNA. Such is the case for the tRNA molecule. Four short segments of the folded tRNA are double-helical, producing a distinctive

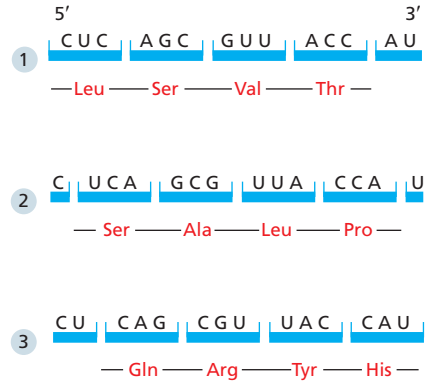


Figure 7–28 In principle, an mRNA molecule can be translated in three possible reading frames. In the process of translating a nucleotide sequence (*blue*) into an amino acid sequence (*red*), the sequence of nucleotides in an mRNA molecule is read from the 5′ to the 3′ end in sequential sets of three nucleotides. In principle, therefore, the same mRNA sequence can specify three completely different amino acid sequences, depending on the nucleotide at which translation begins—that is, on the reading frame used. In reality, however, only one of these reading frames encodes the actual message, as we discuss later.

CRACKING THE GENETIC CODE

By the beginning of the 1960s, the *central dogma* had been accepted as the pathway along which information flows from gene to protein. It was clear that genes encode proteins, that genes are made of DNA, and that mRNA serves as an intermediary, carrying the information from DNA to the ribosome, where the RNA is translated into protein.

Even the general format of the genetic code had been worked out: each of the 20 amino acids found in proteins is represented by a triplet codon in an mRNA molecule. But an even greater challenge remained: biologists, chemists, and even physicists set their sights on breaking the genetic code—attempting to figure out which amino acid each of the 64 possible nucleotide triplets designates. The most straightforward path to the solution would have been to compare the sequence of a segment of DNA or of mRNA with its corresponding polypeptide product. Techniques for sequencing nucleic acids, however, would not be developed for another decade.

So researchers decided that, to crack the genetic code, they would have to synthesize their own simple RNA molecules. If they could feed these RNA molecules to ribosomes—the machines that make proteins—and then analyze the resulting polypeptide product, they would be on their way to deciphering which triplets encode which amino acids.

Losing the cells

Before researchers could test their synthetic mRNAs, they needed to perfect a cell-free system for protein synthesis. This would allow them to translate their messages into polypeptides in a test tube. (Generally speaking, when working in the laboratory, the simpler the system, the easier it is to interpret the results.) To isolate the molecular machinery they needed for such a cell-free translation system, researchers broke open *E. coli* cells and loaded their contents into a centrifuge tube. Spinning these samples at high speed caused the membranes and other large chunks of cellular debris to be dragged to the bottom of the tube; the lighter cellular components required for protein synthesis—including mRNA, the tRNA adaptors, ribosomes, enzymes, and other small molecules—were left floating near the top of the tube (see Panel 4–3, pp. 164–165). Researchers found that simply adding radioactive amino acids to this cell "soup" would trigger the production of radiolabeled polypeptides. By centrifuging this material again, at a higher speed, the researchers could force the ribosomes, and any newly synthesized peptides attached to them, to the bottom of the tube; the labeled polypeptides could then be detected by measuring the radioactivity in the sediment remaining in the tube after the fluid layer above it had been discarded.

The trouble with this particular system was that the proteins it produced were those encoded by the cell's own mRNAs, already present in the extract. But researchers wanted to use their own synthetic messages to direct protein synthesis. This problem was solved when Marshall Nirenberg discovered that he could destroy the cells' mRNA in the extract by adding a small amount of ribonuclease—an enzyme that degrades RNA—to the mix. Now all he needed to do was prepare large quantities of synthetic mRNA, add it to the cell-free system, and see what peptides came out.

Faking the message

Producing a synthetic polynucleotide with a defined sequence was not as simple as it sounds. Again, it would be years before chemists and bioengineers developed machines that could synthesize any given string of nucleic acids quickly and cheaply. Nirenberg decided to use polynucleotide phosphorylase, an enzyme that would join ribonucleotides together in the absence of a template. The sequence of the resulting RNA would then depend entirely on which nucleotides were presented to the enzyme. A mixture of nucleotides would be sewn into a random sequence; but a single type of nucleotide would yield a homogeneous polymer containing only that one nucleotide. Thus Nirenberg, working with his collaborator Heinrich Matthaei, first produced synthetic mRNAs made entirely of uracil—poly U.

Together, the researchers fed this poly U to their cell-free translation system. They then added a single type of radioactively labeled amino acid to the mix. After testing each amino acid—one at a time, in 20 different experiments—they determined that poly U directs the synthesis of a polypeptide containing only phenylalanine (**Figure 7–29**). With this electrifying result, the first word in the genetic code had been deciphered.

Nirenberg and Matthaei then repeated the experiment with poly A and poly C and determined that AAA codes for lysine and CCC for proline. The meaning of poly G could not be ascertained by this method because, as we now know, this polynucleotide forms an aberrant structure that gums up the system.

Feeding ribosomes with synthetic RNA seemed a fruitful technique. But with the single-nucleotide possibilities exhausted, researchers had nailed down only three codons; they had 61 still to go. The other codons, however, were harder to decipher, and a new synthetic approach was needed. In the 1950s, the organic chemist Gobind Khorana had been developing methods for preparing mixed polynucleotides of defined sequence—but his techniques worked only for DNA. When he learned of Nirenberg's work with synthetic RNAs, Khorana directed his energies and skills to producing

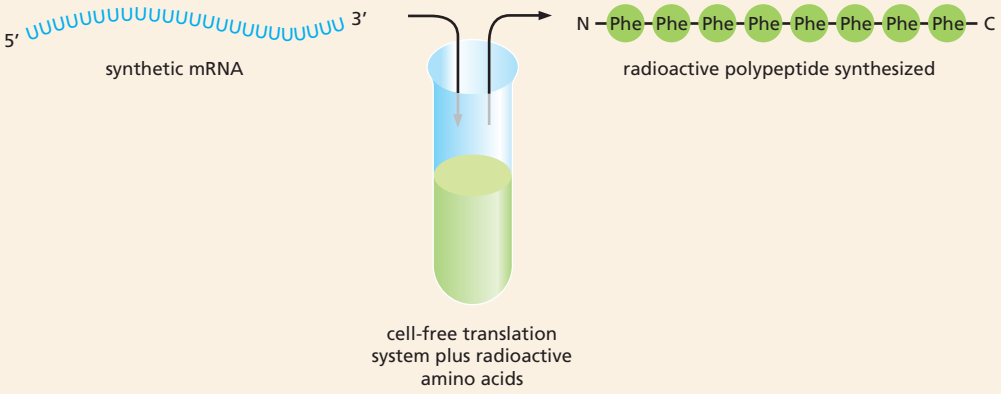


Figure 7–29 UUU codes for phenylalanine. Synthetic mRNAs are fed into a cell-free translation system containing bacterial ribosomes, tRNAs, enzymes, and other small molecules. Radioactive amino acids were added to this mix, one per experiment; when the "correct" amino acid was added, a radioactive polypeptide would be produced. In this case, poly U is shown to encode a polypeptide containing only phenylalanine.

polyribonucleotides. He found that if he started out by making DNAs of a defined sequence, he could then use RNA polymerase to produce RNAs from those. In this way, Khorana prepared a collection of different RNAs of defined repeating sequence: he generated sequences of repeating dinucleotides (such as poly UC), trinucleotides (such as poly UUC), or tetranucleotides (such as poly UAUC).

These mixed polynucleotides, however, yielded results that were much more difficult to decode than the mononucleotide messages that Nirenberg had used. Take poly UG, for example. When this repeating dinucleotide was added to the translation system, researchers discovered that it codes for a polypeptide of alternating cysteines and valines. The RNA, of course, contains two different, alternating codons: UGU and GUG. So the researchers could say that UGU and GUG code for cysteine and valine, although they could not tell which went with which. Thus these mixed messages provided useful information, but they did not definitively reveal which codons specified which amino acids (**Figure 7–30**).

Trapping the triplets

These final ambiguities in the code were resolved when Nirenberg and a young medical graduate named Phil Leder discovered that RNA fragments that were only three nucleotides in length—the size of a single codon—could bind to a ribosome and attract the appropriate amino-acid-containing tRNA molecule. These complexes—containing one ribosome, one mRNA codon, and one radiolabeled aminoacyl-tRNA—could then be captured on a piece of filter paper and the attached amino acid identified.

Their trial run with UUU—the first word—worked splendidly. Leder and Nirenberg primed the usual cell-free translation system with snippets of UUU. These trinucleotides bound to the ribosomes, and Phe-tRNAs bound to the UUU. The new system was up and running, and the researchers had confirmed that UUU codes for phenylalanine.

All that remained was for researchers to produce all 64 possible codons—a task that was quickly accomplished in both Nirenberg's and Khorana's laboratories. Because these small trinucleotides were much simpler to synthesize chemically, and the triplet-trapping tests were easier to perform and analyze than the previous decoding experiments, the researchers were able to work out the complete genetic code within the next year.

MESSAGE	PEPTIDES PRODUCED	CODON ASSIGNMENTS
poly UG	...Cys–Val–Cys–Val...	UGU, GUG — Cys, Val*
poly AG	...Arg–Glu–Arg–Glu...	AGA, GAG — Arg, Glu
poly UUC	...Phe–Phe–Phe... + ...Ser–Ser–Ser... + ...Leu–Leu–Leu...	UUC, UCU, CUU — Phe, Ser, Leu
poly UAUC	...Tyr–Leu–Ser–Ile...	UAU, CUA, UCU, AUC — Tyr, Leu, Ser, Ile

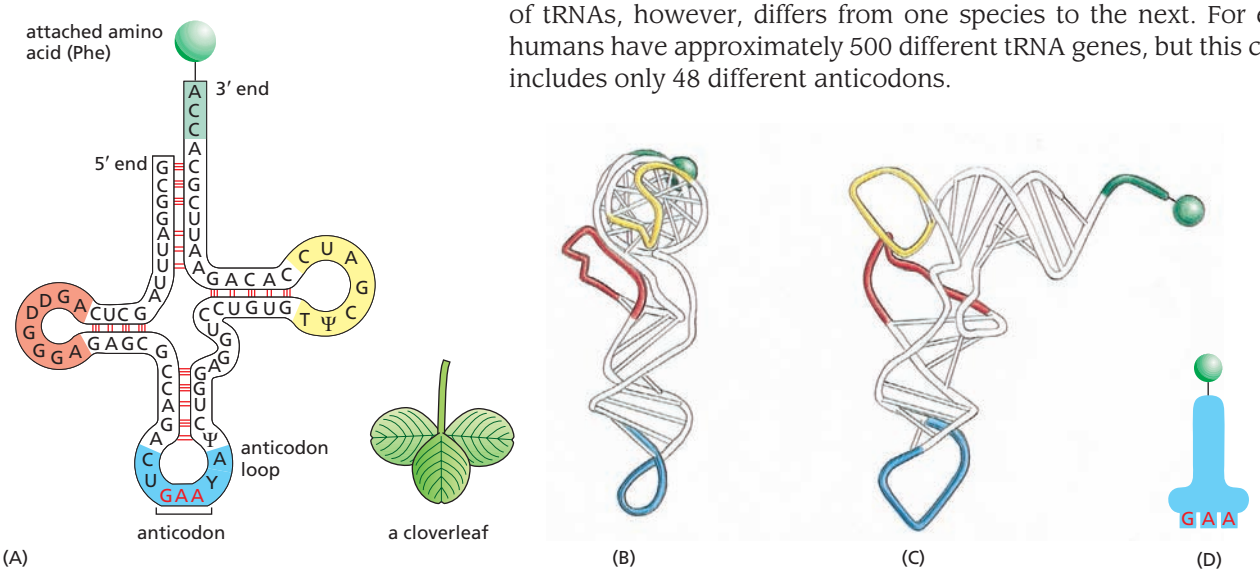
* One codon specifies Cys, the other Val, but which is which? The same ambiguity exists for the other codon assignments shown here.

Figure 7–30 Using synthetic RNAs of mixed, repeating ribonucleotide sequences, scientists further narrowed the coding possibilities. Because these mixed messages produced mixed polypeptides, they did not permit the unambiguous assignment of a single codon to a specific amino acid. For example, the results of the poly-UG experiment cannot distinguish whether UGU or GUG encodes cysteine. As indicated, the same type of ambiguity confounded the interpretation of all the experiments using di-, tri-, and tetranucleotides.

structure that looks like a cloverleaf when drawn schematically (**Figure 7–31A**). As shown in the figure, for example, a 5′-GCUC-3′ sequence in one part of a polynucleotide chain can base-pair with a 5′-GAGC-3′ sequence in another region of the same molecule. The cloverleaf undergoes further folding to form a compact, L-shaped structure that is held together by additional hydrogen bonds between different regions of the molecule (**Figure 7–31B–D**).

Two regions of unpaired nucleotides situated at either end of the L-shaped tRNA molecule are crucial to the function of tRNAs in protein synthesis. One of these regions forms the **anticodon**, a set of three consecutive nucleotides that bind, through base-pairing, to the complementary codon in an mRNA molecule (**Figure 7–31E**). The other is a short, single-stranded region at the 3′ end of the molecule; this is the site where the amino acid that matches the codon is covalently attached to the tRNA.

We saw in the previous section that the genetic code is redundant; that is, several different codons can specify a single amino acid (see Figure 7–27). This redundancy implies either that there is more than one tRNA for many of the amino acids or that some tRNA molecules can base-pair with more than one codon. In fact, both situations occur. Some amino acids have more than one tRNA, and some tRNAs require accurate base-pairing only at the first two positions of the codon and can tolerate a mismatch (or *wobble*) at the third position. This wobble base-pairing explains why so many of the alternative codons for an amino acid differ only in their third nucleotide (see Figure 7–27). Wobble base-pairings make it possible to fit the 20 amino acids to their 61 codons with as few as 31 kinds of tRNA molecules. The exact number of different kinds of tRNAs, however, differs from one species to the next. For example, humans have approximately 500 different tRNA genes, but this collection includes only 48 different anticodons.



5′ GCGGAUUUAGCUCAGDDGGGAGAGCGCCAGACUGAAYAΨCUGGAGGUCCUGUGTΨCGAUCCACAGAAUUCGCACCA 3′

(E) anticodon

Figure 7–31 tRNA molecules are molecular adaptors, linking amino acids to codons. In this series of diagrams, the same tRNA molecule—in this case, a tRNA specific for the amino acid phenylalanine (Phe)—is depicted in various ways. (A) The conventional "cloverleaf" structure shows the complementary base-pairing (*red* lines) that creates the double-helical regions of the molecule. The anticodon loop (*blue*) contains the sequence of three nucleotides (*red* letters) that base-pairs with the Phe codon in mRNA. The amino acid matching the anticodon is attached at the 3′ end of the tRNA. tRNAs contain some unusual bases, which are produced by chemical modification after the tRNA has been synthesized. The bases denoted ψ (for pseudouridine) and D (for dihydrouridine) are derived from uracil. (B and C) Views of the actual L-shaped molecule, based on x-ray diffraction analysis. These two images are rotated 90° with respect to each other. (D) The schematic representation of tRNA that will be used in subsequent figures emphasizes the anticodon. (E) The linear nucleotide sequence of the tRNA molecule, color-coded to match (A), (B), and (C).

Specific Enzymes Couple tRNAs to the Correct Amino Acid

For a tRNA molecule to carry out its role as an adaptor, it must be linked—or charged—with the correct amino acid. How does each tRNA molecule recognize the one amino acid in 20 that is its proper partner? Recognition and attachment of the correct amino acid depend on enzymes called **aminoacyl-tRNA synthetases**, which covalently couple each amino acid to the appropriate set of tRNA molecules. In most organisms, there is a different synthetase enzyme for each amino acid. That means that there are 20 synthetases in all: one attaches glycine to all tRNAs that recognize codons for glycine, another attaches phenylalanine to all tRNAs that recognize codons for phenylalanine, and so on. Each synthetase enzyme recognizes its designated amino acid, as well as nucleotides in the anticodon loop and in the amino-acid-accepting arm that are specific to the correct tRNA (Figure 7–32 and Movie 7.6). The synthetases are thus equal in importance to the tRNAs in the decoding process, because it is the combined action of the synthetases and tRNAs that allows each codon in the mRNA molecule to be correctly matched to its amino acid (Figure 7–33).

The synthetase-catalyzed reaction that attaches the amino acid to the 3′ end of the tRNA is one of many reactions in cells that is coupled to the energy-releasing hydrolysis of ATP (see Figure 3–32). The reaction produces a high-energy bond between the charged tRNA and the amino acid. The energy of this bond is later used to link the amino acid covalently to the growing polypeptide chain.

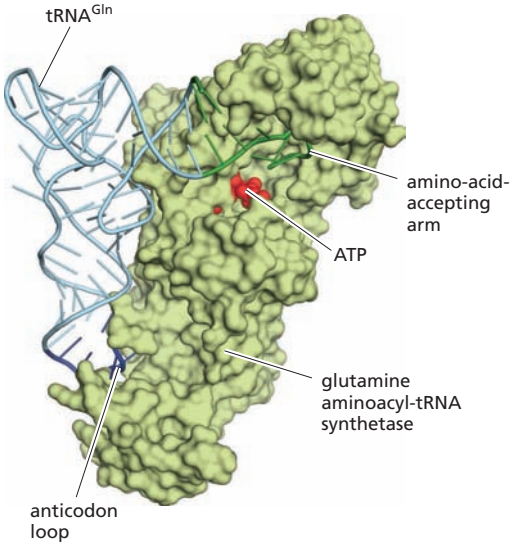


Figure 7–32 Each aminoacyl-tRNA synthetase makes multiple contacts with its tRNA molecule. For this tRNA, which is specific for the amino acid glutamine, nucleotides in both the anticodon loop (*dark blue*) and the amino-acid-accepting arm (*green*) are recognized by the synthetase (*yellow-green*). The ATP molecule that will be hydrolyzed to provide the energy needed to attach the amino acid to the tRNA is shown in *red*.

The mRNA Message Is Decoded on Ribosomes

The recognition of a codon by the anticodon on a tRNA molecule depends on the same type of complementary base-pairing used in DNA replication and transcription. However, accurate and rapid translation of mRNA into protein requires a molecular machine that can latch onto an mRNA, capture and position the correct tRNA molecules, and then covalently link the amino acids that they carry to form a polypeptide chain. In both

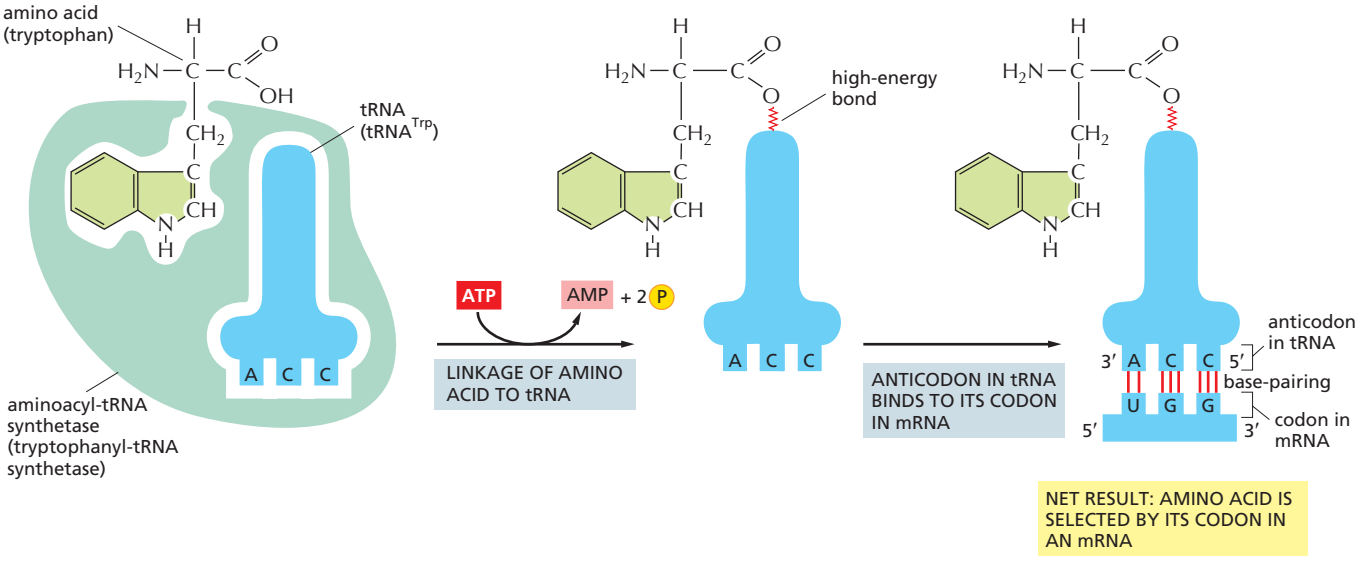


Figure 7–33 The genetic code is translated by aminoacyl-tRNA synthetases and tRNAs. Each synthetase couples a particular amino acid to its corresponding tRNAs, a process called charging. The anticodon on the charged tRNA molecule then forms base pairs with the appropriate codon on the mRNA. An error in either the charging step or the binding of the charged tRNA to its codon will cause the wrong amino acid to be incorporated into a polypeptide chain. In the sequence of events shown, the amino acid tryptophan (Trp) is specified by the codon UGG on the mRNA.

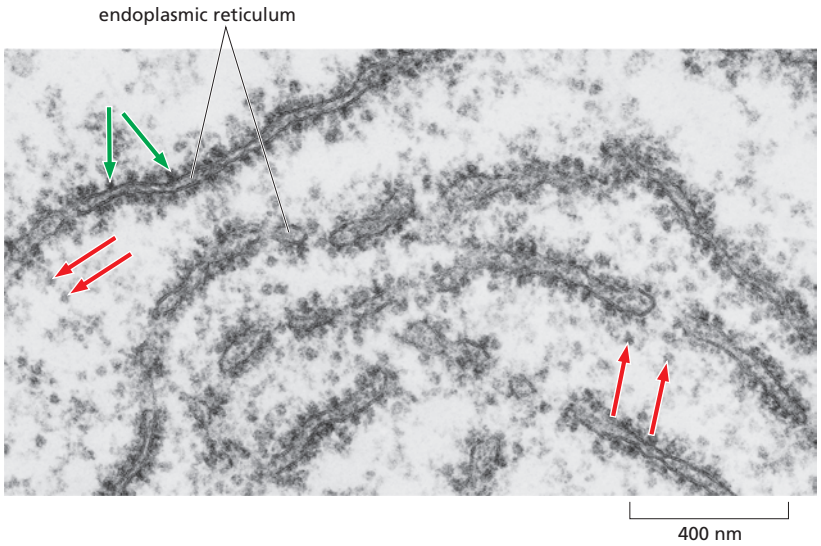


Figure 7–34 Ribosomes are located in the cytoplasm of eukaryotic cells. This electron micrograph shows a thin section of a small region of cytoplasm. The ribosomes appear as small *gray* blobs. Some are free in the cytoplasm (*red* arrows); others are attached to membranes of the endoplasmic reticulum (*green* arrows). (Courtesy of George Palade.)

QUESTION 7–4

In a clever experiment performed in 1962, a cysteine already attached to its tRNA was chemically converted to an alanine. These "hybrid" tRNA molecules were then added to a cell-free translation system from which the normal cysteine-tRNAs had been removed. When the resulting protein was analyzed, it was found that alanine had been inserted at every point in the polypeptide chain where cysteine was supposed to be. Discuss what this experiment tells you about the role of aminoacyl-tRNA synthetases and ribosomes during the normal translation of the genetic code.

prokaryotes and eukaryotes, the machine that gets the job done is the **ribosome**—a large complex made from dozens of small proteins (the *ribosomal proteins*) and several RNA molecules called **ribosomal RNAs** (**rRNAs**). A typical eukaryotic cell contains millions of ribosomes in its cytosol (**Figure 7–34**).

Eukaryotic and prokaryotic ribosomes are very similar in structure and function. Both are composed of one large subunit and one small subunit, which fit together to form a complete ribosome with a mass of several million daltons (**Figure 7–35**); for comparison, an average-sized protein

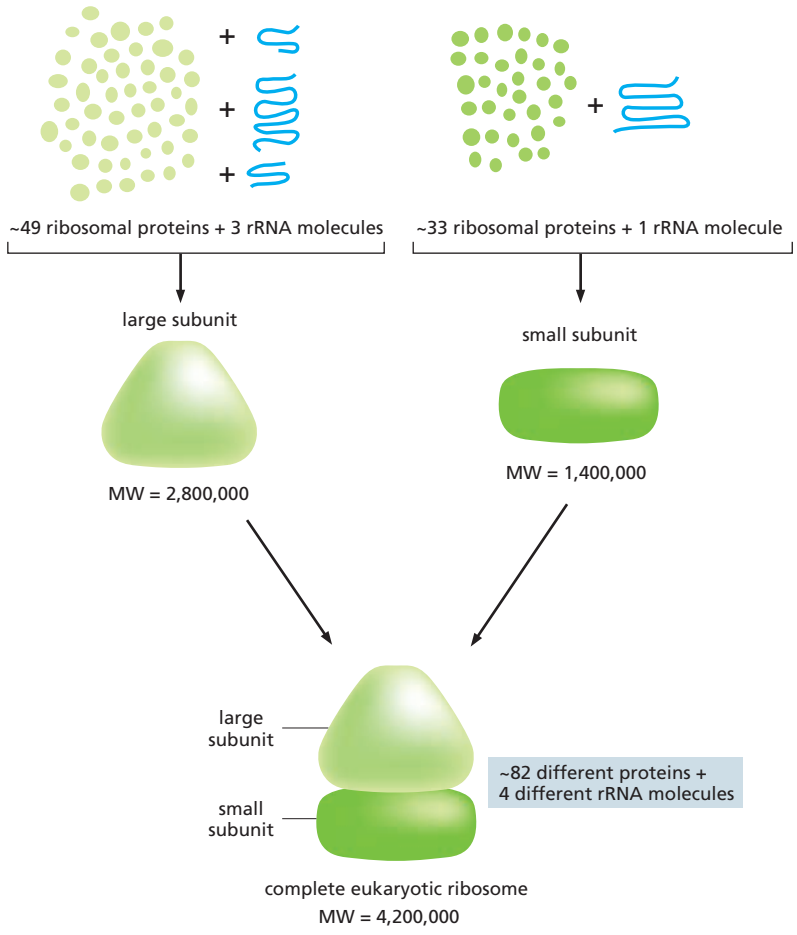


Figure 7–35 The eukaryotic ribosome is a large complex of four rRNAs and more than 80 small proteins. Prokaryotic ribosomes are very similar: both are formed from a large and small subunit, which only come together after the small subunit has bound an mRNA. The RNAs account for most of the mass of the ribosome and give it its overall shape and structure.

has a mass of 30,000 daltons. The small ribosomal subunit matches the tRNAs to the codons of the mRNA, while the large subunit catalyzes the formation of the peptide bonds that covalently link the amino acids together into a polypeptide chain. These two subunits come together on an mRNA molecule near its 5′ end to start the synthesis of a protein. The mRNA is then pulled through the ribosome like a long piece of tape. As the mRNA inches forward in a 5′-to-3′ direction, the ribosome translates its nucleotide sequence into an amino acid sequence, one codon at a time, using the tRNAs as adaptors. Each amino acid is thereby added in the correct sequence to the end of the growing polypeptide chain (**Movie 7.7**). When synthesis of the protein is finished, the two subunits of the ribosome separate. Ribosomes operate with remarkable efficiency: a eukaryotic ribosome adds about 2 amino acids to a polypeptide chain each second; a bacterial ribosome operates even faster, adding about 20 amino acids per second.

How does the ribosome choreograph all the movements required for translation? In addition to a binding site for an mRNA molecule, each ribosome contains three binding sites for tRNA molecules, called the A site, the P site, and the E site (**Figure 7–36**). To add an amino acid to a growing peptide chain, a charged tRNA enters the A site by base-pairing with the complementary codon on the mRNA molecule. Its amino acid is then linked to the growing peptide chain, which is held in place by the tRNA in the neighboring P site. Next, the large ribosomal subunit shifts forward, moving the spent tRNA to the E site before ejecting it (**Figure 7–37**). This cycle of reactions is repeated each time an amino acid is added to the polypeptide chain, with the new protein growing from its amino to its carboxyl end until a stop codon in the mRNA is encountered and the protein is released.

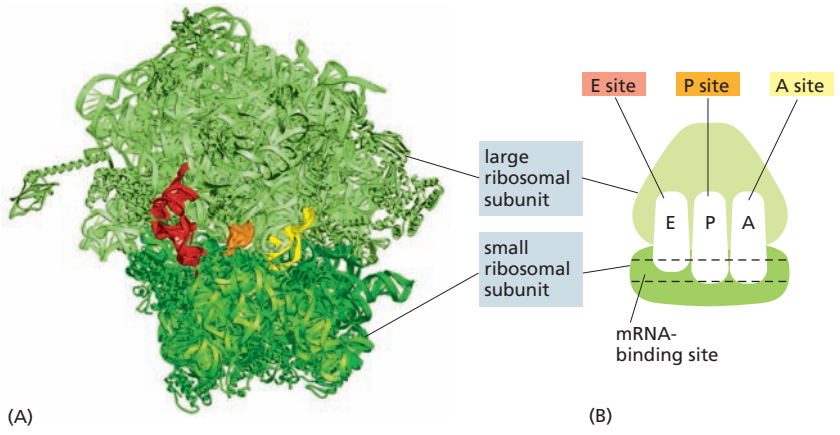


Figure 7–36 Each ribosome has a binding site for an mRNA molecule and three binding sites for tRNAs. The tRNA sites are designated the A, P, and E sites (short for aminoacyl-tRNA, peptidyl-tRNA, and exit, respectively). (A) Three-dimensional structure of a bacterial ribosome, as determined by x-ray crystallography, with the small subunit in *dark green* and the large subunit in *light green*. Both the rRNAs and the ribosomal proteins are shown in *green*. tRNAs are shown bound in the E site (*red*), the P site (*orange*), and the A site (*yellow*). Although all three of the tRNA sites shown here are filled, during protein synthesis only two of these sites are occupied by a tRNA at any one time (see Figure 7–37). (B) Highly schematized representation of a ribosome, in the same orientation as (A), which is used in subsequent figures. Note that both the large and small subunits are involved in forming the A, P, and E sites, while only the small subunit contains the binding site for an mRNA. (A, adapted from M.M. Yusupov et al., *Science* 292:883–896, 2001. Courtesy of Albion A. Bausom and Harry Noller.)

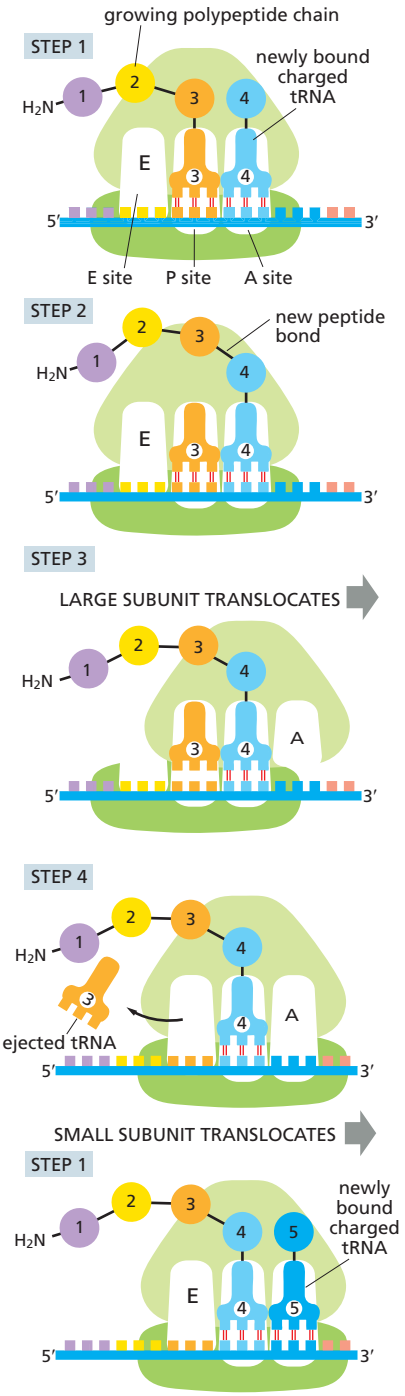


Figure 7–37 Translation takes place in a four-step cycle, which is repeated over and over during the synthesis of a protein. In *step 1*, a charged tRNA carrying the next amino acid to be added to the polypeptide chain binds to the vacant A site on the ribosome by forming base pairs with the mRNA codon that is exposed there. Only a matching tRNA molecule can base-pair with this codon, which determines the specific amino acid added. The A and P sites are sufficiently close together that their two tRNA molecules are forced to form base pairs with codons that are contiguous, with no stray bases in-between. This positioning of the tRNAs ensures that the correct reading frame will be preserved throughout the synthesis of the protein. In *step 2*, the carboxyl end of the polypeptide chain (amino acid 3 in step 1) is uncoupled from the tRNA at the P site and joined by a peptide bond to the free amino group of the amino acid linked to the tRNA at the A site. This reaction is carried out by a catalytic site in the large subunit. In *step 3*, a shift of the large subunit relative to the small subunit moves the two bound tRNAs into the E and P sites of the large subunit. In *step 4*, the small subunit moves exactly three nucleotides along the mRNA molecule, bringing it back to its original position relative to the large subunit. This movement ejects the spent tRNA and resets the ribosome with an empty A site so that the next charged tRNA molecule can bind (**Movie 7.8**).

As indicated, the mRNA is translated in the 5′-to-3′ direction, and the N-terminal end of a protein is made first, with each cycle adding one amino acid to the C-terminus of the polypeptide chain. To watch the translation cycle in atomic detail, see **Movie 7.9**.

The Ribosome Is a Ribozyme

The ribosome is one of the largest and most complex structures in the cell, composed of two-thirds RNA and one-third protein by weight. The determination of the entire three-dimensional structure of its large and small subunits in 2000 was a major triumph of modern biology. The structure confirmed earlier evidence that the rRNAs—not the proteins—are responsible for the ribosome's overall structure and its ability to choreograph and catalyze protein synthesis.

The rRNAs are folded into highly compact, precise three-dimensional structures that form the core of the ribosome (**Figure 7–38**). In contrast to the central positioning of the rRNAs, the ribosomal proteins are generally located on the surface, where they fill the gaps and crevices of the

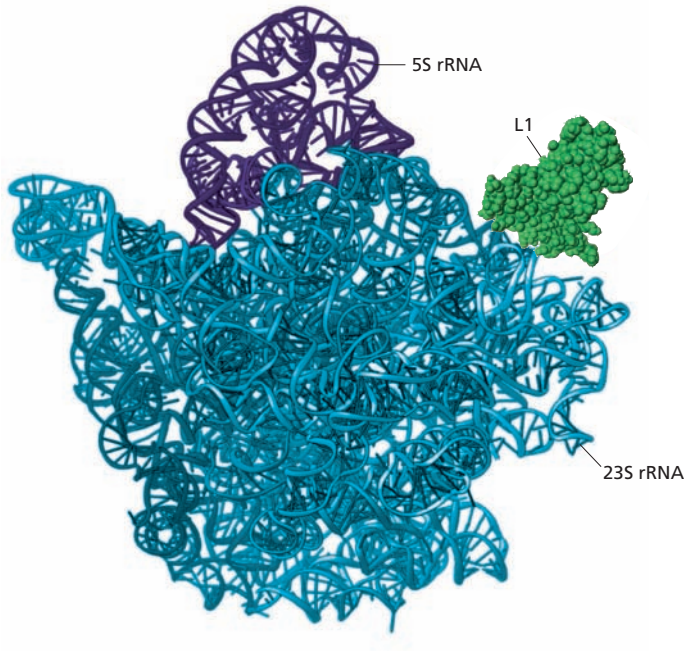


Figure 7–38 Ribosomal RNAs give the ribosome its overall shape. Shown here are the detailed structures of the two rRNAs that form the core of the large subunit of a bacterial ribosome—the 23S rRNA (*blue*) and the 5S rRNA (*purple*). One of the protein subunits of the ribosome (L1) is included as a reference point, as this protein forms a characteristic protrusion on the ribosome surface. Ribosomal RNAs are commonly designated by their "S values," which refer to their rate of sedimentation in an ultracentrifuge. The larger the S value, the larger the size of the molecule. (Adapted from N. Ban et al., *Science* 289: 905–920, 2000.)

folded RNA. The main role of the ribosomal proteins seems to be to help fold and stabilize the RNA core, while permitting the changes in rRNA conformation that are necessary for this RNA to catalyze efficient protein synthesis.

Not only are the three tRNA-binding sites (the A, P, and E sites) on the ribosome formed primarily by the rRNAs, but the catalytic site for peptide bond formation is formed by the 23S rRNA of the large subunit; the nearest ribosomal protein is located too far away to make contact with the incoming amino acid or with the growing polypeptide chain. The catalytic site in this RNA—a peptidyl transferase—is similar in many respects to that found in some protein enzymes: it is a highly structured pocket that precisely orients the two reactants—the elongating polypeptide and the amino acid carried by the incoming tRNA—thereby greatly increasing the likelihood of a productive reaction.

RNA molecules that possess catalytic activity are called **ribozymes**. In the final section of this chapter, we will consider other ribozymes and discuss what the existence of RNA-based catalysis might mean for the early evolution of life on Earth. Here, we need only note that there is good reason to suspect that RNA rather than protein molecules served as the first catalysts for living cells. If so, the ribosome, with its catalytic RNA core, could be viewed as a relic of an earlier time in life's history, when cells were run almost entirely by RNAs.

Specific Codons in an mRNA Signal the Ribosome Where to Start and to Stop Protein Synthesis

In a test tube, ribosomes can be forced to translate any RNA molecule (see How We Know, pp. 246–247). In a cell, however, a specific start signal is required to initiate translation. The site at which protein synthesis begins on an mRNA is crucial, because it sets the reading frame for the entire message. An error of one nucleotide either way at this stage will cause every subsequent codon in the mRNA to be misread, resulting in a nonfunctional protein with a garbled sequence of amino acids (see Figure 7–28). Furthermore, the rate of initiation has a major impact on the overall rate at which the protein is synthesized from the mRNA.

The translation of an mRNA begins with the codon AUG, for which a special charged tRNA is required. This **initiator tRNA** always carries the amino acid methionine (or a modified form of methionine, formylmethionine, in bacteria). Thus newly made proteins all have methionine as the first amino acid at their N-terminal end, the end of a protein that is synthesized first. This methionine is usually removed later by a specific protease.

In eukaryotes, an initiator tRNA, charged with methionine, is first loaded into the P site of the small ribosomal subunit, along with additional proteins called **translation initiation factors** (Figure 7–39). The initiator tRNA is distinct from the tRNA that normally carries methionine. Of all the tRNAs in the cell, only a charged initiator tRNA molecule is capable of binding tightly to the P site in the absence of the large ribosomal subunit.

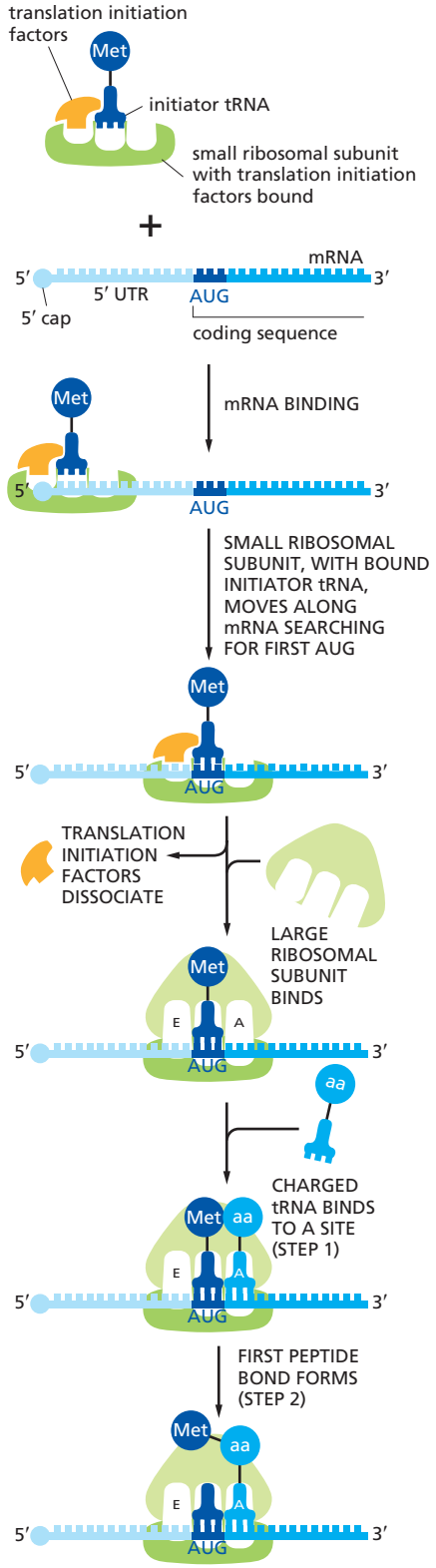
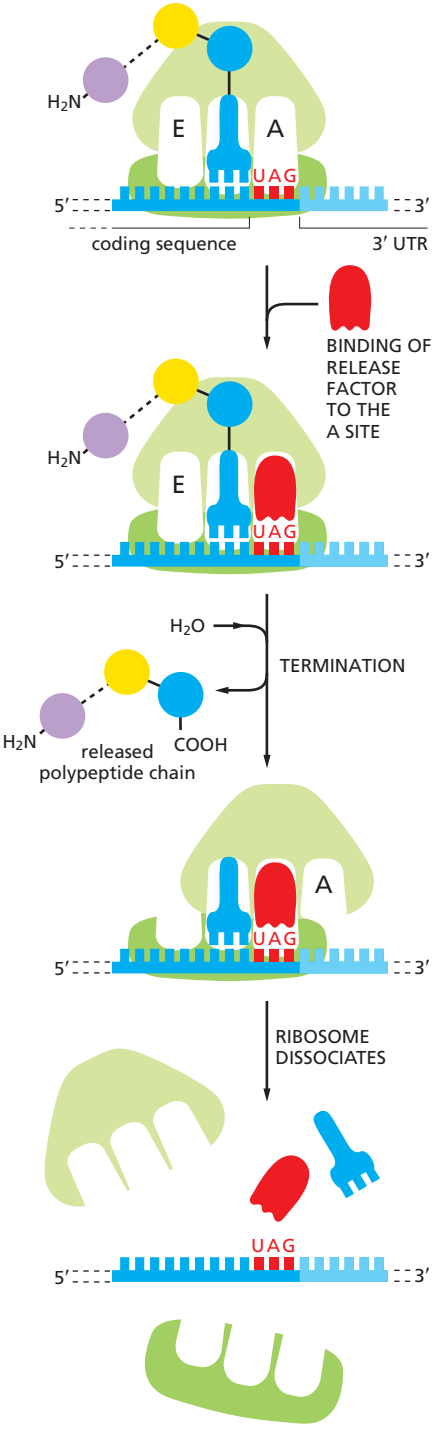


Figure 7–39 Initiation of protein synthesis in eukaryotes requires translation initiation factors and a special initiator tRNA. Although not shown here, efficient translation initiation also requires additional proteins that are bound at the 5′ cap and poly-A tail of the mRNA (see Figure 7–25). In this way, the translation apparatus can ascertain that both ends of the mRNA are intact before initiating translation. Following initiation, the protein is elongated by the reactions outlined in Figure 7–37.

QUESTION 7–5

A sequence of nucleotides in a DNA strand—5′-TTAACGGCTTTTTTC-3′—was used as a template to synthesize an mRNA that was then translated into protein. Predict the C-terminal amino acid and the N-terminal amino acid of the resulting polypeptide. Assume that the mRNA is translated without the need for a start codon.



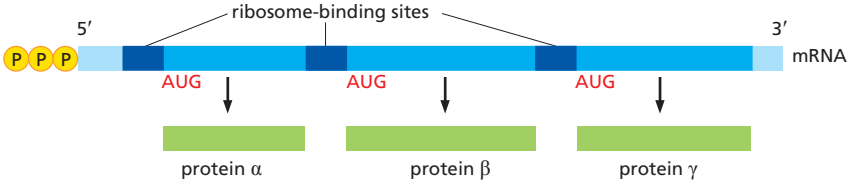


Figure 7–40 A single prokaryotic mRNA molecule can encode several different proteins. In prokaryotes, genes directing the different steps in a process are often organized into clusters (operons) that are transcribed together into a single mRNA. A prokaryotic mRNA does not have the same sort of 5′ cap as a eukaryotic mRNA, but instead has a triphosphate at its 5′ end. Prokaryotic ribosomes initiate translation at ribosome-binding sites (*dark blue*), which can be located in the interior of an mRNA molecule. This feature enables prokaryotes to simultaneously synthesize different proteins from a single mRNA molecule, with each protein made by a different ribosome.

Next, the small ribosomal subunit loaded with the initiator tRNA binds to the 5′ end of an mRNA molecule, which is marked by the 5′ cap that is present on all eukaryotic mRNAs (see Figure 7–17). The small ribosomal subunit then scans the mRNA, in the 5′-to-3′ direction, until it encounters the first AUG. When this AUG is recognized by the initiator tRNA, several of the initiation factors dissociate from the small ribosomal subunit to make way for the large ribosomal subunit to bind and complete ribosomal assembly. Because the initiator tRNA is bound to the P site, protein synthesis is ready to begin with the addition of the next charged tRNA to the A site (see Figure 7–37).

The mechanism for selecting a start codon is different in bacteria. Bacterial mRNAs have no 5′ caps to tell the ribosome where to begin searching for the start of translation. Instead, each mRNA molecule contains a specific ribosome-binding sequence, approximately six nucleotides long, located a few nucleotides upstream of the AUG at which translation is to begin. Unlike a eukaryotic ribosome, a prokaryotic ribosome can readily bind directly to a start codon that lies in the interior of an mRNA, as long as a ribosome-binding site precedes it by several nucleotides. Such ribosome-binding sequences are necessary in bacteria, as prokaryotic mRNAs are often *polycistronic*—that is, they encode several different proteins on the same mRNA molecule; these transcripts contain a separate ribosome-binding site for each protein-coding sequence (**Figure 7–40**). In contrast, a eukaryotic mRNA usually carries the information for a single protein, and so it can rely on the 5′ cap—and the proteins that recognize it—to position the ribosome for its AUG search.

The end of translation in both prokaryotes and eukaryotes is signaled by the presence of one of several codons, called *stop codons*, in the mRNA (see Figure 7–27). The stop codons—UAA, UAG, and UGA—are not recognized by a tRNA and do not specify an amino acid, but instead signal to the ribosome to stop translation. Proteins known as *release factors* bind to any stop codon that reaches the A site on the ribosome; this binding alters the activity of the peptidyl transferase in the ribosome, causing it to catalyze the addition of a water molecule instead of an amino acid to the peptidyl-tRNA (**Figure 7–41**). This reaction frees the carboxyl end of the polypeptide chain from its attachment to a tRNA molecule; because this is the only attachment that holds the growing polypeptide to the

Figure 7–41 Translation halts at a stop codon. In the final phase of protein synthesis, the binding of release factor to an A site bearing a stop codon terminates translation of an mRNA molecule. The completed polypeptide is released, and the ribosome dissociates into its two separate subunits.

ribosome, the completed protein chain is immediately released. At this point, the ribosome also releases the mRNA and dissociates into its two separate subunits, which can then assemble on another mRNA molecule to begin a new round of protein synthesis.

Proteins Are Produced on Polyribosomes

The synthesis of most protein molecules takes between 20 seconds and several minutes. But even during this short period, multiple ribosomes usually bind to each mRNA molecule being translated. If an mRNA is being translated efficiently, a new ribosome will hop onto its 5′ end almost as soon as the preceding ribosome has translated enough of the nucleotide sequence to move out of the way. The mRNA molecules being translated are therefore usually found in the form of *polyribosomes*, also known as *polysomes*. These large cytosolic assemblies are made up of many ribosomes spaced as close as 80 nucleotides apart along a single mRNA molecule (**Figure 7–42**). With multiple ribosomes working simultaneously on a single mRNA, many more protein molecules can be made in a given time than would be possible if each polypeptide had to be completed before the next could be started.

Polysomes operate in both bacteria and eukaryotes, but bacteria can speed up the rate of protein synthesis even further. Because bacterial mRNA does not need to be processed and is also physically accessible to ribosomes while it is being synthesized, ribosomes will typically attach to the free end of a bacterial mRNA molecule and start translating it even before the transcription of that RNA is complete; these ribosomes follow closely behind the RNA polymerase as it moves along DNA.

Inhibitors of Prokaryotic Protein Synthesis Are Used as Antibiotics

The ability to translate mRNAs accurately into proteins is a fundamental feature of all life on Earth. Although the ribosome and other molecules that carry out this complex task are very similar among organisms, we have seen that there are some subtle differences in the way that bacteria and eukaryotes synthesize RNA and proteins. Although they represent a quirk of evolution, these differences form the basis of one of the most important advances in modern medicine.

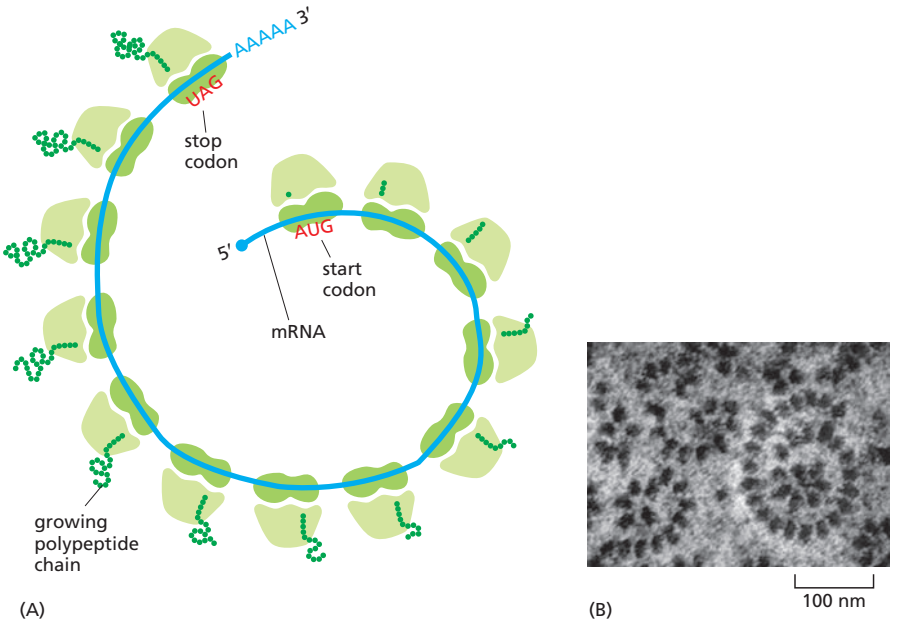


Figure 7–42 Proteins are synthesized on polyribosomes. (A) Schematic drawing showing how a series of ribosomes can simultaneously translate the same mRNA molecule (**Movie 7.10**). (B) Electron micrograph of a polyribosome in the cytosol of a eukaryotic cell. (B, courtesy of John Heuser.)

TABLE 7–3 ANTIBIOTICS THAT INHIBIT BACTERIAL PROTEIN OR RNA SYNTHESIS

Antibiotic	Specific Effect
Tetracycline	blocks binding of aminoacyl-tRNA to A site of ribosome (step 1 in Figure 7–37)
Streptomycin	prevents the transition from initiation complex to chain elongation (see Figure 7–39); also causes miscoding
Chloramphenicol	blocks the peptidyl transferase reaction on ribosomes (step 2 in Figure 7–37)
Cycloheximide	blocks the translocation step in translation (step 3 in Figure 7–37)
Rifamycin	blocks initiation of transcription by binding to and inhibiting RNA polymerase

Many of our most effective antibiotics are compounds that act by inhibiting bacterial, but not eukaryotic, gene expression. Some of these drugs exploit the small structural and functional differences between bacterial and eukaryotic ribosomes, so that they interfere preferentially with bacterial protein synthesis. These compounds can thus be taken in doses high enough to kill bacteria without being toxic to humans. Because different antibiotics bind to different regions of the bacterial ribosome, these drugs often inhibit different steps in protein synthesis. A few of the antibiotics that inhibit bacterial gene expression are listed in **Table 7–3**.

Many common antibiotics were first isolated from fungi. Fungi and bacteria often occupy the same ecological niches, and to gain a competitive edge, fungi have evolved, over time, potent toxins that kill bacteria but are harmless to themselves. Because fungi and humans are both eukaryotes, and are thus much more closely related to each other than either is to bacteria (see Figure 1–29), we have been able to borrow these weapons to combat our own bacterial foes. At the same time, bacteria have unfortunately evolved a resistance to many of these drugs, as we discuss in Chapter 9. Thus it remains a continual challenge for us to remain one step ahead of our microbial foes.

Controlled Protein Breakdown Helps Regulate the Amount of Each Protein in a Cell

After a protein is released from the ribosome, a cell can control its activity and longevity in various ways. The number of copies of a protein in a cell depends, like the number of organisms in a population, not only on how quickly new individuals arise but also on how long they survive. Proteins vary enormously in their lifespan. Structural proteins that become part of a relatively stable tissue such as bone or muscle may last for months or even years, whereas other proteins, such as metabolic enzymes and those that regulate cell growth and division (discussed in Chapter 18), last only for days, hours, or even seconds. But what determines the lifespan of a protein—and how does a protein "die"?

Cells produce many proteins whose job it is to break other proteins down into their constituent amino acids (a process termed *proteolysis*). These enzymes, which degrade proteins, first to short peptides and finally to individual amino acids, are known collectively as **proteases**. Proteases act by cutting (hydrolyzing) the peptide bonds between amino acids (see Panel 2–6, pp. 76–77). One function of proteolytic pathways is to rapidly

degrade those proteins whose lifetime must be kept short. Another is to recognize and remove proteins that are damaged or misfolded. Eliminating improperly folded proteins is critical for an organism, as misfolded proteins tend to aggregate, and protein aggregates can damage cells and even trigger cell death. Eventually, all proteins—even long-lived ones—accumulate damage and are degraded by proteolysis. The amino acids produced by this proteolysis can then be re-used by the cell to make new proteins.

In eukaryotic cells, proteins are broken down by large protein machines called **proteasomes**, present in both the cytosol and the nucleus. A proteasome contains a central cylinder formed from proteases whose active sites face into an inner chamber. Each end of the cylinder is plugged by a large protein complex formed from at least 10 types of protein subunits (**Figure 7–43**). These stoppers bind the proteins destined for degradation and then—using ATP hydrolysis to fuel this activity—unfold the doomed proteins and thread them into the inner chamber of the cylinder. Once the proteins are inside, proteases chop them into short peptides, which are then jettisoned from either end of the proteasome. Housing proteases inside these molecular destruction chambers makes sense, as it prevents the enzymes from running rampant in the cell.

How do proteasomes select which proteins in the cell should be degraded? In eukaryotes, proteasomes act primarily on proteins that have been marked for destruction by the covalent attachment of a small protein called *ubiquitin*. Specialized enzymes tag those proteins that are destined for rapid degradation with a short chain of ubiquitin molecules; these ubiquitylated proteins are then recognized, unfolded, and fed into proteasomes by proteins within the stopper (**Figure 7–44**).

Proteins that are meant to be short-lived often contain a short amino acid sequence that identifies the protein as one to be ubiquitylated and degraded in proteasomes. Damaged or misfolded proteins, as well as proteins containing oxidized or otherwise abnormal amino acids, are also recognized and degraded by this ubiquitin-dependent proteolytic system. The enzymes that add a polyubiquitin chain to such proteins recognize signals that become exposed on these proteins as a result of the misfolding or chemical damage—for example, amino acid sequences or conformational motifs that are typically buried and inaccessible in a "healthy" protein.

There Are Many Steps Between DNA and Protein

We have seen that many steps are required to produce a functional protein from the information contained in a gene. In a eukaryotic cell, mRNAs must be synthesized, processed, and exported to the cytosol

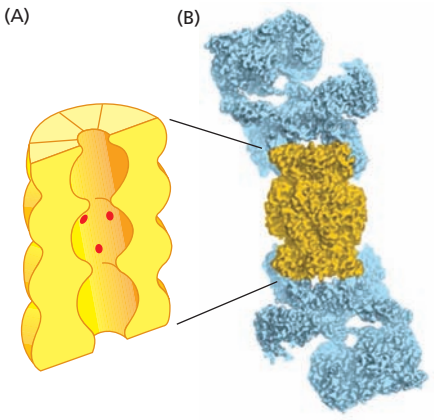


Figure 7–43 Proteins are degraded by the proteasome. The structures depicted here were determined by x-ray crystallography. (A) This drawing shows a cut-away view of the central cylinder of the proteasome, with the active sites of the proteases indicated by *red* dots. (B) The structure of the entire proteasome, in which access to the central cylinder (*yellow*) is regulated by a stopper (*blue*) at each end. (B, from P.C.A. da Fonseca et al., *Mol. Cell* 46:54–66, 2012. With permission from Elsevier.)

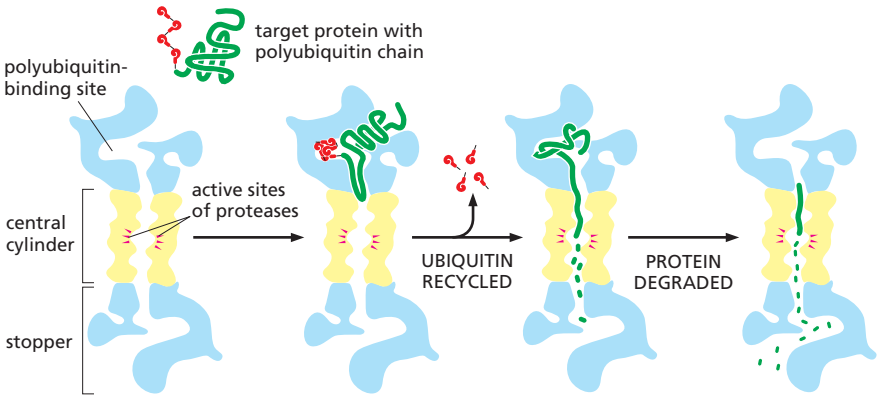


Figure 7–44 Proteins marked by a polyubiquitin chain are degraded by the proteasome. Proteins in the stopper of a proteasome (*blue*) recognize proteins marked by a specific type of polyubiquitin chain (*red*). The stopper unfolds the target protein and threads it into the proteasome's central cylinder (*yellow*), which is lined with proteases that chop the protein to pieces.

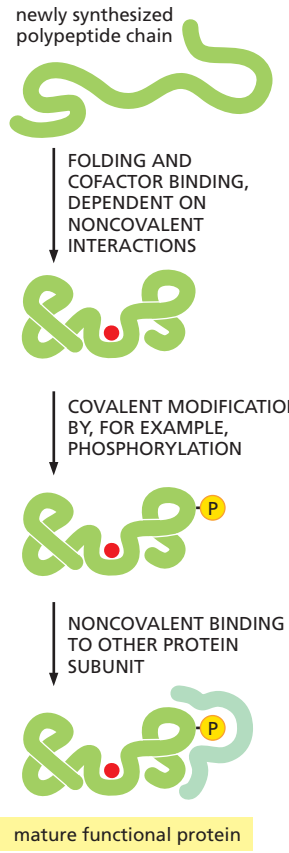


Figure 7–45 Many proteins require post-translational modifications to become fully functional. To be useful to the cell, a completed polypeptide must fold correctly into its three-dimensional conformation and then bind any required cofactors (*red*) and protein partners—all via noncovalent bonding. Many proteins also require one or more covalent modifications to become active—or to be recruited to specific membranes or organelles (not shown). Although phosphorylation and glycosylation are the most common, more than 100 types of covalent modifications of proteins are known.

where they are translated to produce a protein. But the process does not end there. Proteins must then fold into the correct, three-dimensional shape (as we discuss in Chapter 4). Some proteins do so spontaneously, as they emerge from the ribosome. Most, however, require the assistance of *chaperone proteins*, which steer them along productive folding pathways and prevent them from aggregating inside the cell (see Figures 4–8 and 4–9).

In addition to folding properly, many proteins—once they leave the ribosome—require further adjustments before they are useful to the cell. As we discussed in Chapter 4, some proteins are covalently modified—for example, by phosphorylation or glycosylation. Others bind to small-molecule cofactors or associate with additional protein subunits. Such *post-translational modifications* are often needed for a newly synthesized protein to become fully functional (**Figure 7–45**). The final concentration of a protein, therefore, depends on the rate at which each of these steps—from DNA to mature, functional protein—is carried out (**Figure 7–46**).

In principle, any one of these steps can be controlled by cells as they adjust the concentrations of their proteins to suit their needs. However,

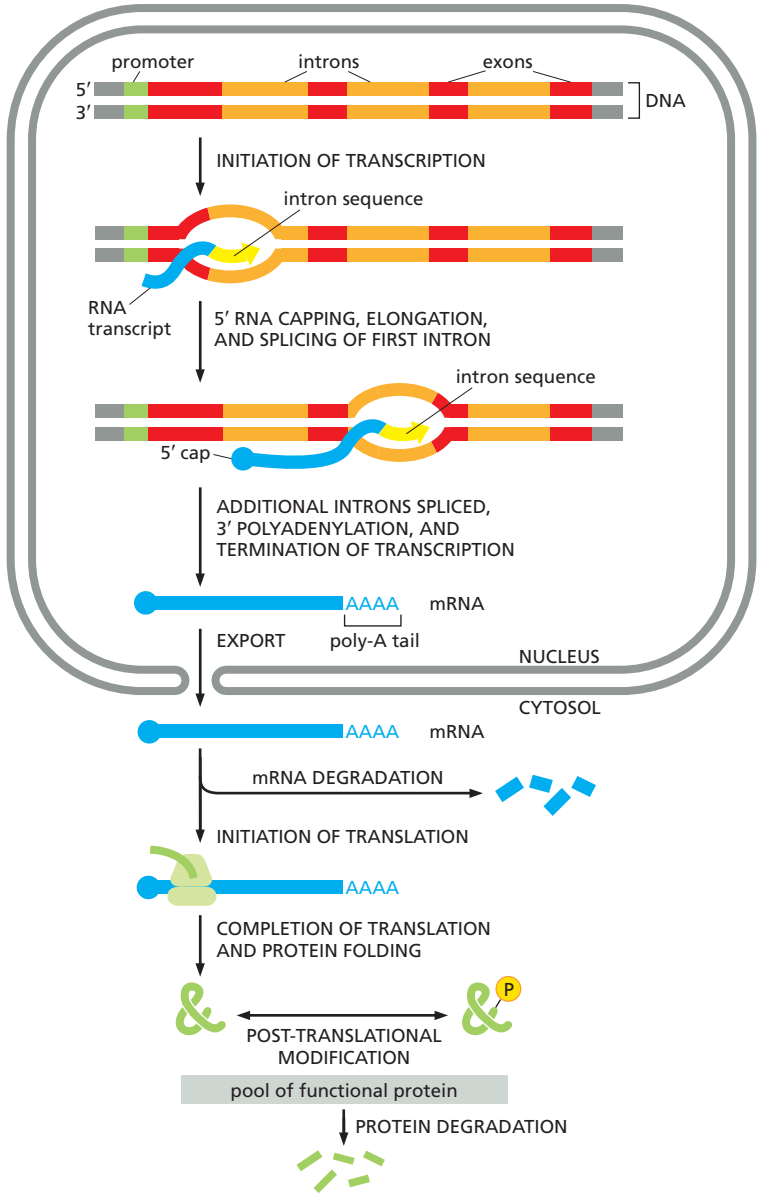


Figure 7–46 Protein production in a eukaryotic cell requires many steps. The final concentration of each protein depends on the rate of each step depicted. Even after an mRNA and its corresponding protein have been produced, their concentrations can be regulated by degradation.

as we will discuss thoroughly in the next chapter, the initiation of transcription is the most common point for a cell to regulate the expression of its genes.

RNA AND THE ORIGINS OF LIFE

The central dogma—that DNA makes RNA, which makes protein—presented evolutionary biologists with a knotty puzzle: if nucleic acids are required to direct the synthesis of proteins, and proteins are required to synthesize nucleic acids, how could this system of interdependent components have arisen? The prevailing view is that an **RNA world** existed on Earth before cells containing DNA and proteins appeared. According to this hypothesis, RNA—which today serves largely as an intermediate between genes and proteins—both stored genetic information and catalyzed chemical reactions in primitive cells. Only later in evolutionary time did DNA take over as the genetic material and proteins become the major catalysts and structural components of cells (**Figure 7–47**). As we have seen, RNA still catalyzes several fundamental reactions in modern cells, including protein synthesis and RNA splicing. These ribozymes are like molecular fossils, holdovers from an earlier RNA world.

Life Requires Autocatalysis

The origin of life requires molecules that possess, if only to a small extent, one crucial property: the ability to catalyze reactions that lead—directly or indirectly—to the production of more molecules like themselves. Catalysts with this self-reproducing property, once they had arisen by chance, would divert raw materials from the production of other substances to make more of themselves. In this way, one can envisage the gradual development of an increasingly complex chemical system of organic monomers and polymers that function together to generate more molecules of the same types, fueled by a supply of simple raw materials in the primitive environment on Earth. Such an *autocatalytic* system would have many of the properties we think of as characteristic of living matter: the system would contain a far-from-random selection of interacting molecules; it would tend to reproduce itself; it would compete with other systems dependent on the same raw materials; and, if deprived of its raw materials or maintained at a temperature that upset the balance of reaction rates, it would decay toward chemical equilibrium and "die."

But what molecules could have had such autocatalytic properties? In present-day living cells, the most versatile catalysts are proteins, which are able to adopt diverse three-dimensional forms that bristle with chemically reactive sites on their surface. However, there is no known way in which a protein can reproduce itself directly. RNA molecules, by contrast, possess properties that—at least, in principle—could be exploited to catalyze their own synthesis.

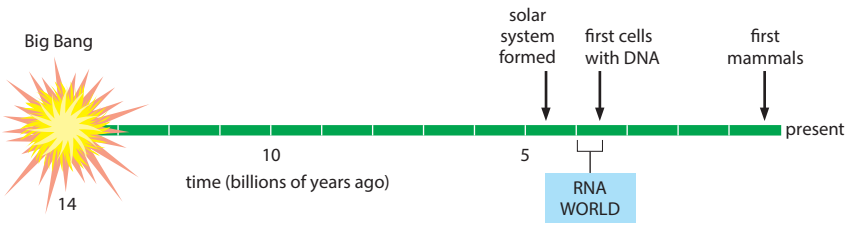


Figure 7–47 An RNA world may have existed before modern cells arose.

RNA Can Store Information and Catalyze Chemical Reactions

We have seen that complementary base-pairing enables one nucleic acid to act as a template for the formation of another. Thus a single strand of RNA or DNA contains the information needed to specify the sequence of a complementary polynucleotide, which, in turn, can specify the sequence of the original molecule, allowing the original nucleic acid to be replicated (**Figure 7–48**). Such complementary templating mechanisms lie at the heart of both DNA replication and transcription in modern-day cells.

But the efficient synthesis of polynucleotides by such complementary templating mechanisms also requires catalysts to promote the polymerization reaction: without catalysts, polymer formation is slow, error-prone, and inefficient. Today, nucleotide polymerization is catalyzed by protein enzymes—such as DNA and RNA polymerases. But how could this reaction be catalyzed before proteins with the appropriate catalytic ability existed? The beginnings of an answer were obtained in 1982, when it was discovered that RNA molecules themselves can act as catalysts.

In present-day cells, RNA is synthesized as a single-stranded molecule, and we have seen that complementary base-pairing can occur between nucleotides in the same chain. This base-pairing, along with nonconventional hydrogen bonds, can cause each RNA molecule to fold up in a unique way that is determined by its nucleotide sequence (see Figure 7–5). Such associations produce complex three-dimensional shapes.

Protein enzymes are able to catalyze biochemical reactions because they have surfaces with unique contours and chemical properties, as we discuss in Chapter 4. In the same way, RNA molecules, with their unique folded shapes, can serve as catalysts (**Figure 7–49**). Catalytic RNAs do not have the same structural and functional diversity as do protein enzymes; they are, after all, built from only four different subunits. Nonetheless, ribozymes can catalyze many types of chemical reactions. Although relatively few catalytic RNAs operate in present-day cells, they play major roles in some of the most fundamental steps in the expression of genetic information—specifically those steps where RNA molecules themselves are spliced or translated into protein. Additional ribozymes, with other catalytic capabilities, have been generated in the laboratory and selected for their activity in a test tube (**Table 7–4**).

RNA, therefore, has all the properties required of an information-containing molecule that could also catalyze its own synthesis (**Figure 7–50**). Although self-replicating systems of RNA molecules have not been found in nature, scientists appear to be well on the way to constructing them in the laboratory. This achievement would not prove that self-replicating RNA molecules were essential to the origin of life on Earth, but it would demonstrate that such a scenario is possible.

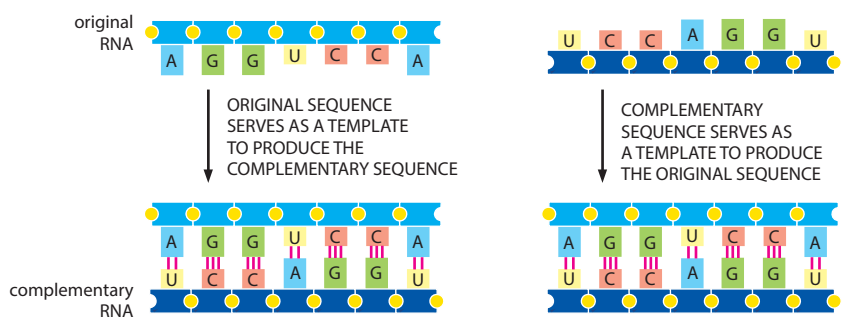


Figure 7–48 An RNA molecule can in principle guide the formation of an exact copy of itself. In the first step, the original RNA molecule acts as a template to produce an RNA molecule of complementary sequence. In the second step, this complementary RNA molecule itself acts as a template to produce an RNA molecule of the original sequence. Since each template molecule can produce many copies of the complementary strand, these reactions can result in the amplification of the original sequence.

TABLE 7–4 BIOCHEMICAL REACTIONS THAT CAN BE CATALYZED BY RIBOZYMES

Activity	Ribozymes
Peptide bond formation in protein synthesis	ribosomal RNA
RNA splicing	small nuclear RNAs (snRNAs), self-splicing RNAs
DNA ligation	*in vitro* selected RNA
RNA polymerization	*in vitro* selected RNA
RNA phosphorylation	*in vitro* selected RNA
RNA aminoacylation	*in vitro* selected RNA
RNA alkylation	*in vitro* selected RNA
C–C bond rotation (isomerization)	*in vitro* selected RNA

RNA Is Thought to Predate DNA in Evolution

If the evolutionary role for RNA proposed above is correct, the first cells on Earth would have stored their genetic information in RNA rather than DNA. And based on the chemical differences between these polynucleotides, it appears that RNA could indeed have arisen before DNA. Ribose (see Figure 7–3A), like glucose and other simple carbohydrates, is readily formed from formaldehyde (HCHO), which is one of the principal products of experiments simulating conditions on the primitive Earth. The sugar deoxyribose is harder to make, and in present-day cells it is produced from ribose in a reaction catalyzed by a protein enzyme, suggesting that ribose predates deoxyribose in cells. Presumably, DNA appeared on the scene after RNA, and then proved better suited than RNA as a permanent repository of genetic information. In particular, the deoxyribose in its sugar–phosphate backbone makes chains of DNA chemically much more stable than chains of RNA, so that DNA can grow to greater lengths without breakage.

The other differences between RNA and DNA—the double-helical structure of DNA and the use of thymine rather than uracil—further enhance DNA stability by making the molecule easier to repair. We saw in Chapter 6 that a damaged nucleotide on one strand of the double helix can be repaired by using the other strand as a template. Furthermore, deamination, one of the most common detrimental chemical changes occurring

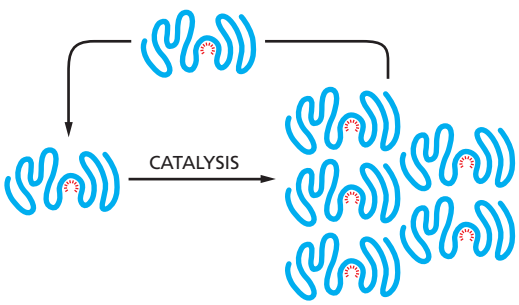


Figure 7–50 Could an RNA molecule catalyze its own synthesis? The process would require that the RNA catalyze the self-templated amplification steps shown in Figure 7–48. The *red* rays represent the active site of this hypothetical ribozyme.

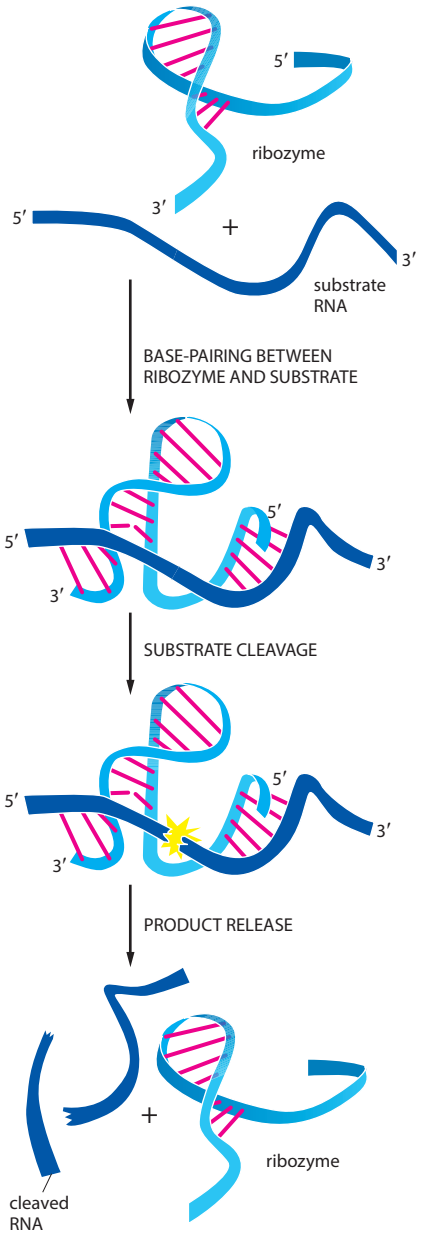


Figure 7–49 A ribozyme is an RNA molecule that possesses catalytic activity. The RNA molecule shown catalyzes the cleavage of a second RNA at a specific site. Such ribozymes are found embedded in large RNA genomes—called viroids—that infect plants, where the cleavage reaction is one step in the replication of the viroid. (Adapted from T.R. Cech and O.C. Uhlenbeck, *Nature* 372:39–40, 1994. With permission from Macmillan Publishers Ltd.)

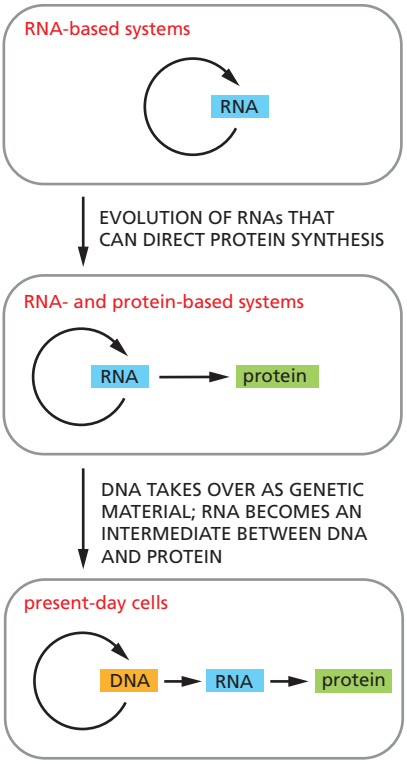


Figure 7–51 RNA may have preceded DNA and proteins in evolution. According to this hypothesis, RNA molecules provided genetic, structural, and catalytic functions in the earliest cells. DNA is now the repository of genetic information, and proteins carry out almost all catalysis in cells. RNA now functions mainly as a go-between in protein synthesis, while remaining a catalyst for a few crucial reactions (including protein synthesis).

QUESTION 7–6

Discuss the following: "During the evolution of life on Earth, RNA lost its glorious position as the first self-replicating catalyst. Its role now is as a mere messenger in the information flow from DNA to protein."

in polynucleotides, is easier to detect and repair in DNA than in RNA (see Figure 6–24). This is because the product of the deamination of cytosine is, by chance, uracil, which already exists in RNA, so that such damage would be impossible for repair enzymes to detect in an RNA molecule. However, in DNA, which has thymine rather than uracil, any uracil produced by the accidental deamination of cytosine is easily detected and repaired.

Taken together, the evidence we have discussed supports the idea that RNA—with its ability to provide genetic, structural, and catalytic functions—preceded DNA in evolution. As cells more closely resembling present-day cells appeared, it is believed that RNAs were relieved of many of the duties they had originally performed: DNA took over the primary storage of genetic information, and proteins became the major catalysts, while RNA remained primarily as the intermediary connecting the two (**Figure 7–51**). With the rise of DNA, cells were able to become more complex, for they could then carry and transmit more genetic information than could be stably maintained by RNA alone. Because of the greater chemical complexity of proteins and the variety of chemical reactions they can catalyze, the shift from RNA to proteins (albeit incomplete) also provided a much richer source of structural components and enzymes, enabling cells to evolve the great diversity of appearance and function that we see today.

ESSENTIAL CONCEPTS

- **The flow of genetic information in all living cells is DNA → RNA → protein. The conversion of the genetic instructions in DNA into RNAs and proteins is termed gene expression.**
- **To express the genetic information carried in DNA, the nucleotide sequence of a gene is first transcribed into RNA. Transcription is catalyzed by the enzyme RNA polymerase, which uses nucleotide sequences in the DNA molecule to determine which strand to use as a template, and where to start and stop transcribing.**
- **RNA differs in several respects from DNA. It contains the sugar ribose instead of deoxyribose and the base uracil (U) instead of thymine (T). RNAs in cells are synthesized as single-stranded molecules, which often fold up into complex three-dimensional shapes.**
- **Cells make several functional types of RNAs, including messenger RNAs (mRNAs), which carry the instructions for making proteins; ribosomal RNAs (rRNAs), which are the crucial components of ribosomes; and transfer RNAs (tRNAs), which act as adaptor molecules in protein synthesis.**
- **To begin transcription, RNA polymerase binds to specific DNA sites called promoters that lie immediately upstream of genes. To initiate transcription, eukaryotic RNA polymerases require the assembly of a complex of general transcription factors at the promoter, whereas bacterial RNA polymerase requires only an additional subunit, called sigma factor.**
- **Most protein-coding genes in eukaryotic cells are composed of a number of coding regions, called exons, interspersed with larger, noncoding regions, called introns. When a eukaryotic gene is transcribed from DNA into RNA, both the exons and introns are copied.**
- **Introns are removed from the RNA transcripts in the nucleus by RNA splicing, a reaction catalyzed by small ribonucleoprotein complexes known as snRNPs. Splicing removes the introns from the RNA and joins together the exons—often in a variety of combinations, allowing multiple proteins to be produced from the same gene.**

- Eukaryotic pre-mRNAs go through several additional RNA processing steps before they leave the nucleus as mRNAs, including 5′ RNA capping and 3′ polyadenylation. These reactions, along with splicing, take place as the pre-mRNA is being transcribed.
- Translation of the nucleotide sequence of an mRNA into a protein takes place in the cytoplasm on large ribonucleoprotein assemblies called ribosomes. As the mRNA moves through the ribosome, its message is translated into protein.
- The nucleotide sequence in mRNA is read in consecutive sets of three nucleotides called codons; each codon corresponds to one amino acid.
- The correspondence between amino acids and codons is specified by the genetic code. The possible combinations of the 4 different nucleotides in RNA give 64 different codons in the genetic code. Most amino acids are specified by more than one codon.
- tRNAs act as adaptor molecules in protein synthesis. Enzymes called aminoacyl-tRNA synthetases covalently link amino acids to their appropriate tRNAs. Each tRNA contains a sequence of three nucleotides, the anticodon, which recognizes a codon in an mRNA through complementary base-pairing.
- Protein synthesis begins when a ribosome assembles at an initiation codon (AUG) in an mRNA molecule, a process that depends on proteins called translation initiation factors. The completed protein chain is released from the ribosome when a stop codon (UAA, UAG, or UGA) in the mRNA is reached.
- The stepwise linking of amino acids into a polypeptide chain is catalyzed by an rRNA molecule in the large ribosomal subunit, which thus acts as a ribozyme.
- The concentration of a protein in a cell depends on the rates at which the mRNA and protein are synthesized and degraded. Protein degradation in the cytosol and nucleus occurs inside large protein complexes called proteasomes.
- From our knowledge of present-day organisms and the molecules they contain, it seems likely that life on Earth began with the evolution of RNA molecules that could catalyze their own replication.
- It has been proposed that RNA served as both the genome and the catalysts in the first cells, before DNA replaced RNA as a more stable molecule for storing genetic information, and proteins replaced RNAs as the major catalytic and structural components. RNA catalysts in modern cells are thought to provide a glimpse into an ancient, RNA-based world.

KEY TERMS

alternative splicing
aminoacyl-tRNA synthetase
anticodon
codon
exon
gene
gene expression
general transcription factors
genetic code
initiator tRNA
intron
messenger RNA (mRNA)
polyadenylation
promoter
protease
proteasome
reading frame
ribosomal RNA (rRNA)
ribosome
ribozyme
RNA
RNA capping
RNA polymerase
RNA processing
RNA splicing
RNA transcript
RNA world
small nuclear RNA (snRNA)
spliceosome
transcription
transfer RNA (tRNA)
translation
translation initiation factor

QUESTIONS

QUESTION 7–7

Which of the following statements are correct? Explain your answers.

A. An individual ribosome can make only one type of protein.

B. All mRNAs fold into particular three-dimensional structures that are required for their translation.

C. The large and small subunits of an individual ribosome always stay together and never exchange partners.

D. Ribosomes are cytoplasmic organelles that are encapsulated by a single membrane.

E. Because the two strands of DNA are complementary, the mRNA of a given gene can be synthesized using either strand as a template.

F. An mRNA may contain the sequence ATTGACCCCGGTCAA.

G. The amount of a protein present in a cell depends on its rate of synthesis, its catalytic activity, and its rate of degradation.

QUESTION 7–8

The Lacheinmal protein is a hypothetical protein that causes people to smile more often. It is inactive in many chronically unhappy people. The mRNA isolated from a number of different unhappy individuals in the same family was found to lack an internal stretch of 173 nucleotides that is present in the Lacheinmal mRNA isolated from happy members of the same family. The DNA sequences of the *Lacheinmal* genes from the happy and unhappy family members were determined and compared. They differed by a single nucleotide substitution, which lay in an intron. What can you say about the molecular basis of unhappiness in this family?

(Hints: [1] Can you hypothesize a molecular mechanism by which a single nucleotide substitution in a gene could cause the observed deletion in the mRNA? Note that the deletion is *internal* to the mRNA. [2] Assuming the 173-base-pair deletion removes coding sequences from the Lacheinmal mRNA, how would the Lacheinmal protein differ between the happy and unhappy people?)

QUESTION 7–9

Use the genetic code shown in Figure 7–27 to identify which of the following nucleotide sequences would code for the polypeptide sequence arginine-glycine-aspartate:

1. 5′-AGA–GGA–GAU-3′
2. 5′-ACA–CCC–ACU-3′
3. 5′-GGG–AAA–UUU-3′
4. 5′-CGG–GGU–GAC-3′

QUESTION 7–10

"The bonds that form between the anticodon of a tRNA molecule and the three nucleotides of a codon in mRNA are _____." Complete this sentence with each of the following options and explain whether each of the resulting statements is correct or incorrect.

A. covalent bonds formed by GTP hydrolysis

B. hydrogen bonds that form when the tRNA is at the A site

C. broken by the translocation of the ribosome along the mRNA

QUESTION 7–11

List the ordinary, dictionary definitions of the terms *replication*, *transcription*, and *translation*. By their side, list the special meaning each term has when applied to the living cell.

QUESTION 7–12

In an alien world, the genetic code is written in pairs of nucleotides. How many amino acids could such a code specify? In a different world, a triplet code is used, but the order of nucleotides is not important; it only matters which nucleotides are present. How many amino acids could this code specify? Would you expect to encounter any problems translating these codes?

QUESTION 7–13

One remarkable feature of the genetic code is that amino acids with similar chemical properties often have similar codons. Thus codons with U or C as the second nucleotide tend to specify hydrophobic amino acids. Can you suggest a possible explanation for this phenomenon in terms of the early evolution of the protein-synthesis machinery?

QUESTION 7–14

A mutation in DNA generates a UGA stop codon in the middle of the mRNA coding for a particular protein. A second mutation in the cell's DNA leads to a single nucleotide change in a tRNA that allows the correct translation of this protein; that is, the second mutation "suppresses" the defect caused by the first. The altered tRNA translates the UGA as tryptophan. What nucleotide change has probably occurred in the mutant tRNA molecule? What consequences would the presence of such a mutant tRNA have for the translation of the normal genes in this cell?

QUESTION 7–15

The charging of a tRNA with an amino acid can be represented by the following equation:

$$\text{amino acid} + \text{tRNA} + \text{ATP} \rightarrow \text{aminoacyl-tRNA} + \text{AMP} + PP_i$$

where PP_i is pyrophosphate (see Figure 3–41). In the aminoacyl-tRNA, the amino acid and tRNA are linked with a high-energy covalent bond; a large portion of the energy derived from the hydrolysis of ATP is thus stored in this bond and is available to drive peptide bond formation during the later stages of protein synthesis. The free-energy change of the charging reaction shown in the equation is close to zero and therefore would not be expected to favor attachment of the amino acid to tRNA. Can you suggest a further step that could drive the reaction to completion?

QUESTION 7–16

A. The average molecular weight of a protein in the cell is about 30,000 daltons. A few proteins, however, are much larger. The largest known polypeptide chain made by any cell is a protein called titin (made by mammalian muscle cells), and it has a molecular weight of 3,000,000 daltons. Estimate how long it will take a muscle cell to translate an mRNA coding for titin (assume the average molecular weight of an amino acid to be 120, and a translation rate of two amino acids per second for eukaryotic cells).

B. Protein synthesis is very accurate: for every 10,000 amino acids joined together, only one mistake is made. What is the fraction of average-sized protein molecules and of titin molecules that are synthesized without any errors? [Hint: the probability P of obtaining an error-free protein is given by $P = (1 - E)^n$, where E is the error frequency and n the number of amino acids.]

C. The combined molecular weight of the eukaryotic ribosomal proteins is about 2.5×10^6 daltons. Would it be advantageous to synthesize them as a single protein?

D. Transcription occurs at a rate of about 30 nucleotides per second. Is it possible to calculate the time required to synthesize a titin mRNA from the information given here?

QUESTION 7–17

Which of the following types of mutations would be predicted to harm an organism? Explain your answers.

A. Insertion of a single nucleotide near the end of the coding sequence.

B. Removal of a single nucleotide near the beginning of the coding sequence.

C. Deletion of three consecutive nucleotides in the middle of the coding sequence.

D. Deletion of four consecutive nucleotides in the middle of the coding sequence.

E. Substitution of one nucleotide for another in the middle of the coding sequence.

QUESTION 7–18

Figure 7–8 shows many molecules of RNA polymerase simultaneously transcribing two adjacent genes on a single DNA molecule. Looking at this figure, label the 5′ and 3′ ends of the DNA template strand and the sets of RNA molecules being transcribed.