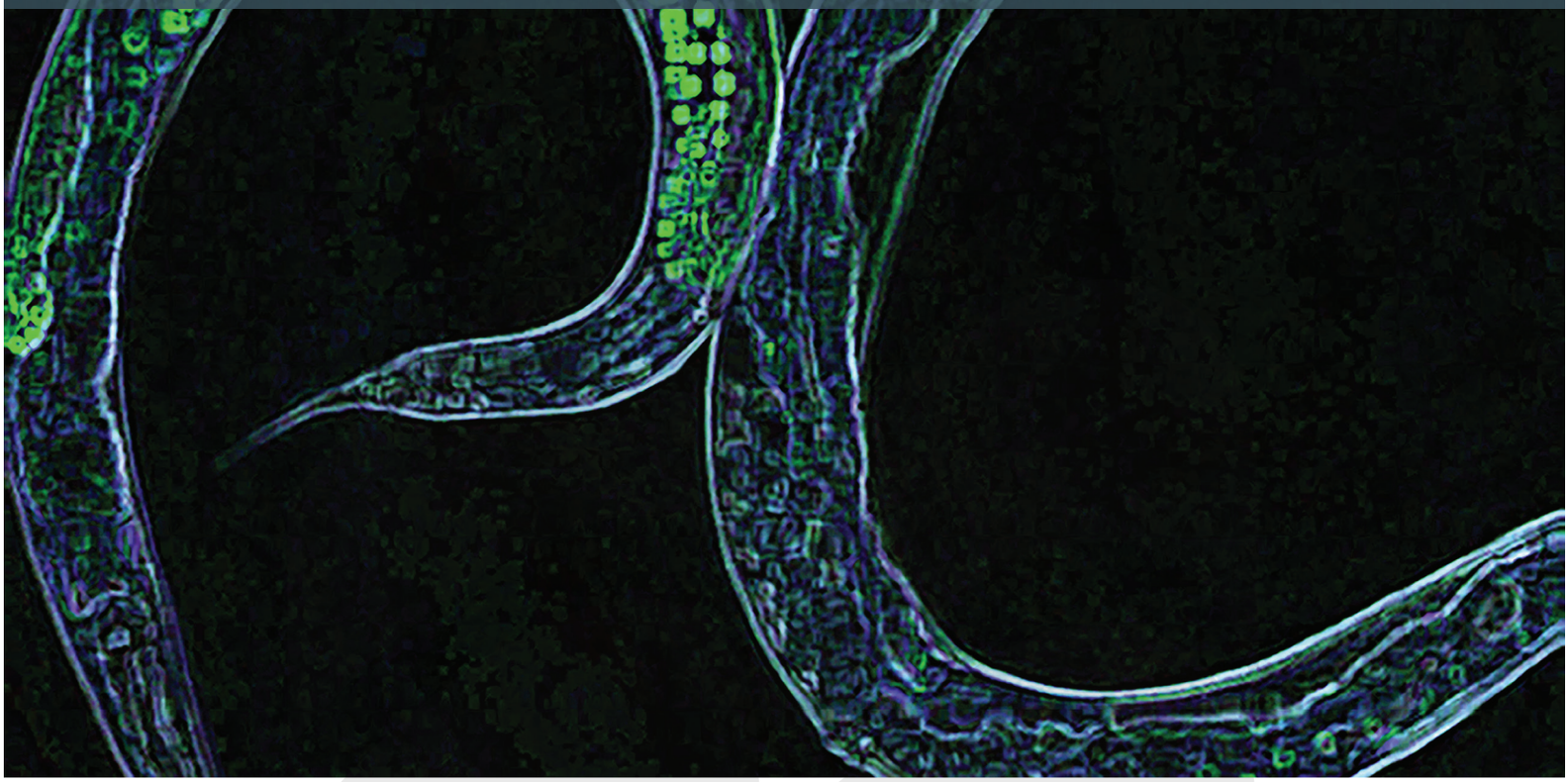
CHAPTER

8

RNA: Transcription, Processing, and Decay



Knowledge of the molecular mechanisms that synthesize and destroy RNAs in cells has led to technologies that allow researchers to control gene expression in precise ways. For example, on the left is a *C. elegans* worm that has been manipulated to express a gene encoding the green fluorescent protein (GFP) in specific cells, and on the right is a genetically identical worm in which GFP expression is silenced. [*Jessica Vasale/Laboratory of Craig Mello.*]

CHAPTER OUTLINE AND LEARNING OBJECTIVES

CHAPTER OBJECTIVE

The broad objective for this chapter is to understand the mechanisms of RNA synthesis, processing, and decay as well as how RNA and protein factors regulate these mechanisms.

In this chapter, we describe how the information stored in DNA is transferred to RNA. The key event in this transfer is **transcription,** which copies the information from one strand of DNA into a strand of RNA. In bacteria, the information in protein-coding RNAs is almost immediately converted into protein by a process called translation (the focus of Chapter 9). In contrast, in eukaryotes, transcription and translation are spatially separated: transcription takes place in the nucleus and translation in the cytoplasm. Furthermore, in eukaryotes, before RNAs are ready to be exported to the cytoplasm for translation, they undergo extensive processing, including deletion of internal nucleotides and addition of special nucleotide structures to the 5′ and 3′ ends.

Both bacteria and eukaryotes also produce other types of RNA that are not translated into protein but instead perform a variety of roles in cells by base pairing to other RNAs, binding proteins, and performing enzymatic reactions. Lastly, the chapter describes how RNAs are eliminated from cells by decay mechanisms that disassemble RNAs into individual nucleotides. **Figure 8-1** provides an overview of the chapter by illustrating the timeline of events that occur in the life cycle of protein-coding RNAs (mRNAs) in bacteria and eukaryotes. Every process described in this chapter relies on molecular interactions that are specified by nucleic acid sequences. RNAs interact with DNA and other RNAs by base pairing of complementary sequences, and proteins interact with DNA and RNA by binding specific sequences. Therefore, mutations in DNA and RNA that disrupt molecular interactions can affect the expression of proteins.

In summary, this chapter focuses on the molecules and mechanisms that produce and destroy RNAs. The molecules and mechanisms are important to geneticists because mutations that affect them change which proteins are expressed, their sequence, and their abundance, and lead to altered phenotypes.

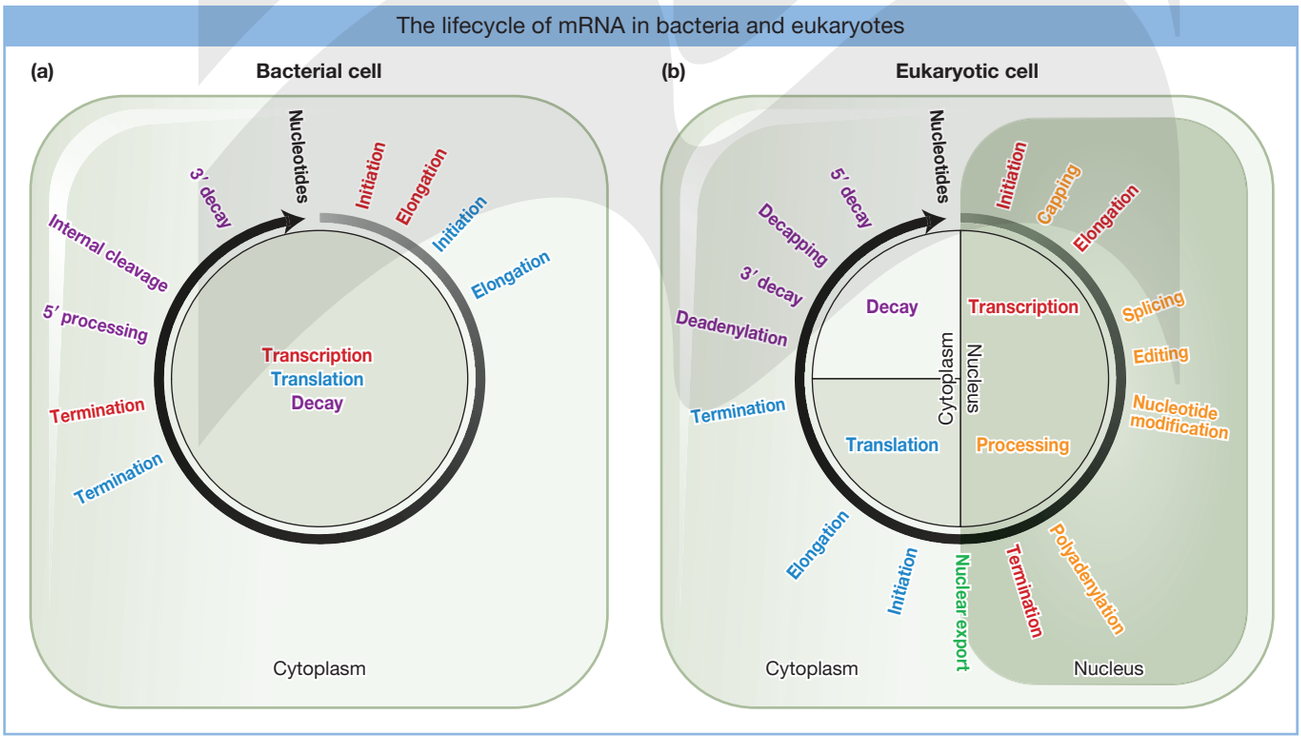


FIGURE 8-1 In both (a) bacteria and (b) eukaryotes, the transfer of information from DNA to protein involves a sequential series of molecular events dedicated to mRNA. Note the similarities and differences in the life cycles of mRNA in bacteria and eukaryotes. Transcription, translation, and decay of mRNA occur in both bacteria and eukaryotes. In contrast, transcription, translation, and decay occur concurrently only in bacteria, and various mRNA processing events and nuclear export are unique to eukaryotes.

KEY CONCEPT The life cycle of an mRNA in bacteria and in eukaryotes includes transcription, translation, and decay. The life cycle of an mRNA in eukaryotes also includes RNA processing and nuclear export.

KEY CONCEPT Central to the transfer of information from DNA to RNA in both bacteria and eukaryotes are molecular interactions that involve base pairing of complementary nucleic acids (DNA and RNA) and binding of proteins to specific nucleic acid sequences.

8.1 RNA STRUCTURE

LO 8.1 Describe how the structure of RNA enables it to function differently from DNA.

RNAs carry out an amazing variety of biological functions, including providing the information for making proteins, regulating translation, processing RNA, and maintaining chromosome ends. The versatility of RNA relative to DNA is due to the ability of single-stranded RNA to form an immense variety of elaborate three-dimensional structures that scaffold the binding of proteins, base pair with other RNAs, and carry out enzymatic reactions. Furthermore, the versatility of RNA as a biomolecule is enhanced by the variety of ways in which RNA function can be regulated, including mechanisms that alter RNA structure, abundance, and cellular localization.

RNA is the information-carrying intermediate between DNA and proteins

Early investigators had good reason for thinking that information is not transferred directly from DNA to protein. In eukaryotic cells, DNA is located in the nucleus, which physically separates it from the protein synthesis machinery in the cytoplasm. Thus, an intermediate is needed that carries the DNA sequence information from the nucleus to the cytoplasm. That intermediate is RNA.

In 1957, Elliot Volkin and Lawrence Astrachan made an observation suggesting that RNA was the intermediate molecule. They found that one of the most striking molecular changes that takes place when *E. coli* is infected with the bacteriophage T2 is a rapid burst of RNA synthesis. Furthermore, this bacteriophage-induced RNA "turns over"; that is, the amount of time it spends in the cell is brief, on the order of minutes. Its rapid appearance and disappearance suggested that RNA might play some role in the synthesis of more T2 phage particles.

Volkin and Astrachan demonstrated the rapid turnover of RNA using a protocol called a pulse–chase experiment. To conduct a pulse–chase experiment, the infected bacteria are first fed (pulsed with) radioactive uracil, a molecule needed for the synthesis of RNA but not DNA. Any RNA synthesized in the bacteria from then on is "labeled" with the readily detectable radioactive uracil. After a short period of incubation, the radioactive uracil is washed away and replaced (chased) by uracil that is not radioactive. This procedure "chases" the label out of RNA because, as the pulse-labeled RNA breaks down, only the unlabeled uracil is available to synthesize new RNA molecules. Volkin and Astrachan found that the RNA recovered shortly after the pulse was labeled, but RNA recovered just a few minutes later was unlabeled, indicating that the RNA has a very short lifetime in bacteria.

KEY CONCEPT Unlike DNA, RNA has a high rate of turnover within cells.

A similar experiment can be done with eukaryotic cells. Cells are first pulsed with radioactive uracil and, after a short time, they are transferred to medium (the liquid they grow in) with unlabeled uracil. In samples taken immediately after the pulse, most of the labeled RNA is in the nucleus. However, in samples taken a few minutes later, the labeled RNA is also found in the cytoplasm (**Figure 8-2**). This indicates that, in eukaryotes, RNA is synthesized in the nucleus and then moves to the cytoplasm, where proteins are synthesized. These data along with other data led to the conclusion that RNA is the information-transfer intermediary between DNA and protein.

KEY CONCEPT In eukaryotic cells, RNA can move from the nucleus to the cytoplasm.

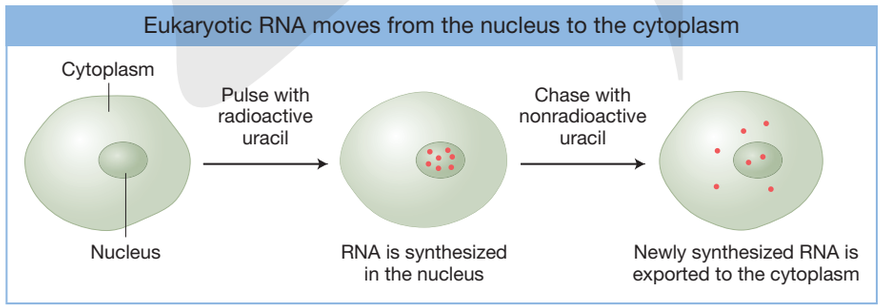


FIGURE 8-2 The pulse–chase experiment showed that RNA moves from the nucleus to the cytoplasm in eukaryotic cells. Cells are grown briefly in medium with radioactive uracil to label newly synthesized RNA (pulse). Cells are washed to remove the radioactive uracil and then grown in medium with nonradioactive uracil (chase). The red dots indicate the location of RNAs containing radioactive uracil over time. The location of radioactive RNAs was not determined by microscopy; rather, it was inferred using a molecular approach. Cells were collected immediately after the pulse as well as after the chase and broken open, nuclei were separated from the cytoplasm by centrifugation, RNA was purified from the nuclear and cytoplasmic fractions (leaving behind radioactive uracil that was not incorporated into RNA), and the amount of radioactivity in the nuclear and cytoplasmic fractions was determined using a scintillation counter (an instrument that measures ionizing radiation).

Consequences of the distinct chemical properties of RNA

Although both RNA and DNA are nucleic acids, the building blocks of RNA differ from those of DNA in three important ways that allow RNA to have greater structural and functional diversity.

1. RNA has **ribose** sugar in its nucleotides, rather than the **deoxyribose** found in DNA (Figure 8-3a). As the names suggest, the sugars differ in the presence or absence of just one oxygen atom at the 2′ carbon. The 2′-OH in RNA reduces the stability of single-stranded RNA relative to single-stranded DNA. RNA cleavage can occur when a deprotonated 2′-OH acts as a nucleophile at a nearby phosphodiester bond. Furthermore, the 2′-OH provides an extra site for hydrogen bonding as well as an extra site for chemical modification such as methylation. Properties of the 2′-OH contribute to the ability of RNAs, which are usually single-stranded, to fold into complex three-dimensional structures.
2. RNA contains the pyrimidine base **uracil** (U), instead of thymine (T) found in DNA (Figure 8-3b). Uracil forms two hydrogen bonds with adenine just like thymine does in DNA (Figure 8-3c). In addition, uracil is capable of base pairing with G in helices of a folded RNA or between two separate RNAs, but not with G in DNA during transcription. The ability of U to base pair with both A and G is a major reason why RNA can form intricate structures.

   Throughout this chapter, we revisit the chemical and structural properties of the ribose 2′-OH and the uracil base because they are critical to the folding, function, and recognition of RNAs in RNA processing events.
3. RNA is usually single-stranded, not double-stranded like DNA. As a consequence, RNA is much more flexible than DNA and can form a greater variety of three-dimensional structures. Base pairing between regions within an RNA (i.e., intramolecular base pairing) is an important determinant of RNA structure. For example, the stem-loop is the fundamental structural element of RNA (Figure 8-3d). Stem-loops are made up of a double-stranded stem of complementary regions of an RNA and a single-stranded loop at the end of the stem.

KEY CONCEPT Unlike DNA, RNA contains 2′-hydroxyls on the ribose sugars, uracil replaces thymine, and it is single-stranded but base pairs to itself to form double-stranded regions.

As exemplified by the terms ribose sugar and uracil base, there is specific terminology for describing the building blocks of RNA. There are terms for each of the four RNA nucleobases (i.e., the base itself), the nucleosides (i.e., the base with a ribose sugar), and the nucleotides (i.e., the base with a ribose sugar and one, two, or three phosphates) (Figure 8-4). For example, uracil is a nucleobase, uridine is a nucleoside, and uridine triphosphate is a nucleotide.

Classes of RNA

RNA molecules can be grouped into two general classes. One class of RNA is **messenger RNA** (**mRNA**) because, like a messenger, it serves as an intermediary that carries information. The information from DNA is transferred to mRNA through the process of transcription, and mRNA passes that information on to proteins through the process of translation. The other class of RNA is **noncoding RNA** (**ncRNA**) because it does not encode proteins. Instead, the ncRNA is the final product whose function is determined by its sequence and three-dimensional structure.

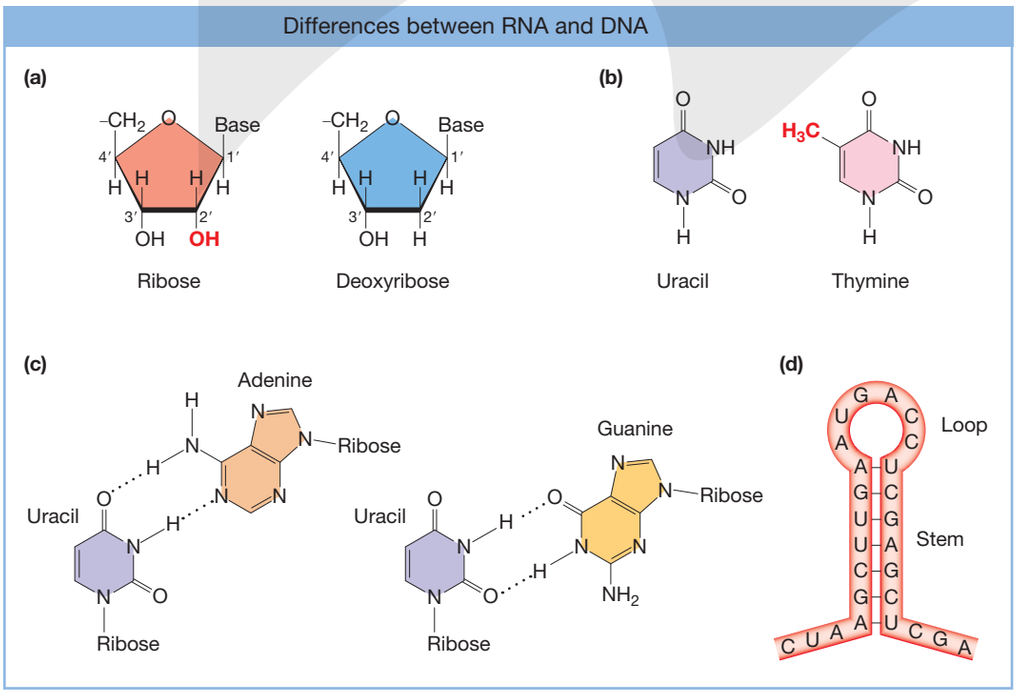


FIGURE 8-3 (a) The 5-carbon sugar in RNA nucleosides (*left*) versus DNA nucleosides (*right*). Ribose carries a hydroxyl group, indicated in red, at the 2′ carbon instead of a hydrogen atom in deoxyribose. (b) The pyrimidine base uracil (*left*) replaces thymine (*right*) in RNA versus DNA, respectively. Uracil differs from thymine by a methyl group, indicated in red. (c) Uracil base pairs via two hydrogen bonds with adenine (*left*) and guanine (*right*). (d) Stem-loops are basic structural features of RNA. Note that U base pairs with both A and G.

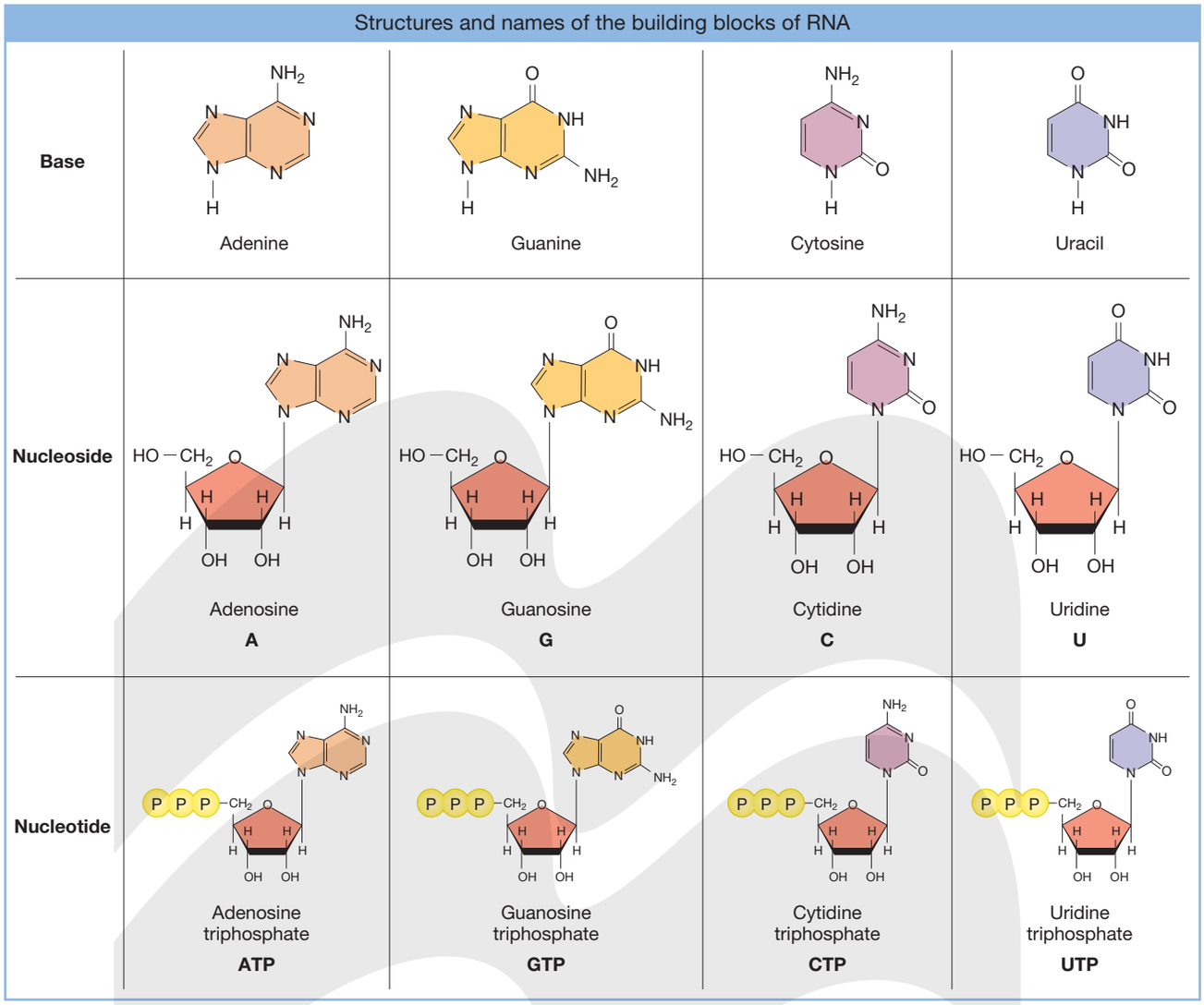


FIGURE 8-4 Each of the four building blocks of RNA has a distinct name for its nucleobase, nucleoside, and nucleotide.

ncRNAs vary in length and function, although they often act similarly by base pairing with another RNA, by serving as a scaffold for binding proteins, or by catalyzing a chemical reaction. In both bacteria and eukaryotes, some ncRNA such as **transfer RNAs** (**tRNAs**) and **ribosomal RNAs** (**rRNAs**) function in translation. Bacteria and eukaryotes also express a variety of ncRNAs that are about 50–300 nucleotides in length and function at multiple steps in gene expression. In eukaryotes, ncRNAs are categorized based on their location in the cell: **small *nuclear* RNAs** (**snRNAs**), small *nucleolar* RNAs (snoRNAs), and small *cytoplasmic* RNAs (scRNAs). The nucleolus is a non–membrane-bound region in the nucleus where ribosomes are produced. Eukaryotes also express **long noncoding RNAs** (**lncRNAs**) that are typically greater than 300 nucleotides in length. Thousands of lncRNAs have been identified in humans, but only a few have been assigned functions, and these are mostly regulators of gene expression. Some eukaryotes also encode **microRNAs** (**miRNAs**) and generate **small interfering RNAs** (**siRNAs**) and other very small RNAs, about 21 nucleotides in length, that suppress the expression of genes and help maintain genome stability.

KEY CONCEPT There are two general classes of RNAs, those that encode proteins (mRNA) and those that do not encode proteins (ncRNA). ncRNAs participate in a variety of cellular processes, including protein synthesis (tRNA and rRNA), RNA processing (snRNA), the regulation of gene expression (siRNA and miRNA), and genome defense (siRNA).

8.2 TRANSCRIPTION AND DECAY OF mRNA IN BACTERIA

LO 8.2 Explain how RNA polymerases are directed to begin and end transcription at specific places in genomes.

The first step in the transfer of information from DNA to protein is to produce an RNA strand whose nucleotide sequence matches the nucleotide sequence of a DNA

segment. Because this process is reminiscent of transcribing (copying) written words, the synthesis of RNA is called *transcription*. The DNA is said to be transcribed into RNA, and the RNA is called a **transcript**. Volkin and Astrachan showed that RNA is transcribed and degraded rapidly within the cell, and later experiments showed that the abundance of a given RNA is regulated by controlling its rate of transcription and its rate of decay. These processes, though chemically simple, are controlled by a variety of factors.

Overview: DNA as transcription template

Information encoded in DNA is transferred to the RNA transcript by the complementary pairing of DNA and RNA bases. Consider the transcription of a chromosomal segment that constitutes a gene. First, the two strands of the DNA double helix separate locally to form a **transcription bubble**. One of the separated strands acts as a template for RNA synthesis and is called the **template strand** (or noncoding strand) and the other strand is called the **non-template strand** (or coding strand) (**Figure 8-5a**). The

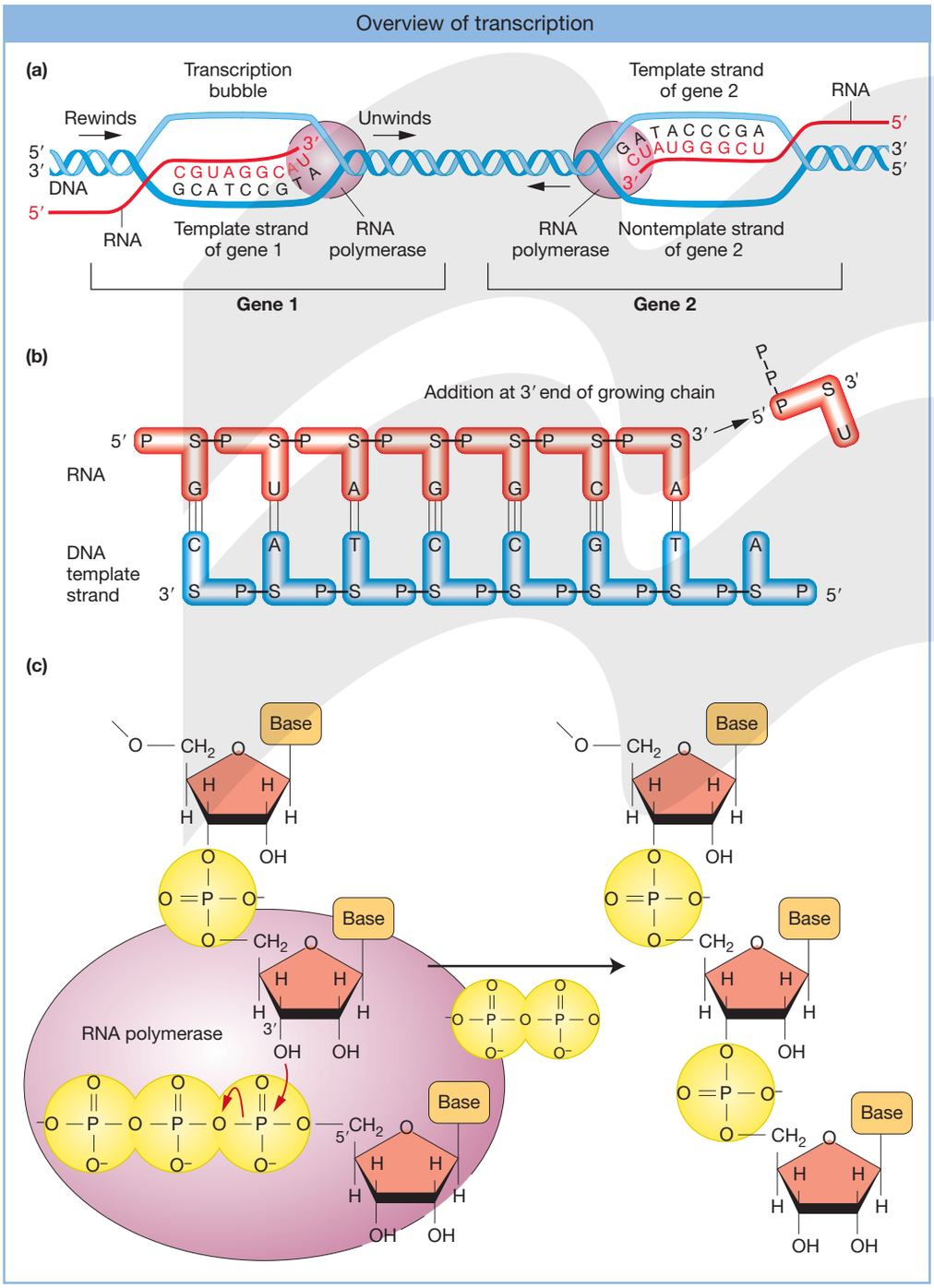


FIGURE 8-5 (a) Only one strand of DNA is the template for gene transcription, but the strand can vary with the gene. RNA is transcribed in the 5′-to-3′ direction using DNA oriented in the 3′-to-5′ direction as a template. Hence, genes transcribed in different directions use opposite strands of DNA as templates. (b) As a gene is transcribed, the 3′-hydroxyl group on the sugar (S) at the end of the growing RNA strand attaches to the 5′-phosphate group (P) on the entering ribonucleotide (A, C, G, or U) that base pairs with the DNA template nucleotide. (c) To form a phosphodiester bond, the 3′ hydroxyl is deprotonated and acts as a nucleophile at the α-phosphate of the entering nucleotide, breaking the bond between the α- and β-phosphates and producing the energy needed to form the new phosphodiester bond.

Visualizing transcription provides insights into transcription mechanisms

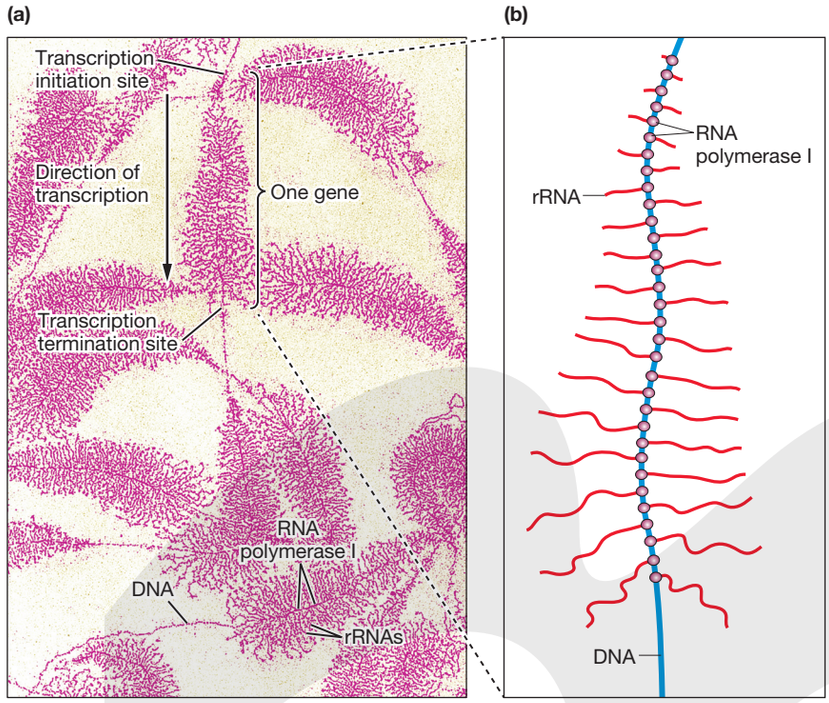


FIGURE 8-6 This picture from an electron microscope shows the transcription of tandemly repeated rRNA genes in the nuclear genome of a newt oocyte (an immature egg). Along each gene, many RNA polymerases (in this case, eukaryotic RNA polymerase I) are transcribing in one direction. Growing rRNA transcripts appear as threads extending outward from the DNA. Transcription starts (initiates) and ends (terminates) at specific places. Shorter transcripts are closer to the start of the gene; longer ones are closer to the end of the gene. Based on their appearance, these structures are called "Christmas trees." They are also called "Miller spreads" after their discoverer Oscar Miller [*W Fawcett Don/Getty Images*.]

resulting RNA sequence is complementary to the template strand and identical (except for the use of uracil in place of thymine) to the non-template strand. When DNA sequence is cited in the scientific literature, the sequence of the non-template strand is almost always given because it is the same as the RNA sequence.

KEY CONCEPT RNA sequence is complementary to the template strand and is the same as the coding (non-template) strand, except it contains U in place of T.

Across the genome, both DNA strands may be used as templates, but in any one gene, only one strand is used (Figure 8-5a). Starting at the 3′ end of the template strand, ribonucleotides form base pairs by hydrogen bonding with their complementary DNA nucleotides. The ribonucleotide A pairs with T in the DNA, C with G, G with C, and U with A. Each ribonucleotide is positioned opposite its complementary nucleotide by the enzyme **RNA polymerase**. This enzyme moves along the DNA template strand in the 3′-to-5′ direction forming **phosphodiester bonds** that covalently link aligned ribonucleotides to build RNA in the 5′-to-3′ direction, as shown in Figures 8-5b and c. As the RNA strand is progressively lengthened, the 5′ end is displaced from the DNA template and the transcription bubble closes behind RNA polymerase. Multiple RNA polymerases, each synthesizing an RNA molecule, can move along a gene at the same time (**Figure 8-6**). Hence, we already see two fundamental mechanisms that bring about the transfer of information from DNA to RNA: base complementarity and protein-nucleic acid interactions.

KEY CONCEPT RNA is transcribed in the 5′-to-3′ direction using a single-stranded DNA template oriented in the 3′-to-5′ direction. Thus, RNAs start with a 5′-triphosphate (5′-ppp) and end with a 3′ hydroxyl (3′-OH).

KEY CONCEPT One gene can be transcribed by multiple RNA polymerase molecules at the same time.

Stages of transcription

Genes are segments of DNA embedded in extremely long DNA molecules (chromosomes). How, then, is a gene accurately transcribed into RNA with a specific beginning and end? Because the DNA of a chromosome is a continuous unit, the transcriptional machinery must be directed to the start of a gene to begin transcription, continue transcribing the length of the gene, and finally stop transcribing at the end of the gene. These three distinct stages of transcription are called **initiation, elongation,** and **termination,** respectively. Although the overall process of transcription is remarkably similar in bacteria and eukaryotes, there are important differences. For this reason, we will follow the three stages first in bacteria by using the gut bacterium *E. coli* as an example, and then we will repeat the process in eukaryotes, emphasizing the similarities and differences between bacteria and eukaryotes.

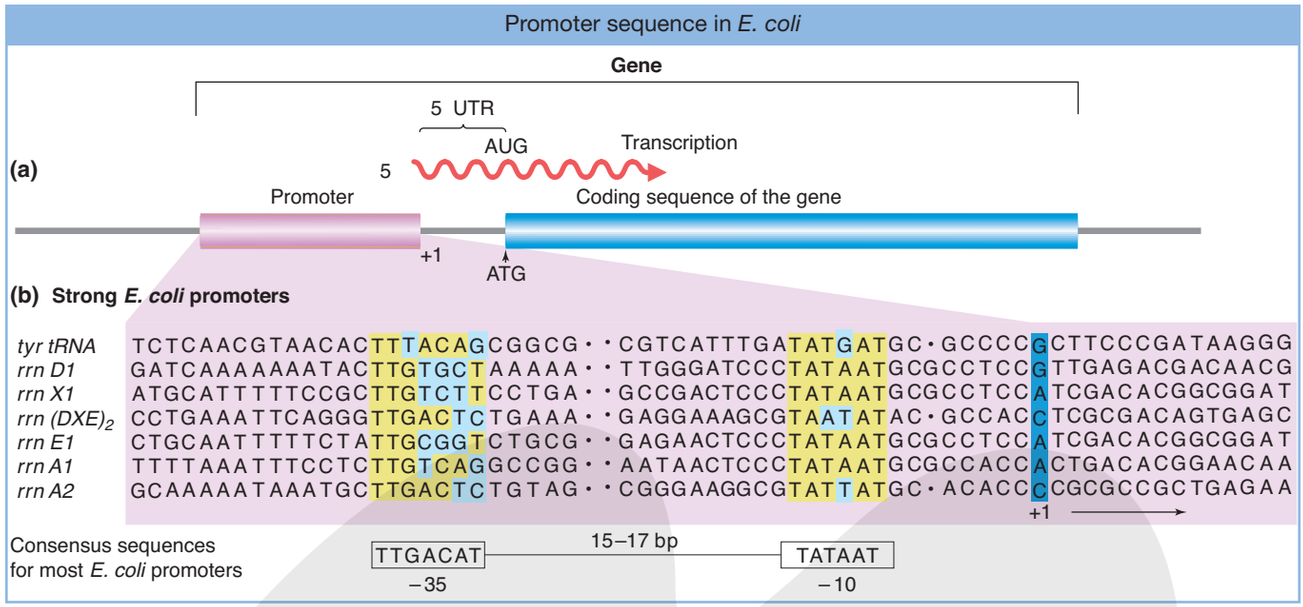


FIGURE 8-7 (a) The promoter lies "upstream" (i.e., 5′) of the transcription start site (+1) and protein-coding sequences. (b) Promoters have regions of similar sequences, as indicated by the yellow shading in seven different promoter sequences in *E. coli*. Spaces (dots) are inserted in the sequence to optimize the alignment of common sequences. Numbers refer to the number of bases before (−) or after (+) the transcription start site.

ANIMATED ART SaplingPlus
Transcription in bacteria

Transcription initiation in bacteria How does RNA polymerase find the correct starting point for transcription? In bacteria, RNA polymerase usually binds to a specific DNA sequence called a **promoter**, located close to the start of the transcribed region. Promoters are an important part of the transcriptional regulatory region of a gene (Figure 8-7). The first transcribed base is called the *initiation site* or the **transcription start site**. The promoter is referred to as **upstream** of the initiation site because it is located ahead of the initiation site (5′ of the gene). A **downstream** site is located later in the direction of transcription. Nucleotide positions upstream of the initiation site are indicated by a negative (−) sign and those downstream by a positive (+) sign. By convention, the first DNA base to be transcribed is numbered +1.

KEY CONCEPT Nucleotides in genes are numbered relative to the transcription start site; those before the start site have negative numbers and are called upstream and those after the start site have positive numbers and are called downstream.

Figure 8-7 shows the promoter sequences of seven different genes in *E. coli*. Because the same RNA polymerase binds to the promoter sequences of these different genes, similarities among the promoters are not surprising. In particular, two regions of great similarity appear in virtually every case. These regions have been termed the −35 (minus 35) and −10 (minus 10) regions because they are located 35 base pairs and 10 base pairs, respectively, upstream of the first transcribed base. They are shown in yellow in Figure 8-7. As you can see, the −35 and −10 regions from different genes do not have to be identical to perform a similar function. Nonetheless, it is possible to arrive at a sequence of nucleotides, called a **consensus sequence**, that is in agreement with most sequences. The *E. coli* promoter consensus sequence is shown at the bottom of Figure 8-7. An RNA polymerase holoenzyme (see the next paragraph) binds to the DNA at this point, then unwinds the DNA double helix and begins the synthesis of an RNA molecule. Note in Figure 8-7 that the protein-coding part of the gene usually begins at an AUG sequence in the mRNA (i.e., ATG in DNA), but the transcription start site, where transcription begins, is usually well upstream of this sequence. The region between the start of transcription and the start of translation is referred to as the **5′ untranslated region (5′ UTR)**.

The bacterial RNA polymerase that scans the DNA for a promoter sequence is called the **RNA polymerase holoenzyme** (Figure 8-8a). This multi-protein complex is composed of a five-subunit **core enzyme** (two subunits of α, one of β, one of β′, and one of ω) plus a subunit called **sigma factor (σ)**. The two α subunits help assemble the enzyme and promote interactions with regulatory proteins, the β subunit is active in catalysis, the β′ subunit binds DNA, and the ω subunit has roles in assembly of the holoenzyme and the regulation of gene expression. The σ subunit binds to the −10 and −35 regions, thus positioning the holoenzyme to initiate transcription correctly at the start site (Figure 8-8a). The σ subunit also has a role in separating the DNA strands around the −10 region so that the core enzyme can bind tightly to the DNA in preparation for RNA synthesis. After the holoenzyme is bound, transcription initiates and the σ subunit dissociates (Figure 8-8b). The core enzyme then elongates through the gene (Figure 8-8c).

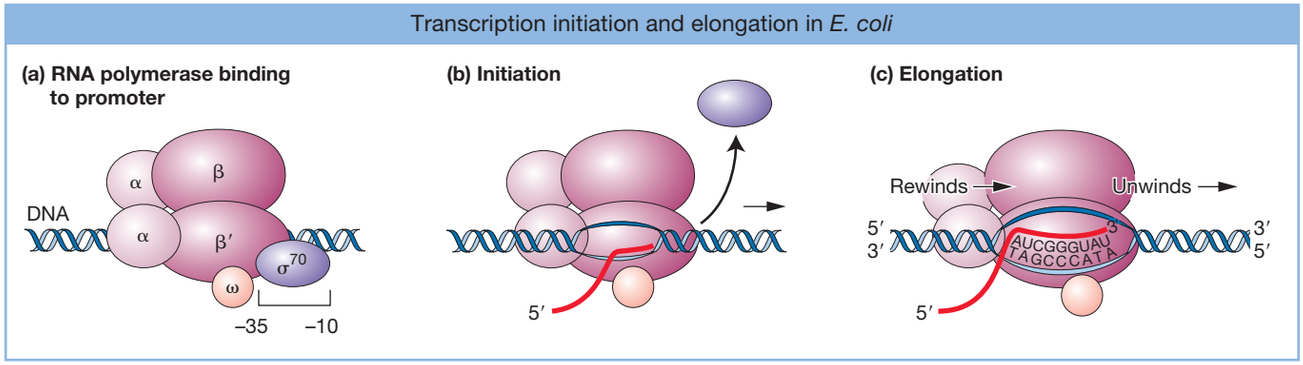


FIGURE 8-8 (a) Binding of the σ subunit to the −10 and −35 regions positions the RNA polymerase holoenzyme for correct initiation. (b) Shortly after RNA synthesis begins, the σ subunit dissociates from the core enzyme, which continues transcription. (c) Synthesis of an RNA strand complementary to the single-strand region of the DNA template is in the 5′-to-3′ direction. DNA that is unwound ahead of RNA polymerase is rewound after it is transcribed.

ANIMATED ART SaplingPlus
Transcription in bacteria

E. coli, like most other bacteria, has several different σ factors. One, called σ^{70} because its mass in kilodaltons (kDa) is 70, is the primary σ factor used to initiate transcription of the vast majority of *E. coli* genes. Alternative σ factors recognize promoters with different consensus sequences. Thus, by associating with different σ factors, the same core RNA polymerase enzyme can transcribe different genes to respond to stresses, changes in cell shape, and nitrogen uptake. This is discussed in greater detail in Chapter 11.

KEY CONCEPT A sequence called a promoter controls where RNA polymerase begins transcription. In bacteria, promoters are bound by RNA polymerase σ factors.

Transcription elongation in bacteria As the RNA polymerase moves along the DNA, it unwinds the DNA ahead of it and rewinds the DNA that has already been transcribed (Figure 8-8c). In this way, it maintains a region of single-stranded DNA, called a transcription bubble, within which the template strand is exposed. In the bubble, RNA polymerase monitors the binding of a free ribonucleoside triphosphate to the next exposed base on the DNA template and, if there is a complementary match, adds it to the chain (Figure 8-5b). Energy for the addition of a nucleotide is derived from breaking a phosphate bond. RNA polymerase synthesizes RNA at a rate of 50 to 90 nucleotides per second. Within this range, slower rates of synthesis may provide time for the RNA to fold properly and for translation to synchronize with transcription.

Inside the transcription bubble, the last 8 to 9 nucleotides added to the RNA chain form an RNA–DNA hybrid by complementary base pairing with the template strand. As the RNA chain lengthens at its 3′ end, the 5′ end is further extruded from the polymerase. The complementary base pairs are broken at the point of exit, leaving the extruded region of RNA single-stranded.

Transcription termination in bacteria Transcription continues beyond the protein-coding segment of a gene, creating a **3′ untranslated region (3′ UTR)** at the end of the transcript. Elongation proceeds until RNA polymerase recognizes special nucleotide sequences that act as a signal to stop transcription and release RNA polymerase and the nascent (i.e., newly synthesized) RNA from the template. There are two major types of termination mechanisms in *E. coli* (and other bacteria), **factor-independent termination** (also called intrinsic or rho-independent) and **Rho-dependent termination** (also called factor-dependent) (Figure 8-9).

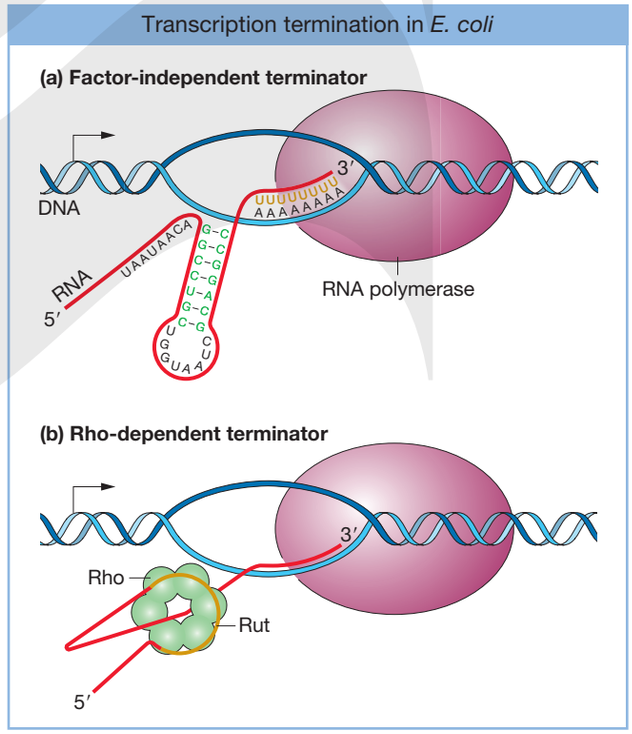


FIGURE 8-9 Transcription termination occurs by two mechanisms that involve different sequences in the RNA. (a) Factor-independent termination depends on a terminator signal consisting of a GC-rich stem-loop structure followed by a stretch of U's. (b) Rho-dependent termination depends on a C-rich binding site for the protein Rho (called a Rut site) that is located upstream of the termination site.

Factor-independent termination occurs after the transcription of a GC-rich stretch followed by an A-rich stretch in the template strand. In the RNA, the GC-rich sequence is self-complementary and forms a 7- to 20-base-pair stem-loop followed by a 7- to 8-nucleotide U-rich tract. Stem-loops are also called hairpins because they resemble the metal clips used to hold a person's hair in place (Figure 8-9a). Normally, in the course of transcription elongation, RNA polymerase will pause if the short RNA–DNA hybrid in the transcription bubble is weak and will backtrack to stabilize the hybrid. The strength of the hybrid is determined by the relative number of 3-hydrogen bond G–C and C–G base pairs compared with 2-hydrogen bond A–T and U–A base pairs. In the factor-independent mechanism, the polymerase is believed to pause after synthesizing the U's (U–A forms a weak RNA–DNA hybrid). However, the backtracking polymerase encounters the hairpin. This roadblock sets off the release of RNA from the polymerase and the polymerase from the DNA template.

In the Rho-dependent termination mechanism, a protein called Rho factor recognizes nucleotide sequences in the RNA that act as a termination signal for RNA polymerase. RNAs with Rho-dependent termination signals do not have the string of U residues at their 3′ end and usually do not have a hairpin (Figure 8-9b). Instead, they have a sequence of about 50–90 nucleotides that is rich in C residues and poor in G residues and includes an upstream segment called a Rut (Rho utilization) site. Rut sites are located upstream (recall that upstream means 5′ of) from sequences at which the RNA polymerase tends to pause. Rho is a homo-hexamer consisting of six identical subunits that has **helicase** activity. Helicases use energy from ATP hydrolysis to move along a nucleic acid and unwind nucleic acid helices. Once bound at the Rut site, Rho travels toward the 3′ end of the transcript. When Rho encounters a paused RNA polymerase, it unwinds the RNA-DNA hybrid within the transcription bubble, dissociating the RNA and terminating transcription. Thus, Rho-dependent termination entails binding of Rho to the Rut site, pausing of RNA polymerase, and Rho-mediated release of the RNA from RNA polymerase.

KEY CONCEPT Special sequences within the 3′ UTR of an mRNA direct transcription termination using mechanisms that are either factor-independent or Rho-dependent.

mRNA decay in bacteria

RNA abundance in cells is determined not only by transcription mechanisms that control the synthesis of RNA by RNA polymerase, but also by decay mechanisms that carry out the destruction of RNAs. RNA destruction, which is commonly called RNA degradation or **decay,** is carried out by ribonucleases. Bacteria have about 25 different ribonucleases. Some are involved in decay and others function in the precise processing of RNA precursors such as cutting the long rRNA precursor into individual functional rRNAs. The measure of decay is **half-life** (also denoted $t_{1/2}$), which is the amount of time it takes for half of the pool of an RNA molecule to be decayed. In bacteria, mRNAs typically have a half-life of less than two minutes. Rapid mRNA decay is thought to allow bacteria to quickly alter gene expression in response to changing nutritional and environmental conditions.

KEY CONCEPT The abundance of an RNA in cells is determined by transcription and decay.

Decay of mRNA in bacteria is commonly initiated by an **endonuclease** that cuts the mRNA into pieces, followed by digestion of the RNA pieces by **exonucleases** that remove nucleotides one at a time from the 3′ end (**Figure 8-10**). In *E. coli,* decay often begins with conversion of the triphosphate at the 5′ end of the RNA to a monophosphate via removal of pyrophosphate (PP_i) by an RNA pyrophosphohydrolase. The 5′-monophosphate serves as a binding site for the main endonuclease RNase E, which cuts single-stranded RNA. The RNase E products are then digested by 3′-to-5′ exonucleases. Because access to the mRNA by RNase E is critical to decay, the presence of ribosomes can affect the half-life of an mRNA. Remember that translation in bacteria occurs while mRNAs are being transcribed. Inefficient translation initiation presumably increases the distance between translating ribosomes, providing greater opportunity for cleavage by RNase E and decreasing the half-life of an mRNA.

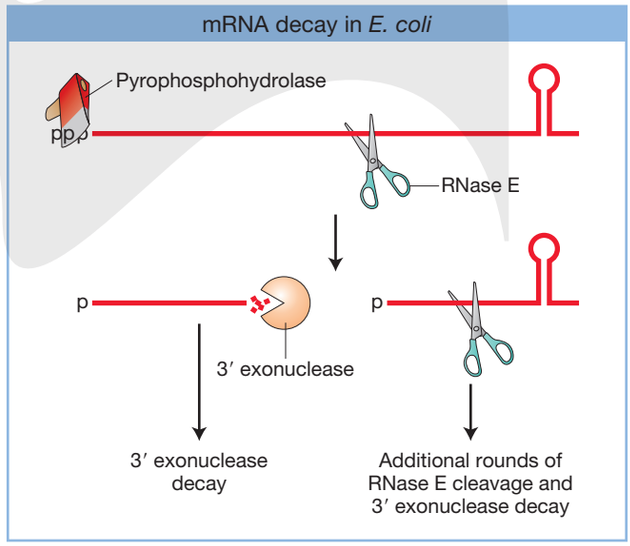


FIGURE 8-10 mRNA decay in *E. coli* is carried out by the sequential activities of an endonuclease (RNase E), which generates RNA fragments, and 3′-to-5′ exonucleases. RNase E recognizes a 5′-monophosphate, which is generated by pyrophosphohydrolase. Exonucleases are commonly drawn as the Pac-Man video game character, since exonucleases eat nucleotides from the ends of RNAs analogous to the way Pac-Man eats through a maze of Pac-Dots.

KEY CONCEPT The first step of decay in bacteria is carried out by an endonuclease, which cuts an RNA into two pieces, and the next step is carried out by exonucleases, which digest the RNA pieces into single nucleotides starting at their 3′ end.

8.3 TRANSCRIPTION IN EUKARYOTES

LO 8.3 Describe how mRNA transcription and decay mechanisms in eukaryotes are similar to those in bacteria.

Transcription in eukaryotic organisms, including humans, is similar to transcription in bacteria in that they retain many of the events associated with initiation, elongation, and termination. However, transcription in eukaryotes is more complex in four main ways.

1. **Eukaryotes have many genes that are spaced far apart.** The larger eukaryotic genomes have many more genes to be recognized and transcribed. Whereas bacteria usually have a few thousand genes, eukaryotes can have tens of thousands of genes. Furthermore, there is much more non-transcribed DNA in eukaryotes. Non-transcribed DNA originates by a variety of mechanisms that will be discussed in Chapter 16. So, even though eukaryotes have more genes than bacteria, their genes are, on average, farther apart. For example, whereas the average gene density in the bacteria *E. coli* is 1 gene per 1400 base pairs, that number drops in eukaryotic organisms to 1 gene per 9000 base pairs for the fruit fly *Drosophila melanogaster* and 1 gene per 100,000 base pairs for humans.
2. **Eukaryotes have three RNA polymerases.** In contrast to bacteria that have one RNA polymerase that transcribes all genes in the cellular genome, eukaryotes have three RNA polymerases that work with distinct initiation, elongation, and termination factors. **RNA polymerase I** transcribes rRNA, excluding 5S rRNA. **RNA polymerase II** transcribes all mRNAs and some ncRNAs, including miRNAs and some snRNAs. **RNA polymerase III** transcribes a variety of ncRNAs, including tRNAs, 5S rRNA, and some snRNAs. Each RNA polymerase is a multisubunit enzyme composed of about 12 proteins, some of which are identical or similar among the polymerases and others of which are unique to one polymerase.

   Unique features of the RNA polymerases carry out polymerase-specific functions. For example, the largest protein in RNA polymerase II contains a unique sequence called the **carboxy-terminal domain (CTD)** that helps coordinate RNA processing events that are exclusive to its transcripts. In addition, since eukaryotic RNA polymerases cannot bind promoters or initiate transcription on their own, each polymerase functions along with its own set of **general transcription factors (GTFs)** that are required to bind promoters and recruit the RNA polymerase to the transcription start site. The roles of GTFs and their interactions with RNA polymerases will be described in the section on eukaryotic transcription initiation.
3. **Transcription in eukaryotes takes place in the nucleus.** An important cellular difference between bacteria and eukaryotes is the presence of a nucleus in eukaryotes (see Figure 8-2). Because bacteria lack a nucleus, the information in RNA is almost immediately translated into protein, as described in Chapter 9. In eukaryotes, the nuclear membrane spatially separates transcription and translation—transcription takes place in the nucleus and translation in the cytoplasm. It also means that mechanisms exist to export RNAs from the nucleus to the cytoplasm. Additionally, before RNAs leave the nucleus, they are modified in several ways. Both ends of an mRNA are chemically modified to protect against degradation: capping at the 5′ end and polyadenylation at the 3′ end. These modifications and others are collectively referred to as **RNA processing**. Newly synthesized RNAs that are not yet processed are called primary transcripts or **precursor RNAs (pre-RNAs)**, for example, pre-mRNA and pre-rRNA. RNA processing often occurs co-transcriptionally, that is, while the RNA is being transcribed. Thus, RNA polymerases synthesize RNA while simultaneously coordinating a variety of processing events.
4. **DNA in eukaryotes is packaged with proteins into chromatin.** The template for transcription, genomic DNA, is tightly wrapped around proteins to form chromatin in eukaryotes, whereas DNA is less compacted in bacteria. The structure of chromatin can affect transcription initiation, elongation, and termination by all three RNA polymerases as well as the processing of their transcripts. These chromatin-based mechanisms will be covered in Chapter 12.

KEY CONCEPT Differences in transcription between eukaryotes and bacteria are related to (1) larger eukaryotic genomes with genes that are spaced further apart, (2) the division of transcription in eukaryotes among three RNA polymerases, (3) the nuclear membrane in eukaryotic cells that decouples transcription and translation and necessitates nuclear RNA export, and (4) the tight packaging of eukaryotic genomic DNA into chromatin.

Transcription initiation in eukaryotes

RNA polymerases I, II, and III cannot recognize promoter sequences on their own. However, unlike bacteria, where promoters are recognized by σ factor as an integral part of the RNA polymerase holoenzyme, eukaryotic promoters are recognized by GTFs that first bind specific sequences in the promoter and then bind the RNA polymerase. Nevertheless, the mechanisms in bacteria and eukaryotes are conceptually similar. In both cases, the information that defines a promoter is provided by short DNA sequences located near the transcription start site, and

KEY CONCEPT RNA polymerase I, II, and III genes have unique promoters that direct transcription initiation. Promoters are first recognized by RNA polymerase-specific general transcription factors (GTFs). One of the main functions of GTFs is to recruit a specific RNA polymerase and position it to begin RNA synthesis at the transcription start site.

the sequences are bound by proteins that associate with RNA polymerase and position it at the correct site to start transcription.

RNA polymerase I promoters and GTFs In eukaryotic organisms, ribosomal RNA (rRNA) is transcribed by RNA polymerase I from hundreds of near-identical copies of rDNA genes that are tandemly repeated in the genome and reside in the **nucleolus**, a non–membrane-bound region in the nucleus where rRNA transcripts are synthesized, processed, and assembled with proteins into ribosomes (see Figure 8.6). Each rDNA gene encodes a single rRNA transcript that contains 18S, 5.8S, and 28S rRNAs along with an external transcribed spacer (ETS) and internal transcribed spacers (ITSs) (**Figure 8-11**, top). After transcription, spacer regions are removed by processing enzymes, 18S rRNA is assembled with ribosomal proteins to form the 40S small ribosomal subunit, and 5.8S, 28S, and 5S rRNAs are assembled with ribosomal proteins to form the 60S ribosomal subunit. RNA polymerase III transcribes 5S rRNA from tandem arrays of hundreds of gene copies located at different places in the genome from those of the RNA polymerase I genes. Note that the "S" in 18S, 5.8S, 28S, 5S, 40S, and 60S stands for Svedberg units, which is a measure of a molecule's size that is based on its rate of sedimentation upon centrifugation.

Between the tandemly repeated rDNA genes are intergenic spacers (IGSs) that contains two promoter elements important for transcription initiation. A Core element is located at the transcription start site, and an Upstream Control Element (UCE) is located 100 to 150 base pairs upstream of the transcription start site (Figure 8-11, bottom). In humans, the Core element is bound by a multi-protein complex called Selectivity Factor 1 (SL1), which contains TATA-binding protein (TBP), and the UCE is bound by Upstream Binding Factor (UBF). In addition to SL1 and UBF, the protein TIF-1A (Transcription Initiation Factor 1A) is also required for recruitment of RNA polymerase I to the transcription start site of rDNA genes. Unlike SL1 and UBF, TIF-1A does not bind DNA, but instead it functions through protein-protein interactions and forms a bridge between SL1 and RNA polymerase I.

Every time a cell divides, the number of ribosomes must be doubled to keep the number of ribosomes in the two daughter cells equal to that in the parent cell. Mammalian cells have one to two million ribosomes. Therefore, every cell division involves the production of one to two million rRNA transcripts by RNA polymerase I. Consequently, the RNA polymerase I transcription mechanism has evolved to be exquisitely sensitive to environmental conditions that promote or inhibit cell proliferation (i.e., an increase in the number of cells). Conditions that affect cell proliferation can act on transcription initiation by RNA polymerase I by altering the activities of SLI, UBF, and TIF-1A. For example, conditions that promote cell proliferation lead to phosphorylation of TIF-1A on a specific serine residue, increasing its ability to recruit RNA polymerase I and trigger transcription initiation.

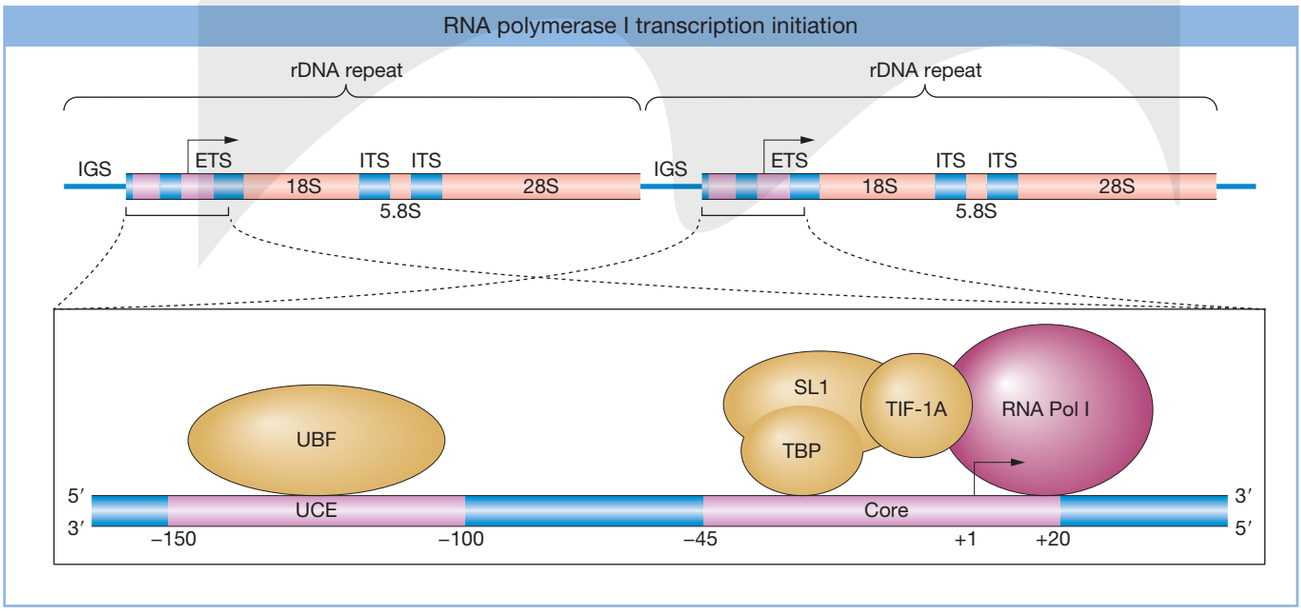


FIGURE 8-11 The information that controls transcription initiation by RNA polymerase I is contained in rDNA promoter sequences (UCE and Core) that are located upstream of the transcription start site (+1) and are bound by protein factors (UBF and SL1). A third factor, TIF-1A, does not directly bind DNA but is important for the recruitment and function of RNA polymerase I. UCE = Upstream Control Element, UBF = Upstream Binding Factor, TBP = TATA Binding Protein, SL1 = Selectivity Factor 1, TIF-1A = Transcription Initiation Factor 1A.

KEY CONCEPT In the nucleolus, initiation of rDNA gene transcription by RNA polymerase I is regulated by UCE and Core promoter elements and UBF, SL1, and TIF-1A regulatory factors.

RNA polymerase II promoters and GTFs RNA polymerase II is responsible for transcribing all mRNAs as well as numerous ncRNAs, including snRNAs involved in splicing and miRNAs involved in mRNA decay and inhibition of translation. Transcription by RNA polymerase II is the most complex transcriptional system in eukaryotes because of the large number of gene targets with unique expression patterns.

RNA polymerase II promoters, which are somewhat arbitrarily defined as sequences located within 100 base pairs of the transcription start site, contain a variety of promoter elements, a few of which are relatively common (**Figure 8-12**). About 25 percent of promoters in yeast and humans contain a **TATA box**, an sequence element so-named because the nucleotide sequence TATA appears in the consensus sequence TATAAAA. In animals, the TATA box is located about 30 nucleotides upstream of the transcription start site (i.e., −30), but in yeast its location is more variable (between −50 and −125). Another common promoter element is the initiator (Inr), which is located right at the transcription start site in about 40 percent of genes.

Collectively, only about 50 percent of RNA polymerase II genes contain a TATA box and/or an Inr. This predicts the existence of other promoter elements. Computational analyses that searched for common sequences surrounding the transcription start sites of RNA polymerase II genes have identified additional promoter elements, including the downstream promoter element (DPE), which is located at about +25, and the TFIIB recognition element (BRE), which is located at about −40. However, since many genes lack all of the known promoter elements, it is likely that promoter elements remain to be discovered.

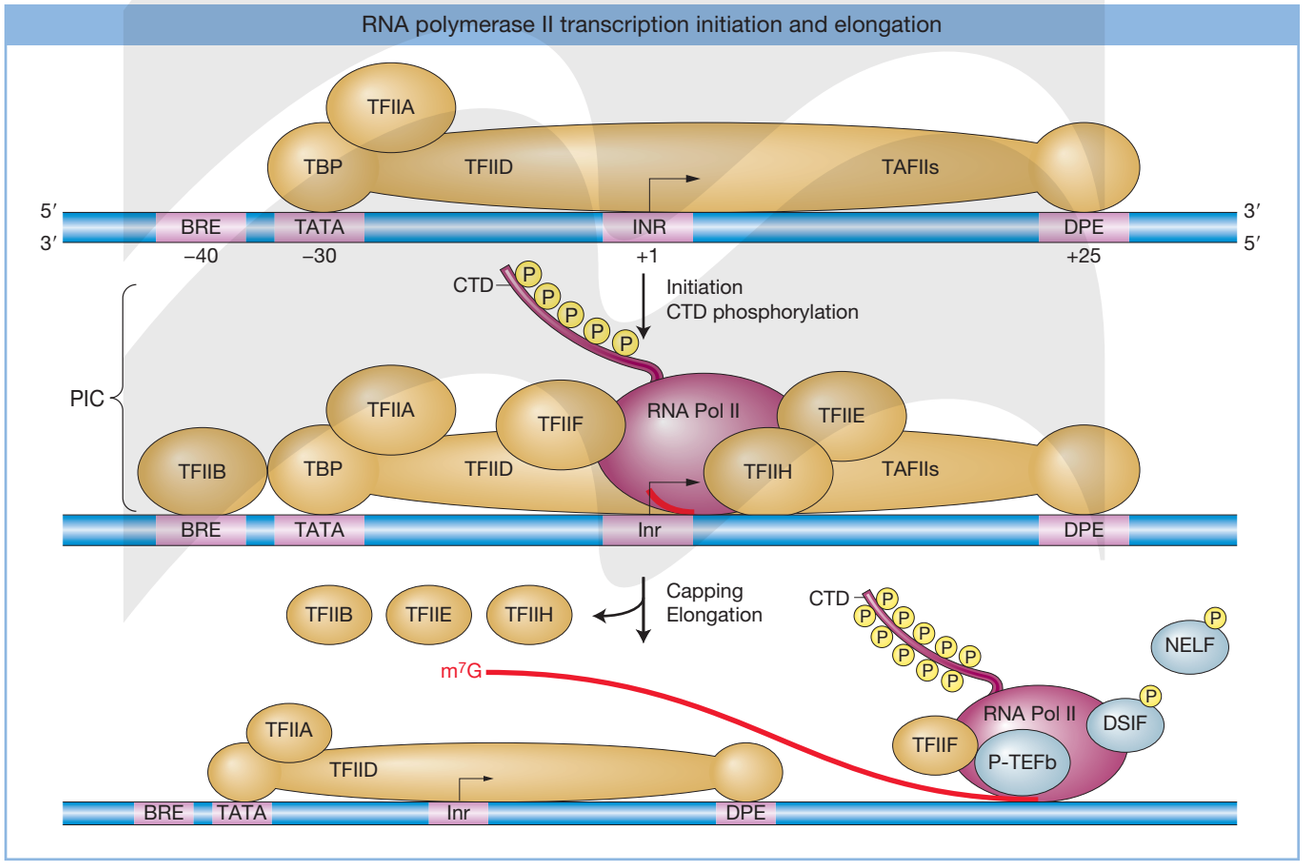


FIGURE 8-12 Transcription initiation by RNA polymerase II is directed by a variety of promoter elements, including the BRE, TATA box, Inr, and DPE, located within 100 base pairs upstream or downstream of the transcription start site (+1). Assembly of the PIC occurs in a sequential manner, starting with TFIID, which contains proteins (TBP and TAFs) that bind promoter elements. Assembly of TFIID leads to recruitment of the other GTFs and RNA polymerase II. TFIIH is required during initiation for opening the transcription bubble and phosphorylation of the CTD. Shortly after initiation, the pre-mRNA is capped and elongation is promoted by the P-TEFb kinase, which phosphorylates the RNA polymerase II CTD, DSIF, and NELF.

ANIMATED ART SaplingPlus
Transcription in eukaryotes

KEY CONCEPT About half of RNA polymerase II genes contain TATA box and/or Inr promoter elements. The other half contain less common promoter elements, some of which remain to be defined.

All of the proteins that bind RNA polymerase II promoter elements are subunits of GTFs. Transcription factor IIB (TFIIB) binds the BRE, and TFIID binds the other promoter elements. The TFIID complex contains TBP (the same protein involved in RNA polymerase I transcription) and about 15 TBP-associated factors (TBP). TBP binds the TATA box, and TAFs bind the Inr and DPE. Binding of TFIID at a promoter is the first step in the sequential assembly of other GTFs and RNA polymerase II. TFIID binding instructs assembly of TFIIA and TFIIB, followed by TFIIF and RNA polymerase II as a pre-assembled complex and ending with addition of TFIIE and TFIIH. The assemblage of GTFs and RNA polymerase II constitutes the **preinitiation complex** (**PIC**), which serves to position RNA polymerase II at the transcription start site, generate the transcription bubble, and position the DNA in the active site of RNA polymerase II. TFIIA stabilizes the binding of TFIIB and TFIID at the promoter. TFIIH, which is recruited to the promoter by TFIIE, contains proteins with helicase activity that unwind the DNA into two strands to form the transcription bubble. Lastly, TFIIF places the promoter DNA in a position in RNA polymerase II that is appropriate for DNA unwinding and initiation of transcription at the start site. After transcription has been initiated, RNA polymerase II dissociates from most of the GTFs to elongate the RNA transcript. Some GTFs, including TFIID, remain at the promoter to attract the next RNA polymerase II. In this way, multiple RNA polymerase II molecules can simultaneously synthesize transcripts from a single gene.

KEY CONCEPT The general transcription factors (GTFs) TFIIB and TFIID recruit RNA polymerase II and other GTFs to the promoter, forming a pre-initiation complex (PIC) and a transcription bubble.

RNA polymerase III promoters and GTFs RNA polymerase III transcribes noncoding RNAs (ncRNAs) shorter than 300 nucleotides. The RNA polymerase III gene targets are classified into three types based on their promoter elements. The 5S rRNA genes are Type 1 and contain three promoter elements, Box A, intermediate element (IE), and Box C, that are all located downstream of the transcription start site (Figure 8-13a).

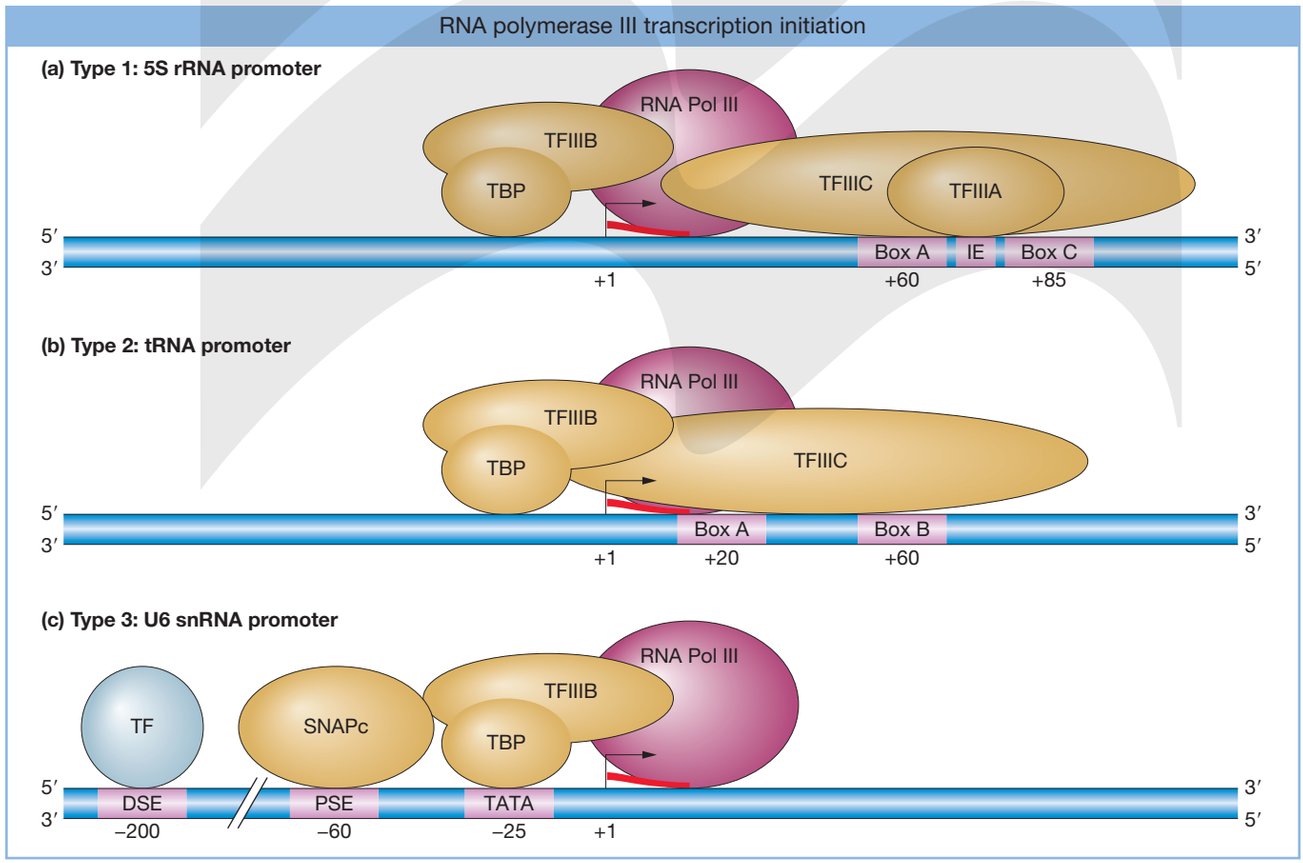


FIGURE 8-13 RNA polymerase III transcribes genes with three types of promoters, Types 1, 2, and 3 (a–c). TFIIIA, TFIIIC, and $SNAP_c$ bind promoter elements (IE, Box B, and PSE, respectively) that are unique to each type of RNA polymerase III gene and can be viewed as the specificity factors for the gene type. The main function of these factors is to recruit TFIIIB to the promoter, which then leads to recruitment of RNA polymerase III and transcription initiation.

tRNA genes are Type 2 and contain the same Box A element as 5S genes, but instead of a Box C they have a Box B (Figure 8-13b). Lastly, Type 3 genes such as the U6 snRNA gene that is involved in pre-mRNA splicing contain promoter elements that are located upstream of the transcription start site and look very similar to promoter elements found in RNA polymerase II genes, including a TATA box (Figure 8-13c).

The GTFs for RNA polymerase III transcription are designated TFIIIA, TFIIIB, TFIIIC (transcription factor for RNA polymerase III), and $SNAP_c$ (snRNA activating protein complex). TFIIIB is required for the transcription of all three types of RNA polymerase III genes and functions to recruit RNA polymerase III as well as open the transcription bubble. TFIIIA binds promoter elements in Type 1 genes and helps recruit TFIIIC, which binds promoter elements in Type 1 and 2 genes. Note that since TFIIIA and TFIIIC bind downstream of the transcription start site, they need to be temporarily displaced as RNA polymerase III transcribes through the promoter DNA. For Type 3 genes, recruitment of RNA polymerase III is assisted by binding of both the TBP subunit of TFIIIB to the TATA box and of $SNAP_c$ to the proximal sequence element (PSE). Binding of $SNAP_c$ to the PSE is helped by transcription factors that bind to the distal sequence element (DSE). Transcription factors are discussed in detail in Chapter 12.

KEY CONCEPT Genes transcribed by RNA polymerase III are divided into three types, based on their promoter elements. Type 1 and Type 2 have promoter elements downstream of the transcription start site. Type 3 promoter elements include a TATA box and are positioned upstream of the transcription start site.

RNA polymerase II transcription elongation

Shortly after transcription initiation, phosphorylation of RNA polymerase II by a protein kinase in TFIIH helps coordinate the processing of mRNAs as they are being transcribed. The carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase II contains the sequence YSPTSPS (tyrosine-serine-proline-threonine-serine-proline-serine) tandemly repeated 26 times in yeast and 52 times in humans. Phosphorylation of the serine in position 5 of the repeat (S5) by TFIIH serves as a signal for the binding of enzymes that cap the 5′ end of the mRNA (discussed in the next section) (**Figure 8-14**). The CTD is located near the site where nascent RNA emerges from RNA polymerase II, so it is in an ideal place to orchestrate the binding and release of proteins needed to process the nascent transcript while RNA synthesis continues. Post-translational modification of S5 and other amino acids in the CTD change as RNA polymerase II transcribes through a gene, creating different binding sites for other processing factors as well as factors that regulate transcription elongation and termination.

KEY CONCEPT During elongation, the CTD of RNA polymerase II is chemically modified to serve as a binding site for other proteins involved in transcription and RNA processing.

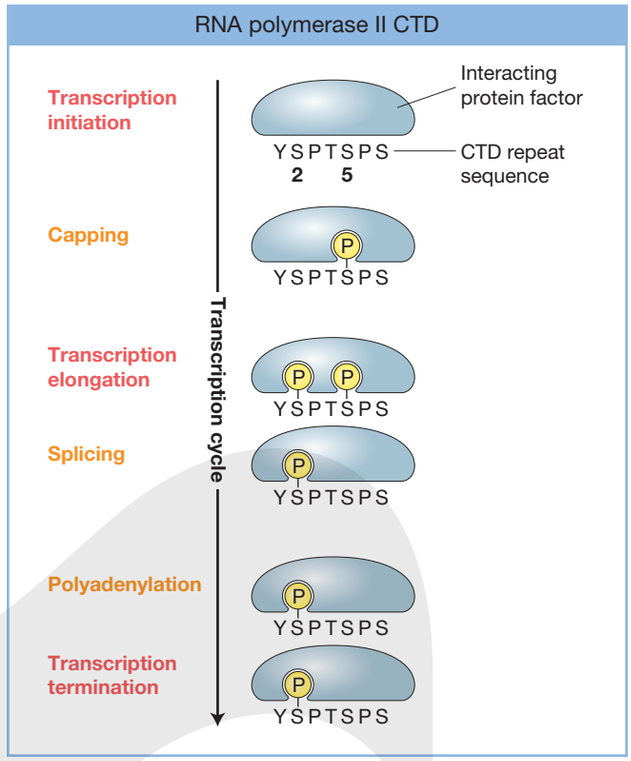


FIGURE 8-14 The pattern of amino acid modifications within the RNA polymerase II CTD changes as the polymerase transcribes through a gene. The modifications, including phosphorylation of serines 2 and 5 by kinases and dephosphorylation by phosphatases, create a code that directs the binding of factors at specific stages of transcription. Factors that bind the CTD include regulators of transcription elongation and termination as well as RNA processing events, including capping, splicing, and cleavage and polyadenylation.

Transcription initiation is not a green light that liberates RNA polymerase II to completely transcribe a gene unimpeded. In fact, transcription continues to be regulated all along the length of a gene. For example, for a large fraction of human genes, transcription elongation is temporarily stopped (i.e., paused) about 50 base pairs downstream of the transcription start site. Pausing of this type is caused by the protein factors NELF (negative elongation factor) and DSIF (DRB sensitivity-inducing factor) and relieved by P-TEFb (positive transcription elongation factor b) (see Figure 8-12). To release paused RNA polymerase II into productive elongation, P-TEFb phosphorylates NELF and DSIF. NELF dissociates from the elongation complex, and DSIF travels along with RNA polymerase II and functions as a positive elongation factor. P-TEFb also phosphorylates the RNA polymerase II CTD on serine 2 (S2) within the YSPTSPS repeats (see Figure 8-14), which serves as a signal for the binding of factors involved in processing of the pre-mRNA and transcription termination.

Transcription termination in eukaryotes

Transcription termination for the three RNA polymerases occurs by different mechanisms. Elongating RNA polymerase I is stopped by protein factors bound at specific DNA sequences called terminator elements and is released from DNA by other factors. In contrast, RNA polymerase III terminates elongation

and dissociates from DNA after the synthesis of a poly(U) stretch, similar to factor-independent termination in bacteria.

Two models have been proposed for transcription termination by RNA polymerase II—the torpedo model and the allosteric model. The models are conceptually similar to Rho-dependent and factor-independent mechanisms, respectively, in *E. coli* (see Figure 8-9), but different factors are involved. Both RNA polymerase II termination models couple 3′-end formation to termination. As described below in the section on polyadenylation, the 3′ ends of mRNAs are determined by cleavage of the pre-mRNA and addition of a poly(A) tail to the new 3′ end.

In the **torpedo termination model** (Figure 8-15a), RNA polymerase II continues to transcribe past the site of cleavage, the pre-mRNA is cleaved, and the new 5′-monophosphorylated end that is formed is a substrate for a 5′-to-3′ exonuclease called Xrn2, which digests the RNA one nucleotide at a time until eventually reaching RNA polymerase II and causing it to dissociate from DNA. Xrn2 is positioned to act in termination through its association with the CTD phosphorylated on serine 2 (see Figure 8-14).

In the **allosteric termination model** (Figure 8-15b), transcription through the site of cleavage causes elongation factors to dissociate, leading to a conformational change within the active site of RNA polymerase II and its release from DNA. In this model, it remains to be determined how RNA polymerase II senses passage through the site of cleavage and how this leads to dissociation of elongation factors.

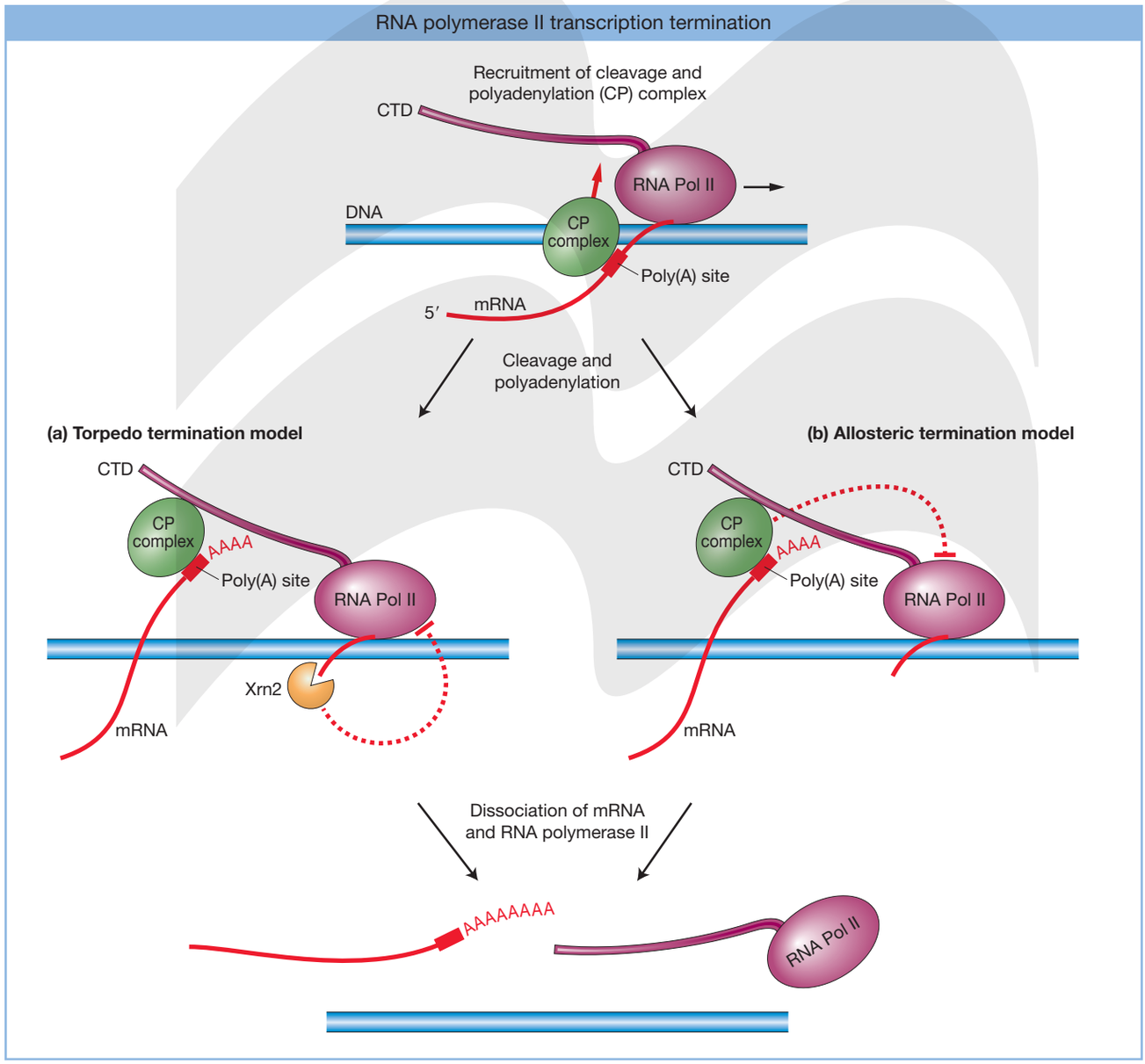


FIGURE 8-15 Current data support two models for how transcription of mRNAs by RNA polymerase II is terminated. (a) In the torpedo model, the piece of RNA that stays associated with RNA polymerase II and continues to be synthesized after cleavage is a substrate for the 5′-to-3′ exonuclease Xrn2 that degrades the RNA to elicit termination. (b) In the allosteric model, upon encountering cleavage and polyadenylation signals, RNA polymerase II undergoes a conformation change that commits it to termination.

KEY CONCEPT Transcription termination by RNA polymerases I, II, and III occurs by different mechanisms. Termination of mRNA transcription by RNA polymerase II may occur by allosteric or torpedo mechanisms that are analogous to factor-independent and Rho-dependent mechanisms, respectively, in *E. coli* and are directed by 3′-end formation sequences in the mRNA.

8.4 PROCESSING OF mRNA IN EUKARYOTES

LO 8.4 Explain how mRNA processing, editing, and modification occur and can affect the abundance and sequence of proteins in eukaryotes.

Unlike bacterial mRNAs, eukaryotic mRNAs undergo numerous processing events that affect their structure and function. Many of these events occur at the same time as transcription (i.e., co-transcriptionally) and are coordinated with transcription initiation, elongation, or termination as well as with one another. Some themes emerge upon comparison of the processing events: (1) sequence elements within mRNAs often direct where processing occurs, and (2) sequence elements are bound by proteins or ncRNAs that are themselves enzymes or that recruit enzymes to carry out the processing.

Capping

RNAs synthesized by RNA polymerase II, including mRNAs and snRNAs, are modified at their 5′ end by addition of a methylated guanine nucleotide, **7-methylguanosine** ($\mathbf{m^7G}$), more commonly referred to as a **cap** because it covers the "head" of the RNA (Figure 8-16a). The 5′ cap is added during transcription when the RNA is about 25 nucleotides long and has just emerged from the exit channel of RNA polymerase II. The process of adding a cap (capping) involves the sequential action of three enzymes

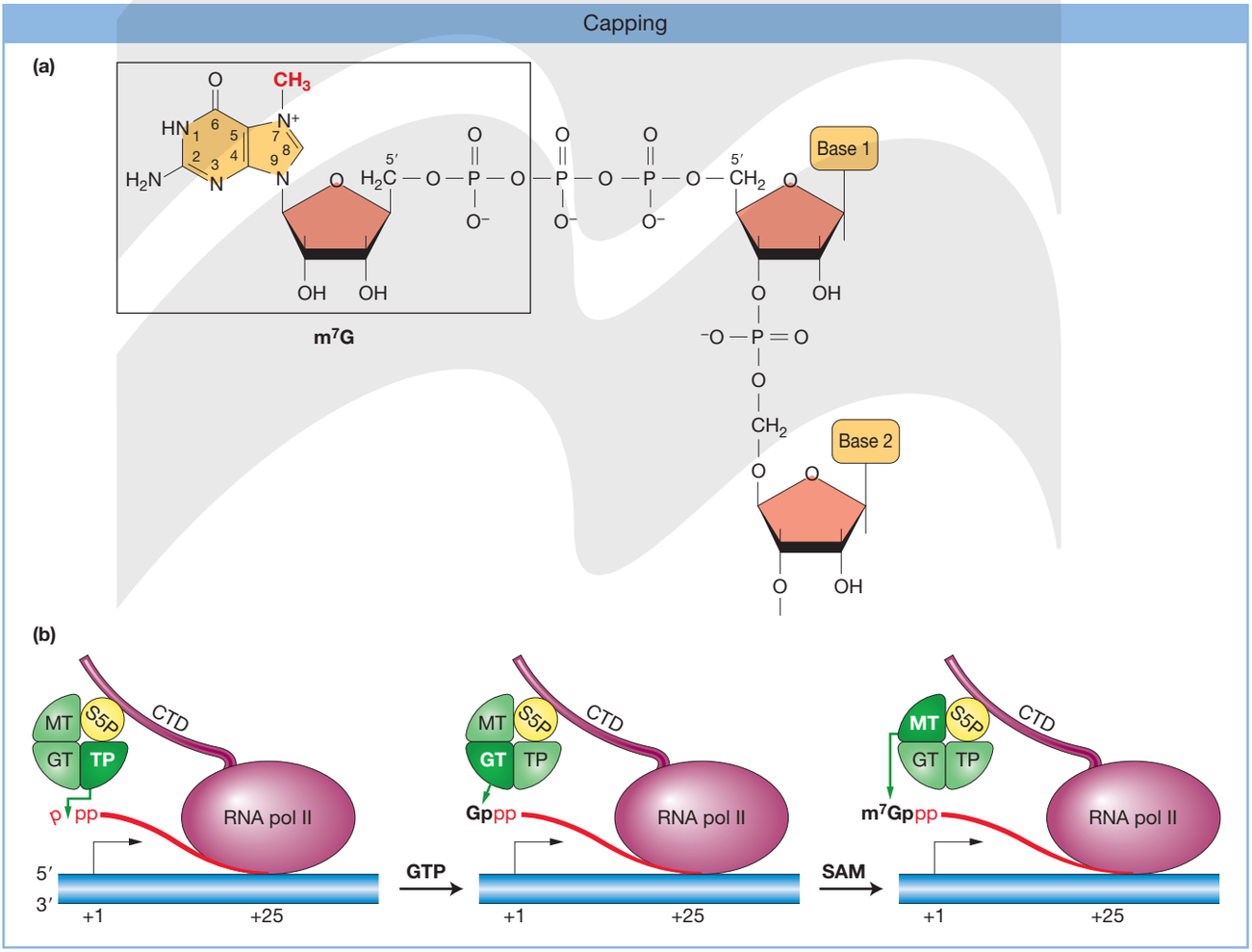


FIGURE 8-16 (a) The first nucleotide in an mRNA is modified by addition of an m^7G cap. (b) Capping is initiated by a triphosphatase (TP) enzyme that removes the terminal phosphate from the first nucleotide. Then a guanylyltransferase (GT) adds a guanine monophosphate nucleotide, and a methyltransferase (MT) adds a methyl group (CH_3). All three enzymes are positioned to act on the mRNA shortly after it emerges from the exit tunnel of RNA polymerase II because of their association with the CTD that is phosphorylated on serine 5 (S5P) of repeat sequences.

ANIMATED ART SaplingPlus Transcription in eukaryotes

(Figure 8-16b): RNA triphosphatase removes the gamma phosphate from the first nucleotide in the RNA chain, guanylyltransferase uses GTP as a substrate and links GMP to the first nucleotide by an unusual 5′,5′-triphosphate linkage (GpppN, where N is the first nucleotide in the mRNA), and 7-methyltransferase adds a methyl group (CH_3) from *S*-adenosylmethionine (SAM) to the N-7 position of the newly added guanine base. Through their interaction with the CTD of RNA polymerase II, these enzymes are in position to act on RNAs early in the transcription process (see Figure 8-14).

Caps serve multiple functions. They protect RNAs from decay by exonucleases, which often require 5′-phosphates to recognize their substrates. Caps on mRNAs also serve as a binding site for proteins such as the cap binding complex (CBC) that mediate subsequent events, including splicing, polyadenylation, and nuclear export, by interacting with processing and export factors. The CBC is also critical for the first round of translation, while another cap binding protein, eIF4E, is required for subsequent rounds of translation, as described in Chapter 9.

KEY CONCEPT The 5′ end of a eukaryotic mRNA is modified to prevent decay and to serve as a binding site for factors that mediate mRNA processing and export. Capping of mRNAs is programmed to occur early in transcription through the association of capping enzymes with phosphorylated serine 5 on the CTD of RNA polymerase II.

Polyadenylation

Like the 5′ end, the 3′ end of an mRNA is also processed to protect it from decay and to promote translation. Processing at the 3′ end consists of two events: cleavage, which cuts the mRNA away from the transcribing RNA polymerase II, and **polyadenylation**, which adds 50–250 adenosine (A) residues to the end of the cleaved mRNA. Sequence elements within the 3′ UTR determine where cleavage occurs.

In humans, the highly conserved six-nucleotide (hexanucleotide) sequence AAUAAA is located 10–30 nucleotides upstream of the cleavage site, also known as the poly(A) site (**Figure 8-17a**). A less well conserved U-rich or GU-rich downstream sequence element (DSE) is located 20–40 nucleotides downstream of the poly(A) site. Cleavage at the poly(A) site often occurs after a CA or UA (Figure 8-17b). The AAUAAA is important for both cleavage and polyadenylation because it is bound by a protein complex called Cleavage and Polyadenylation Specificity Factor (CPSF), which contains the endonuclease enzyme that executes the cleavage step. CPSF also recruits **poly(A) polymerase (PAP),** the enzyme that uses ATP as a substrate to add a string of A's onto the 3′-OH of the mRNA, which is referred to as a **poly(A) tail.** PAP is an unusual RNA polymerase because, unlike DNA polymerases and other RNA polymerases, it does not copy a nucleic acid template. In addition to CPSF, the cleavage step involves CstF (Cleavage Stimulatory Factor), which binds the DSE and helps determine the site of cleavage that is bound by Cleavage Factors I and II (CFI and CFII). During its synthesis, the poly(A) tail is bound by poly(A) binding protein (PABP), which in the cytoplasm protects the mRNA from decay by exonucleases and promotes translation by interacting with the translation machinery.

KEY CONCEPT The 3′ end of mRNAs is modified by addition of a long stretch of adenosine nucleotides, which protects the mRNA from decay and supports translation. The poly(A) tail is added by a special type of RNA polymerase following mRNA cleavage at a site that is determined by protein factors that bind sequence elements in the mRNA.

The discovery of splicing

Transcription copies the DNA sequence of protein-coding genes into mRNA, yet sequence comparison of most pairs of human mRNAs and genes shows that they are different: large stretches of DNA sequence are transcribed into RNA and later removed from the RNA. In 1977, the laboratories of Philip Sharp and Richard Roberts independently discovered this process of mRNA **splicing**, which removes segments of mRNA called **introns** and links together the remaining segments called **exons**. The cutting out of introns and the joining of exons is called splicing because it resembles the way in which movie film is cut and rejoined to delete a specific segment.

KEY CONCEPT The sequence of an mRNA is not always identical to its gene sequence because, as pre-mRNAs are transcribed, introns are removed and the exons that remain are joined together in the process of splicing.

The number and size of introns varies from gene to gene and from organism to organism. For example, only about 5 percent of the genes in yeast (*S. cerevisiae*) have introns. Intron-containing genes in yeast almost always have a single intron that ranges in length from 50 to 1000 nucleotides, with an average length of 250 nucleotides. In contrast, 85 percent of human genes have at least one intron, and an average human gene has eight introns and nine exons. An extreme example is the gene that is mutated in Duchenne muscular dystrophy, which has 78 introns and 79 exons spread across 2.3 million base pairs. Human introns vary in length from 50 to 300,000 nucleotides, with an average of 6000 nucleotides, whereas the average exon length is 300 nucleotides. The 20-fold-larger average size of introns relative to exons means that introns account for a much greater fraction of the human genome than exons.

The splicing mechanism

After the discovery of exons and introns, researchers turned their attention to the mechanism of mRNA splicing. Because splicing must occur with single nucleotide precision to maintain the information that directs translation, the intron-containing mRNA precursor (pre-mRNA) must hold the information that points the splicing machine called

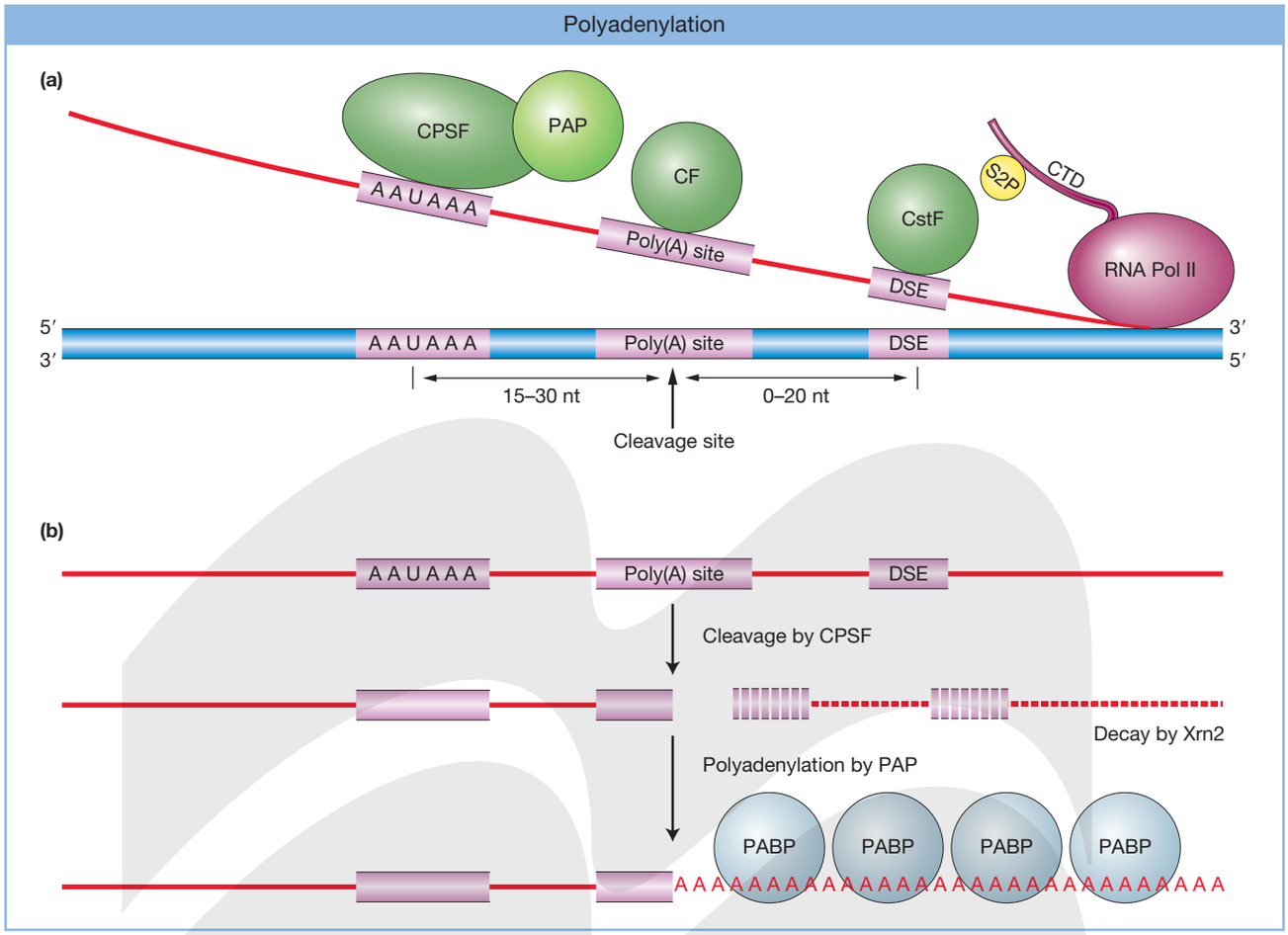


FIGURE 8-17 The 3′ end of an mRNA is generated by consecutive cleavage and polyadenylation reactions. (a) In humans, the site of cleavage is directed by three sequences in the 3′ UTR of an mRNA (i.e., AAUAAA, poly(A) site, and DSE), each of which is bound by protein factors (i.e., CPSF, CF, and CstF, respectively) that are recruited to the ends of genes by association with the RNA polymerase II CTD phosphorylated on serine 2 (S2P) of repeat sequences. (b) Cleavage at the poly(A) site by CPSF is followed by addition of a poly(A) tail to the new 3′ end by PAP and by association of PABP with the poly(A) tail.

the **spliceosome** where to act. Researchers hypothesized that the information would be provided by sequences at the boundaries between exons and introns. In fact, alignments of boundary sequences for many pre-mRNAs revealed that almost all introns begin with GU and end with AG (**Figure 8-18a**). In addition, there is high conservation of intron and exon nucleotides adjacent to the GU and AG. The GU and AG sequence elements define the **5′ splice site** and **3′ splice site,** respectively, where cuts are made by the spliceosome to remove the intron. In addition, a third conserved sequence called the **branch point** is located 15–45 nucleotides upstream of the 3′ splice site. An invariant adenosine within the branch point participates in the first catalytic step of splicing. The existence of conserved nucleotide sequences at splice sites and the branch point suggested that components of the spliceosome are directed to act at specific places in pre-mRNAs by binding to these sequences.

A serendipitous finding in the laboratory of Joan Steitz led to the discovery of components of the spliceosome. Patients with a variety of autoimmune diseases, including systemic lupus erythematosus, produce antibodies against their own proteins. In the course of analyzing blood samples from patients with lupus, Steitz and colleagues identified antibodies that bound nuclear RNA-protein complexes called **small nuclear ribonucleoproteins (snRNPs),** pronounced "snurps," that are comprised of a small nuclear RNA (snRNA) 100–200 nucleotides in length that serves as a scaffold for binding several proteins. They observed that the sequence at the 5′ end of the snRNA named U1 has extensive complementarity to the sequence at 5′ splice sites, suggesting that the U1 snRNA identifies 5′ splice sites by base pairing (Figure 8-18b). To test this hypothesis, the laboratory of Alan Weiner performed a mutational analysis. They found that splicing was dramatically reduced by mutations in a 5′ splice site sequence that partially disrupted base pairing with the U1 snRNA (Figure 8-18c). Moreover, they found that splicing of the mutant pre-mRNA was recovered by mutations in the 5′ end of the U1 snRNA that restored base pairing. This

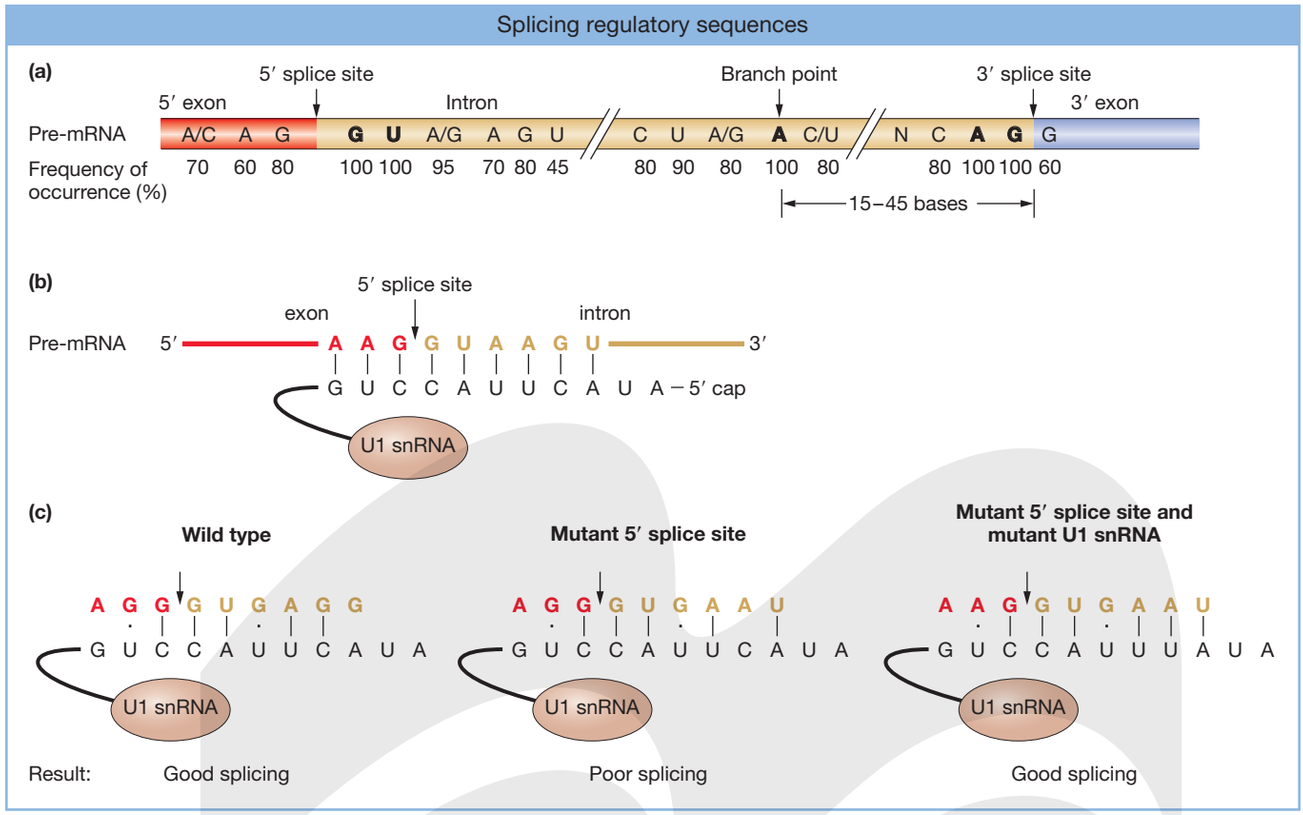


FIGURE 8-18 (a) Conserved nucleotide sequences are present at the junctions of exons and introns (i.e., 5′ and 3′ splice sites) as well as at the branch point, which is near the 3′ splice site. Invariant nucleotides (GU at the 5′ splice site, A at the branch point, and AG at the 3′ splice site) are shown in bold font, and N represents any base. (b) The U1 snRNA recognizes 5′ splice sites by base pairing. Sequences near the 5′ end of the U1 snRNA form eight consecutive base pairs with a consensus 5′ splice site. (c) The efficiency of splicing is affected by the strength of U1 snRNA base pairing at the 5′ splice site. Mutations in the 5′ splice site that reduce the number of hydrogen bonds lead to a decrease in the efficiency of splicing (compare *left* and *middle*); however, splicing efficiency can be restored by compensatory mutations in the U1 snRNA (*right*).

"compensatory mutation" analysis demonstrated that base pairing between snRNAs and the pre-mRNA is important for the selection of splice sites.

KEY CONCEPT snRNAs facilitate splicing by base pairing with conserved sequences in the pre-mRNA.

In addition to the U1 snRNP, the spliceosome contains U2, U4, U5, and U6 snRNPs as well as many proteins that have conserved functions in eukaryotes from yeast to humans. The splicing reaction begins with stepwise recognition of pre-mRNA sequence elements (**Figure 8-19a**). First, U1 binds the 5′ splice site and U2 binds the branch point, with the U2 snRNA base pairing to nucleotides across the branch point, except for the key adenosine. Spliceosome assembly is completed by entry of the U4, U5, and U6 snRNPs as a preassembled tri-snRNP complex. At this point, the spliceosome undergoes several conformational changes to become catalytically active. The U1 and U4 snRNPs are released from the spliceosome, the U6 snRNP base pairs to the 5′ splice site, and the U5 snRNP base pairs to both exon sequences, placing the splice sites in close proximity.

Splicing, then, takes place by means of two transesterification reactions (Figure 8-19b). The first step of the reaction involves nucleophilic attack by the 2′-OH of the unpaired branch point adenosine at the phosphodiester bond at the 5′ splice site, which cuts the pre-mRNA between the 5′ exon and the intron and produces an intron with a loop structure called a lariat because it resembles the shape of a cowboy's lariat (lasso). The second step of the reaction involves nucleophilic attack by the 3′-OH of the 5′ exon at the phosphodiester bond at the 3′ splice site, which covalently links together the 5′ and 3′ exons and frees the intron as a lariat. Lastly, the U2, U5, and U6 snRNPs are released from the excised lariat and participate in another cycle of splicing along with previously released U1 and U4 snRNPs. This process is repeated for each intron in a pre-mRNA.

KEY CONCEPT Splicing is a two-step reaction. The first step is cleavage at the 5′ splice, and the second step is cleavage at the 3′ splice site, which results in removal of the intron and joining of the exons.

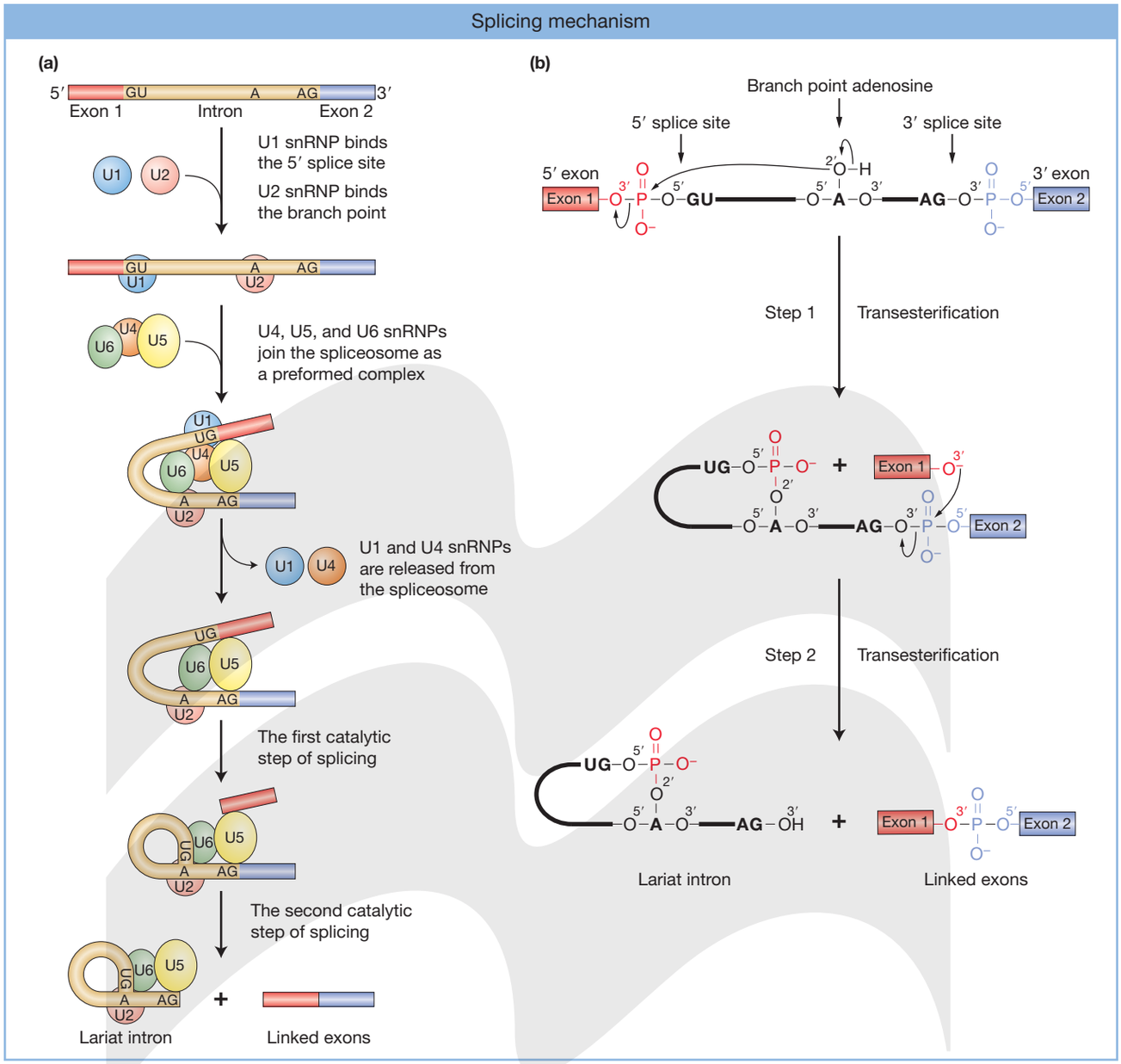


FIGURE 8-19 (a) The spliceosome is formed by sequential assembly of five snRNPs (U1, U2, U4, U5, and U6) onto an intron. Conformational changes that lead to release of the U1 and U4 snRNPs position the U2 and U6 snRNAs and the branch point adenosine near the 5′ splice site for the first catalytic step of splicing. (b) In the first step of splicing, the 2′-hydroxyl of the branch point adenosine attacks the phosphodiester bond at the 5′ splice site. This produces two intermediates, a 5′ exon and an intron–3′ exon in a lariat structure. In the second step of splicing, the 3′-hydroxyl of the 5′ exon attacks the phosphodiester bond at the 3′ splice site, producing linked exons and the released lariat intron.

ANIMATED ART SaplingPlus Mechanism of mRNA splicing

snRNAs in the spliceosome may carry out the catalytic steps of splicing

Researchers initially assumed that proteins in the spliceosome carry out the catalytic reaction, but in 1981 studies by Thomas Cech's laboratory raised the possibility that pre-mRNA splicing by the spliceosome is catalyzed by the snRNAs. Cech and co-workers reported that precursor ribosomal RNA (pre-rRNA) from the ciliated protozoan *Tetrahymena thermophila* could splice a 413-nucleotide intron from itself without the help of proteins, thus demonstrating that RNA can function as an enzyme, a **ribozyme.** There are two distinct classes of self-splicing introns, called Group I and II, that are found in bacteria and bacterial viruses as well as some nuclear-encoded mitochondrial and chloroplast genes in fungi, algae, and plants. Since the structure of base-paired pre-mRNA, U2 snRNA, and U6 snRNA in the active site of the spliceosome is similar to the secondary

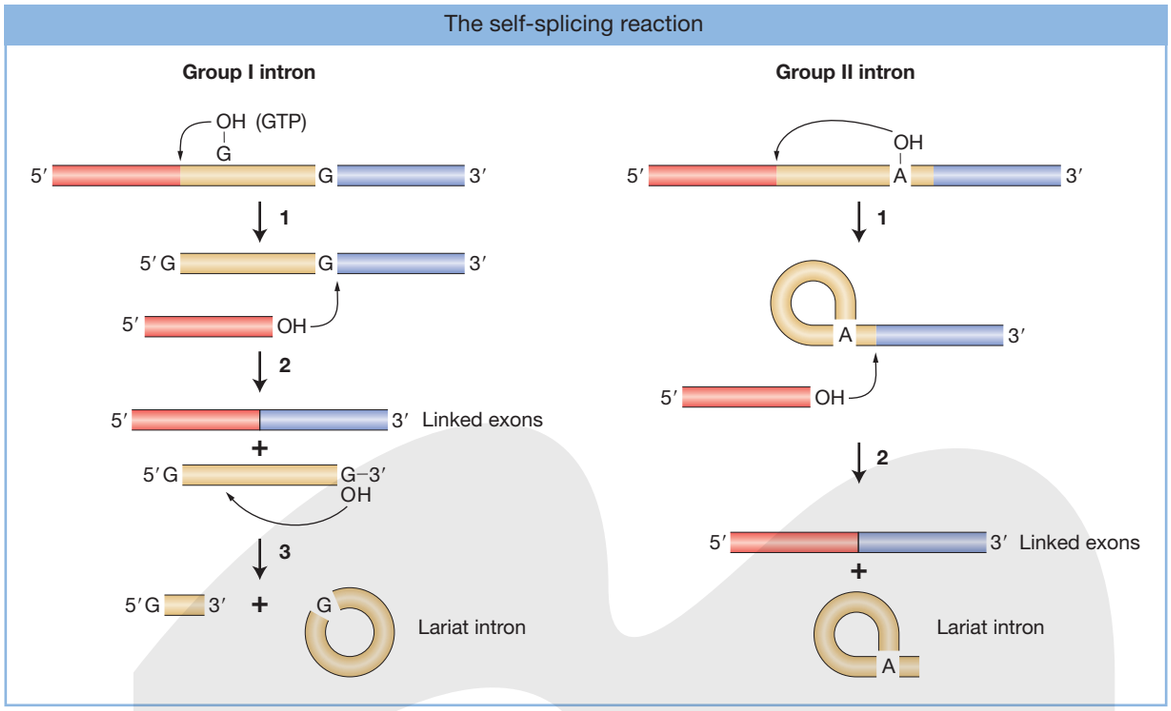


FIGURE 8-20 Similar to splicing by the spliceosome, RNA-catalyzed self-splicing of Group I and Group II introns involves two transesterification reactions. In Group I splicing, the first reaction is carried out by GTP, whereas in Group II splicing and spliceosomal splicing. The first reaction is carried out by the branch point adenosine.

structure of Group II introns and the chemistry of splicing by the spliceosome is the same as that carried out by Group II introns (**Figure 8-20**), it has been hypothesized that spliceosomal and Group II splicing mechanisms are evolutionarily related and that the spliceosome is a ribozyme.

KEY CONCEPT Intron removal and exon joining are catalyzed by RNA molecules. In eukaryotes, the snRNAs of the spliceosome catalyze the removal of introns from pre-mRNA. Some introns are self-splicing; in these cases, the intron catalyzes its own removal. RNAs capable of catalysis are called ribozymes.

Alternative splicing can expand the proteome

A major rationale for having genes with introns is that introns provide a mechanism to encode different proteins (called protein **isoforms**) from a single gene. Through the process of **alternative splicing,** exons in a pre-mRNA can be joined together in different combinations to produce different mature mRNAs that encode protein isoforms. An extreme example is the *Dscam* gene in *Drosophila* that via alternative splicing can produce 38,016 different Dscam proteins.

Alternative splicing can produce protein isoforms with different functional domains. This is illustrated by *FGFR2,* a human gene that encodes a receptor that binds fibroblast growth factors and then transduces a signal inside the cell (**Figure 8-21**). The FGFR2 protein is made up of several domains, including an extracellular ligand-binding domain. Alternative splicing results in two protein isoforms that differ in their extracellular domains. Because of this difference, each isoform binds different growth factors. In addition, as illustrated by the *α-tropomyosin* gene (**Figure 8-22**), protein isoforms can be produced in particular cells by cell type-specific alternative splicing, and they can also be produced at different stages of development. Thus, alternative splicing expands the **proteome** (the set of all proteins that can be expressed) of eukaryotic organisms.

In humans, about 95 percent of intron-containing genes undergo alternative splicing to encode two or more protein isoforms. There are four general types of alternative splicing, the most common of which is exon skipping, where an exon is either included or excluded in the mature mRNA (**Figure 8-23**). The other types of alternative splicing are alternative 3′ splice sites (i.e., one 5′ splice site and a choice of two 3′ splice sites), alternative 5′ splice sites (i.e., one 3′ splice site and a choice of two 5′ splice sites), and mutually exclusive exons (i.e., only one of several exons is included in the mature mRNA, as illustrated by *FGFR2* in Figure 8-21).

A key feature of alternative splicing mechanisms is that 5′ and 3′ splice site sequences differ among exons. Alternative exons tend to have weak splice site sequences that have lower affinity for spliceosome components than splice sites associated with constitutive exons (i.e., exons that are always spliced into the mature mRNA). Weak splice sites are subject

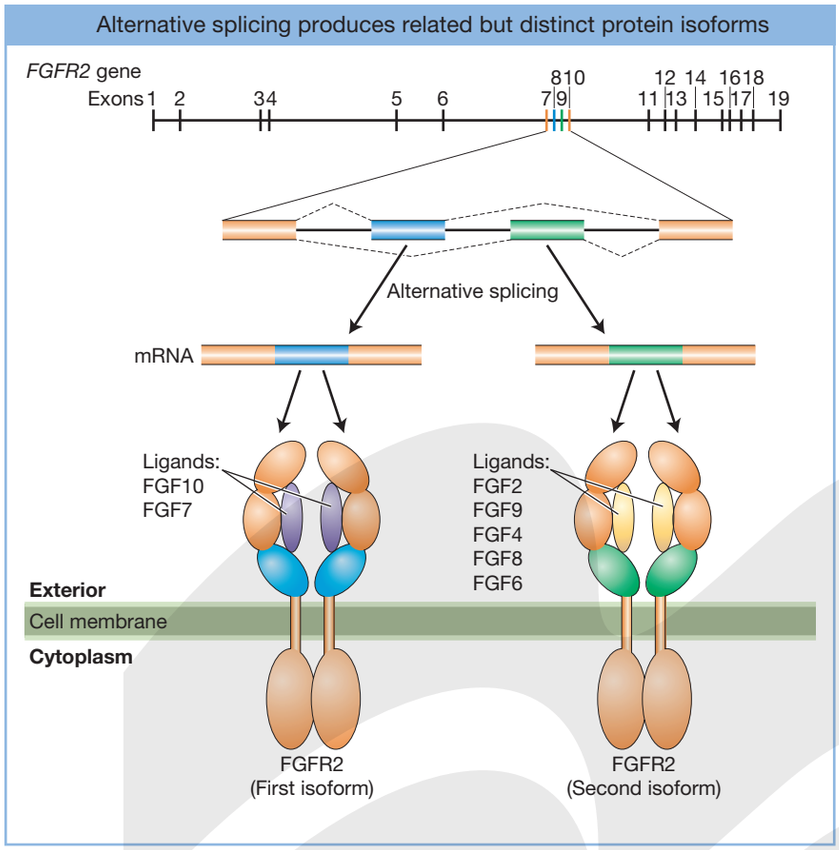


FIGURE 8-21 Alternative splicing of mutually exclusive exons in the *FGFR2* pre-mRNA produces two protein isoforms that bind different FGF proteins.

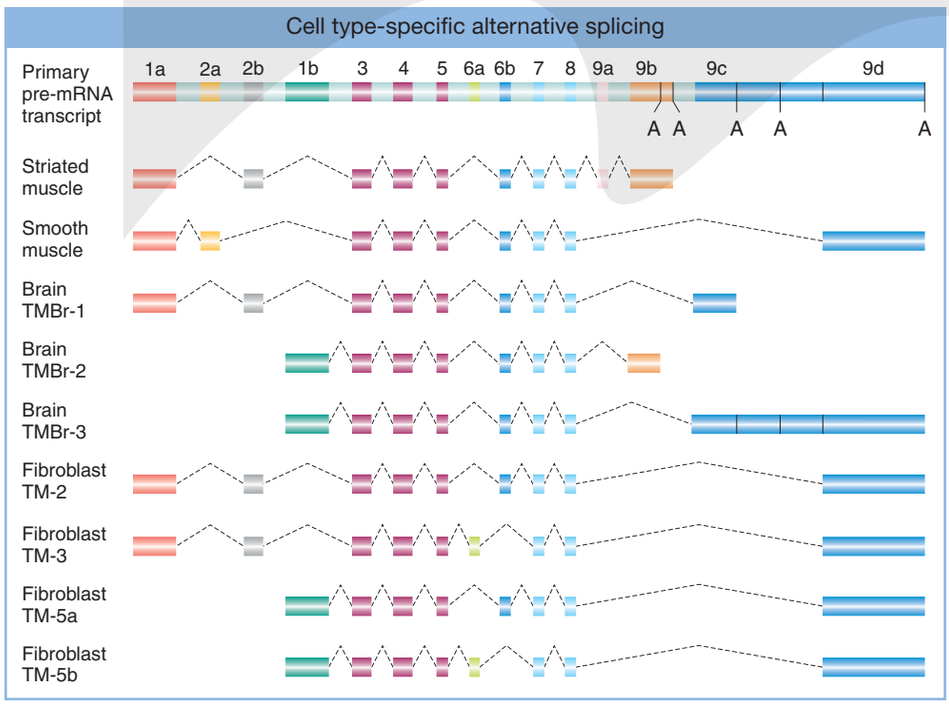


FIGURE 8-22 The rat α-*tropomyosin* gene is alternatively spliced in different patterns in different cell types. Light blue boxes represent introns; other colors represent exons. Note that in addition to alternative splicing, the α-*tropomyosin* gene undergoes alternative transcription initiation (starting transcription at the beginning of either the peach or the green exons) and alternative polyadenylation (occurring at five sites indicated by A's). Dashed lines indicate introns that have been removed by splicing. TM, tropomyosin.

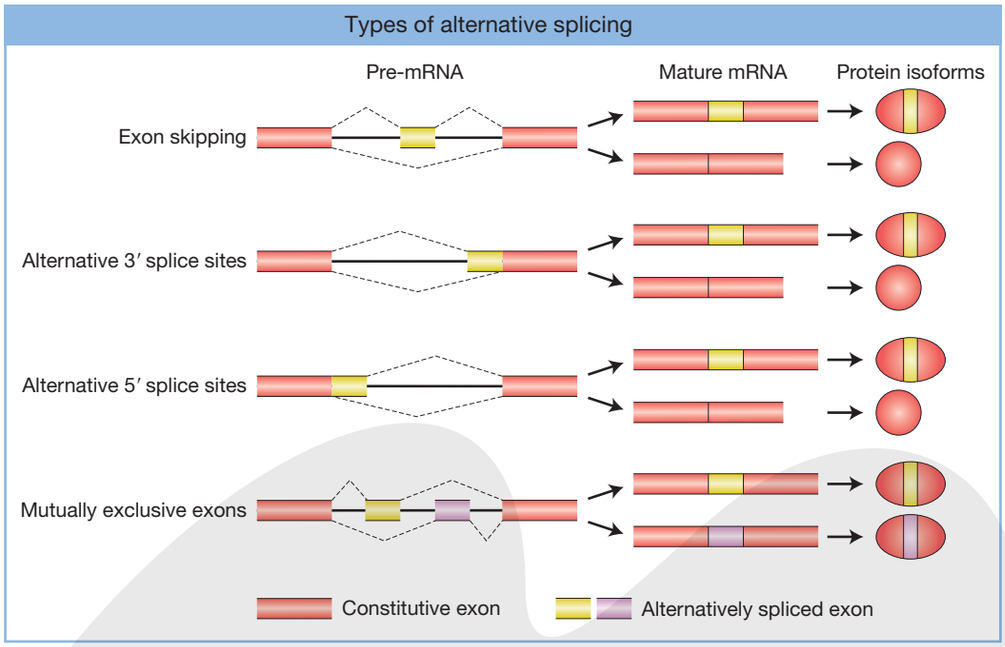


FIGURE 8-23 A single pre-mRNA that contains exons and introns can be spliced in different patterns to produce mature mRNAs that encode different proteins. There are four major types of alternative splicing. Dotted lines on the top and bottom of each pre-mRNA show how the 5′ and 3′ splice sites can be alternatively joined to produce different mature mRNAs.

to alternative splicing because their use by the spliceosome is not optimal and can thus be enhanced or suppressed by regulatory factors. Alternative splicing can also be affected by the rate of transcription elongation. Central to this mechanism is the fact that most splicing events take place during transcription. Spliceosomes assemble on introns as soon as they are transcribed. Slow elongation by RNA polymerase II provides a longer window of opportunity for the use of an alternative weak splice site before a strong splice site is transcribed.

KEY CONCEPT The joining of exons in different patterns via alternative splicing greatly expands the number of proteins encoded in the human genome and other eukaryotic genomes.

RNA editing

RNA sequences encoded in eukaryotic genomes are not only changed by RNA processing events such as splicing but also by RNA editing. **RNA editing** is a general term that describes molecular processes through which nucleotide sequences in RNAs are changed after transcription. Editing events include insertion and deletion of nucleotides as well as base substitution. Many types of RNA, including tRNAs, rRNAs, mRNAs, and snRNAs, are edited.

In animals, the most common type of editing is adenosine-to-inosine (A-to-I) editing, which converts adenosine to inosine by deamination. A-to-I editing is catalyzed by double-strand RNA-binding enzymes called adenosine deaminase acting on RNAs (ADARs) (Figure 8-24a). Inosine is a non-canonical nucleoside that can base pair with cytidine, so during translation inosine is read as a guanosine, rather than adenosine (Figures 8-24b and 8-24c), changing the amino acid sequence of a protein.

A-to-I editing can also affect regulatory elements in RNAs that function by base pairing to another RNA or are bound by a protein. For example, conversion of a stable A-U base pair into a less stable I-U base pair can alter splicing by affecting base pairing between snRNAs and pre-mRNAs. High-throughput RNA sequencing methods have identified over two million A-to-I edited sites in the human **transcriptome** (the set of all RNAs that can be expressed in humans). The physiological consequence of A-to-I editing at most of these sites is yet to be determined, but global effects on A-to-I editing caused by mutation of ADARs leads to behavioral and locomotion abnormalities in *Drosophila* and seizures and early death in mice, highlighting the importance of A-to-I editing.

RNA nucleotide modification

The structure and function of RNAs can be altered by post-transcriptional chemical modifications. Most RNA modifications consist of the addition of a methyl group (CH_3) to a nucleoside base such as N^6-methyladenosine (m^6A) (Figure 8-24d), to a ribose sugar such as 2′-*O*-methyladenosine (Am) (Figure 8-24e), or to both such as N^6, 2′-*O*-dimethyladenosine (m^6Am). More than 100 different chemical modifications of RNA have been identified. Each modification can have distinct effects on RNA structure and interactions with other RNAs and proteins, which can affect all aspects of RNA metabolism, including the processing, stability, and translation of mRNAs.

m^6A is the most common modification in human mRNAs. Of the more than 20,000 m^6A sites that have been identified in humans, 70 percent occur in the last exon of a

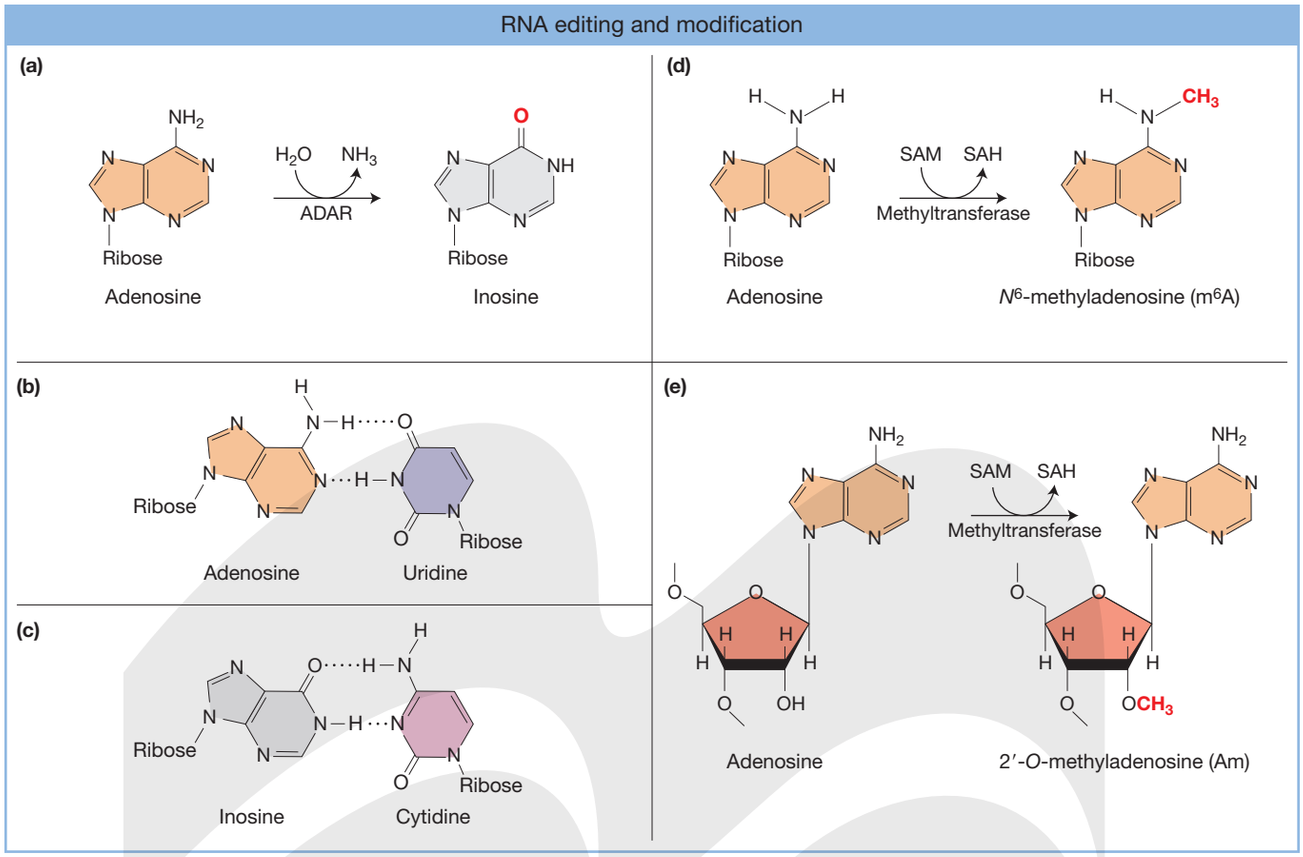


FIGURE 8-24 The information contained in RNA can be changed by editing and modification. (a) A-to-I editing is carried out by ADAR enzymes. (b) Adenosine base pairs with uridine, but (c) inosine base pairs with cytidine. (d) The chemical structure of ribonucleoside bases and (e) ribose sugars can be changed by methylation. These reactions are catalyzed by different methyltransferases, but both reactions use SAM as the methyl-group donor.

transcript and nearly half occur in the 3′ UTR. m^6A modifications are added during transcription and prior to the completion of splicing. A major function of m^6A is to destabilize mRNAs, as demonstrated by the finding that knocking out the methyltransferase that writes m^6A into mRNA results in longer half-lives of m^6A-containing mRNAs. Newly developed technologies that map the sites of chemical modifications in RNAs are making it possible for researchers to identify the writers, readers, and erasers of modifications as well as determine their molecular and biological functions.

KEY CONCEPT RNAs are subject to editing and modification. Editing can change the protein sequence encoded by an mRNA, and both editing and modification can create new signals in mRNAs and ncRNAs that change their structure, function, and stability.

RNA export from the nucleus

Many eukaryotic RNAs that are transcribed in the nucleus from the nuclear genome spend some part of their life in the cytoplasm. For example, mRNAs are exported from the nucleus to the cytoplasm where they are translated into proteins, and snRNAs involved in splicing are produced in the nucleus, exported to the cytoplasm for assembly with proteins, and then returned to the nucleus. Export of mRNAs and snRNAs from the nucleus occurs by different mechanisms, but both mechanisms involve adaptor proteins that bind the RNAs early in their biogenesis and escort them to the cytoplasm through channels in the nuclear membrane called nuclear pores. In human cells, mRNAs are transported out of the nucleus by the TREX (transcription export) complex, whereas snRNA are transported by PHAX (phosphorylated adaptor for RNA export). Both TREX and PHAX interact with their RNA cargo during transcription through binding to the cap binding complex (CBC).

KEY CONCEPT Mechanisms exist in eukaryotic organisms to transport and localize RNAs to particular places in cells.

8.5 DECAY OF mRNA IN EUKARYOTES

LO 8.5 Describe how siRNAs regulate the abundance of specific RNAs and play a role in maintaining genome integrity in eukaryotes.

As in bacteria, decay counterbalances transcription to regulate the abundance of mRNAs in eukaryotes. The half-life of

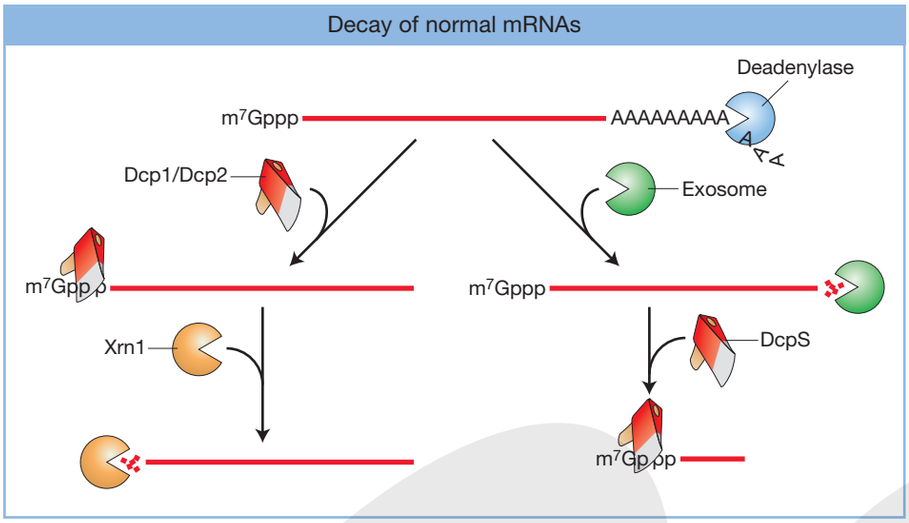


FIGURE 8-25 mRNA decay is normally initiated by removal of the poly(A) tail by a deadenylase enzyme. Deadenylation is followed either by decapping by Dcp1/Dcp2 and 5′-to-3′ decay by the exonuclease Xrn1 or by 3′-to-5′ decay by the exosome and decapping by DcpS.

eukaryotic mRNAs can vary widely. For example, the *β-globin* mRNA, which encodes a subunit of hemoglobin in blood cells, has a half-life of 20–24 hours. In contrast, the *c-Myc* mRNA, which encodes a transcription factor that regulates the cell cycle, has a half-life of 20–30 minutes. Mechanisms that control mRNA half-life help ensure that proteins are only present in cells when they are needed. mRNAs with a long half-life, like *β-globin,* tend to encode proteins with structural or metabolic functions, and mRNAs with a short half-life, like *c-Myc,* tend to encode proteins with regulatory functions.

The half-life of an mRNA is experimentally determined by turning off all RNA polymerase II transcription and then measuring how long it takes for half of the existing mRNA molecules to be degraded. The most common method involves treating cells with Actinomycin D, an inhibitor of RNA polymerase II elongation, and quantifying mRNA abundance by Northern blot or reverse transcription–PCR analysis (discussed in Chapter 10).

mRNA decay mechanisms

mRNA decay typically occurs in the cytoplasm and requires translation. Most mRNA decay occurs by two general pathways that are both initiated via removal of the poly(A) tail by a **deadenylase**, a special type of exonuclease that specifically cleaves phosphodiester bonds between adenosine nucleotides one at a time in the 3′-to-5′ direction (**Figure 8-25**). Deadenylation is sometimes followed by removal of the 5′ m^7G cap by the **decapping enzyme** Dcp1/Dcp2. This enzyme cuts off 5′-m^7Gpp (see Figure 8-16a), leaving behind an mRNA with a 5′-monophosphate that serves as a substrate for complete digestion by a 5′-to-3′ exonuclease Xrn1. Alternatively, deadenylation is followed by digestion by a 3′-to-5′ exonuclease called the exosome. Following decay of the mRNA body by the exosome, a different decapping enzyme called the scavenger decapping enzyme (DcpS) catalyzes the hydrolysis of the 5′ m^7G cap, releasing 5′-m^7Gp. The major difference between decay in bacteria and decay in eukaryotes is that most decay in eukaryotes is initiated by an exonuclease, not an endonuclease (compare Figures 8-10 and 8-25).

The efficiency of decay pathways can be enhanced or suppressed by RNAs and proteins that bind specific sequences within 3′ UTRs of mRNAs and affect the recruitment of decay factors. For example, base pairing of miRNAs to sequences within the 3′ UTR not only inhibits translation (discussed in Chapter 9) but also enhances decay by recruiting deadenylase and decapping enzymes. Similarly, sequences in the 3′ UTR that are rich in adenosine and uridine known as AU-rich elements (AREs) serve as binding sites for RNA-binding proteins that enhance or suppress decay by affecting recruitment of decay factors.

Related but distinct pathways detect and rapidly decay particular types of abnormal mRNAs during their translation and prevent the production of truncated or erroneous proteins. Nonsense-mediated decay (NMD) detects mRNAs that have a premature translation stop site, non-stop decay (NSD) detects mRNAs that lack a translation stop site, and no-go decay (NGD) detects mRNAs that contain sequences or structures such as strong stem-loops that stall translation elongation. Decay of abnormal mRNAs involves many of the same enzymes as decay of normal mRNAs, including decapping enzymes, deadenylases, and exonucleases, but it can also involve endonucleases.

KEY CONCEPT After removal of the poly(A) tail, mRNA decay by specialized enzymes occurs in both the 5′-to-3′ and 3′-to-5′ direction.

The discovery of RNA interference (RNAi)

In 2002, one of the leading science journals, *Science* magazine, named "Small RNA" as their Breakthrough of the Year. The RNAs to which they were referring were not the previously described small RNAs such as snRNAs or tRNAs, which are considered to have housekeeping roles and, as such, are synthesized all the time (i.e., constitutively). Instead, these other small RNAs are synthesized in response to changes in a cell's developmental state or its surroundings. We now know that small RNAs are critically important for

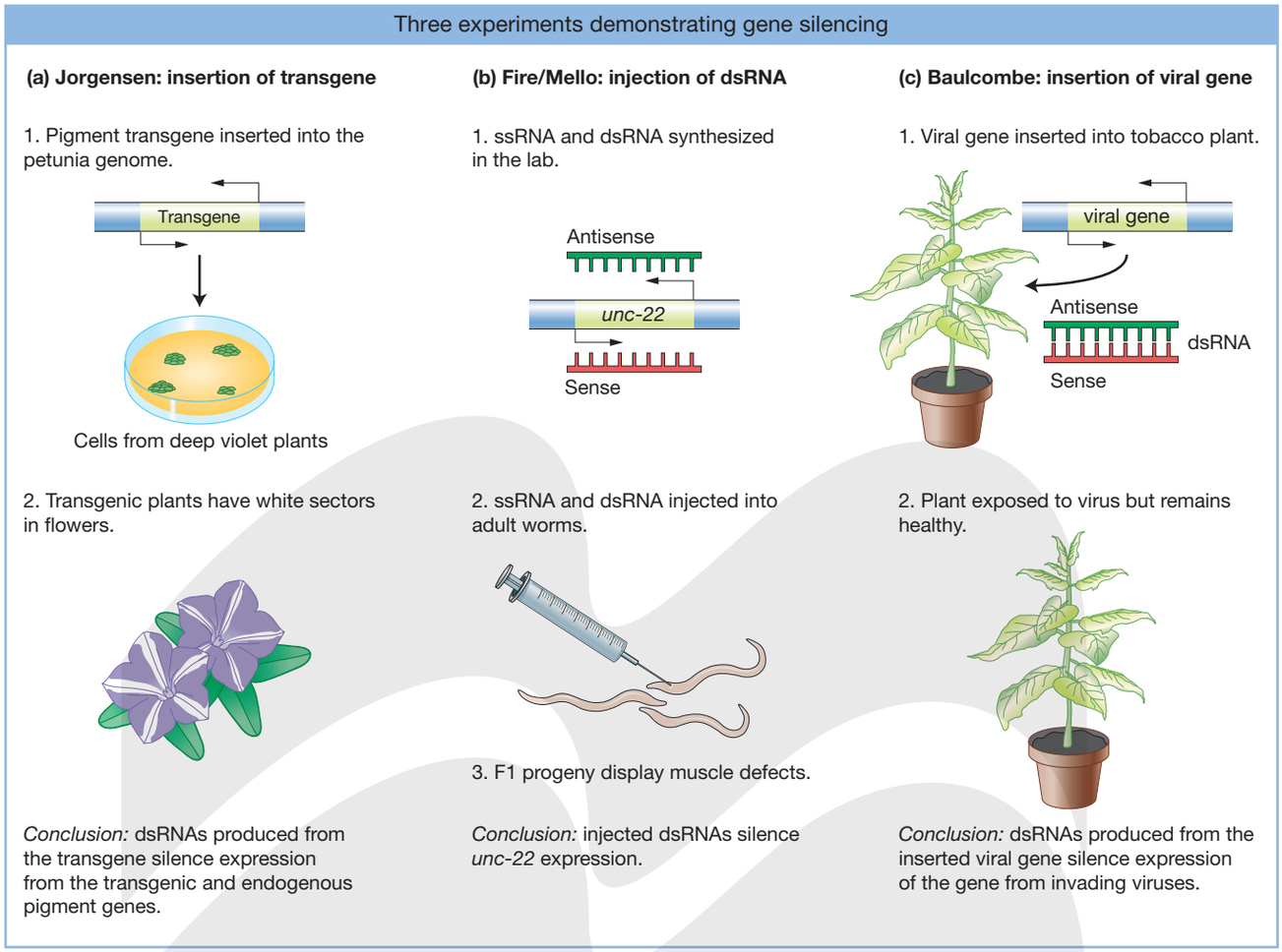


FIGURE 8-26 Three experiments reveal key features of gene silencing. (a) Jorgensen discovered that a transgene can silence an endogenous petunia gene necessary for floral color. (b) Fire and Mello demonstrated that dsRNA can selectively silence genes in *C. elegans*. (c) Baulcombe showed that plants with a copy of a viral transgene were resistant to viral infection and produced siRNAs complementary to the viral genome.

the regulation of gene expression and the maintenance of a stable genome. Studies that led to the discovery of one class of small RNA called small interfering RNA (siRNA) were preceded by several reports of unanticipated changes in gene expression when RNAs were injected into an organism or expressed from a **transgene** (a gene introduced by researchers into the chromosomes of an organism).

One of the greatest joys of doing scientific research is observing a completely unexpected result. In 1990, this is precisely what occurred for Richard Jorgensen in his studies of plant coloration. To increase pigmentation in petunia plants that had pale pink flowers, Jorgensen inserted a petunia gene that encodes an enzyme necessary for the synthesis of purple-blue flower pigment (**Figure 8-26a**). As a control, he inserted the same gene into plants with deep violet flowers. He expected that relative to the parental plants, the flower color of the pale pink transgenic plants would be darker, but the deep violet transgenic plants would be unchanged. However, none of the transgenic plants had darker flowers. In fact, all the transgenic plants produced flowers that were either pure white or a variety of white patterns (**Figure 8-27**). In a totally unexpected outcome, Jorgensen found that the transgene suppressed the expression of its own mRNA as well as the mRNA produced from the endogenous pigment gene (i.e., the one that is normally in the petunia genome). Therefore, he called the phenomenon co-suppression.

The mechanism underlying co-suppression remained a mystery until 1998, when Andrew Fire and Craig Mello used the roundworm *C. elegans* to test the hypothesis that **double-stranded RNA (dsRNA)** was the agent that triggered co-suppression. They injected *C. elegans* with RNAs that were identical in sequence to an endogenous gene that when mutated causes adult worms to twitch (Figure 8-26b). If the injected RNAs triggered co-suppression, they expected to see the twitcher phenotype even though the endogenous gene was intact. Indeed, they found that, relative to injection of single-stranded RNA, injection of dsRNA caused a much stronger twitcher phenotype, demonstrating that dsRNA mediates suppression of endogenous gene expression in a process that is now called **RNA interference (RNAi)**. The study by Fire and Mello also uncovered other remarkable features of RNAi: (1) RNAi is very specific, and only RNAs

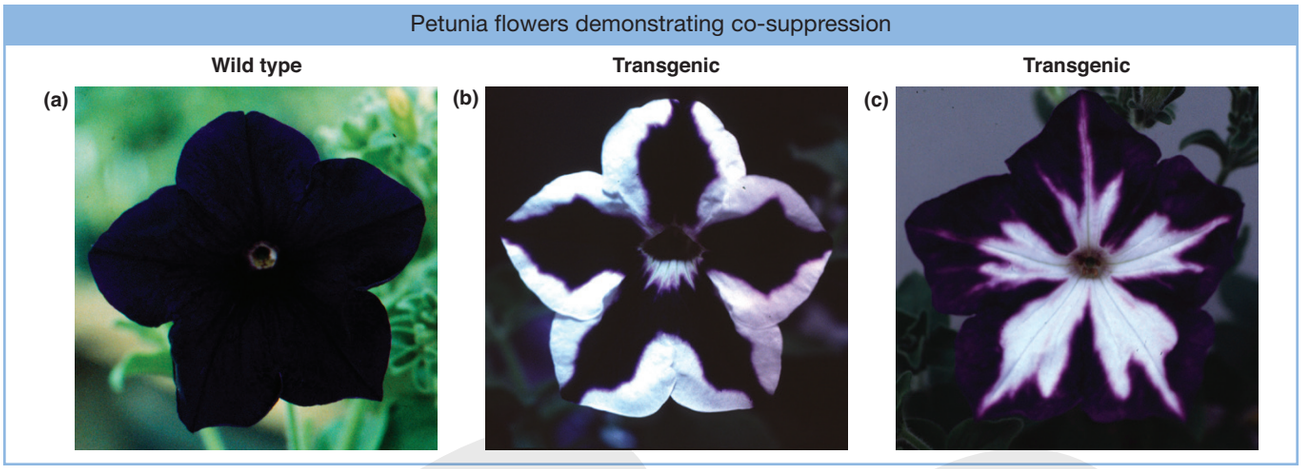


FIGURE 8-27 (a) The wild-type (no transgene) phenotype. (b and c) So-called co-suppression phenotypes resulting from insertion of a transgene that controls pigmentation into the genome of a wild-type petunia. In white regions of the petals, both the transgene and the endogenous chromosomal copy of the same gene have been inactivated. [*Richard Jorgensen, Department of Plant Biology, Carnegie Institution for Science.*]

with perfect complementarity to the dsRNA are affected; (2) RNAi is extremely potent, as only a few dsRNA molecules are required per cell to inhibit expression of the targeted gene, indicating that the process is catalytic; (3) RNAi can affect cells and tissues that are far removed from the site of introduction, indicating that there is an RNA transport mechanism; and (4) RNAi affects the progeny of injected animals, indicating that the targeting information is heritable.

Many labs continue to study the mechanism underlying RNAi, which is discussed in the next section. Nevertheless, even without a complete understanding of the mechanism, RNAi has had a tremendous impact on almost all fields of biology research through its use as a tool to perform loss-of-function experiments. Researchers have developed creative methods to introduce or express dsRNAs in cells and whole organisms to reduce the expression of a specific gene and determine its necessity for molecular, cellular, and organismal processes. The terms *knockdown* and *silencing* are used in conjunction with RNAi because the reduction in the abundance of targeted RNAs is rarely complete; instead, it is knocked down or silenced.

RNAi technologies have also been developed to perform genome-wide screens for genes involved in cellular processes. In this approach, libraries of dsRNAs are generated that target all of the protein-coding genes in an organism, and screens are performed to identify the few dsRNAs that produce a desired phenotype. Thus, RNAi has made it much easier to perform genetic studies in many organisms and cell culture systems for which there had been no simple method to manipulate gene expression.

siRNA-mediated RNA decay and transcriptional silencing

RNAi silences gene expression by targeting RNAs for decay in the cytoplasm of cells. The RNAi decay mechanism involves three main components: (1) small interfering RNAs (siRNAs) that provide the specificity of RNAi by base pairing to target RNAs, (2) Dicer, an endonuclease that precisely cuts dsRNAs into siRNAs, and (3) Argonaute (Ago), an RNA endonuclease that is programmed to cut RNAs that base pair to bound siRNAs.

siRNAs are approximately 21-nucleotide dsRNAs. Each strand of an siRNA has a 5′-monophosphate, a 3′-hydroxyl, and a two-nucleotide 3′ overhang beyond the core base-paired region of 19 nucleotides (**Figure 8-28**). These features of siRNAs are important for their recognition by proteins that carry out RNAi. Dicer uses its PAZ domain to bind the 3′ overhang and generate an siRNA from a hairpin or long dsRNA via its two endonuclease domains (**Figure 8-29**). Once bound to the end of a dsRNA, the endonuclease domains are positioned to make cuts in the strands that are 21 nucleotides away and staggered by two nucleotides. Dicer can repeat this process, producing multiple siRNAs from a single dsRNA. One of the 3′ overhangs of siRNAs is then bound by the PAZ domain of Ago, which is part of a multi-protein complex called

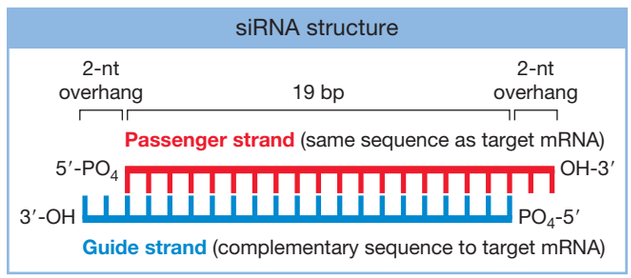


FIGURE 8-28 Small interfering RNAs (siRNAs) that are produced from hairpin or long dsRNAs by Dicer have specific features that are important for recognition by Ago. siRNAs are 19–21 base pairs in length and each strand has a 5′-phosphate and a 2-nucleotide 3′ overhang with a 3′-hydroxyl.

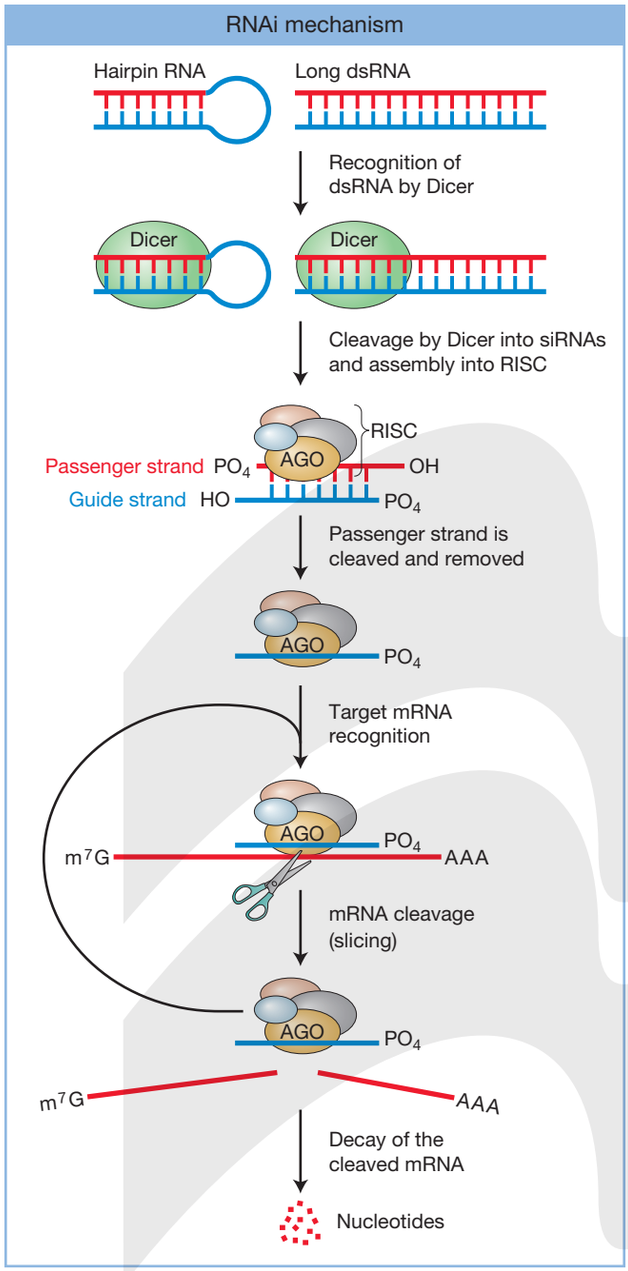


FIGURE 8-29 In the RNAi pathway, Dicer produces siRNAs from dsRNA, and siRNAs are bound by Ago-containing RISC. The siRNA guide strand targets RISC to specific RNAs by base pairing. The endonuclease activity of Ago cuts the target RNA into two fragments that are decayed to nucleotides by normal decay pathways (see Figure 8-25). RNAi results in silencing (i.e., knock down) of the expression of targeted genes.

RNA-induced silencing complex (RISC). Ago also uses its PIWI endonuclease domain to cut and displace one of the siRNA strands called the passenger strand, leaving behind a single-stranded siRNA guide strand. Perfect base pairing between an siRNA and a target mRNA stimulates cleavage of the target by the Ago PIWI domain. The resulting pieces of target RNA are decayed by the normal pathways, involving deadenylation, decapping, and exonucleolytic cleavage from the 5′ and 3′ ends.

KEY CONCEPT Dicer cuts dsRNA to produce 21-bp siRNAs with 2-nt overhangs on each end. siRNAs are bound by RISC, which contains Ago, an endonuclease that cuts the passenger strand, leaving the guide strand intact. When the guide strand base pairs to a complementary segment of mRNA, Ago cleaves the mRNA, triggering its degradation.

In some cases, siRNAs can enter the nucleus and block transcription of target genes by inducing heterochromatin formation (discussed in Chapter 12). For example, in *C. elegans*, siRNAs are transported from the cytoplasm to the nucleus by an Ago protein that lacks endonuclease activity. The Ago-siRNA complex localizes to a specific gene, presumably by base pairing of the siRNA to an mRNA during its synthesis. This localization leads to recruitment of a methyltransferase enzyme that methylates histone H3 on lysine 9 (H3K9me) to generate heterochromatin and turn off transcription of the gene.

RNAi protects the genome from foreign DNA

The function of RNAi is clearly not to shut off genes at the whim of scientists. In fact, in plants, RNAi is a form of antiviral defense. This discovery came from experiments conducted by David Baulcombe and co-workers who had engineered the genome of tobacco plants to express a viral gene (see Figure 8-26c). He found that plants engineered with a viral transgene but not plants that lacked the transgene were resistant to subsequent infection by the virus. Furthermore, he found that that resistant plants, and only the resistant plants, produced large amounts of siRNAs complementary to the viral genome. Thus, RNAi serves as an antiviral defense system.

RNAi also silences the expression of endogenous repeated sequences such as transposons, which make up a large part of many eukaryotic genomes (Chapter 16). When translated into proteins, transposons facilitate insertion of their DNA into new sites in the genome. Transposon mobilization can disrupt host genes and promote chromosomal rearrangements, leading to diseases such as cancer. However, some transposons contain inverted repeat sequence or antisense promoters that can produce dsRNA and trigger RNAi (see Figure 16-28). Thus, one of the normal functions of RNAi is to protect against invading sources of nucleic acids such as viruses and transposons that threaten the integrity of the host genome.

KEY CONCEPT Many eukaryotic organisms use siRNA-mediated RNAi to silence the expression of foreign genes. Researchers have taken advantage of the endogenous RNAi machinery to knock down the expression of a specific gene by introducing into cells dsRNA that is identical in sequence to the target gene.

SUMMARY

RNAs serve numerous and varied purposes in bacteria and eukaryotic cells. mRNA garners much of the attention in this book and elsewhere because it is the template for the synthesis of proteins, which perform the vast majority of structural and enzymatic roles in cells. Nevertheless, ncRNAs are also important. For example, the translation of mRNA into proteins cannot take place without tRNAs and rRNAs, and ncRNAs function at many steps in gene expression. ncRNAs operate in three general ways: they interact with other RNAs and DNA by base pairing, they serve as scaffolds for the assembly of proteins, and they carry out enzymatic reactions. The activities of ncRNAs are integral to transcription and RNA processing as well as other events in gene expression, including translation (Chapter 9). For example, in ribosomes (the protein synthesis machines), sequences in tRNAs base pair to sequences in mRNA, rRNAs serve as scaffolds for assembly of ribosomal proteins, and rRNAs catalyze peptide bond formation between amino acids. The variety of RNA functions is possible because of RNA's unique chemical features.

In bacteria, all RNAs are synthesized by a single RNA polymerase, whereas in eukaryotic cells there is a division of labor among three RNA polymerases (I, II, and III). Regardless of the organism or the RNA polymerase, the sites of transcription initiation are marked by conserved promoter elements that are located near the transcription start site and are bound by proteins. In bacteria, a subunit of RNA polymerase binds promoter elements, but in eukaryotes, GTFs bind promoter elements and recruit a specific RNA polymerase. After RNA polymerase recruitment, the DNA is locally unwound and RNA polymerase begins incorporating ribonucleotides in the 5′-to-3′ direction that are complementary to the template DNA strand. As transcription transitions from the initiation to the elongation phase, the factors associated with RNA polymerase change. This is exemplified by sigma factors in bacteria and factors that associate with different modified forms of the RNA polymerase II CTD in eukaryotes. In the case of the CTD, associated factors are involved in processing of the pre-mRNA and can alter the elongation rate of RNA polymerase II, including causing RNA polymerase II to pause at particular sites. In both bacteria and eukaryotes, RNA polymerase terminates transcription after transcribing termination signals into a nascent mRNA, and multiple mechanisms are involved in recognizing the termination signals and releasing the mRNA and RNA polymerase from DNA.

In eukaryotic organisms, RNAs undergo extensive processing, often while they are being transcribed. For example, mRNAs are modified with a cap at the 5′ end and a poly(A) tail at the 3′ end, and introns are removed. Modifications at the ends increase the stability of an mRNA and assist translation. Sequences within a pre-mRNA together with snRNA or protein factors that bind them define sequences as intron or exon for splicing as well as dictate the site of cleavage and polyadenylation. Additional factors that bind other RNA sequence elements can enhance or suppress the use of particular sites for splicing. This leads to alternative splicing, which can increase the types of proteins encoded by a gene from one to, in some cases, thousands. In bacteria, ribosomes associate with mRNAs as they are being transcribed, whereas in eukaryotes ribosome association and translation can take place only after mRNAs are exported from the nucleus to the cytoplasm.

The last step in the life cycle of an RNA is decay. mRNA decay in both bacteria and eukaryotes occurs via defined pathways that begin with recruitment of specific enzymes. In bacteria an endonuclease is recruited by interacting with a 5′-monophosphate, and in eukaryotes an exonuclease is recruited by interacting with proteins that associate with the 3′ UTR. The initiating steps of decay generate recognition sites for further decay by other enzymes. Additionally, in some eukaryotes, very short RNAs such as siRNAs base pair to mRNAs and bring along an endonuclease that initiates decay. One of the normal functions of siRNAs is to silence the expression of repetitive genes in genomes such as transposons. Researchers have taken advantage of this decay activity to perform targeted loss-of-function experiments.

The recent development of technologies that detect low-abundance RNAs has led to the discovery of thousands of RNAs of unknown function. However, researchers have a head start in figuring out how these RNAs are transcribed, processed, transported, and decayed as well as how they function, because it is likely that aspects of the mechanisms are shared with mRNAs, tRNAs, rRNAs, snRNAs, or siRNAs.

KEY TERMS

3′ splice site (p. 285)
3′ untranslated region (3′ UTR) (p. 275)
7-methylguanosine (m^7G) (p. 283)
5′ splice site (p. 285)
5′ untranslated region (5′ UTR) (p. 274)
allosteric termination model (p. 282)
alternative splicing (p. 288)
branch point (p. 285)
cap (p. 283)
carboxy-terminal domain (CTD) (p. 277)
consensus sequence (p. 274)
deadenylase (p. 292)
decapping enzyme (p. 292)
decay (p. 276)
deoxyribose (p. 270)
double-stranded RNA (dsRNA) (p. 293)

downstream (p. 274)
elongation (p. 273)
endonuclease (p. 276)
exon (p. 284)
exonuclease (p. 276)
factor-independent termination (p. 275)
general transcription factor (GTF) (p. 277)
half-life (p. 276)
helicase (p. 276)
initiation (p. 273)
intron (p. 284)
isoform (p. 288)
long noncoding RNA (lncRNA) (p. 271)
messenger RNA (mRNA) (p. 270)
microRNA (miRNA) (p. 271)
noncoding RNA (ncRNA) (p. 270)
non-template strand (coding strand) (p. 272)
nucleolus (p. 278)
phosphodiester bond (p. 273)
poly(A) polymerase (PAP) (p. 284)
poly(A) tail (p. 284)
polyadenylation (p. 284)
precursor RNA (pre-RNA) (p. 277)
preinitiation complex (PIC) (p. 280)
promoter (p. 274)
proteome (p. 288)
Rho-dependent termination (p. 275)
ribose (p. 270)
ribosomal RNA (rRNA) (p. 271)
ribozyme (p. 287)
RNA editing (p. 290)
RNA interference (RNAi) (p. 293)
RNA polymerase (p. 273)
RNA polymerase I (p. 277)
RNA polymerase II (p. 277)
RNA polymerase III (p. 277)
RNA polymerase core enzyme (p. 274)
RNA polymerase holoenzyme (p. 274)
RNA processing (p. 277)
sigma factor (σ) (p. 274)
small interfering RNA (siRNA) (p. 271)
small nuclear ribonucleoprotein (snRNP) (p. 285)
small nuclear RNA (snRNA) (p. 271)
spliceosome (p. 285)
splicing (p. 284)
TATA box (p. 279)
template strand (noncoding strand) (p. 272)
termination (p. 273)
torpedo termination model (p. 282)
transcript (p. 272)
transcription (p. 268)
transcription bubble (p. 272)
transcription start site (p. 274)
transcriptome (p. 290)
transfer RNA (tRNA) (p. 271)
transgene (p. 293)
upstream (p. 274)
uracil (U) (p. 270)

PROBLEMS

Visit SaplingPlus for supplemental content. Problems with the icon are available for review/grading.

WORKING WITH THE FIGURES

(The first 30 questions require inspection of text figures.)

1. In Figures 8-1a and b, draw a generic mRNA at each stage in the life cycle.
2. In Figure 8-2, if the chase was continued for a longer period of time, how would the distribution of radioactive RNAs change, and why?
3. In Figure 8-3, draw the ribonucleotide uridine-5′-monophosphate (UMP) base paired to adenosine-5′-monophosphate (AMP).
4. In Figure 8-4, what would have to happen in a cell to convert uracil into uridine triphosphate?
5. In Figure 8-5a, place an arrow at the location of the transcriptional promoter for each gene.
6. In Figure 8-5c, put a circle around 5′ ribose carbons and a square around 3′ ribose carbons.
7. In Figure 8-6, are adjacent rDNA genes transcribed in the same direction, or in different directions?
8. In Figure 8-7b, write the sequence of the first 10 nucleotides of the *rrn D1* transcript.
9. In Figure 8-8, what change occurs to RNA polymerase as it initiates transcription?
10. In Figure 8-9, write the sequence of the DNA template and non-template strands that encode the factor-independent mRNA stem-loop termination signal.
11. In Figure 8-10, why is pyrophosphohydrolase required only for the first endonucleolytic cleavage by RNase E and not for subsequent cleavage events?
12. In Figure 8-11, why are RNA polymerase II or III not recruited to rDNA promoters?
13. In Figure 8-12, what GTF is most likely to recognize promoters that lack a BRE, TATA box, Inr, and DPE? Provide a rationale for your answer.
14. In Figure 8-13, which GTF could be mutated to block transcription of all three types of RNA polymerase III genes?
15. In Figure 8-14, what happens to the CTD between the third and fourth steps in the model? What type of enzyme carries out this reaction?
16. In Figure 8-15, 3′-end formation is “coupled” to transcription termination. Explain what this means in the context of the allosteric and torpedo models.
17. In Figure 8-16a, does the identity of the first nucleotide in the RNA chain (A, C, G, or U) affect capping? Why or why not?

18. In Figure 8-17, which of the sequence elements that regulate cleavage and polyadenylation are retained in the mRNA after the reaction is complete?
19. In Figure 8-18, mutations of which five intron nucleotides is most likely to block splicing? What data support your hypothesis?
20. In Figure 8-19a, what snRNPs are in the spliceosome when the catalytic steps of splicing occur?
21. In Figure 8-20, what is the primary difference between Group I and Group II self-splicing?
22. In Figure 8-21, what other mRNAs could be produced by alternative splicing of the blue and green exons?
23. In Figure 8-22, which exons are spliced by a mutually exclusive mechanism?
24. In Figure 8-23, for each mechanism, which alternatively spliced product is expected to be more frequently produced if transcription is slow?
25. In Figure 8-24, draw N^6, 2′-*O*-dimethyladenosine (m^6Am).
26. In Figure 8-25, is the 5′-to-3′ decay pathway on the left or right?
27. In Figure 8-26, what is the source of the foreign dsRNA in each of the experiments?
28. In Figures 8-27b and c, in what part of the flower is co-suppression (RNAi) occurring?
29. In Figure 8-28, draw the location of the Ago PAZ and PIWI domains on the siRNA.
30. In Figure 8-29, why are only a few siRNA molecules needed to knock down hundreds or even thousands of copies of an mRNA?

BASIC PROBLEMS

31. Draw the longest continuous base-pairing interaction between the following RNAs:

5′-AAUGCCGGUAACGAUUAACGCCCGAUAUCCG-3′

5′-GAGCUUCCAUAUCGGGCGUUGGUGAUUCGAA-3′

32. What role does the branch point ribose 2′-OH play in the splicing reaction?
33. How are the ends of an mRNA protected to prevent decay?
34. Why might a mutation in a 3′ UTR affect the rate of decay of an mRNA?
35. How does Rho in bacteria function similarly to Xrn2 in eukaryotes to terminate transcription?
36. What problem is encountered by 3′-to-5′ exonucleases that might block the complete decay of excised introns?
37. What is the primary function of the sigma factor in bacteria? Is there a factor in eukaryotes that is functionally analogous to the sigma factor?
38. Write the sequence of the template and non-template strands of DNA that encode the following fragment of a bacterial mRNA:

pppGUUCACUGGGACUAAAGCCCGGGAACUAGG

39. Write the sequence of the template and non-template strands of DNA that encode the following eukaryotic mRNA, where the underlined sequence is the poly(A) tail:

m^7GpppGUUCACUGGGACUGAAUAAAGGGAAC-
UAGGA$\underline{\text{AAAAAAAAAAAAA}}_{(n = 150)}$

40. Draw the possible alternative splicing products of the following pre-mRNA, where the white boxes are constitutive exons and the shaded boxes are alternative exons:

41. Develop a consensus sequence for the following six RNA sequences:

UCGGUAGAUCCC
CCGCGAGGUUCC
CCGAAAGACCCC
UCGCGAGACUCC
UCGACAGGCUCC
CCGUAAGGUCCC

42. Draw the base-pairing interaction between the following 5′ splice site (the exon is underlined and the intron is not underlined) and the U1 snRNA, and count the number of hydrogen bonds:

5′-$\underline{\text{CAG}}$GUGACU-3′

43. A researcher repeated the pulse–chase experiment shown in Figure 8-2 with UTP that was radioactively labeled on the gamma (γ) phosphate and were unable to detect radioactive RNA in the cells. Why?
44. In addition to phosphorylation of serines 2 and 5 in the CTD repeats of RNA polymerase II, phosphorylation also occurs on serine 7. What additional "codes" are possible for factor recruitment?
45. How often does a random 21-nucleotide siRNA sequence appear in the human genome? How might the answer explain the specificity of RNAi?
46. Can a bacterial promoter direct transcription initiation in a eukaryotic cell? Why or why not?
47. A researcher found that the abundance of an mRNA increased between normal and stressed conditions. What two processes might be affected by the stress?
48. If you knew the sequence of an mRNA and its genes, how would you determine where the introns were located in the pre-mRNA?
49. If you isolated an mRNA from a eukaryotic cell, what features would it have at its 5′ and 3′ ends if it is full length?

50. If you had the sequence of an mRNA and the genome of a new organism, how would you determine the location in the genome of the transcription start site of the gene that encodes the mRNA?

51. Describe two functions for ATP and GTP in the production of RNA.

52. Draw base-paired passenger and guide strands of an siRNA that could be used to knock down the following mRNA:

5′-AAGUCCGGCAAUGCGACCAAGUCGUAAGCU-
UUAGGCGUCUUGGCAAAGA-3′

53. In bacteria and eukaryotes, describe what else is happening to an mRNA while RNA polymerase is synthesizing it from the DNA template.

54. Based on the experiment used to test the requirement for base pairing between the U1 snRNA and the 5′ splice site (Figure 8-18c), how would you test the requirement of base pairing in the stem-loop structure for Rho-dependent termination (Figure 8-9a)?

55. In Figure 8-29, propose how the 5′ mRNA fragment produced by RNAi is decayed to nucleotides.

56. What makes poly(A) polymerase an unusual nucleic acid polymerase?

57. A researcher sequenced an RNA and found that there was a G in a position where there was an A in the genome. What is likely to have happened to the RNA?

58. Which of the types of RNA polymerase III genes is most likely to also be transcribed by RNA polymerase II, and why?

59. Why is single-stranded RNA less stable in a test tube than single-stranded DNA?

60. List four similarities and four differences between eukaryotic mRNAs and ncRNAs.

CHALLENGING PROBLEMS

61. What information argues for and against the possibility that DNA is directly used as a template for translation?

62. The following data represent the base compositions of double-stranded DNA from two different bacterial species and their RNA products obtained in experiments conducted in vitro:

Species	$\frac{(A+T)}{(G+C)}$	$\frac{(A+U)}{(G+C)}$	$\frac{(A+G)}{(U+C)}$
Bacillus subtilis	1.36	1.30	1.02
E. coli	1.00	0.98	0.80

a. From these data, determine whether the RNA of these species is copied from a single strand or from both strands of the DNA. Draw a diagram to show how you solve this problem.

b. How can you tell if the RNA itself is single-stranded or double-stranded?

63. Researchers performed a genetic screen for genes that increase the lifespan of *C. elegans*. They sequenced the complete genome of a mutant with a longer lifespan and found a single A-to-T base change. List the possible ways in which the A-to-T change could alter gene expression to produce the longer lifespan phenotype. For example, the A-to-T could change the amino acid sequence of a protein.

GENETICS AND SOCIETY

In 2018, the first therapy based on RNAi (RNA interference) was approved by the US Food and Drug Administration (FDA). A pharmaceutical company used RNAi to treat hereditary transthyretin amyloidosis, a progressive and often fatal disease caused by an autosomal dominant mutation in the *transthyretin* gene that makes a toxic form of the transthyretin protein. The drug silences expression of the *transthyretin* gene using an siRNA (small interfering RNA) that targets the transthyretin mRNA for destruction. Based on your knowledge of how RNAi works, why do you think that this first RNAi drug has provided great optimism that RNAi-based therapeutics will become a widespread approach to address genetic diseases?