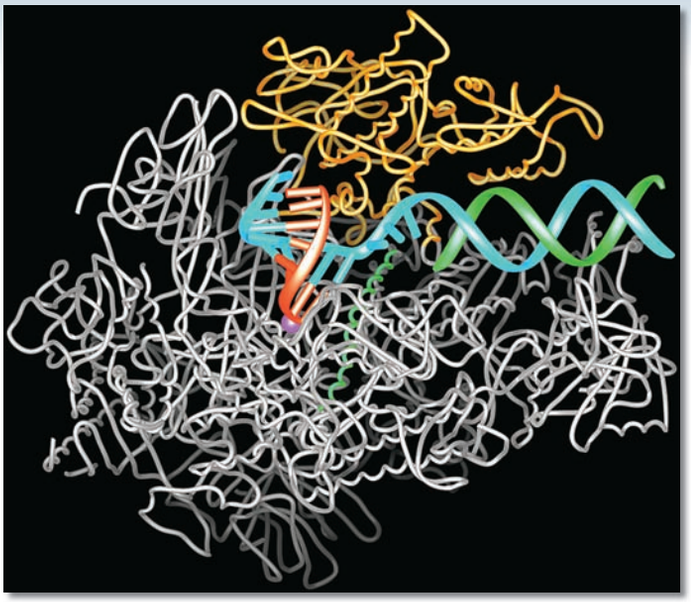
CHAPTER OUTLINE

12.1 Overview of Transcription
12.2 Transcription in Bacteria
12.3 Transcription in Eukaryotes
12.4 RNA Modification



A molecular model showing the enzyme RNA polymerase in the act of sliding along the DNA and synthesizing a copy of RNA.

12 GENE TRANSCRIPTION AND RNA MODIFICATION

The function of the genetic material is that of a blueprint. It stores the information necessary to create a living organism. The information is contained within units called genes. At the molecular level, a **gene** is defined as a segment of DNA that is used to make a functional product, either an RNA molecule or a polypeptide. How is the information within a gene accessed? The first step in this process is called **transcription,** which literally means the act or process of making a copy. In genetics, this term refers to the process of synthesizing RNA from a DNA sequence (**Figure 12.1**). The structure of DNA is not altered as a result of transcription. Rather, the DNA base sequence has only been accessed to make a copy in the form of RNA. Therefore, the same DNA can continue to store information. DNA replication, which was discussed in Chapter 11, provides a mechanism for copying that information so it can be transmitted from cell to cell and from parent to offspring.

Structural genes encode the amino acid sequence of a polypeptide. When a structural gene is transcribed, the first product is an RNA molecule known as **messenger RNA (mRNA).**

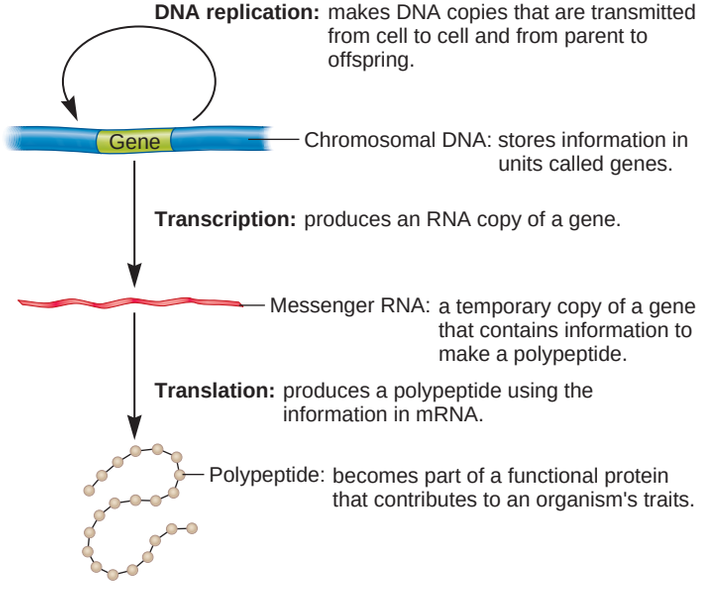


FIGURE 12.1 The central dogma of genetics. The usual flow of genetic information is from DNA to mRNA to polypeptide. Note: The direction of informational flow shown in this figure is the most common direction found in living organisms, but exceptions occur. For example, RNA viruses and certain transposable elements use an enzyme called reverse transcriptase to make a copy of DNA from RNA.

During polypeptide synthesis—a process called **translation**—the sequence of nucleotides within the mRNA determines the sequence of amino acids in a polypeptide. One or more polypeptides then assemble into a functional protein. The synthesis of functional proteins ultimately determines an organism's traits. The model depicted in Figure 12.1, which is called the **central dogma of genetics** (also called the central dogma of molecular biology), was first enunciated by Francis Crick in 1958. It forms a cornerstone of our understanding of genetics at the molecular level. The flow of genetic information occurs from DNA to mRNA to polypeptide.

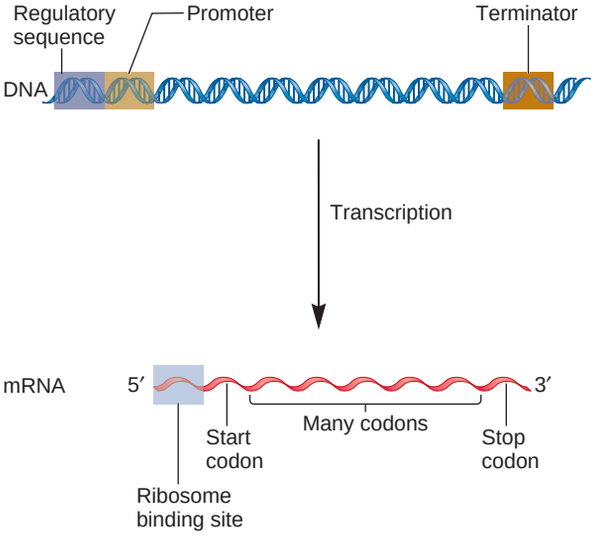
In this chapter, we begin to study the molecular steps in gene expression, with an emphasis on transcription and the modifications that may occur to an RNA transcript after it has been made. Chapter 13 will examine the process of translation, and Chapters 14 and 15 will focus on how the level of gene expression is regulated at the molecular level.

12.1 OVERVIEW OF TRANSCRIPTION

One key concept important in the process of transcription is that short base sequences define the beginning and ending of a gene and also play a role in regulating the level of RNA synthesis. In this section, we begin by examining the sequences that determine where transcription starts and ends, and also briefly consider DNA sequences, called regulatory sites, that influence whether a gene is turned on or off. The functions of regulatory sites will be examined in greater detail in Chapters 14 and 15. A second important concept is the role of proteins in transcription. DNA sequences, in and of themselves, just exist. For genes to be actively transcribed, proteins must recognize particular DNA sequences and act on them in a way that affects the transcription process. In the later part of this section, we will consider how proteins participate in the general steps of transcription and the types of RNA transcripts that can be made.

Gene Expression Requires Base Sequences That Perform Different Functional Roles

At the molecular level, **gene expression** is the overall process by which the information within a gene is used to produce a functional product, such as a polypeptide. Along with environmental factors, the molecular expression of genes determines an organism's traits. For a gene to be expressed, a few different types of base sequences perform specific roles. **Figure 12.2** shows a common organization of base sequences needed to create a structural gene that functions in a bacterium such as *E. coli*. Each type of base sequence performs its role during a particular stage of gene expression. For example, the promoter and terminator are base sequences used during gene transcription. Specifically, the **promoter** provides a site to begin transcription, and the **terminator** specifies the end of transcription. These two sequences cause RNA synthesis to occur within a defined location. As shown in Figure 12.2, the DNA is transcribed into RNA from the end of the promoter to the



DNA:

- **Regulatory sequences:** site for the binding of regulatory proteins; the role of regulatory proteins is to influence the rate of transcription. Regulatory sequences can be found in a variety of locations.
- **Promoter:** site for RNA polymerase binding; signals the beginning of transcription.
- **Terminator:** signals the end of transcription.

mRNA:

- **Ribosomal binding site:** site for ribosome binding; translation begins near this site in the mRNA. In eukaryotes, the ribosome scans the mRNA for a start codon.
- **Start codon:** specifies the first amino acid in a polypeptide sequence, usually a formylmethionine (in bacteria) or a methionine (in eukaryotes).
- **Codons:** 3-nucleotide sequences within the mRNA that specify particular amino acids. The sequence of codons within mRNA determines the sequence of amino acids within a polypeptide.
- **Stop codon:** specifies the end of polypeptide synthesis.
- Bacterial mRNA may be polycistronic, which means it encodes two or more polypeptides.

FIGURE 12.2 **Organization of sequences of a bacterial gene and its mRNA transcript.** This figure depicts the general organization of sequences that are needed to create a functional gene that encodes an mRNA.

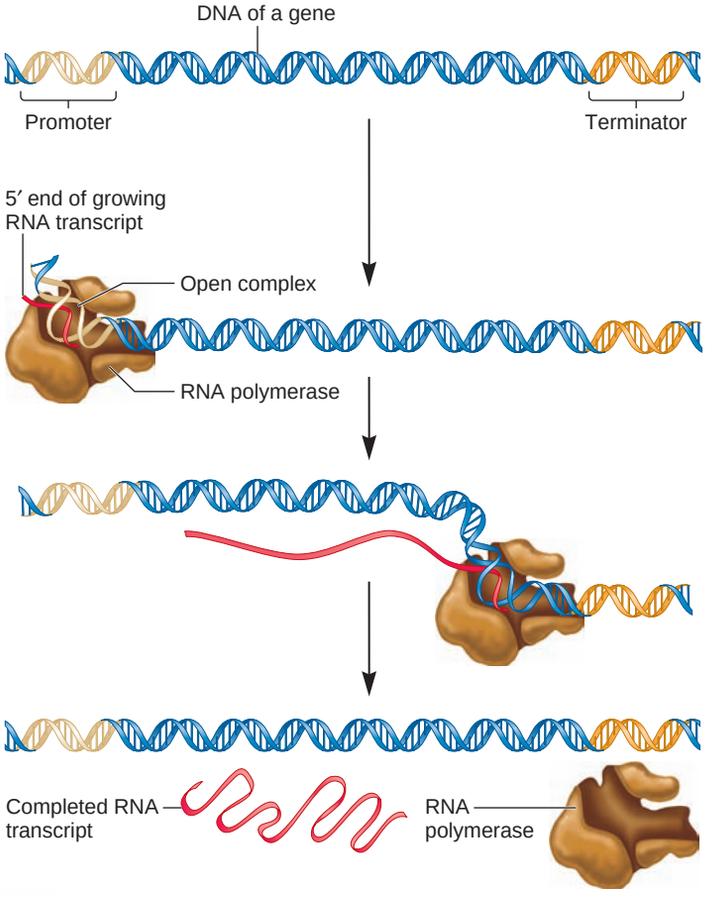
terminator. As described later, the base sequence in the RNA transcript is complementary to the **template strand** of DNA. The opposite strand is the **nontemplate strand.** For structural genes, the nontemplate strand is also called the **coding strand** because its sequence is the same as the transcribed mRNA that encodes a polypeptide, except that the DNA has T's in places where the mRNA contains U's.

A category of proteins called **transcription factors** recognizes base sequences in the DNA and controls transcription. Some transcription factors bind directly to the promoter and facilitate transcription. Other transcription factors recognize **regulatory sequences,** or **regulatory elements**—short stretches of DNA involved in the regulation of transcription. Certain transcription factors bind to such regulatory sequences and increase the rate of transcription while others inhibit transcription.

Base sequences within an mRNA are used during the translation process. In bacteria, a short sequence within the mRNA, the **ribosome-binding site,** provides a location for the ribosome to bind and begin translation. The bacterial ribosome recognizes this site because it is complementary to a sequence in ribosomal RNA. In addition, mRNA contains a series of **codons,** read as groups of three nucleotides, which contain the information for a polypeptide's sequence. The first codon, which is very close to the ribosome-binding site, is the **start codon.** This is followed by many more codons that dictate the sequence of amino acids within the synthesized polypeptide. Finally, a **stop codon** signals the end of translation. Chapter 13 will examine the process of translation in greater detail.

The Three Stages of Transcription Are Initiation, Elongation, and Termination

Transcription occurs in three stages: **initiation; elongation,** or synthesis of the RNA transcript; and **termination** (Figure 12.3). These steps involve protein-DNA interactions in which proteins such as **RNA polymerase,** the enzyme that synthesizes RNA, interact with DNA sequences. What causes transcription to begin? The initiation stage in the transcription process is a recognition step. The sequence of bases within the promoter region is recognized by transcription factors. The specific binding of transcription factors to the promoter sequence identifies the starting site for transcription.



Initiation: The promoter functions as a recognition site for transcription factors (not shown). The transcription factor(s) enables RNA polymerase to bind to the promoter. Following binding, the DNA is denatured into a bubble known as the open complex.

Elongation/synthesis of the RNA transcript: RNA polymerase slides along the DNA in an open complex to synthesize RNA.

Termination: A terminator is reached that causes RNA polymerase and the RNA transcript to dissociate from the DNA.

ONLINE ANIMATION

FIGURE 12.3 Stages of transcription.

Genes→Traits The ability of genes to produce an organism's traits relies on the molecular process of gene expression. Transcription is the first step in gene expression. During transcription, the gene's sequence within the DNA is used as a template to make a complementary copy of RNA. In Chapter 13, we will examine how the sequence in mRNA is translated into a polypeptide chain. After polypeptides are made within a living cell, they fold into functional proteins that govern an organism's traits.

Transcription factors and RNA polymerase first bind to the promoter region when the DNA is in the form of a double helix. For transcription to occur, the DNA strands must be separated. This allows one of the two strands to be used as a template for the synthesis of a complementary strand of RNA. This synthesis occurs as RNA polymerase slides along the DNA, forming a small bubble-like structure known as the open promoter complex, or simply as the **open complex.** Eventually, RNA polymerase reaches a terminator, which causes both RNA polymerase and the newly made RNA transcript to dissociate from the DNA.

RNA Transcripts Have Different Functions

Once they are made, RNA transcripts play different functional roles (**Table 12.1**). Well over 90% of all genes are structural genes, which are transcribed into mRNA. For structural genes, mRNAs are made first, but the final, functional products are polypeptides that are components of proteins. The remaining types of RNAs described in Table 12.1 are never translated. The RNA transcripts from such nonstructural genes have various important cellular functions. For nonstructural genes, the functional product is the RNA. In some cases, the RNA transcript becomes part of a complex that contains both protein subunits and one or more RNA molecules. Examples of protein-RNA complexes include ribosomes, signal recognition particles, RNaseP, spliceosomes, and telomerase.

TABLE **12.1**

Functions of RNA Molecules

Type of RNA	Description
mRNA	Messenger RNA (mRNA) encodes the sequence of amino acids within a polypeptide. In bacteria, some mRNAs encode a single polypeptide. Other mRNAs are polycistronic—a single mRNA encodes two or more polypeptides. In most species of eukaryotes, each mRNA usually encodes a single polypeptide. However, in some species, such as *Caenorhabditis elegans* (a nematode worm), polycistronic mRNAs are relatively common.
tRNA	Transfer RNA (tRNA) is necessary for the translation of mRNA. The structure and function of transfer RNA are outlined in Chapter 13.
rRNA	Ribosomal RNA (rRNA) is necessary for the translation of mRNA. Ribosomes are composed of both rRNAs and protein subunits. The structure and function of ribosomes are examined in Chapter 13.
MicroRNA	MicroRNAs (miRNAs) are short RNA molecules that are involved in gene regulation in eukaryotes (see Chapter 15).
scRNA	Small cytoplasmic RNA (scRNA) is found in the cytoplasm of bacteria and eukaryotes. In bacteria, scRNA is needed for protein secretion. An example in eukaryotes is 7S RNA, which is necessary in the targeting of proteins to the endoplasmic reticulum. It is a component of a complex known as signal recognition particle (SRP), which is composed of 7S RNA and six different protein subunits.
RNA of RNaseP	RNaseP is a catalyst necessary in the processing of tRNA molecules. The RNA is the catalytic component. RNaseP is composed of a 350- to 410-nucleotide RNA and one protein subunit.
snRNA	Small nuclear RNA (snRNA) is necessary in the splicing of eukaryotic pre-mRNA. snRNAs are components of a spliceosome, which is composed of both snRNAs and protein subunits. The structure and function of spliceosomes are examined later in this chapter.
Telomerase RNA	The enzyme telomerase, which is involved in the replication of eukaryotic telomeres, is composed of an RNA molecule and protein subunits.
snoRNA	Small nucleolar RNA (snoRNA) is necessary in the processing of eukaryotic rRNA transcripts. snoRNAs are also associated with protein subunits. In eukaryotes, snoRNAs are found in the nucleolus, where rRNA processing and ribosome assembly occur.
Viral RNAs	Some types of viruses use RNA as their genome, which is packaged within the viral capsid.

12.2 TRANSCRIPTION IN BACTERIA

Our molecular understanding of gene transcription initially came from studies involving bacteria and bacteriophages. Several early investigations focused on the production of viral RNA after bacteriophage infection. The first suggestion that RNA is derived from the transcription of DNA was made by Elliot Volkin and Lazarus Astrachan in 1956. When the researchers exposed *E. coli* cells to T2 bacteriophage, they observed that the RNA made immediately after infection had a base composition substantially different from the base composition of RNA prior to infection. Furthermore, the base composition after infection was very similar to the base composition in the T2 DNA, except that the RNA contained uracil instead of thymine. These results were consistent with the idea that the bacteriophage DNA is used as a template for the synthesis of bacteriophage RNA.

In 1960, Matthew Meselson and François Jacob found that proteins are synthesized on ribosomes. One year later, Jacob and his colleague Jacques Monod proposed that a certain type of RNA acts as a genetic messenger (from the DNA to the ribosome) to provide the information for protein synthesis. They hypothesized that this RNA, which they called messenger RNA (mRNA), is transcribed from the sequence within DNA and then directs the synthesis of particular polypeptides. In the early 1960s, this proposal was remarkable, considering that it was made before the actual isolation and characterization of the mRNA molecules in vitro. In 1961, the hypothesis was confirmed by Sydney Brenner in collaboration with Jacob and Meselson. They found that when a virus infects a bacterial cell, a virus-specific RNA is made that rapidly associates with preexisting ribosomes in the cell.

Since these pioneering studies, a great deal has been learned about the molecular features of bacterial gene transcription. Much of our knowledge comes from studies of *E. coli*. In this section, we will examine the three steps in the gene transcription process as they occur in bacteria.

A Promoter Is a Short Sequence of DNA That Is Necessary to Initiate Transcription

The type of DNA sequence known as the promoter gets its name from the idea that it "promotes" gene expression. More precisely, this sequence of bases directs the exact location for the initiation of RNA transcription. Most of the promoter region is located just ahead of or upstream from the site where transcription of a gene actually begins. By convention, the bases in a promoter sequence are numbered in relation to the **transcriptional start site** (Figure 12.4). This site is the first base used as a template for RNA transcription and is denoted +1. The bases preceding this site are numbered in a negative direction. No base is numbered zero. Therefore, most of the promoter region is labeled with negative numbers that describe the number of bases preceding the beginning of transcription.

Although the promoter may encompass a region several dozen nucleotides in length, short **sequence elements** are particularly critical for promoter recognition. By comparing the sequence of DNA bases within many promoters, researchers have learned that certain sequences of bases are necessary to create a functional promoter. In many promoters found in *E. coli* and similar species, two sequence elements are important. These are located at approximately the –35 and –10 sites in the promoter region (see Figure 12.4). The sequence in the top DNA strand at the –35 region is 5′–TTGACA–3′, and the one at the –10 region is 5′–TATAAT–3′. The TATAAT sequence is called the **Pribnow box** after David Pribnow, who initially discovered it in 1975.

The sequences at the –35 and –10 sites can vary among different genes. For example, Figure 12.5 illustrates the sequences found in several different *E. coli* promoters. The most commonly occurring bases within a sequence element form the **consensus sequence.** This sequence is efficiently recognized by proteins that initiate transcription. For many bacterial genes, a strong correlation is found between the maximal rate of RNA transcription and the degree to which the –35 and –10 regions agree with their consensus sequences.

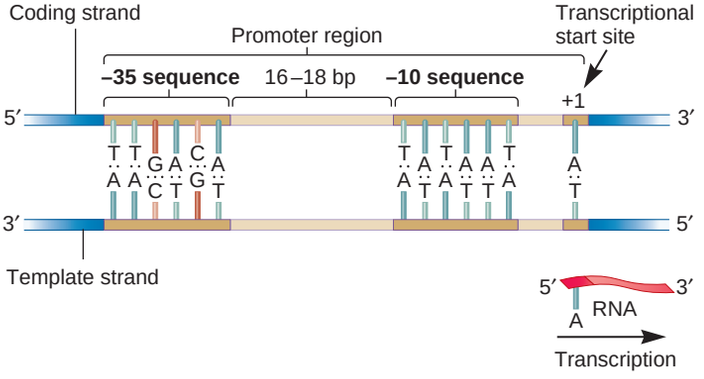


FIGURE 12.4 **The conventional numbering system of promoters.** The first nucleotide that acts as a template for transcription is designated +1. The numbering of nucleotides to the left of this spot is in a negative direction, whereas the numbering to the right is in a positive direction. For example, the nucleotide that is immediately to the left of the +1 nucleotide is numbered –1, and the nucleotide to the right of the +1 nucleotide is numbered +2. There is no zero nucleotide in this numbering system. In many bacterial promoters, sequence elements at the –35 and –10 regions play a key role in promoting transcription.

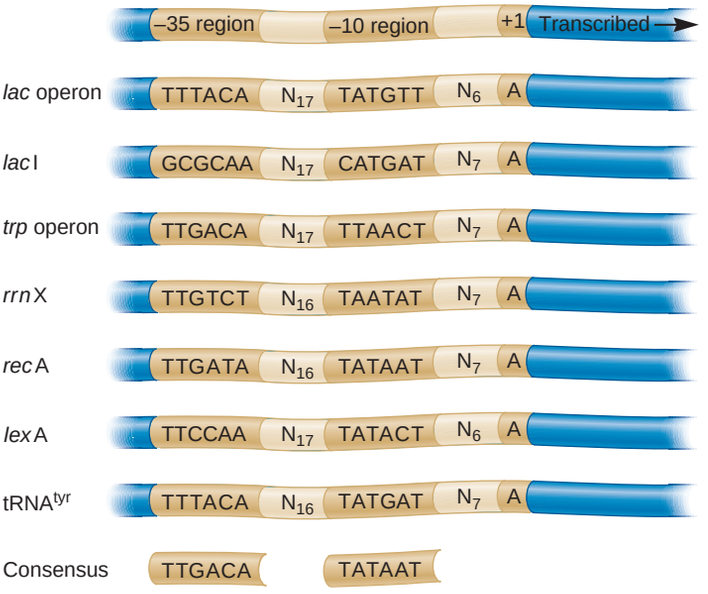


FIGURE 12.5 **Examples of –35 and –10 sequences within a variety of bacterial promoters.** This figure shows the –35 and –10 sequences for one DNA strand found in seven different bacterial and bacteriophage promoters. The consensus sequence is shown at the bottom. The spacer regions contain the designated number of nucleotides between the –35 and –10 region or between the –10 region and the transcriptional start site. For example, N_{17} means there are 17 nucleotides between the end of the –35 region and the beginning of the –10 region.

Bacterial Transcription Is Initiated When RNA Polymerase Holoenzyme Binds at a Promoter Sequence

Thus far, we have considered the DNA sequences that constitute a functional promoter. Let's now turn our attention to the proteins that recognize those sequences and carry out the transcription process. The enzyme that catalyzes the synthesis of RNA is RNA polymerase. In *E. coli*, the **core enzyme** is composed of five subunits, $\alpha_2\beta\beta'\omega$. The association of a sixth subunit, **sigma (σ) factor,** with the core enzyme is referred to as RNA polymerase **holoenzyme.** The different subunits within the holoenzyme play distinct functional roles. The two α subunits are important in the proper assembly of the holoenzyme and in the process of binding to DNA. The β and β′ subunits are also needed for binding to the DNA and carry out the catalytic synthesis of RNA. The ω (omega) subunit is important for the proper assembly of the core enzyme. The holoenzyme is required to initiate transcription; the primary role of σ factor is to recognize the promoter. Proteins, such as σ factor, that influence the function of RNA polymerase are types of transcription factors.

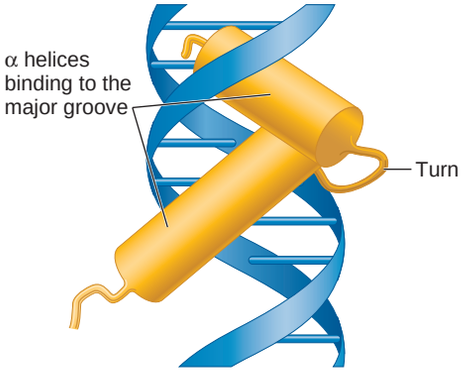


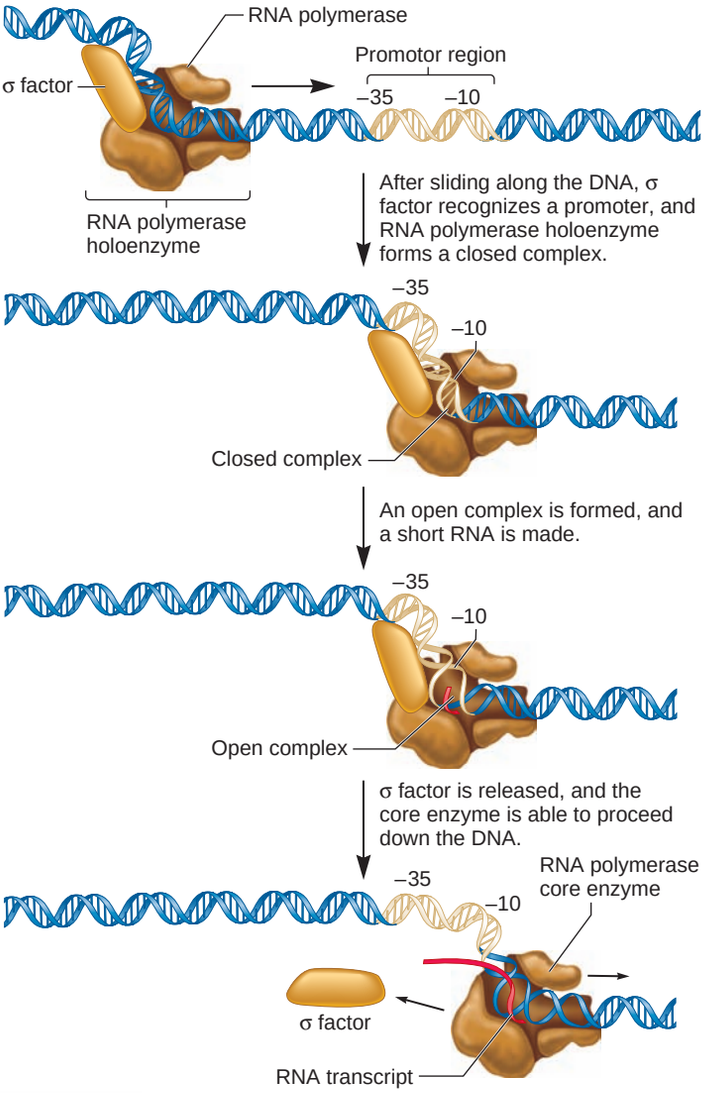
FIGURE 12.6 The binding of *σ* factor protein to the DNA double helix. In this example, the protein contains two α helices connected by a turn, termed a helix-turn-helix motif. Two α helices of the protein can fit within the major groove of the DNA. Amino acids within the α helices form hydrogen bonds with the bases in the DNA.

After RNA polymerase holoenzyme is assembled into its six subunits, it binds loosely to the DNA and then slides along the DNA, much as a train rolls down the tracks. How is a promoter identified? When the holoenzyme encounters a promoter sequence, σ factor recognizes the bases at both the –35 and –10 regions. σ factor protein contains a structure called a **helix-turn-helix motif** that can bind tightly to these regions. Alpha (α) helices within the protein fit into the major groove of the DNA double helix and form hydrogen bonds with the bases. This phenomenon of molecular recognition is shown in **Figure 12.6**. Hydrogen bonding occurs between nucleotides in the –35 and –10 regions of the promoter and amino acid side chains in the helix-turn-helix structure of σ factor.

As shown in **Figure 12.7**, the process of transcription is initiated when σ factor within the holoenzyme has bound to the promoter region to form the **closed complex.** For transcription to begin, the double-stranded DNA must then be unwound into an open complex. This unwinding first occurs at the TATAAT sequence in the –10 region, which contains only AT base pairs, as shown in Figure 12.4. AT base pairs form only two hydrogen bonds, whereas GC pairs form three. Therefore, DNA in an AT-rich region is more easily separated because fewer hydrogen bonds must be broken. A short strand of RNA is made within the open complex, and then σ factor is released from the core enzyme. The release of σ factor marks the transition to the elongation phase of transcription. The core enzyme may now slide down the DNA to synthesize a strand of RNA.

The RNA Transcript Is Synthesized During the Elongation Stage

After the initiation stage of transcription is completed, the RNA transcript is made during the elongation stage. During the synthesis of the RNA transcript, RNA polymerase moves along the DNA, causing it to unwind (**Figure 12.8**). As previously mentioned, the DNA strand used as a template for RNA synthesis

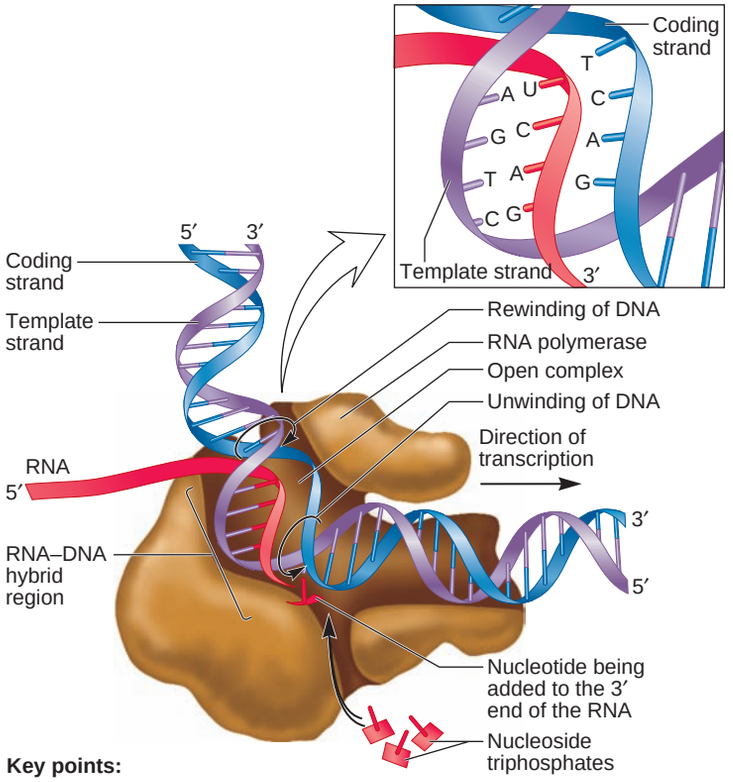


ONLINE ANIMATION

FIGURE 12.7 The initiation stage of transcription in bacteria. The σ factor subunit of the RNA polymerase holoenzyme recognizes the –35 and –10 regions of the promoter. The DNA unwinds in the –10 region to form an open complex, and a short RNA is made. σ factor then dissociates from the holoenzyme, and the RNA polymerase core enzyme can proceed down the DNA to transcribe RNA, forming an open complex as it goes.

is called the template, or antisense, strand. The opposite DNA strand is the coding, or sense, strand; it has the same sequence as the RNA transcript except that T in the DNA corresponds to U in the RNA. Within a given gene, only the template strand is used for RNA synthesis, whereas the coding strand is never used. As it moves along the DNA, the open complex formed by the action of RNA polymerase is approximately 17 bp long. On average, the rate of RNA synthesis is about 43 nucleotides per second! Behind the open complex, the DNA rewinds back into a double helix.

As described in Figure 12.8, the chemistry of transcription by RNA polymerase is similar to the synthesis of DNA via DNA polymerase, which is discussed in Chapter 11. RNA polymerase



Key points:

- RNA polymerase slides along the DNA, creating an open complex as it moves.
- The DNA strand known as the template strand is used to make a complementary copy of RNA as an RNA–DNA hybrid.
- RNA polymerase moves along the template strand in a 3′ to 5′ direction, and RNA is synthesized in a 5′ to 3′ direction using nucleoside triphosphates as precursors. Pyrophosphate is released (not shown).
- The complementarity rule is the same as the AT/GC rule except that U is substituted for T in the RNA.

ONLINE ANIMATION

FIGURE 12.8 Synthesis of the RNA transcript.

always connects nucleotides in the 5′ to 3′ direction. During this process, RNA polymerase catalyzes the formation of a bond between the 5′ phosphate group on one nucleotide and the 3′–OH group on the previous nucleotide. The complementarity rule is similar to the AT/GC rule, except that uracil substitutes for thymine in the RNA. In other words, RNA synthesis obeys an $A_{DNA}U_{RNA}$/ $T_{DNA}A_{RNA}$/$G_{DNA}C_{RNA}$/$C_{DNA}G_{RNA}$ rule.

When considering the transcription of multiple genes within a chromosome, the direction of transcription and the DNA strand used as a template varies among different genes. **Figure 12.9** shows three genes adjacent to each other within a chromosome. Genes *A* and *B* are transcribed from left to right, using the bottom DNA strand as a template. By comparison, gene *C* is transcribed from right to left and uses the top DNA strand as a template. Note that in all three cases, the template strand is read in the 3′ to 5′ direction, and the synthesis of the RNA transcript occurs in a 5′ to 3′ direction.

Transcription Is Terminated by Either an RNA-Binding Protein or an Intrinsic Terminator

The end of RNA synthesis is referred to as termination. Prior to termination, the hydrogen bonding between the DNA and RNA within the open complex is of central importance in preventing dissociation of RNA polymerase from the template strand. Termination occurs when this short RNA-DNA hybrid region is forced to separate, thereby releasing RNA polymerase as well as the newly made RNA transcript. In *E. coli*, two different mechanisms for termination have been identified. For certain genes, an RNA-binding protein known as **ρ (rho)** is responsible for terminating transcription, in a mechanism called ρ-dependent termination. For other genes, termination does not require the involvement of the ρ protein. This is referred to as ρ-independent termination.

In **ρ-dependent termination,** the termination process requires two components. First, a sequence upstream from the

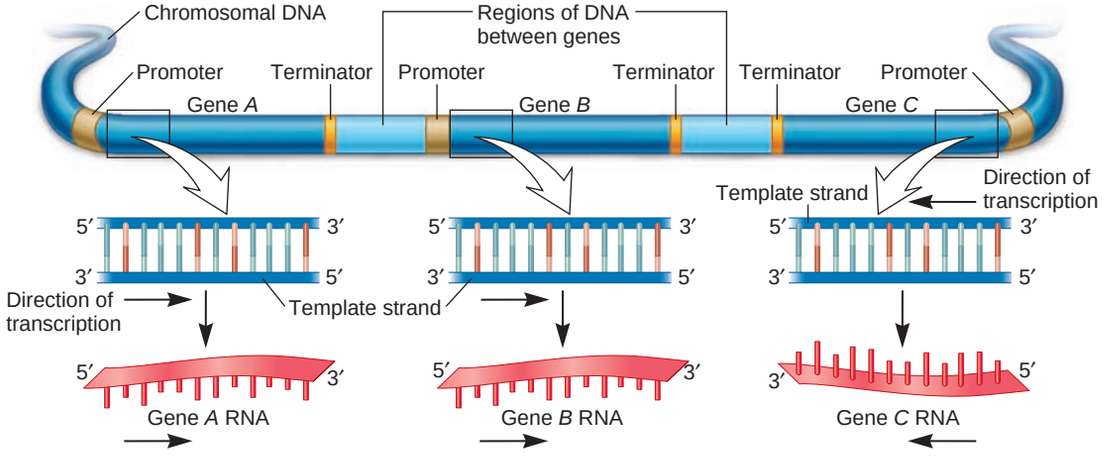
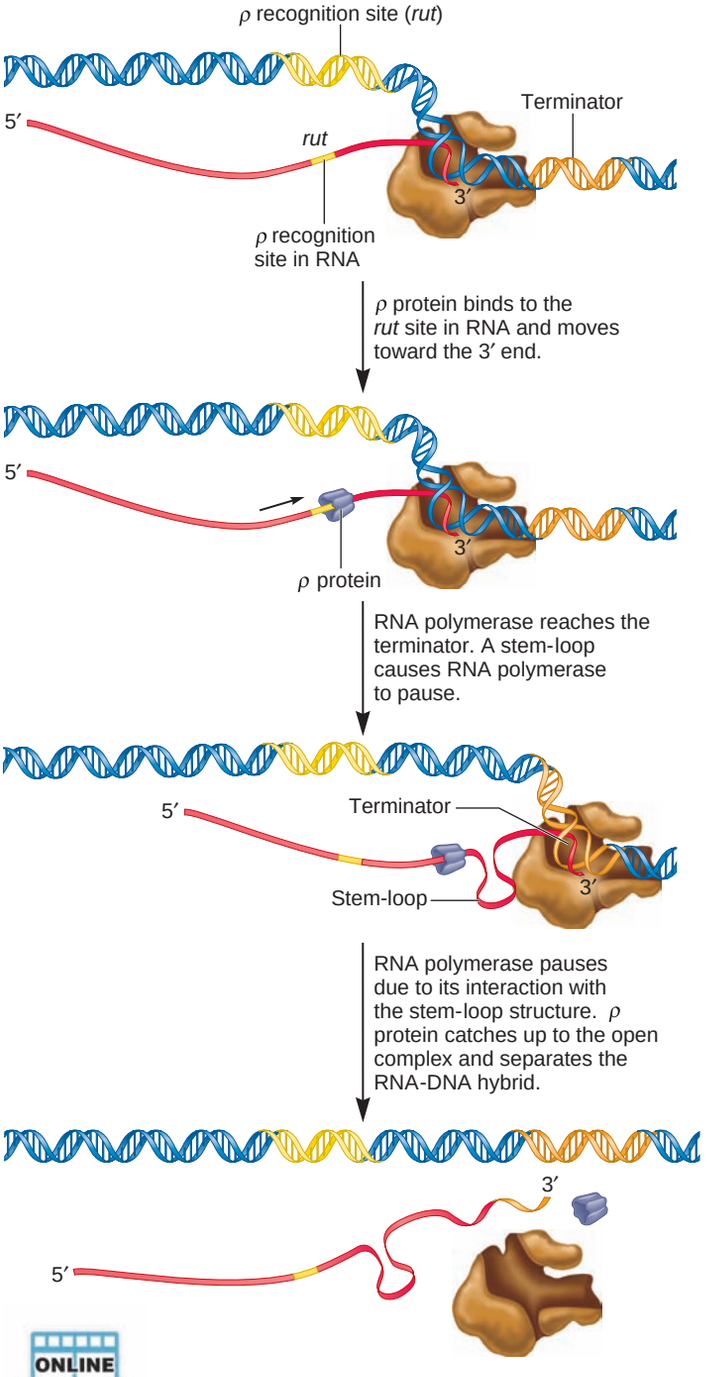


FIGURE 12.9 The transcription of three different genes found in the same chromosome. RNA polymerase synthesizes each RNA transcript in a 5′ to 3′ direction, sliding along a DNA template strand in a 3′ to 5′ direction. However, the use of the template strand varies from gene to gene. For example, genes *A* and *B* use the bottom strand, but gene *C* uses the top strand.

terminator, called the *rut* site for rho utilization site, acts as a recognition site for the binding of the ρ protein (**Figure 12.10**). How does ρ protein facilitate termination? The ρ protein functions as a helicase, an enzyme that can separate RNA-DNA hybrid regions. After the *rut* site is synthesized in the RNA, ρ protein binds to the RNA and moves in the direction of RNA polymerase. The second component of ρ-dependent termination is the site where termination actually takes place. At this terminator site, the DNA encodes an RNA sequence containing several GC base pairs that form a stem-loop structure. RNA synthesis terminates several nucleotides beyond this stem-loop. As discussed in Chapter 9, a stem-loop structure, also called a hairpin, can form due to complementary sequences within the RNA (refer back to Figure 9.22). This stem-loop forms almost immediately after the RNA sequence is synthesized and quickly binds to RNA polymerase. This binding results in a conformational change that causes RNA polymerase to pause in its synthesis of RNA. The pause allows ρ protein to catch up to the stem-loop, pass through it, and break the hydrogen bonds between the DNA and RNA within the open complex. When this occurs, the completed RNA strand is separated from the DNA along with RNA polymerase.

Let's now turn our attention to **ρ-independent termination,** a process that does not require the ρ protein. In this case, the terminator is composed of two adjacent nucleotide sequences that function within the RNA (**Figure 12.11**). One is a uracil-rich sequence located at the 3′ end of the RNA. The second sequence is adjacent to the uracil-rich sequence and promotes the formation of a stem-loop structure. As shown in Figure 12.11, the formation of the stem-loop causes RNA polymerase to pause in its synthesis of RNA. This pausing is stabilized by other proteins that bind to RNA polymerase. For example, a protein called NusA, which is bound to RNA polymerase, promotes pausing at stem-loop sequences. At the precise time RNA polymerase pauses, the



ONLINE ANIMATION

FIGURE 12.10 **ρ-Dependent termination.**

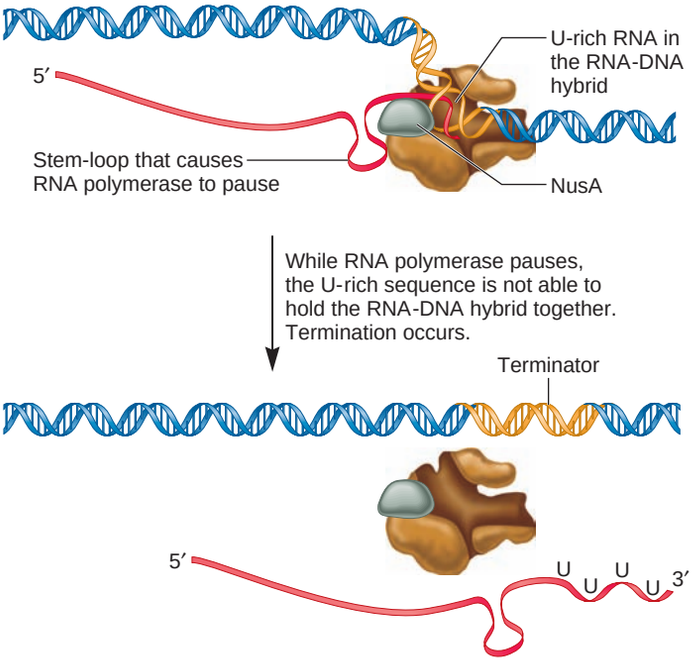


FIGURE 12.11 **ρ-Independent or intrinsic termination.** When RNA polymerase reaches the end of the gene, it transcribes a uracil-rich sequence. As this uracil-rich sequence is transcribed, a stem-loop forms just upstream from the open complex. The formation of this stem-loop causes RNA polymerase to pause in its synthesis of the transcript. This pausing is stabilized by NusA, which binds near the region where RNA exits the open complex. While it is pausing, the RNA in the RNA-DNA hybrid is a uracil-rich sequence. Because hydrogen bonds between U and A are relatively weak interactions, the transcript and RNA polymerase dissociate from the DNA.

uracil-rich sequence in the RNA transcript is bound to the DNA template strand. As previously mentioned, the hydrogen bonding of RNA to DNA keeps RNA polymerase clamped onto the DNA. However, the binding of this uracil-rich sequence to the DNA template strand is relatively weak, causing the RNA transcript to spontaneously dissociate from the DNA and cease further transcription. Because this process does not require a protein (the ρ protein) to physically remove the RNA transcript from the DNA, it is also referred to as **intrinsic termination.** In *E. coli*, about half of the genes show intrinsic termination, and the other half are terminated by ρ protein.

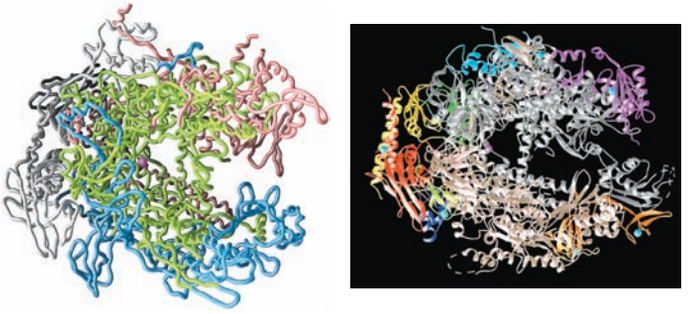
12.3 TRANSCRIPTION IN EUKARYOTES

Many of the basic features of gene transcription are very similar in bacterial and eukaryotic species. Much of our understanding of transcription has come from studies in *Saccharomyces cerevisiae* (baker's yeast) and other eukaryotic species, including mammals. In general, gene transcription in eukaryotes is more complex than that of their bacterial counterparts. Eukaryotic cells are larger and contain a variety of compartments known as organelles. This added level of cellular complexity dictates that eukaryotes contain many more genes encoding cellular proteins. In addition, most eukaryotic species are multicellular, being composed of many different cell types. Multicellularity adds the requirement that genes be transcribed in the correct type of cell and during the proper stage of development. Therefore, in any given species, the transcription of the thousands of different genes that an organism possesses requires appropriate timing and coordination. In this section, we will examine the basic features of gene transcription in eukaryotes. We will focus on the proteins that are needed to make an RNA transcript. In addition, an important factor that affects eukaryotic gene transcription is chromatin structure. Eukaryotic gene transcription requires changes in the positions and structures of nucleosomes. However, because these changes are important for regulating transcription, they are described in Chapter 15 rather than this chapter.

Eukaryotes Have Multiple RNA Polymerases That Are Structurally Similar to the Bacterial Enzyme

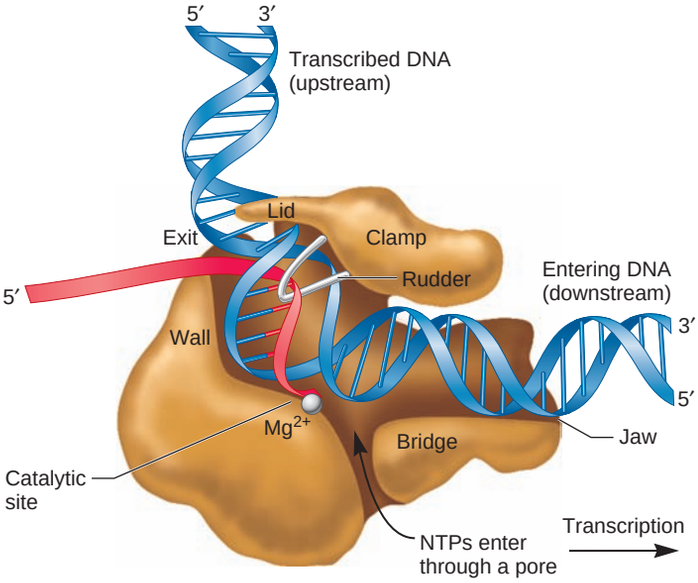
The genetic material within the nucleus of a eukaryotic cell is transcribed by three different RNA polymerase enzymes, designated RNA polymerase I, II, and III. What are the roles of these enzymes? Each of the three RNA polymerases transcribes different categories of genes. RNA polymerase I transcribes all of the genes that encode ribosomal RNA (rRNA) except for the 5S rRNA. RNA polymerase II plays a major role in cellular transcription because it transcribes all of the structural genes. It is responsible for the synthesis of all mRNA and also transcribes certain snRNA genes, which are needed for pre-mRNA splicing. RNA polymerase III transcribes all tRNA genes and the 5S rRNA gene.

All three RNA polymerases are structurally very similar and are composed of many subunits. They contain two large catalytic subunits similar to the β and β′ subunits of bacterial RNA polymerase. The structures of RNA polymerase from a few different species have been determined by X-ray crystallography. A remarkable similarity exists between the bacterial enzyme and its eukaryotic counterparts. **Figure 12.12a** compares the



(a) Structure of a bacterial RNA polymerase

Structure of a eukaryotic RNA polymerase II (yeast)



(b) Schematic structure of RNA polymerase

FIGURE 12.12 Structure and molecular function of RNA polymerase. **(a)** A comparison of the crystal structures of a bacterial RNA polymerase (left) to a eukaryotic RNA polymerase II (right). The bacterial enzyme is from *Thermus aquaticus*. The eukaryotic enzyme is from *Saccharomyces cerevisiae*. **(b)** A mechanism for transcription based on the crystal structure. In this diagram, the direction of transcription is from left to right. The double-stranded DNA enters the polymerase along a bridge surface that is between the jaw and clamp. At a region termed the wall, the RNA-DNA hybrid is forced to make a right-angle turn, which enables nucleotides to bind to the template strand. Mg^{2+} is located at the catalytic site. Nucleoside triphosphates (NTPs) enter the catalytic site via a pore region and bind to the template DNA. At the catalytic site, the nucleotides are covalently attached to the 3′ end of the RNA. As RNA polymerase slides down the template, a small region of the protein termed the rudder separates the RNA-DNA hybrid. The single-stranded RNA then exits under a small lid.

structures of a bacterial RNA polymerase with RNA polymerase II from yeast. As seen here, both enzymes have a very similar structure. Also, it is very exciting that this structure provides a way to envision how the transcription process works. As seen in **Figure 12.12b**, DNA enters the enzyme through the jaw and lies on a surface within RNA polymerase termed the bridge. The part of the enzyme called the clamp is thought to control the movement of the DNA through RNA polymerase. A wall in the enzyme forces the RNA-DNA hybrid to make a right-angle turn. This bend facilitates the ability of nucleotides to bind to the template strand. Mg^{2+} is located at the catalytic site, which is precisely at the 3′ end of the growing RNA strand. Nucleoside triphosphates (NTPs) enter the catalytic site via a pore region. The correct nucleotide binds to the template DNA and is covalently attached to the 3′ end. As RNA polymerase slides down the template, a rudder, which is about 9 bp away from the 3′ end of the RNA, forces the RNA-DNA hybrid apart. The single-stranded RNA then exits under a small lid.

Eukaryotic Structural Genes Have a Core Promoter and Regulatory Elements

In eukaryotes, the promoter sequence is more variable and often more complex than that found in bacteria. For structural genes, at least three features are found in most promoters: regulatory elements, a TATA box, and a transcriptional start site. **Figure 12.13** shows a common pattern of sequences found within the promoters of eukaryotic structural genes. The **core promoter** is a relatively short DNA sequence that is necessary for transcription to take place. It consists of a TATAAA sequence called the **TATA box** and the transcriptional start site, where transcription begins. The TATA box, which is usually about 25 bp upstream from a transcriptional start site, is important in determining the precise starting point for transcription. If it is missing from the core promoter, the transcription start site point becomes undefined, and transcription may start at a variety of different locations. The core promoter, by itself, produces a low level of transcription. This is termed **basal transcription.**

Regulatory elements are short DNA sequences that affect the ability of RNA polymerase to recognize the core promoter and begin the process of transcription. These elements are recognized by transcription factors—proteins that bind to regulatory elements and influence the rate of transcription. There are two categories of regulatory elements. Activating sequences, known as **enhancers,** are needed to stimulate transcription. In the absence of enhancer sequences, most eukaryotic genes have very low levels of basal transcription. Under certain conditions, it may also be necessary to prevent transcription of a given gene. This occurs via **silencers**—DNA sequences that are recognized by transcription factors that inhibit transcription. As seen in Figure 12.13, a common location for regulatory elements is the –50 to –100 region. However, the locations of regulatory elements vary considerably among different eukaryotic genes. These elements can be far away from the core promoter yet strongly influence the ability of RNA polymerase to initiate transcription.

DNA sequences such as the TATA box, enhancers, and silencers exert their effects only over a particular gene. They are called ***cis*-acting elements.** The term *cis* comes from chemistry nomenclature meaning "next to." *Cis*-acting elements, though possibly far away from the core promoter, are always found within the same chromosome as the genes they regulate. By comparison, the regulatory transcription factors that bind to such elements are called ***trans*-acting factors** (the term *trans* means "across from"). The transcription factors that control the expression of a gene are themselves encoded by genes; regulatory genes that encode transcription factors may be far away from the genes they control. When a gene encoding a *trans*-acting factor is expressed, the transcription factor protein that is made can diffuse throughout the cell and bind to its appropriate *cis*-acting element. Let's now turn our attention to the function of such proteins.

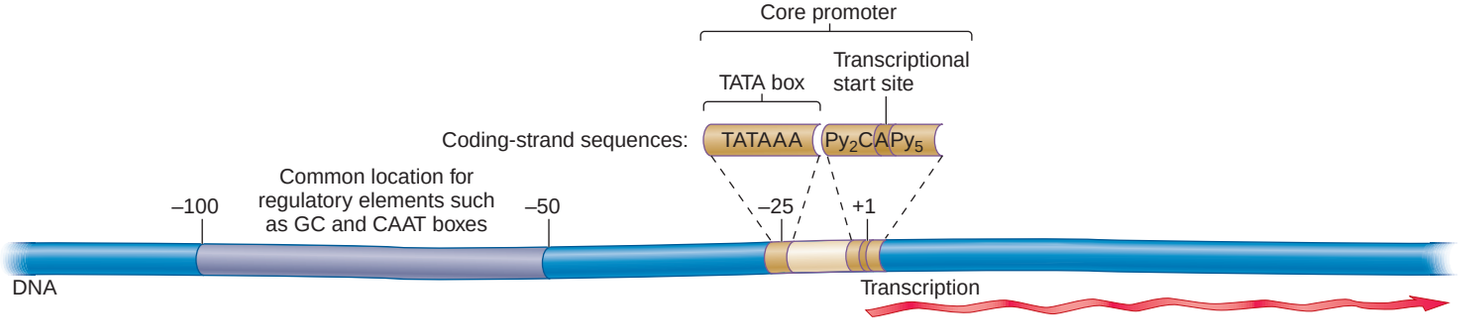


FIGURE 12.13 A common pattern found for the promoter of structural genes recognized by RNA polymerase II. The start site usually occurs at adenine; two pyrimidines (Py: cytosine or thymine) and a cytosine precede this adenine, and five pyrimidines (Py) follow it. A TATA box is approximately 25 bp upstream. However, the sequences that constitute eukaryotic promoters are quite diverse, and not all structural genes have a TATA box. Regulatory elements, such as GC or CAAT boxes, vary in their locations but are often found in the –50 to –100 region. The core promoters for RNA polymerase I and III are quite different. A single upstream regulatory element is involved in the binding of RNA polymerase I to its promoter, whereas two regulatory elements, called A and B boxes, facilitate the binding of RNA polymerase III.

Transcription of Eukaryotic Structural Genes Is Initiated When RNA Polymerase II and General Transcription Factors Bind to a Promoter Sequence

Thus far, we have considered the DNA sequences that play a role in the promoter region of eukaryotic structural genes. By studying transcription in a variety of eukaryotic species, researchers have discovered that three categories of proteins are needed for basal transcription at the core promoter: RNA polymerase II, general transcription factors, and mediator (Table 12.2).

Five different proteins called **general transcription factors (GTFs)** are always needed for RNA polymerase II to initiate transcription of structural genes. Figure 12.14 describes the assembly of GTFs and RNA polymerase II at the TATA box. As shown here, a series of interactions leads to the formation of the open complex. Transcription factor IID (TFIID) first binds to the TATA box and thereby plays a critical role in the recognition of the core promoter. TFIID is composed of several subunits, including TATA-binding protein (TBP), which directly binds to the TATA box, and several other proteins called TBP-associated factors (TAFs). After TFIID binds to the TATA box, it associates with TFIIB. TFIIB promotes the binding of RNA polymerase II and TFIIF to the core promoter. Lastly, TFIIE and TFIIH bind to the complex. This completes the assembly of proteins to form a closed complex, also known as a **preinitiation complex.**

TFIIH plays a major role in the formation of the open complex. TFIIH has several subunits that perform different functions.

TABLE **12.2**

Proteins Needed for Transcription via the Core Promoter of Eukaryotic Structural Genes

RNA polymerase II: The enzyme that catalyzes the linkage of ribonucleotides in the 5′ to 3′ direction, using DNA as a template. Essentially all eukaryotic RNA polymerase II proteins are composed of 12 subunits. The two largest subunits are structurally similar to the β and β′ subunits found in *E. coli* RNA polymerase.

General transcription factors:

- **TFIID:** Composed of TATA-binding protein (TBP) and other TBP-associated factors (TAFs). Recognizes the TATA box of eukaryotic structural gene promoters.
- **TFIIB:** Binds to TFIID and then enables RNA polymerase II to bind to the core promoter. Also promotes TFIIF binding.
- **TFIIF:** Binds to RNA polymerase II and plays a role in its ability to bind to TFIIB and the core promoter. Also plays a role in the ability of TFIIE and TFIIH to bind to RNA polymerase II.
- **TFIIE:** Plays a role in the formation or the maintenance (or both) of the open complex. It may exert its effects by facilitating the binding of TFIIH to RNA polymerase II and regulating the activity of TFIIH.
- **TFIIH:** A multisubunit protein that has multiple roles. First, certain subunits act as helicases and promote the formation of the open complex. Other subunits phosphorylate the carboxyl terminal domain (CTD) of RNA polymerase II, which releases its interaction with TFIIB, thereby allowing RNA polymerase II to proceed to the elongation phase.

Mediator: A multisubunit complex that mediates the effects of regulatory transcription factors on the function of RNA polymerase II. Though mediator typically has certain core subunits, many of its subunits vary, depending on the cell type and environmental conditions. The ability of mediator to affect the function of RNA polymerase II is thought to occur via the CTD of RNA polymerase II. Mediator can influence the ability of TFIIH to phosphorylate CTD, and subunits within mediator itself have the ability to phosphorylate CTD. Because CTD phosphorylation is needed to release RNA polymerase II from TFIIB, mediator plays a key role in the ability of RNA polymerase II to switch from the initiation to the elongation stage of transcription.

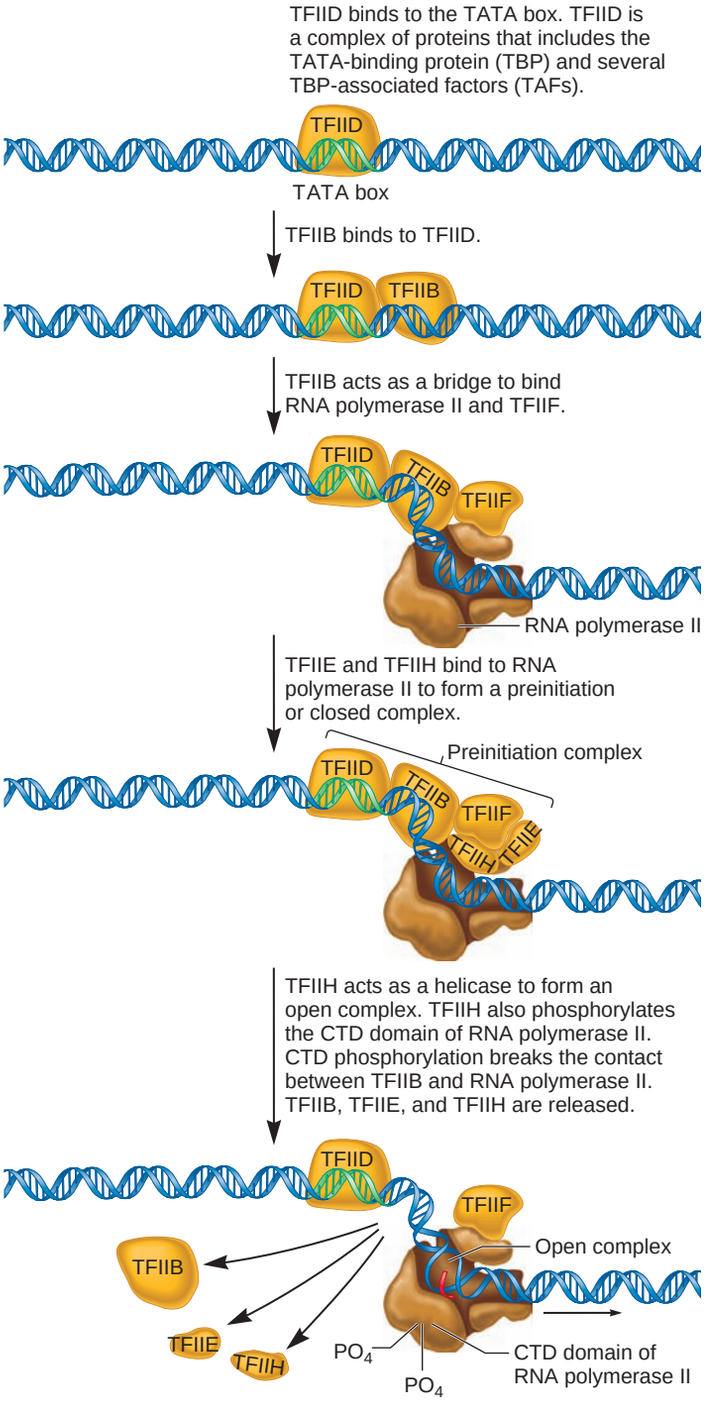


FIGURE 12.14 **Steps leading to the formation of the open complex.**

Certain subunits act as helicases, which break the hydrogen bonding between the double-stranded DNA and thereby promote the formation of the open complex. Another subunit hydrolyzes ATP and phosphorylates a domain in RNA polymerase II known as the carboxyl terminal domain (CTD). Phosphorylation of the CTD releases the contact between RNA polymerase II and TFIIB. Next, TFIIB, TFIIE, and TFIIH dissociate, and RNA polymerase II is free to proceed to the elongation stage of transcription.

In vitro, when researchers mix together TFIID, TFIIB, TFIIF, TFIIE, TFIIH, RNA polymerase II, and a DNA sequence containing a TATA box and transcriptional start site, the DNA is transcribed into RNA. Therefore, these components are referred to as the **basal transcription apparatus.** In a living cell, however, additional components regulate transcription and allow it to proceed at a reasonable rate.

In addition to GTFs and RNA polymerase II, another component required for transcription is a large protein complex termed mediator. This complex was discovered by Roger Kornberg and colleagues in 1990. In 2006, Kornberg was awarded the Nobel Prize in chemistry for his studies regarding the molecular basis of eukaryotic transcription. **Mediator** derives its name from the observation that it mediates interactions between RNA polymerase II and regulatory transcription factors that bind to enhancers or silencers. It serves as an interface between RNA polymerase II and many diverse regulatory signals. The subunit composition of mediator is quite complex and variable. The core subunits form an elliptically shaped complex that partially wraps around RNA polymerase II. Mediator itself may phosphorylate the CTD of RNA polymerase II, and it may regulate the ability of TFIIH to phosphorylate the CTD. Therefore, it can play a pivotal role in the switch between transcriptional initiation and elongation. The function of mediator during eukaryotic gene regulation is explored in greater detail in Chapter 15.

Transcriptional Termination of RNA Polymerase II Occurs After the 3′ End of the Transcript Is Cleaved Near the PolyA Signal Sequence

As discussed later in this chapter, eukaryotic pre-mRNAs are modified by cleavage near their 3′ end and the subsequent attachment of a string of adenine nucleotides (look ahead at Figure 12.24). This processing, which is called polyadenylation, requires a polyA signal sequence that directs the cleavage of the pre-mRNA. Transcription via RNA polymerase II typically terminates about 500 to 2000 nucleotides downstream from the polyA signal.

Figure 12.15 shows a simplified scheme for the transcriptional termination of RNA polymerase II. After RNA polymerase II has transcribed the polyA signal sequence, the RNA is cleaved just downstream from this sequence. This cleavage occurs before transcriptional termination. Two models have been proposed for transcriptional termination. According to the allosteric model, RNA polymerase II becomes destabilized after it has transcribed the polyA signal sequence, and it eventually dissociates from the DNA. This destabilization may be caused by the loss of proteins that function as elongation factors or by the binding of proteins that function as termination factors. A second model, called the torpedo model, suggests that RNA polymerase II is physically removed from the DNA. According to this model, the region of RNA that is downstream from the polyA signal sequence is cleaved by an exonuclease that degrades the transcript in the 5′ to 3′ direction. When the exonuclease catches up to RNA polymerase II, this causes RNA polymerase II to dissociate from the DNA.

Which of these two models is correct? Additional research is needed, but the results of studies over the past few years have provided evidence that the two models are not mutually exclusive. Therefore, both mechanisms may play a role in transcriptional termination.

12.4 RNA MODIFICATION

During the 1960s and 1970s, studies in bacteria established the physical structure of the gene. The analysis of bacterial genes showed that the sequence of DNA within the coding strand corresponds to the sequence of nucleotides in the mRNA, except that T is replaced with U. During translation, the sequence of codons in the mRNA is then read, providing the instructions for the correct amino acid sequence in a polypeptide. The one-to-one correspondence between the sequence of codons in the DNA coding strand and the amino acid sequence of the polypeptide has been termed the **colinearity** of gene expression.

The situation dramatically changed in the late 1970s, when the tools became available to study eukaryotic genes at the molecular level. The scientific community was astonished by the discovery that eukaryotic structural genes are not always colinear with their functional mRNAs. Instead, the coding sequences within many eukaryotic genes are separated by DNA sequences that are not translated into protein. The coding sequences are found within **exons,** which are regions that are contained within mature RNA. By comparison, the sequences that are found between the exons are called **intervening sequences,** or **introns.** During transcription, an RNA is made corresponding to the entire gene sequence. Subsequently, as it matures, the sequences in the RNA that correspond to the introns are removed and the exons are connected, or spliced, together. This process is called **RNA splicing.** Since the 1970s, research has revealed that splicing is a common genetic phenomenon in eukaryotic species. Splicing occurs occasionally in bacteria as well.

Aside from splicing, research has also shown that RNA transcripts can be modified in several other ways. **Table 12.3** describes the general types of RNA modifications. For example, rRNAs and tRNAs are synthesized as long transcripts that are processed into smaller functional pieces. In addition, most eukaryotic mRNAs have a cap attached to their 5′ end and a tail attached at their 3′ end. In this section, we will examine the molecular mechanisms that account for several types of RNA modifications and consider why they are functionally important.

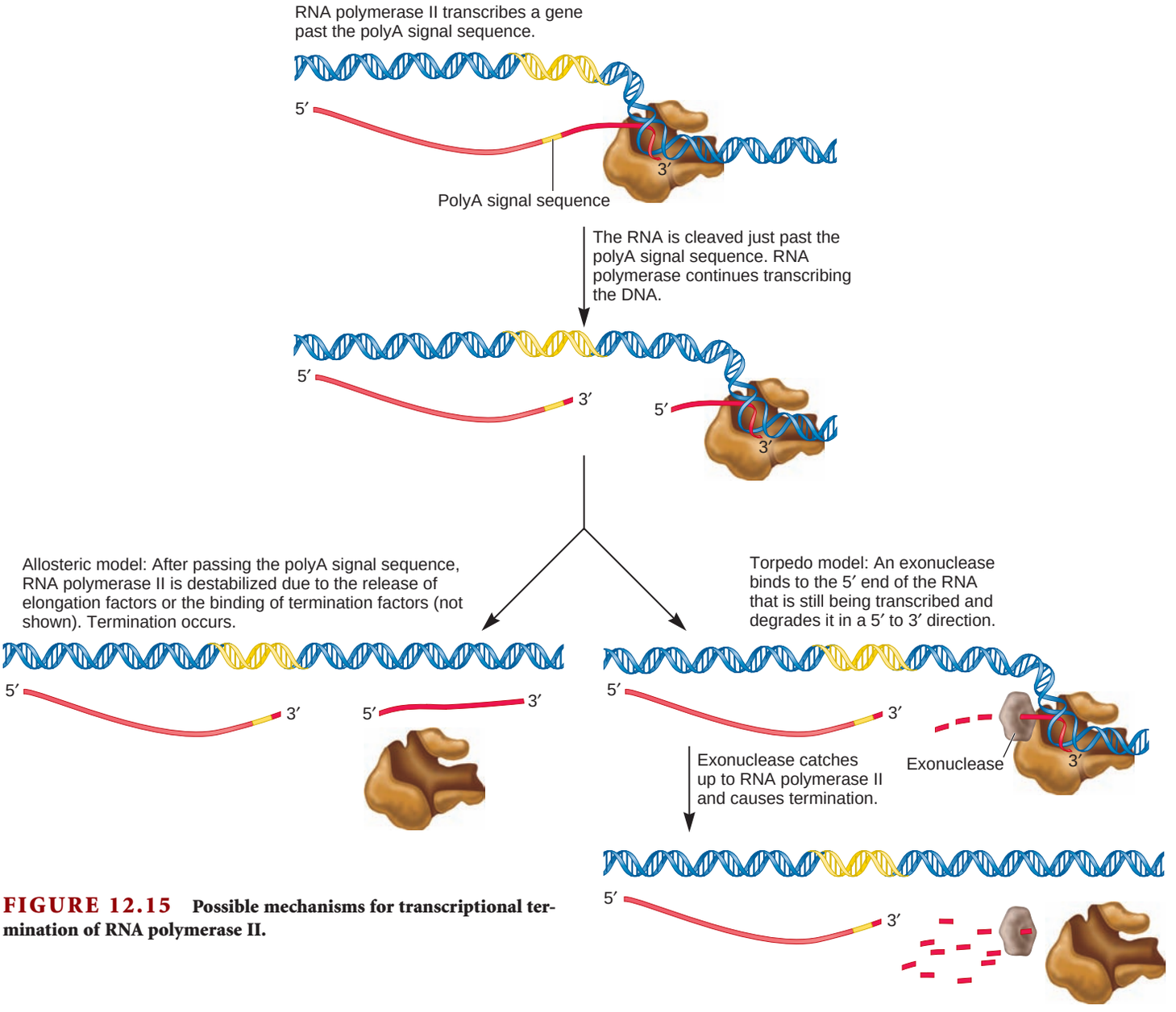


FIGURE 12.15 **Possible mechanisms for transcriptional termination of RNA polymerase II.**

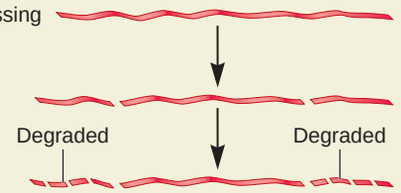
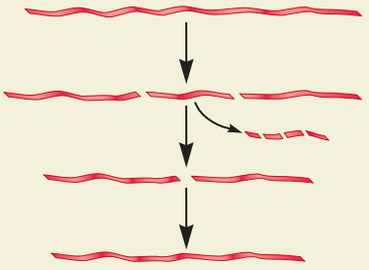
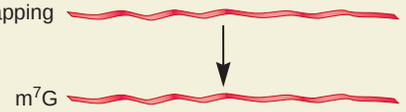
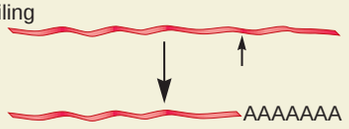
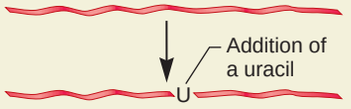
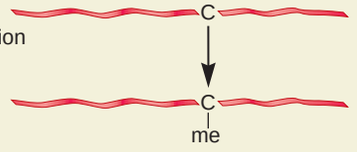
Some Large RNA Transcripts Are Cleaved into Smaller Functional Transcripts

For many nonstructural genes, the RNA transcript initially made during gene transcription is processed or cleaved into smaller pieces. As an example, **Figure 12.16** shows the processing of mammalian ribosomal RNA. The ribosomal RNA gene is transcribed by RNA polymerase I to make a long primary transcript, known as 45S rRNA. The term 45S refers to the sedimentation characteristics of this transcript in Svedberg units. Following the synthesis of the 45S rRNA, cleavage occurs at several points to produce three fragments, termed 18S, 5.8S, and 28S rRNA. These are functional rRNA molecules that play a key role in forming the structure of the ribosome. In eukaryotes, the cleavage of 45S rRNA into smaller rRNAs and the assembly of ribosomal subunits occur in a structure within the cell nucleus known as the **nucleolus.**

The production of tRNA molecules requires processing via exonucleases and endonucleases. An **exonuclease** is a type of enzyme that cleaves a covalent bond between two nucleotides at one end of a strand. Starting at one end, an exonuclease can digest a strand, one nucleotide at a time. Some exonucleases can begin this digestion only from the 3′ end, traveling in the 3′ to 5′ direction, whereas others can begin only at the 5′ end and digest in the 5′ to 3′ direction. By comparison, an **endonuclease** can cleave the bond between two adjacent nucleotides within a strand.

Like ribosomal RNA, tRNAs are synthesized as large precursor tRNAs that must be cleaved to produce mature, functional tRNAs that bind to amino acids. This processing has

TABLE 12.3
Modifications That May Occur to RNAs

Modification	Description	Occurrence
Processing 	The cleavage of a large RNA transcript into smaller pieces. One or more of the smaller pieces becomes a functional RNA molecule. Processing occurs for rRNA and tRNA transcripts.	Occurs in both prokaryotes and eukaryotes.
Splicing 	Splicing involves both cleavage and joining of RNA molecules. The RNA is cleaved at two sites, which allows an internal segment of RNA, known as an intron, to be removed. After the intron is removed, the two ends of the RNA molecules are joined together.	Splicing is common among eukaryotic pre-mRNAs, and it also occurs occasionally in rRNAs, tRNAs, and a few bacterial RNAs.
5′ capping 	The attachment of a 7-methylguanosine cap (m^7G) to the 5′ end of mRNA. The cap plays a role in the splicing of introns, the exit of mRNA from the nucleus, and the binding of mRNA to the ribosome.	Capping occurs on eukaryotic mRNAs.
3′ polyA tailing 	The attachment of a string of adenine-containing nucleotides to the 3′ end of mRNA at a site where the mRNA is cleaved (see upward arrow). It is important for RNA stability and translation in eukaryotes.	Occurs on eukaryotic mRNAs and occasionally occurs on bacterial RNAs.
RNA editing 	The change of the base sequence of an RNA after it has been transcribed.	Occurs occasionally among eukaryotic RNAs.
Base modification 	The covalent modification of a base within an RNA molecule.	Base modification commonly occurs in tRNA molecules found in both prokaryotes and eukaryotes. C—me indicates methylation of cytosine.

been studied extensively in *E. coli*. **Figure 12.17** shows the processing of a precursor tRNA, which involves the action of two endonucleases and one exonuclease. The precursor tRNA is recognized by RNaseP, which is an endonuclease that cuts the precursor tRNA. The action of RNaseP produces the correct 5′ end of the mature tRNA. A different endonuclease cleaves the precursor tRNA to remove a 170-nucleotide segment from the 3′ end. Next, an exonuclease, called RNaseD, binds to the 3′ end and digests the RNA in the 3′ to 5′ direction. When it reaches an ACC sequence, the exonuclease stops digesting the precursor tRNA molecule. Therefore, all tRNAs in *E. coli* have an ACC sequence at their 3′ ends. Finally, certain bases in tRNA molecules may be covalently modified to alter their structure. The functional importance of modified bases in tRNAs is discussed in Chapter 13.

As researchers studied tRNA processing, they discovered certain features that were very unusual and exciting, changing the way biologists view the actions of catalysts. RNaseP has been found to be a catalyst that contains both RNA and protein subunits. In 1983, Sidney Altman and colleagues made the surprising

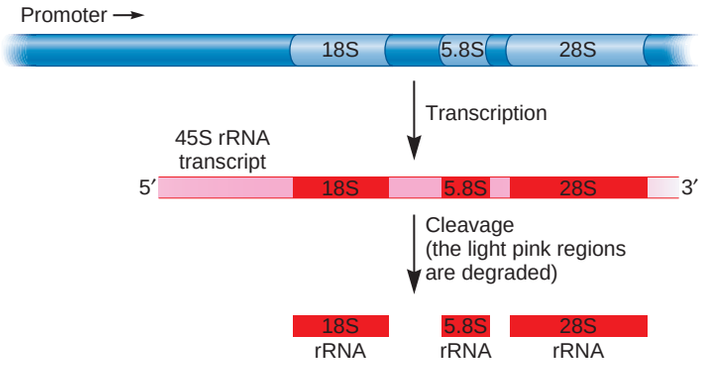


FIGURE 12.16 The processing of ribosomal RNA in eukaryotes. The large ribosomal RNA gene is transcribed into a long 45S rRNA primary transcript. This transcript is cleaved to produce 18S, 5.8S, and 28S rRNA molecules, which become associated with protein subunits in the ribosome. This processing occurs within the nucleolus of the cell.

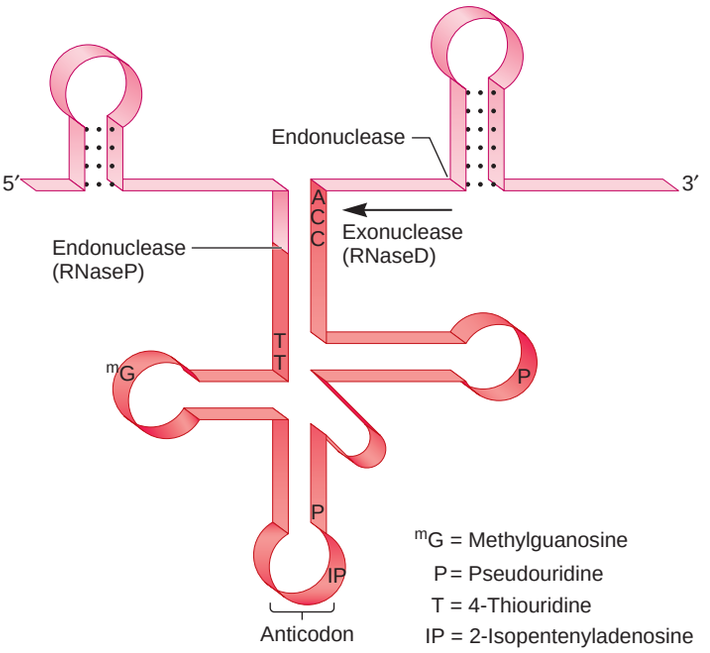


FIGURE 12.17 The processing of a precursor tRNA molecule in *E. coli*. RNaseP is an endonuclease that makes a cut that creates the 5′ end of the mature tRNA. To produce the 3′ end of mature tRNA, an endonuclease makes a cut, and then the exonuclease RNaseD removes nine nucleotides at the 3′ end. In addition to these cleavage steps, several bases within the tRNA molecule are modified to other bases as schematically indicated. A similar type of precursor tRNA processing occurs in eukaryotes.

discovery that the RNA portion of RNaseP, not the protein subunit, contains the catalytic ability to cleave the precursor tRNA. RNaseP is an example of a **ribozyme,** an RNA molecule with catalytic activity. Prior to the study of RNaseP and the identification of self-splicing RNAs (discussed later), biochemists had staunchly believed that only protein molecules could function as biological catalysts.

EXPERIMENT 12A

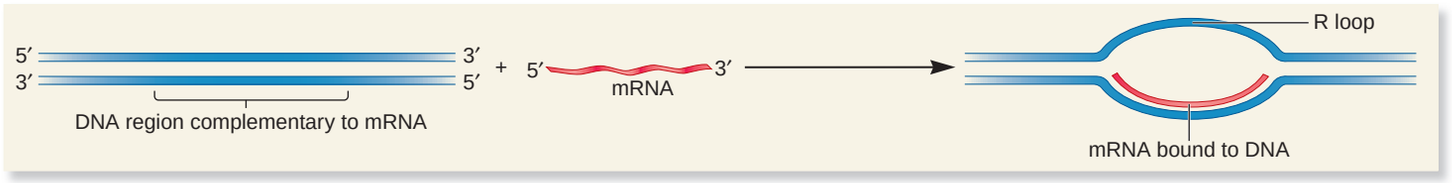
Introns Were Experimentally Identified via Microscopy

Although the discovery of ribozymes was very surprising, the observation that tRNA and rRNA transcripts are processed to a smaller form did not seem unusual to geneticists and biochemists, because the cleavage of RNA was similar to the cleavage that can occur for other macromolecules such as DNA and proteins. In sharp contrast, when splicing was detected in the 1970s, it was a novel concept. Splicing involves cleavage at two sites. An intron is removed, and—in a unique step—the remaining fragments are hooked back together again.

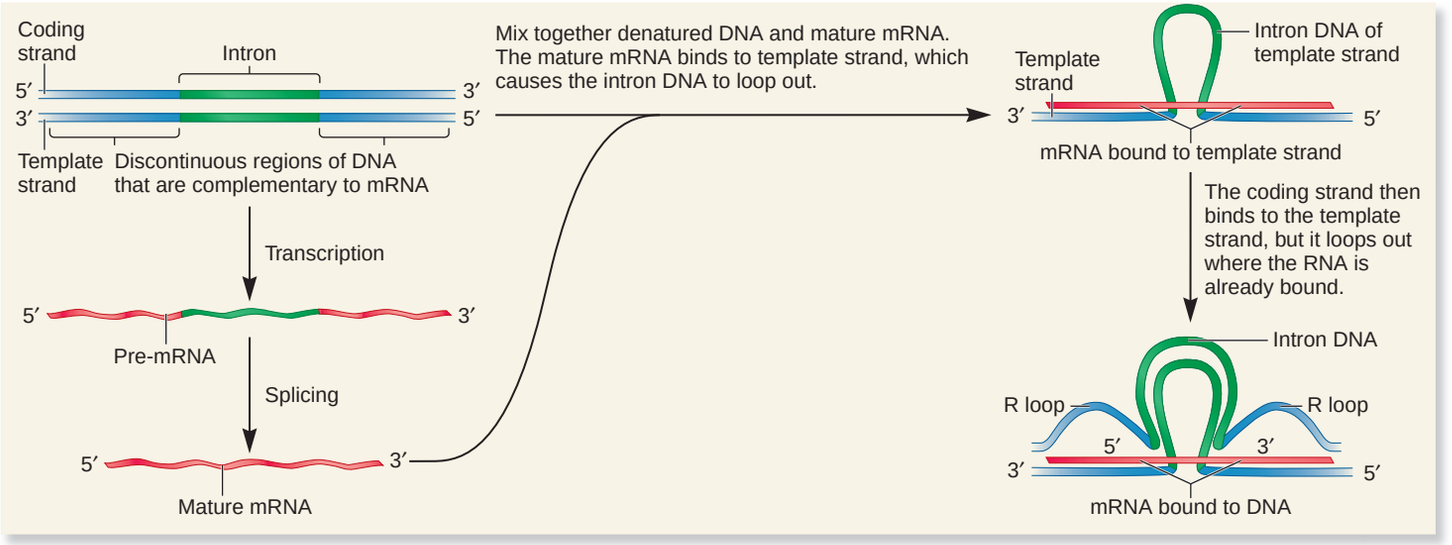
Eukaryotic introns were first detected by comparing the base sequence of viral genes and their mRNA transcripts during viral infection of mammalian cells by adenovirus. This research was carried out in 1977 by two groups headed by Philip Sharp and Richard Roberts. This pioneering observation led to the next question: Are introns a peculiar phenomenon that occurs only in viral genes, or are they found in eukaryotic genes as well?

In the late 1970s, several research groups, including those of Pierre Chambon, Bert O'Malley, and Philip Leder, investigated the presence of introns in eukaryotic structural genes. The experiments of Leder used electron microscopy to identify introns in the β-globin gene. β globin is a polypeptide that is a subunit of hemoglobin, the protein that carries oxygen in red blood cells. To detect introns within the gene, Leder considered the possible effects of mRNA binding to a gene. **Figure 12.18a** considers the situation in which a gene does not contain an intron. In this experiment, a segment of double-stranded chromosomal DNA containing a gene was first denatured and mixed with mature mRNA encoded by that gene. Because the mRNA is complementary to the template strand of the DNA, the template strand and the mRNA bind to each other to form a hybrid molecule. This event is called **hybridization.** Later, when the DNA is allowed to renature, the binding of the mRNA to the template strand of DNA prevents the two strands of DNA from forming a double helix. In the absence of any introns, the single-stranded DNA forms a loop. Because the RNA has displaced one of the DNA strands, this structure is known as an RNA displacement loop, or **R loop,** as shown in Figure 12.18a.

In contrast, Leder and colleagues realized that a different type of R loop structure would form if a gene contained an intron (**Figure 12.18b**). When mRNA is hybridized to a region of a gene containing one intron and then the other DNA strand



(a) No introns in the DNA



(b) One intron in the DNA. The intron in the pre-mRNA is spliced out.

FIGURE 12.18 **Hybridization of mRNA to double-stranded DNA.** In this experiment, the DNA is denatured and then allowed to renature under conditions that favor an RNA-DNA hybrid. **(a)** If the DNA does not contain an intron, the binding of the mRNA to the template strand of DNA prevents the two strands of DNA from forming a double helix. The single-stranded region of DNA will form an R loop. **(b)** When mRNA hybridizes to a gene containing one intron, two single-stranded R loops will form that are separated by a double-stranded DNA region. The intervening double-stranded region occurs because an intron has been spliced out of the mRNA and the mRNA cannot hybridize to this segment of the gene.

is allowed to renature, two single-stranded R loops form that are separated by a double-stranded DNA region. The intervening double-stranded region occurs because an intron has been spliced out of the mature mRNA, so the mRNA cannot hybridize to this segment of the gene.

As shown in steps 1 through 4 of **Figure 12.19**, this hybridization approach was used to identify introns within the *β*-globin gene. Following hybridization, the samples were placed on a microscopy grid, shadowed with heavy metal, and then observed by electron microscopy.

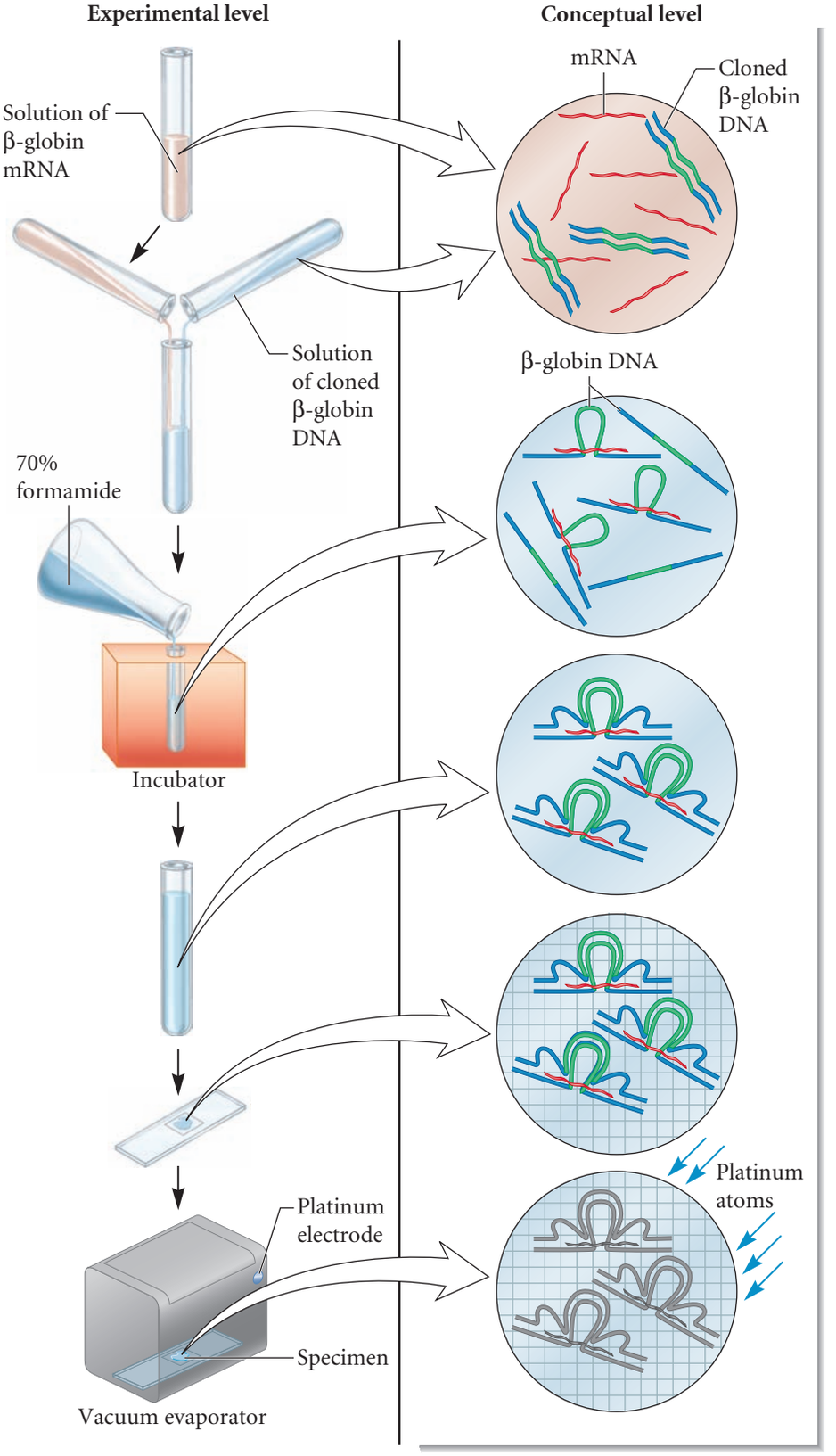
THE HYPOTHESIS

The mouse β-globin gene contains one or more introns.

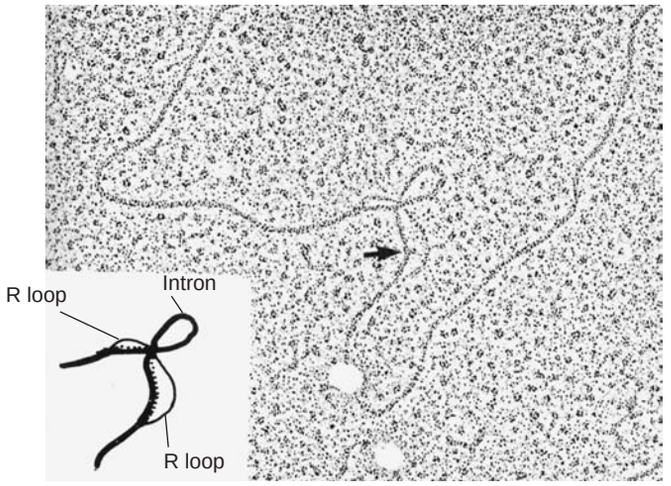
TESTING THE HYPOTHESIS — FIGURE 12.19 RNA hybridization to the β-globin gene reveals an intron.

Starting material: A cloned fragment of chromosomal DNA that contains the mouse β-globin gene.

1. Isolate mature mRNA for the mouse β-globin gene. Note: Globin mRNA is abundant in reticulocytes, which are immature red blood cells.
2. Mix together the β-globin mRNA and cloned DNA of the β-globin gene.
3. Separate the double-stranded DNA and allow the mRNA to hybridize. This is done using 70% formamide, at 52°C, for 16 hours.
4. Dilute the sample to decrease the formamide concentration. This allows the DNA to re-form a double-stranded structure. Note: The DNA cannot form a double-stranded structure in regions where the mRNA has already hybridized.
5. Spread the sample onto a microscopy grid.
6. Stain with uranyl acetate and shadow with heavy metal. Note: The technique of electron microscopy is described in the Appendix.
7. View the sample under the electron microscope.



THE DATA



Data from: Tilghman, S.M., Tiemeier, D.C., Seidman, J.G., Peterlin, B.M., Sullivan, M., Maizel, J.V., and Leder, P. (1978) Intervening sequence of DNA identified in the structural portion of a mouse beta-globin gene. *Proc. Natl. Acad. Sci. USA* 75:725–729.

INTERPRETING THE DATA

As seen in the electron micrograph, the β-globin mRNA hybridized to the DNA of the β-globin gene, which resulted in the formation of two R loops separated by a double-stranded DNA region. These data were consistent with the idea that the DNA of the β-globin gene contains an intron. Similar results were obtained by Chambon and O'Malley for other structural genes. Since these initial discoveries, introns have been found in many eukaryotic genes. The prevalence and biological significance of introns are discussed later in this chapter.

Since the late 1970s, DNA sequencing methods have permitted an easier and more precise way of detecting introns. Researchers can clone a fragment of chromosomal DNA that contains a particular gene. This is called a **genomic clone.** In addition, mRNA can be used as a starting material to make a copy of DNA known as **complementary DNA (cDNA).** The cDNA does not contain introns, because the introns have been previously removed during RNA splicing. In contrast, if a gene contains introns, a genomic clone for a eukaryotic gene also contains introns. Therefore, a comparison of the DNA sequences from genomic and cDNA clones can provide direct evidence that a particular gene contains introns. Compared with genomic DNA, the cDNA is missing base sequences that were removed during splicing.

A self-help quiz involving this experiment can be found at www.mhhe.com/brookergenetics4e.

Different Splicing Mechanisms Can Remove Introns

Since the original discovery of introns, the investigations of many research groups have shown that most structural genes in complex eukaryotes contain one or more introns. Less commonly, introns can occur within tRNA and rRNA genes. At the molecular level, different RNA splicing mechanisms have been identified. In the three examples shown in **Figure 12.20**, splicing leads to the removal of the intron RNA and the covalent connection of the exon RNA.

The splicing among **group I** and **group II introns** occurs via **self-splicing**—splicing that does not require the aid of other catalysts. Instead, the RNA functions as its own ribozyme. Group I and II differ in the ways that the introns are removed and the exons are connected. Group I introns that occur within the rRNA of *Tetrahymena* (a protozoan) have been studied extensively by Thomas Cech and colleagues. In this organism, the splicing process involves the binding of a single guanosine to a guanosine-binding site within the intron (Figure 12.20a). This guanosine breaks the bond between the first exon and the intron and becomes attached to the 5′ end of the intron. The 3′—OH group of exon 1 then breaks the bond next to a different nucleotide (in this example, a G) that lies at the boundary between the end of the intron and exon 2; exon 1 forms a phosphoester bond with the 5′ end of exon 2. The intron RNA is subsequently degraded. In this example, the RNA molecule functions as its own ribozyme, because it splices itself without the aid of a catalytic protein.

In group II introns, a similar splicing mechanism occurs, except the 2′—OH group on ribose found in an adenine nucleotide already within the intron strand begins the catalytic process (Figure 12.20b). Experimentally, group I and II self-splicing can occur in vitro without the addition of any proteins. However, in a living cell, proteins known as **maturases** often enhance the rate of splicing of group I and II introns.

In eukaryotes, the transcription of structural genes produces a long transcript known as **pre-mRNA,** which is located within the nucleus. This pre-mRNA is usually altered by splicing and other modifications before it exits the nucleus. Unlike group I and II introns, which may undergo self-splicing, pre-mRNA splicing requires the aid of a multicomponent structure known as the **spliceosome.** As discussed shortly, this is needed to recognize the intron boundaries and to properly remove it.

Table 12.4 describes the occurrence of introns among the genes of different species. The biological significance of group I and II introns is not understood. By comparison, pre-mRNA splicing is a widespread phenomenon among complex eukaryotes. In mammals and flowering plants, most structural genes have at least one intron that can be located anywhere within the gene. For example, an average human gene has about eight introns. In some cases, a single gene can have many introns. As an extreme example, the human dystrophin gene, which, when mutated, causes Duchenne muscular dystrophy, has 79 exons punctuated by 78 introns.

Pre-mRNA Splicing Occurs by the Action of a Spliceosome

As noted previously, the spliceosome is a large complex that splices pre-mRNA in eukaryotes. It is composed of several

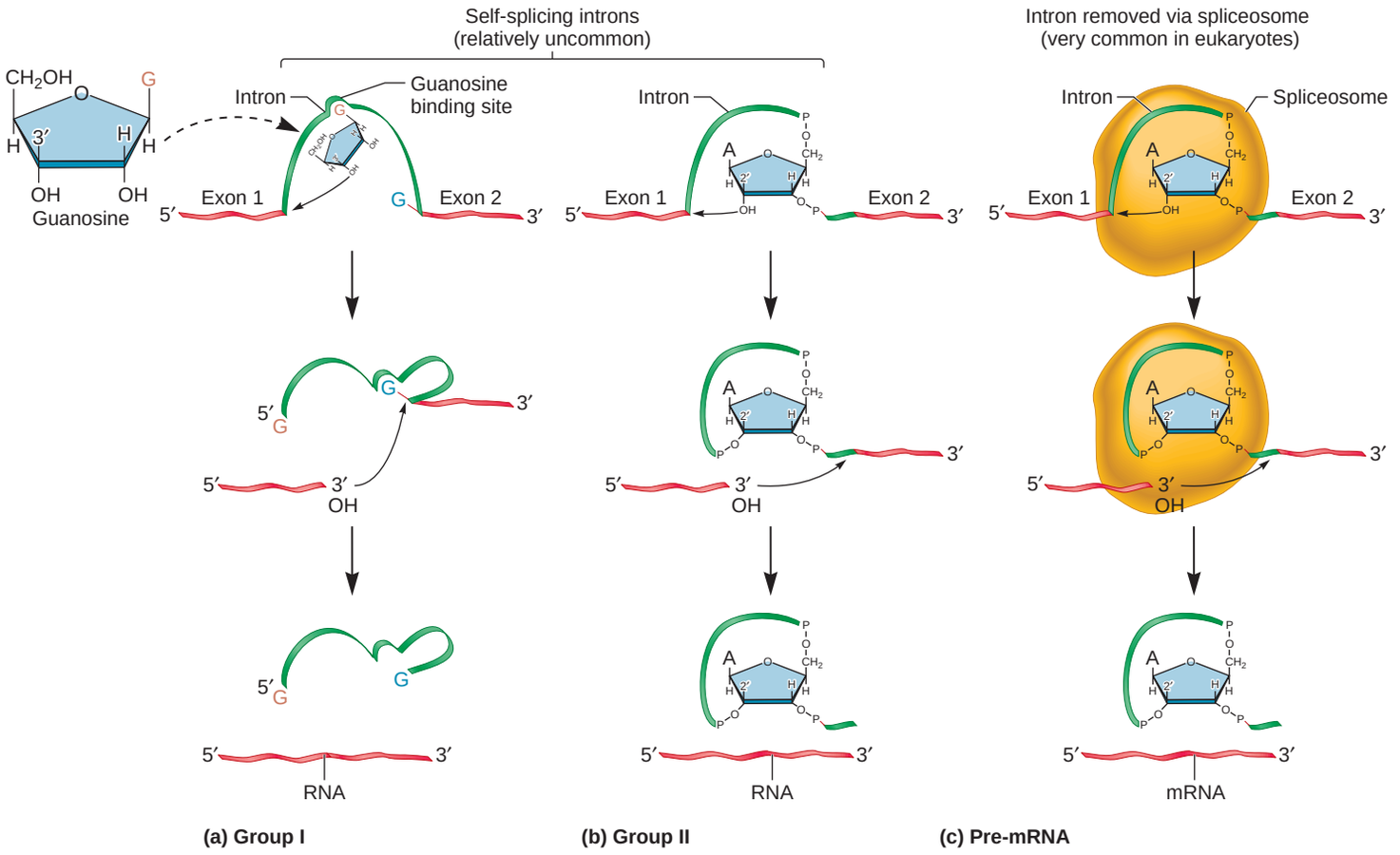


FIGURE 12.20 **Mechanisms of RNA splicing.** Group I and II introns are self-splicing. **(a)** The splicing of group I introns involves the binding of a free guanosine to a site within the intron, leading to the cleavage of RNA at the 3′ end of exon 1. The bond between a different guanine nucleotide (in the intron strand) and the 5′ end of exon 2 is cleaved. The 3′ end of exon 1 then forms a covalent bond with the 5′ end of exon 2. **(b)** In group II introns, a similar splicing mechanism occurs, except that the 2′—OH group on an adenine nucleotide (already within the intron) begins the catalytic process. **(c)** Pre-mRNA splicing requires the aid of a multicomponent structure known as the spliceosome.

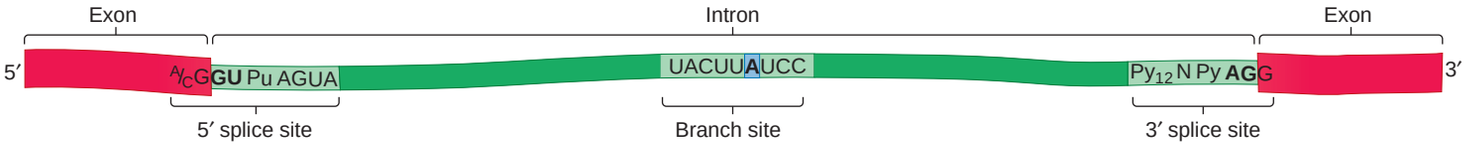
TABLE **12.4**

Occurrence of Introns

Type of Intron	Mechanism of Removal	Occurrence
Group I	Self-splicing	Found in rRNA genes within the nucleus of *Tetrahymena* and other simple eukaryotes. Found in a few structural, tRNA, and rRNA genes within the mitochondrial DNA (fungi and plants) and in chloroplast DNA. Found very rarely in tRNA genes within bacteria.
Group II	Self-splicing	Found in a few structural, tRNA, and rRNA genes within the mitochondrial DNA (fungi and plants) and in chloroplast DNA. Also found rarely in bacterial genes.
Pre-mRNA	Spliceosome	Very commonly found in structural genes within the nucleus of eukaryotes.

subunits known as **snRNPs** (pronounced "snurps"). Each snRNP contains small nuclear RNA and a set of proteins. During splicing, the subunits of a spliceosome carry out several functions. First, spliceosome subunits bind to an intron sequence and precisely recognize the intron-exon boundaries. In addition, the spliceosome must hold the pre-mRNA in the correct configuration to ensure the splicing together of the exons. And finally, the spliceosome catalyzes the chemical reactions that cause the introns to be removed and the exons to be covalently linked.

Intron RNA is defined by particular sequences within the intron and at the intron-exon boundaries. The consensus sequences for the splicing of mammalian pre-mRNA are shown in **Figure 12.21**. These sequences serve as recognition sites for the binding of the spliceosome. The bases most commonly found at these sites—those that are highly conserved evolutionarily—are shown in bold. The 5′ and 3′ splice sites occur at the ends of the intron, whereas the branch site is somewhere in the middle. These sites are recognized by components of the spliceosome.



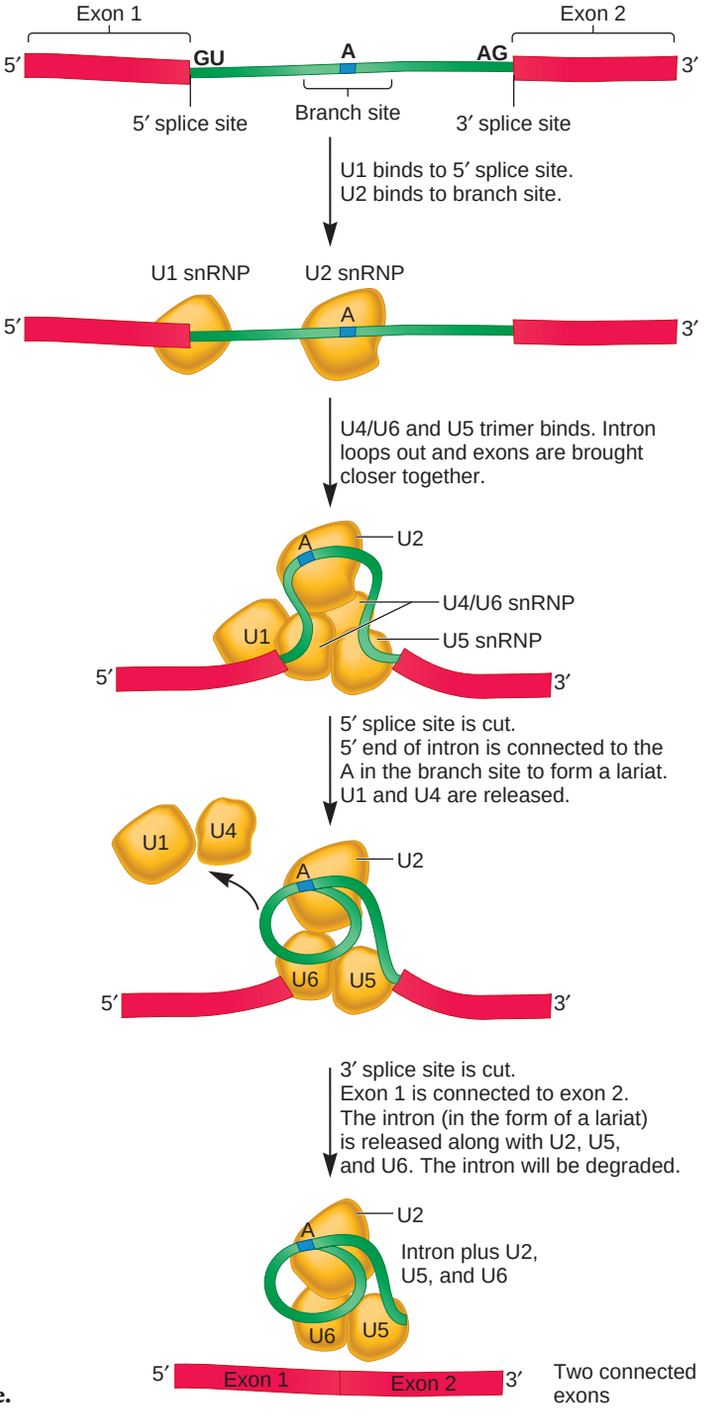
ONLINE ANIMATION

FIGURE 12.21 **Consensus sequences for pre-mRNA splicing in complex eukaryotes.** Consensus sequences exist at the intron-exon boundaries and at a branch site found within the intron itself. The adenine nucleotide shown in blue in this figure corresponds to the adenine nucleotide at the branch site in Figure 12.22. The nucleotides shown in bold are highly conserved. Designations: A/C = A or C, Pu = purine, Py = pyrimidine, N = any of the four bases.

The molecular mechanism of pre-mRNA splicing is depicted in **Figure 12.22**. The snRNP designated U1 binds to the 5′ splice site, and U2 binds to the branch site. This is followed by the binding of a trimer of three snRNPs: a U4/U6 dimer plus U5. The intron loops outward, and the two exons are brought closer together. The 5′ splice site is then cut, and the 5′ end of the intron becomes covalently attached to the 2′—OH group of a specific adenine nucleotide in the branch site. U1 and U4 are released. In the final step, the 3′ splice site is cut, and then the exons are covalently attached to each other. The three snRNPs—U2, U5, and U6—remain attached to the intron, which is in a lariat configuration. Eventually, the intron is degraded, and the snRNPs are used again to splice other pre-mRNAs.

The chemical reactions that occur during pre-mRNA splicing are not completely understood. Though further research is needed, evidence is accumulating that certain snRNA molecules within the spliceosome may play a catalytic role in the removal of introns and the connection of exons. In other words, snRNAs may function as ribozymes that cleave the RNA at the exon-intron boundaries and connect the remaining exons. Researchers have speculated that RNA molecules within U2 and U6 may have this catalytic function.

Before ending our discussion of pre-mRNA splicing, let's consider why it may be an advantage for a species to have genes that contain introns. One benefit is a phenomenon called **alternative splicing.** When a pre-mRNA has multiple introns, variation may occur in the pattern of splicing, so the resulting mRNAs contain alternative combinations of exons. The variation in splicing may happen in different cell types or during different stages of development. The biological advantage of alternative splicing is that two or more different proteins can be derived from a single gene. This allows an organism to carry fewer genes in its genome. The molecular mechanism of alternative splicing is examined in Chapter 15. It involves the actions of proteins (not shown in Figure 12.22) that influence whether or not U1 and U2 can begin the splicing process.



ONLINE ANIMATION

FIGURE 12.22 **Splicing of pre-mRNA via a spliceosome.**

The Ends of Eukaryotic Pre-mRNAs Have a 5′ Cap and a 3′ Tail

In addition to splicing, pre-mRNAs in eukaryotes are also subjected to modifications at their 5′ and 3′ ends. At their 5′ end, most mature mRNAs have a 7-methylguanosine covalently attached—an event known as **capping.** Capping occurs while the pre-mRNA is being made by RNA polymerase II, usually when the transcript is only 20 to 25 nucleotides in length. As shown in **Figure 12.23**, it is a three-step process. The nucleotide at the 5′ end of the transcript has three phosphate groups. First, an enzyme called RNA 5′-triphosphatase removes one of the phosphates, and then a second enzyme, guanylyltransferase, uses guanosine triphosphate (GTP) to attach a guanosine monophosphate (GMP) to the 5′ end. Finally, a methyltransferase attaches a methyl group to the guanine base.

What are the functions of the 7-methylguanosine cap? The cap structure is recognized by cap-binding proteins, which perform various roles. For example, cap-binding proteins are required for the proper exit of most mRNAs from the nucleus. Also, the cap structure is recognized by initiation factors that are needed during the early stages of translation. Finally, the cap structure may be important in the efficient splicing of introns, particularly the first intron located nearest the 5′ end.

Let's now turn our attention to the 3′ end of the RNA molecule. Most mature mRNAs have a string of adenine nucleotides, referred to as a **polyA tail,** which is important for mRNA stability and in the synthesis of polypeptides. The polyA tail is not encoded in the gene sequence. Instead, it is added enzymatically after the pre-mRNA has been completely transcribed—a process termed **polyadenylation.**

The steps required to synthesize a polyA tail are shown in **Figure 12.24**. To acquire a polyA tail, the pre-mRNA contains a polyadenylation signal sequence near its 3′ end. In mammals, the consensus sequence is AAUAAA. This sequence is downstream (toward the 3′ end) from the stop codon in the pre-mRNA. An endonuclease recognizes the signal sequence and cleaves the pre-mRNA at a location that is about 20 nucleotides beyond the 3′ end of the AAUAAA sequence. The fragment beyond the 3′ cut is degraded. Next, an enzyme known as polyA-polymerase attaches many adenine-containing nucleotides. The length of the polyA tail varies among different mRNAs, from a few dozen to several hundred adenine nucleotides. As discussed in Chapter 15, a long polyA tail increases the stability of mRNA and plays a role during translation.

The Nucleotide Sequence of RNA Can Be Modified by RNA Editing

The term **RNA editing** refers to a change in the nucleotide sequence of an RNA molecule that involves additions or deletions of particular bases, or a conversion of one type of base to another, such as a cytosine to a uracil. In the case of mRNAs, editing can have various effects, such as generating start codons, generating stop codons, and changing the coding sequence for a polypeptide.

The phenomenon of RNA editing was first discovered in trypanosomes, the protists that cause sleeping sickness. As with the

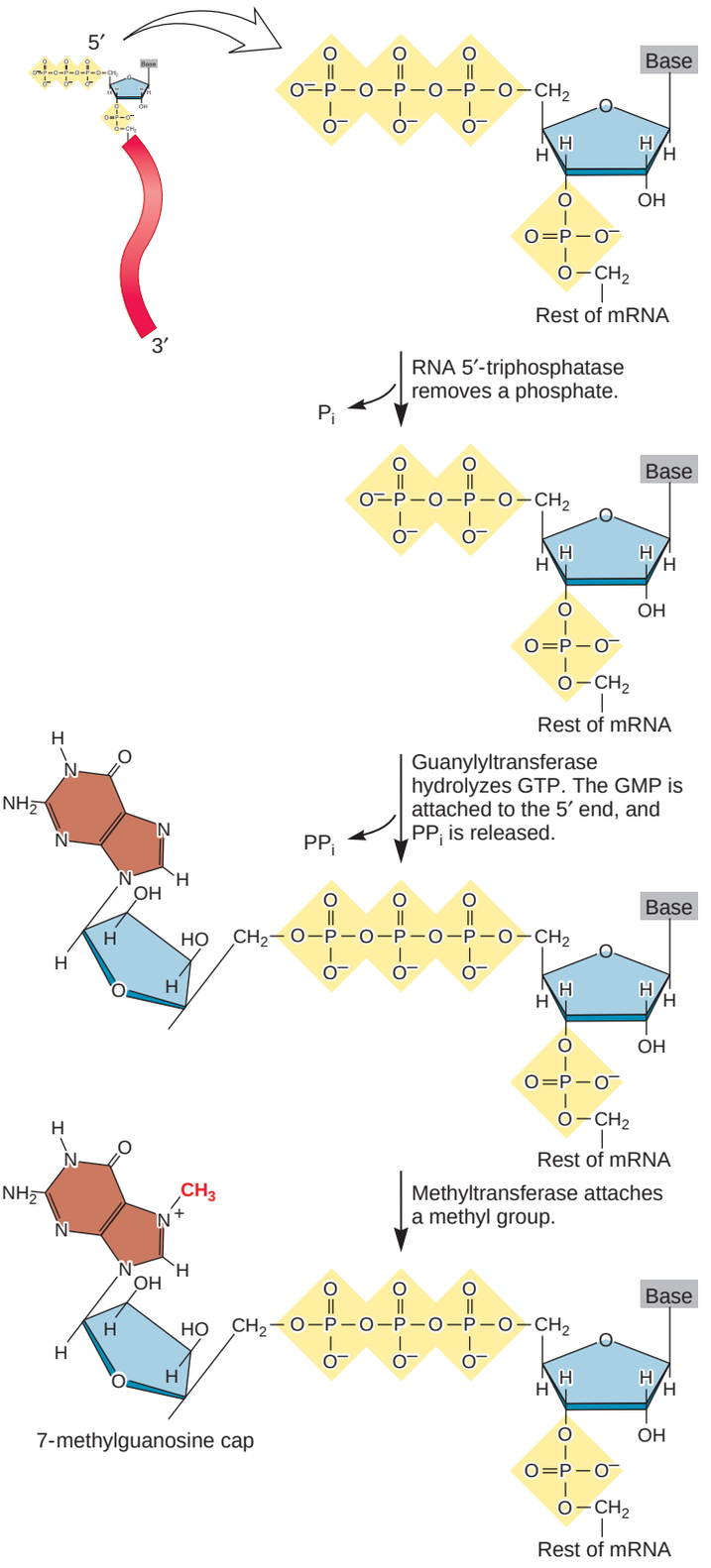


FIGURE 12.23 Attachment of a 7-methylguanosine cap to the 5′ end of mRNA. When the transcript is about 20 to 25 nucleotides in length, RNA 5′-triphosphatase removes one of the three phosphates, and then a second enzyme, guanylyltransferase, attaches GMP to the 5′ end. Finally, a methyltransferase attaches a methyl group to the guanine base.

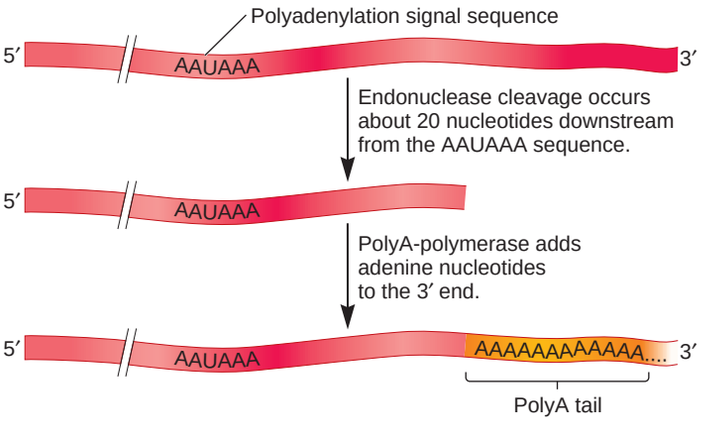


FIGURE 12.24 **Attachment of a polyA tail.** First, an endonuclease cuts the RNA at a location that is 11 to 30 nucleotides after the AAUAAA polyadenylation sequence, making the RNA shorter at its 3′ end. Adenine-containing nucleotides are then attached, one at a time, to the 3′ end by the enzyme polyA-polymerase.

discovery of RNA splicing, the initial finding of RNA editing was met with great skepticism. Since that time, however, RNA editing has been shown to occur in various organisms and in a variety of ways, although its functional significance is slowly emerging (**Table 12.5**). In the specific case of trypanosomes, the editing process involves the addition or deletion of one or more uracil nucleotides.

A more widespread mechanism for RNA editing involves changes of one type of base to another. In this form of editing, a base in the RNA is deaminated—an amino group is removed from the base. When cytosine is deaminated, uracil is formed, and when adenine is deaminated, inosine is formed (**Figure 12.25**). Inosine is recognized as guanine during translation.

An example of RNA editing occurs in mammals involving an mRNA that encodes a protein called apolipoprotein B. In the liver, the RNA editing process produces apolipoprotein B-100, a protein that is essential for the transport of cholesterol in the blood. In intestinal cells, the mRNA may be edited so that a single C is changed to a U. What is the significance of this base substitution? This change converts a glutamine codon (CAA) to a stop codon (UAA) and thereby results in a shorter apolipoprotein. In this case, RNA editing produces an apolipoprotein B with an altered structure. Therefore, RNA editing can produce two proteins from the same gene, much like the phenomenon of alternative splicing.

How widespread is RNA editing that involves C to U and A to I substitutions? In invertebrates such as *Drosophila*, researchers estimate that 50 to 100 RNAs are edited for the purpose of changing the RNA coding sequence. In mammals, the RNAs from less than 25 genes are known to be edited.

TABLE **12.5**

Examples of RNA Editing

Organism	Type of Editing	Found In
Trypanosomes (protozoa)	Primarily additions but occasionally deletions of uracil nucleotides	Many mitochondrial mRNAs
Land plants	C-to-U conversion	Many mitochondrial and chloroplast mRNAs, tRNAs, and rRNAs
Slime mold	C additions	Many mitochondrial mRNAs
Mammals	C-to-U conversion	Apoliproprotein B mRNA, and NFI mRNA, which encodes a tumor-suppressor protein
	A-to-I conversion	Glutamate receptor mRNA, many tRNAs
Drosophila	A-to-I conversion	mRNA for calcium and sodium channels

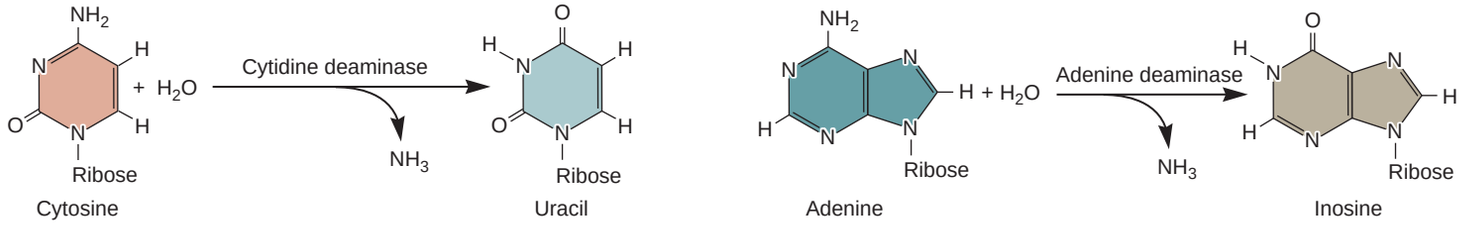


FIGURE 12.25 **RNA editing by deamination.** A cytidine deaminase can remove an amino group from cytosine, thereby creating uracil. An adenine deaminase can remove an amino group from adenine to make inosine.

KEY TERMS

Page 299. gene, transcription, structural genes, messenger RNA (mRNA)
Page 300. translation, central dogma of genetics, gene expression, promoter, terminator
Page 301. template strand, nontemplate strand, coding strand, transcription factors, regulatory sequences, regulatory elements, ribosome-binding site, codons, start codon, stop codon, initiation, elongation, termination, RNA polymerase
Page 302. open complex
Page 303. transcriptional start site, sequence elements, Pribnow box, consensus sequence, core enzyme, sigma (σ) factor, holoenzyme
Page 304. helix-turn-helix motif, closed complex
Page 305. ρ (rho), ρ-dependent termination
Page 306. ρ-independent termination
Page 307. intrinsic termination

Page 308. core promoter, TATA box, basal transcription, enhancers, silencers, *cis*-acting elements, *trans*-acting factors
Page 309. general transcription factors (GTFs), preinitiation complex
Page 310. basal transcription apparatus, mediator, colinearity, exons, intervening sequences, introns, RNA splicing
Page 311. nucleolus, exonuclease, endonuclease
Page 313. ribozyme, hybridization, R loop
Page 316. genomic clone, complementary DNA (cDNA), group I introns, group II introns, self-splicing, maturases, pre-mRNA, spliceosome
Page 317. snRNPs
Page 318. alternative splicing
Page 319. capping, polyA tail, polyadenylation, RNA editing

CHAPTER SUMMARY

- According to the central dogma of genetics, DNA is transcribed into mRNA, and mRNA is translated into a polypeptide. DNA replication allows the DNA to be passed from cell to cell and from parent to offspring (see Figure 12.1).

12.1 Overview of Transcription

- A gene is an organization of DNA sequences. A promoter signals the start of transcription, and a terminator signals the end. Regulatory sequences control the rate of transcription. For genes that encode polypeptides, the gene sequence also specifies a start codon, a stop codon, and many codons in between. Bacterial genes also specify a ribosomal binding sequence (see Figure 12.2).
- Transcription occurs in three phases called initiation, elongation, and termination (see Figure 12.3).
- RNA transcripts have several different functions (see Table 12.1).

12.2 Transcription in Bacteria

- Many bacterial promoters have sequence elements at the –35 and –10 regions. The transcriptional start site is at +1 (see Figures 12.4, 12.5).
- During the initiation phase of transcription in *E. coli*, sigma (σ) factor, which is bound to RNA polymerase, binds into the major groove of DNA and recognizes sequence elements at the promoter. This process forms a closed complex. Following the formation of an open complex, *σ* factor is released (see Figures 12.6, 12.7).
- During the elongation phase of transcription, RNA polymerase slides along the DNA and maintains an open complex as it goes. RNA is made in the 5′ to 3′ direction according to complementary base pairing (see Figure 12.8).
- In a given chromosome, the use of the template strand varies from gene to gene (see Figure 12.9).
- Transcriptional termination in *E. coli* occurs by a rho (ρ)-dependent or ρ-independent mechanism (see Figures 12.10, 12.11).

12.3 Transcription in Eukaryotes

- Eukaryotes use RNA polymerase I, II, and III to transcribe different categories of genes. Prokaryotic and eukaryotic RNA polymerases have similar structures (see Figure 12.12).
- Eukaryotic promoters consist of a core promoter and regulatory elements such as enhancers and silencers (see Figure 12.13).
- Transcription of structural genes in eukaryotes requires RNA polymerase II, five general transcription factors, and mediator. The five general transcription factors and RNA polymerase assemble together to form an open complex (see Table 12.2, Figure 12.14).
- Transcriptional termination of RNA polymerase II occurs while the 3′ end of the transcript is being processed (see Figure 12.15).

12.4 RNA Modification

- RNA transcripts can be modified in a variety of ways, which include processing, splicing, 5′ capping, 3′ polyA tailing, RNA editing, and base modification (see Table 12.3).
- Certain RNA molecules such as ribosomal RNAs and precursor tRNAs are processed to smaller, functional molecules via cleavage steps (see Figures 12.16, 12.17).
- Leder and colleagues identified introns in a globin gene using microscopy (see Figures 12.18, 12.19).
- Group I and II introns are removed by self-splicing. Pre-mRNA introns are removed via a spliceosome (see Table 12.4, Figure 12.20).
- The spliceosome is a multicomponent structure that recognizes intron sequences and removes them from pre-mRNA (see Figures 12.21, 12.22).
- In eukaryotes, mRNA is given a methylguanosine cap at the 5′ end and polyA tail at the 3′ end (see Figures 12.23, 12.24).
- RNA editing changes the base sequence of an RNA after it has been synthesized (see Table 12.5, Figure 12.25).

PROBLEM SETS & INSIGHTS

Solved Problems

S1. Describe the important events that occur during the three stages of gene transcription in bacteria. What proteins play critical roles in the three stages?

Answer: The three stages are initiation, elongation, and termination.

Initiation: RNA polymerase holoenzyme slides along the DNA until σ factor recognizes a promoter. σ factor binds tightly to this sequence, forming a closed complex. The DNA is then denatured to form a bubble-like structure known as the open complex.

Elongation: RNA polymerase core enzyme slides along the DNA, synthesizing RNA as it goes. The *α* subunits of RNA polymerase keep the enzyme bound to the DNA, while the β subunits are responsible for binding and for the catalytic synthesis of RNA. The ω (omega) subunit is also important for the proper assembly of the core enzyme. During elongation, RNA is made according to the AU/GC rule, with nucleotides being added in the 5′ to 3′ direction.

Termination: RNA polymerase eventually reaches a sequence at the end of the gene that signals the end of transcription. In ρ-independent termination, the properties of the termination sequences in the DNA are sufficient to cause termination. In ρ-dependent termination, the ρ (rho) protein recognizes a sequence within the RNA, binds there, and travels toward RNA polymerase. When the formation of a stem-loop structure causes RNA polymerase to pause, ρ catches up and separates the RNA-DNA hybrid region, releasing the RNA polymerase.

S2. What is the difference between a structural gene and a nonstructural gene?

Answer: Structural genes encode mRNAs that are translated into polypeptide sequences. Nonstructural genes encode RNAs that are never translated. Products of nonstructural genes include tRNA and rRNA, which function during translation; microRNA, which is involved in gene regulation; 7S RNA, which is part of a complex known as SRP; scRNA, small cytoplasmic RNA found in bacteria; the RNA of RNaseP; telomerase RNA, which is involved in telomere replication; snoRNA, which is involved in rRNA trimming; and snRNA, which is a component of spliceosome. In many cases, the RNA from nonstructural genes becomes part of a complex composed of RNA molecules and protein subunits.

S3. When RNA polymerase transcribes DNA, only one of the two DNA strands is used as a template. Take a look at Figure 12.4 and explain how RNA polymerase determines which DNA strand is the template strand.

Answer: The binding of σ factor and RNA polymerase depends on the sequence of the promoter. RNA polymerase binds to the promoter in such a way that the –35 sequence TTGACA and the –10 sequence TATAAT are within the coding strand, whereas the –35 sequence AACTGT and the –10 sequence ATATTA are within the template strand.

S4. The process of transcriptional termination is not as well understood in eukaryotes as it is in bacteria. Nevertheless, current evidence suggests several different mechanisms exist for eukaryotic termination. The termination of structural genes appears to occur via the release of elongation factors and/or an RNA-binding protein that functions as an exonuclease. Another type of mechanism is found for the termination of rRNA genes by RNA polymerase I. In this case, a protein known as TTFI (transcription termination factor I) binds to the DNA downstream from the termination site. Discuss how the binding of a protein downstream from the termination site could promote transcriptional termination.

Answer: First, the binding of TTFI could act as a roadblock to the movement of RNA polymerase I. Second, TTFI could promote the dissociation of the RNA transcript and RNA polymerase I from the DNA; it may act like a helicase. Third, it could cause a change in the structure of the DNA that prevents RNA polymerase from moving past the termination site. Though multiple effects are possible, the third effect seems the most likely because TTFI is known to cause a bend in the DNA when it binds to the termination sequence.

Conceptual Questions

C1. Genes may be structural genes that encode polypeptides, or they may be nonstructural genes.

A. Describe three examples of genes that are not structural genes.

B. For structural genes, one DNA strand is called the template strand, and the complementary strand is called the coding strand. Are these two terms appropriate for nonstructural genes? Explain.

C. Do nonstructural genes have a promoter and terminator?

C2. In bacteria, what event marks the end of the initiation stage of transcription?

C3. What is the meaning of the term consensus sequence? Give an example. Describe the locations of consensus sequences within bacterial promoters. What are their functions?

C4. What is the consensus sequence of the following six DNA sequences?

GGCATTGACT
GCCATTGTCA
CGCATAGTCA
GGAAATGGGA
GGCTTTGTCA
GGCATAGTCA

C5. Mutations in bacterial promoters may increase or decrease the level of gene transcription. Promoter mutations that increase transcription are termed up-promoter mutations, and those that decrease transcription are termed down-promoter mutations. As shown in Figure 12.5, the sequence of the –10 region of the promoter for the *lac* operon is TATGTT. Would you expect the following mutations to be up-promoter or down-promoter mutations?

A. TATGTT to TATATT

B. TATGTT to TTTGTT

C. TATGTT to TATGAT

C6. According to the examples shown in Figure 12.5, which positions of the –35 sequence (i.e., first, second, third, fourth, fifth, or sixth) are more tolerant of changes? Do you think that these positions play a more or less important role in the binding of σ factor? Explain why.

C7. In Chapter 9, we considered the dimensions of the double helix (see Figure 9.16). In an α helix of a protein, there are 3.6 amino acids per complete turn. Each amino acid advances the α helix by 0.15 nm; a complete turn of an α helix is 0.54 nm in length. As shown in Figure 12.6, two α helices of a transcription factor occupy the major groove of the DNA. According to Figure 12.6, estimate the number of amino acids that bind to this region. How many complete turns of the α helices occupy the major groove of DNA?

C8. A mutation within a gene sequence changes the start codon to a stop codon. How will this mutation affect the transcription of this gene?

C9. What is the subunit composition of bacterial RNA polymerase holoenzyme? What are the functional roles of the different subunits?

C10. At the molecular level, describe how σ factor recognizes bacterial promoters. Be specific about the structure of σ factor and the type of chemical bonding.

C11. Let's suppose a DNA mutation changes the consensus sequence at the –35 location so that σ factor does not bind there as well. Explain how a mutation could inhibit the binding of σ factor to the DNA. Look at Figure 12.5 and describe two specific base substitutions you think would inhibit the binding of σ factor. Explain why you think your base substitutions would have this effect.

C12. What is the complementarity rule that governs the synthesis of an RNA molecule during transcription? An RNA transcript has the following sequence:

5′–GGCAUGCAUUACGGCAUCACACUAGGGAUC–3′

What is the sequence of the template and coding strands of the DNA that encodes this RNA? On which side (5′ or 3′) of the template strand is the promoter located?

C13. Describe the movement of the open complex along the DNA.

C14. Describe what happens to the chemical bonding interactions when transcriptional termination occurs. Be specific about the type of chemical bonding.

C15. Discuss the differences between ρ-dependent and ρ-independent termination.

C16. In Chapter 11, we discussed the function of DNA helicase, which is involved in DNA replication. The structure and function of DNA helicase and ρ protein are rather similar to each other. Explain how the functions of these two proteins are similar and how they are different.

C17. Discuss the similarities and differences between RNA polymerase (described in this chapter) and DNA polymerase (described in Chapter 11).

C18. Mutations that occur at the end of a gene may alter the sequence of the gene and prevent transcriptional termination.

A. What types of mutations would prevent ρ-independent termination?

B. What types of mutations would prevent ρ-dependent termination?

C. If a mutation prevented transcriptional termination at the end of a gene, where would gene transcription end? Or would it end?

C19. If the following RNA polymerases were missing from a eukaryotic cell, what types of genes would not be transcribed?

A. RNA polymerase I

B. RNA polymerase II

C. RNA polymerase III

C20. What sequence elements are found within the core promoter of structural genes in eukaryotes? Describe their locations and specific functions.

C21. For each of the following transcription factors, how would eukaryotic transcriptional initiation be affected if it were missing?

A. TFIIB

B. TFIID

C. TFIIH

C22. Describe the allosteric and torpedo models for transcriptional termination of RNA polymerase II. Which model is more similar to ρ-dependent termination in bacteria and which model is more similar to ρ-independent termination?

C23. Which eukaryotic transcription factor(s) shown in Figure 12.14 plays an equivalent role to σ factor found in bacterial cells?

C24. The initiation phase of eukaryotic transcription via RNA polymerase II is considered an assembly and disassembly process. Which types of biochemical interactions—hydrogen bonding, ionic bonding, covalent bonding, and/or hydrophobic interactions—would you expect to drive the assembly and disassembly process? How would temperature and salt concentration affect assembly and disassembly?

C25. A eukaryotic structural gene contains two introns and three exons: exon 1–intron 1–exon 2–intron 2–exon 3. The 5′ splice site at the boundary between exon 2 and intron 2 has been eliminated by a small deletion in the gene. Describe how the pre-mRNA encoded by this mutant gene would be spliced. Indicate which introns and exons would be found in the mRNA after splicing occurs.

C26. Describe the processing events that occur during the production of tRNA in *E. coli*.

C27. Describe the structure and function of a spliceosome. Speculate why the spliceosome subunits contain snRNA. In other words, what do you think is/are the functional role(s) of snRNA during splicing?

C28. What is the unique feature of ribozyme function? Give two examples described in this chapter.

C29. What does it mean to say that gene expression is colinear?

C30. What is meant by the term self-splicing? What types of introns are self-splicing?

C31. In eukaryotes, what types of modification occur to pre-mRNA?

C32. What is alternative splicing? What is its biological significance?

C33. The processing of ribosomal RNA in eukaryotes is shown in Figure 12.16. Why is this called cleavage or processing but not splicing?

C34. In the splicing of group I introns shown in Figure 12.20, does the 5′ end of the intron have a phosphate group? Explain.

C35. According to the mechanism shown in Figure 12.22, several snRNPs play different roles in the splicing of pre-mRNA. Identify the snRNP that recognizes the following sites:

A. 5′ splice site

B. 3′ splice site

C. Branch site

C36. After the intron (which is in a lariat configuration) is released during pre-mRNA splicing, a brief moment occurs before the two exons are connected to each other. Which snRNP(s) holds the exons in place so they can be covalently connected to each other?

C37. A lariat contains a closed loop and a linear end. An intron has the following sequence: 5′–GUPuAGUA–60 nucleotides–UACUUAUCC–100 nucleotides–Py_{12}NPyAG–3′. Which sequence would be found within the closed loop of the lariat, the 60-nucleotide sequence or the 100-nucleotide sequence?

Experimental Questions

E1. A research group has sequenced the cDNA and genomic DNA from a particular gene. The cDNA is derived from mRNA, so it does not contain introns. Here are the DNA sequences.

cDNA:

5′-ATTGCATCCAGCGTATACTATCTCGGGCCCAATTAATGCCA-
GCGGCCAGACTATCACCCAACTCGGTTACCTACTAGTATATC-
CCATATACTAGCATATATTTTACCCATAATTTGTGTGTGGGTATA-
CAGTATAATCATATA-3′

Genomic DNA (contains one intron):

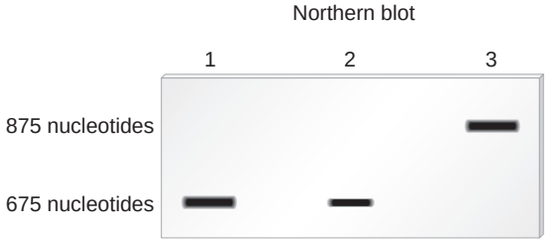
5′-ATTGCATCCAGCGTATACTATCTCGGGCCCAATTAATGC-
CAGCGGCCAGACTATCACCCAACTCGGCCCACCCCCCAGGTTTA-
CACAGTCATACCATACATACAAAAATCGCAGTTACTTATCCCA-
AAAAAACCTAGATACCCCACATACTATTAACTCTTTCTTTCTAG-
GTTACCTACTAGTATATCCCATATACTAGCATATATTTTAC-
CCATAATTTGTGTGTGGGTATACAGTATAATCATATA-3′

Indicate where the intron is located. Does the intron contain the normal consensus splice site sequences based on those described in Figure 12.21? Underline the splice site sequences, and indicate whether or not they fit the consensus sequence.

E2. What is an R loop? In an R loop experiment, to which strand of DNA does the mRNA bind, the coding strand or the template strand?

E3. If a gene contains three introns, draw what it would look like in an R loop experiment.

E4. Chapter 18 describes a technique known as Northern blotting that can be used to detect RNA transcribed from a particular gene. In this method, a specific RNA is detected using a short segment of cloned DNA as a probe. The DNA probe, which is radioactive, is complementary to the RNA that the researcher wishes to detect. After the radioactive probe DNA binds to the RNA, the RNA is visualized as a dark (radioactive) band on an X-ray film. As shown here, the method of Northern blotting can be used to determine the amount of a particular RNA transcribed in a given cell type. If one type of cell produces twice as much of a particular mRNA as occurs in another cell, the band will appear twice as intense. Also, the method can distinguish if alternative RNA splicing has occurred to produce an RNA that has a different molecular mass.



Lane 1 is a sample of RNA isolated from nerve cells.

Lane 2 is a sample of RNA isolated from kidney cells. Nerve cells produce twice as much of this RNA as do kidney cells.

Lane 3 is a sample of RNA isolated from spleen cells. Spleen cells produce an alternatively spliced version of this RNA that is about 200 nucleotides longer than the RNA produced in nerve and kidney cells.

Let's suppose a researcher was interested in the effects of mutations on the expression of a particular structural gene in eukaryotes. The gene has one intron that is 450 nucleotides long. After this intron is removed from the pre-mRNA, the mRNA transcript is 1100 nucleotides in length. Diploid somatic cells have two copies of this gene. Make a drawing that shows the expected results of a Northern blot using mRNA from the cytosol of somatic cells, which were obtained from the following individuals:

Lane 1: A normal individual

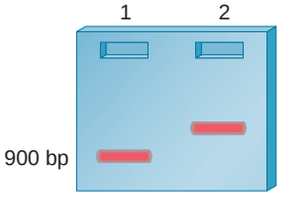
Lane 2: An individual homozygous for a deletion that removes the −50 to −100 region of the gene that encodes this mRNA

Lane 3: An individual heterozygous in which one gene is normal and the other gene had a deletion that removes the −50 to −100 region

Lane 4: An individual homozygous for a mutation that introduces an early stop codon into the middle of the coding sequence of the gene

Lane 5: An individual homozygous for a three-nucleotide deletion that removes the AG sequence at the 3′ splice site

E5. A gel retardation assay can be used to study the binding of proteins to a segment of DNA. This method is described in Chapter 18. When a protein binds to a segment of DNA, it retards the movement of the DNA through a gel, so the DNA appears at a higher point in the gel (see the following).



Lane 1: 900-bp fragment alone

Lane 2: 900-bp fragment plus a protein that binds to the 900-bp fragment

In this example, the segment of DNA is 900 bp in length, and the binding of a protein causes the DNA to appear at a higher point in the gel. If this 900-bp fragment of DNA contains a eukaryotic promoter for a structural gene, draw a gel that shows the relative locations of the 900-bp fragment under the following conditions:

Lane 1: 900 bp plus TFIID

Lane 2: 900 bp plus TFIIB

Lane 3: 900 bp plus TFIID and TFIIB

Lane 4: 900 bp plus TFIIB and RNA polymerase II

Lane 5: 900 bp plus TFIID, TFIIB, and RNA polymerase II/TFIIF

E6. As described in Chapter 18 and in experimental question E5, a gel retardation assay can be used to determine if a protein binds to DNA. This method can also determine if a protein binds to RNA. In the combinations described here, would you expect the migration of the RNA to be retarded due to the binding of a protein?

A. mRNA from a gene that is terminated in a ρ-independent manner plus ρ protein

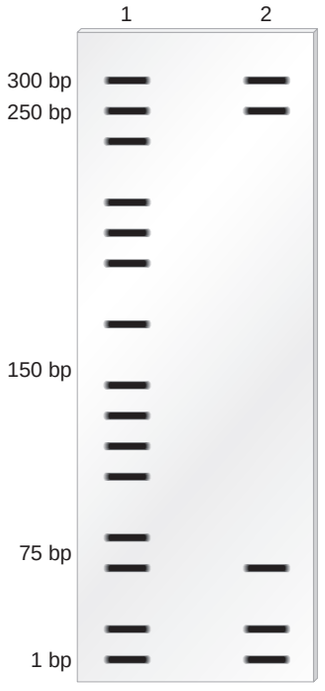
B. mRNA from a gene that is terminated in a ρ-dependent manner plus ρ protein

C. pre-mRNA from a structural gene that contains two introns plus the snRNP called U1

D. Mature mRNA from a structural gene that contains two introns plus the snRNP called U1

E7. The technique of DNA footprinting is described in Chapter 18. If a protein binds over a region of DNA, it will protect chromatin in that region from digestion by DNase I. To carry out a DNA footprinting experiment, a researcher has a sample of a cloned DNA fragment. The fragments are exposed to DNase I in the presence and absence of a DNA-binding protein. Regions of the DNA fragment not covered by the DNA-binding protein will be digested by DNase I, and this will produce a series of bands on a gel. Regions of the DNA fragment not digested by DNase I (because a DNA-binding protein is preventing DNase I from gaining access to the DNA) will be revealed, because a region of the gel will not contain any bands.

In the DNA footprinting experiment shown here, a researcher began with a sample of cloned DNA 300 bp in length. This DNA contained a eukaryotic promoter for RNA polymerase II. For the sample loaded in lane 1, no proteins were added. For the sample loaded in lane 2, the 300-bp fragment was mixed with RNA polymerase II plus TFIID and TFIIB.



A. How long of a region of DNA is "covered up" by the binding of RNA polymerase II and the transcription factors?

B. Describe how this binding would occur if the DNA was within a nucleosome structure. (Note: The structure of nucleosomes is described in Chapter 10.) Do you think that the DNA is in a nucleosome structure when RNA polymerase and transcription factors are bound to the promoter? Explain why or why not.

E8. As described in Table 12.1, several different types of RNA are made, especially in eukaryotic cells. Researchers are sometimes interested in focusing their attention on the transcription of structural genes in eukaryotes. Such researchers want to study mRNA. One method that is used to isolate mRNA is column chromatography. (Note: See the Appendix for a general description of chromatography.) Researchers can covalently attach short pieces of DNA that contain stretches of thymine (i.e., TTTTTTTTTTTT) to the column matrix. This is called a poly-dT column. When a cell extract is poured over the column, mRNA binds to the column, but other types of RNA do not.

A. Explain how you would use a poly-dT column to obtain a purified preparation of mRNA from eukaryotic cells. In your description, explain why mRNA binds to this column and what you would do to release the mRNA from the column.

B. Can you think of ways to purify other types of RNA, such as tRNA or rRNA?

Questions for Student Discussion/Collaboration

1. Based on your knowledge of introns and pre-mRNA splicing, discuss whether or not you think alternative splicing fully explains the existence of introns. Can you think of other possible reasons to explain the existence of introns?

2. Discuss the types of RNA transcripts and the functional roles they play. Why do you think some RNAs form complexes with protein subunits?

Note: All answers appear at the website for this textbook; the answers to even-numbered questions are in the back of the textbook.

www.mhhe.com/brookergenetics4e

Visit the website for practice tests, answer keys, and other learning aids for this chapter. Enhance your understanding of genetics with our interactive exercises, quizzes, animations, and much more.

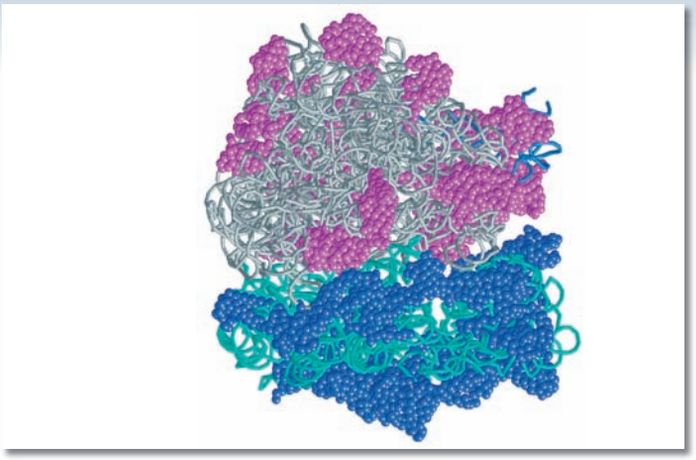
CHAPTER OUTLINE

13.1 The Genetic Basis for Protein Synthesis

13.2 Structure and Function of tRNA

13.3 Ribosome Structure and Assembly

13.4 Stages of Translation



A molecular model for the structure of a ribosome. *This is a model of ribosome structure based on X-ray crystallography. Ribosomes are needed to synthesize polypeptides, using mRNA as a template. A detailed description of this model is discussed later in Figure 13.15.*

13 TRANSLATION OF mRNA

The synthesis of cellular proteins occurs via the translation of the sequence of codons within mRNA into a sequence of amino acids that constitute a polypeptide. The general steps that occur in this process were already outlined in Chapter 1. In this chapter, we will explore the current state of knowledge regarding translation, with an eye toward the specific molecular interactions responsible for this process. During the past few decades, the concerted efforts of geneticists, cell biologists, and biochemists have profoundly advanced our understanding of translation. Even so, many questions remain unanswered, and this topic continues to be an exciting area of investigation.

We will begin by considering classic experiments that revealed the purpose of some genes is to encode proteins that function as enzymes. Next, we examine how the genetic code is used to decipher the information within mRNA to produce a polypeptide with a specific amino acid sequence. The rest of this chapter is devoted to an understanding of translation at the molecular level as it occurs in living cells. This will involve an examination of the cellular components—including many different proteins, RNAs, and small molecules—needed for the translation process. We will consider the structure and function of tRNA molecules, which act as the translators of the genetic information within mRNA, and then examine the composition of ribosomes. Finally, we will explore the differences between translation in bacterial cells and eukaryotic cells.

13.1 THE GENETIC BASIS FOR PROTEIN SYNTHESIS

Proteins are critically important as active participants in cell structure and function. The primary role of DNA is to store the information needed for the synthesis of all the proteins that an organism makes. As we discussed in Chapter 12, genes that encode an amino acid sequence are known as **structural genes.** The RNA transcribed from structural genes is called **messenger RNA (mRNA).** The main function of the genetic material is to encode the production of cellular proteins in the correct cell, at the proper time, and in suitable amounts. This is an extremely complicated task because living cells make thousands of different proteins. Genetic analyses have shown that a typical bacterium can make a few thousand different proteins, and estimates for eukaryotes range from several thousand in simple eukaryotic