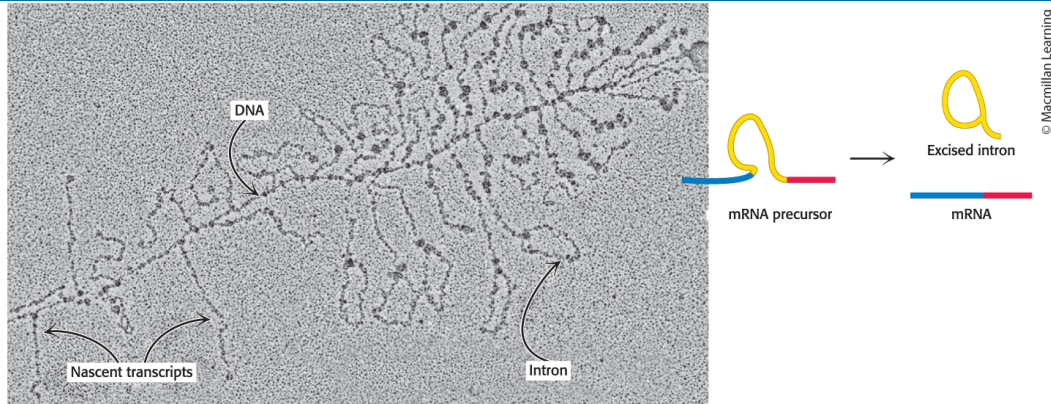


# CHAPTER 29

## RNA Functions, Biosynthesis, and Processing



RNA synthesis is a key step in the expression of genetic information. This electron micrograph shows a DNA template from a fruit fly in the genus *Drosophila* (the center “stem”) being actively transcribed into RNAs (the “branches”). In eukaryotic organisms such as fruit flies, the initial RNA transcript often contains sections that do not encode protein sequence (introns). These sections are removed by splicing; proteins associated with RNA splicing can be seen as dark blobs on the mRNA transcripts in the image on the left.

Data from *Does the pachytene checkpoint, a feature of meiosis, filter out mistakes in double-strand DNA break repair and as a side-effect strongly promote adaptive speciation?* Foe, Victoria E, *Integrative Organismal Biology*, 10 Apr 2022, Vol. 4, Issue 1 Oxford University Press, Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>)



### OUTLINE

#### [29.1 RNA Molecules Play Different Roles, Primarily in Gene Expression](#)

[29.2 RNA Polymerases Catalyze Transcription](#)

[29.3 Transcription Is Highly Regulated](#)

[29.4 Some RNA Transcription Products Are Processed](#)

[29.5 The Discovery of Catalytic RNA Revealed a Unique Splicing Mechanism](#)

---

## LEARNING GOALS



*By the end of this chapter, you should be able to:*

1. Discuss the primary function of RNA polymerases, the reaction they catalyze, and the chemical mechanism of that reaction.
2. Describe transcription, including the processes of initiation, elongation, and termination.
3. Compare the roles of eukaryotic RNA polymerases I, II, and III in producing ribosomal, transfer, and messenger RNAs.
4. Recognize the significance of transcription factors and enhancers in the regulation of transcription in eukaryotes.
5. Describe the process of RNA splicing, including the roles of the spliceosome and self-splicing RNA molecules.
6. Understand some of the differences between transcription in bacteria and in eukaryotes.

---

DNA stores genetic information in a stable form that can be readily replicated. The expression of this genetic information requires its flow from DNA to RNA and, usually, to protein, as was introduced in [Chapter 8](#). This chapter examines transcription, which, you will recall, is the process of synthesizing an RNA transcript from a DNA template, transferring the sequence information within the DNA to the new RNA molecule. We begin with a brief discussion of the diverse types of RNA molecules; then we will turn to RNA polymerases, the large and complex enzymes that carry out the

synthetic process. This will lead into a discussion of transcription in bacteria and focus on the three stages of transcription: promoter binding and initiation, elongation of the nascent RNA transcript, and termination. We then examine transcription in eukaryotes, focusing on the distinctions between bacterial and eukaryotic transcription.

# 29.1 RNA Molecules Play Different Roles, Primarily in Gene Expression

While the function of some RNAs has been known for some time, other classes of RNAs have only recently been discovered. The investigation of some of these RNA molecules has been one of the most productive areas of biochemical research in recent years.

## RNAs play key roles in protein biosynthesis

As we will explore in [Chapter 30](#), the long-known ribosomal and transfer RNA molecules, along with messenger RNAs, are central to protein synthesis. Ribosomal RNAs are critical components of ribosomes (the sites of protein synthesis), and transfer RNAs play a role in delivering amino acids (the building blocks of proteins) to the ribosome. Messenger RNAs carry the information that ribosomes use for the production of specific protein sequences.

## Some RNAs can guide modifications of themselves or other RNAs

In eukaryotes, one of the most striking examples of RNA modification is the splicing of mRNA precursors, a process that is catalyzed by large complexes composed of both proteins and small nuclear RNAs. These small nuclear RNAs play a crucial role in guiding the splicing of messenger RNAs.

Remarkably, some RNA molecules can splice themselves in the absence of other proteins and RNAs. This landmark discovery of self-splicing introns, which we will discuss later in this chapter, revealed that RNA molecules can serve as catalysts, which greatly influenced our view of molecular evolution. Many other types of RNAs, such as small regulatory RNAs and long noncoding RNAs, have been discovered more recently, and while their functions are still under active investigation, our understanding is rapidly expanding.

## Some viruses have RNA genomes

While DNA is the genetic material in most organisms, some viruses have genomes made of RNA ([Section 8.4](#)). RNA viruses are responsible for several diseases, such as influenza, polio, mumps, Ebola, the common cold, and — of particular note — COVID-19. The coronavirus SARS-CoV-2, which causes COVID-19, belongs to a family of coronaviruses that have an unusually large, single-stranded RNA genome. Coronaviruses cause a variety of diseases in mammals and birds that have a wide range of symptoms, including respiratory distress in humans that can potentially be lethal.

RNA viruses vary in their genome organization. They have either single- or double-stranded RNA molecules arranged in single or multiple fragments. Inside their hosts, RNA viruses replicate their genomes with a virus-encoded RNA polymerase that uses RNA as a template. In some viruses, a complementary RNA molecule is made from the single-stranded viral RNA genome, while in other viruses, a double-stranded DNA copy is made that can integrate itself into the host's genome. Viral RNA polymerases do not have the same proofreading ability of other polymerases, which leads to high mutation rates of RNA viruses.

# Messenger RNA vaccines provide protection against diseases



Messenger RNA (mRNA) vaccines take advantage of the fact that cells can be tricked into making proteins they don't usually make, even those from other organisms such as viruses. Unlike other vaccines that use parts of a weakened or inactivated pathogen to trigger an immune response, mRNA vaccines — including those available for the SARS-CoV-2 virus — use sections of pathogenic (usually viral) mRNA that have been generated in the laboratory. When human cells are injected with this mRNA, they will produce the corresponding protein from this set of instructions. This “foreign” protein will be presented on the surface of the cells, trigger an immune response, and provide some level of protection if and when the actual pathogen invades. Even though research on mRNA vaccines had been going on for many years, the COVID-19 pandemic accelerated the development, approval, and distribution of mRNA vaccines to the public.

---

## SELF-CHECK QUESTION

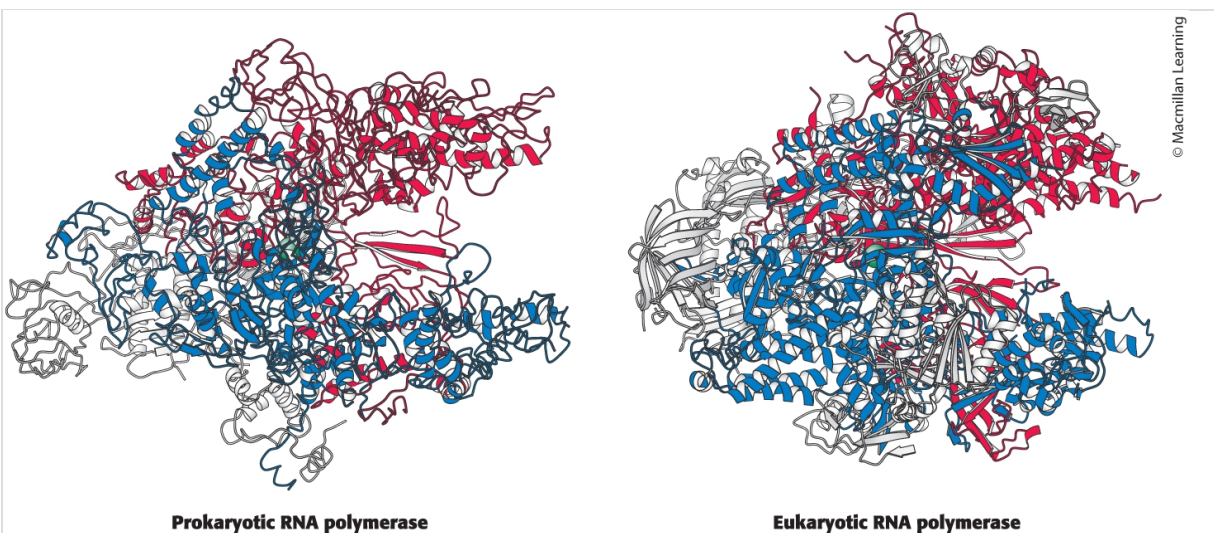


Compare the biological roles of RNA and DNA. What aspects of the structure and chemistry of RNA make it so versatile?

---


## 29.2 RNA Polymerases Catalyze Transcription

**Transcription**, the synthesis of RNA molecules from a DNA template, is catalyzed by large enzymes called **RNA polymerases**. The basic biochemistry of RNA synthesis is shared by all organisms, a commonality that has been beautifully illustrated by the three-dimensional structures of representative RNA polymerases from prokaryotes and eukaryotes (**Figure 29.1**). Despite substantial differences in size and number of polypeptide subunits, the overall structures of these enzymes are quite similar, revealing a common evolutionary origin.



**FIGURE 29.1 RNA polymerase structures in prokaryotes and eukaryotes resemble each other.** In these three-dimensional models of RNA polymerases from a prokaryote (*Thermus aquaticus*) and a eukaryote (*Saccharomyces cerevisiae*), the two largest subunits for each structure are shown in dark red and dark blue. Both structures contain a central metal ion (green) in their active sites, near a large cleft on the right. The similarity of these structures reveals that these enzymes have the same evolutionary origin and have many mechanistic features in common.

[Drawn from 116V.pdb and 116H.pdb.]

INTERACT with these models in  Achieve



RNA polymerases are very large, complex enzymes. For example, the core of the RNA polymerase of *E. coli* consists of five kinds of subunits with the composition  $\alpha_2\beta\beta'\omega$  ([Table 29.1](#)). A typical eukaryotic RNA polymerase is larger and more complex, having 12 subunits and a total molecular mass of more than 500 kDa. Despite this complexity, the detailed structures of RNA polymerases have been determined by x-ray crystallography in work pioneered by Roger Kornberg and Seth Darst. The structures of many additional RNA polymerase complexes have been determined by cryo-electron microscopy.

**TABLE 29.1 Subunits of RNA polymerase from *E. coli***

Subunit	Gene	Number	Mass (kDa)
$\alpha$	<i>rpoA</i>	2	37
$\beta$	<i>rpoB</i>	1	151
$\beta'$	<i>rpoC</i>	1	155
$\omega$	<i>rpoZ</i>	1	10
$\sigma^{70}$	<i>rpoD</i>	1	70



# RNA synthesis comprises three stages: initiation, elongation, and termination

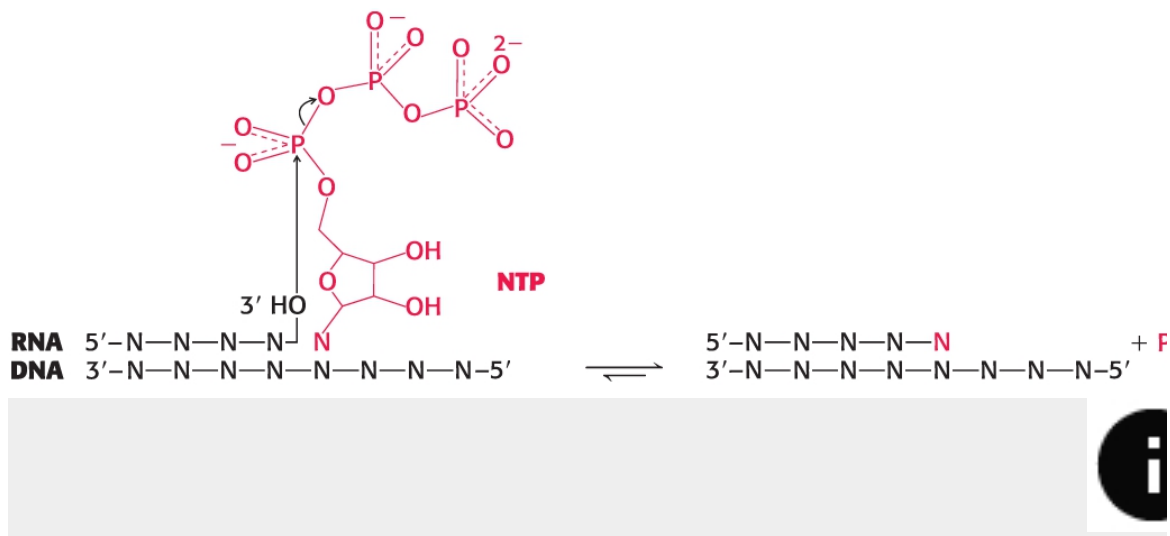
RNA synthesis, like all biological polymerization reactions, takes place in three stages: *initiation*, *elongation*, and *termination*. RNA polymerases perform multiple functions in this process:

1. They search DNA for initiation sites, also called *promoter sites* or simply **promoters**. For instance, *E. coli* DNA has about 2000 promoters in its  $4.6 \times 10^6$  bp genome.
2. They unwind a short stretch of double-helical DNA to produce single-stranded DNA templates from which the sequence of bases can be easily read out.
3. They select the correct ribonucleoside triphosphate and catalyze the formation of a phosphodiester bond. This process is repeated many times as the enzyme moves along the DNA template. RNA polymerase is completely processive — a transcript is synthesized from start to end by a single RNA polymerase molecule.
4. They detect termination signals that specify where a transcript ends.
5. Their activity is regulated by activator and repressor proteins that interact with the promoter and modulate the ability of the RNA polymerase to initiate transcription. Gene expression is controlled substantially at the level of transcription, as will be discussed in detail in [Chapter 31](#).

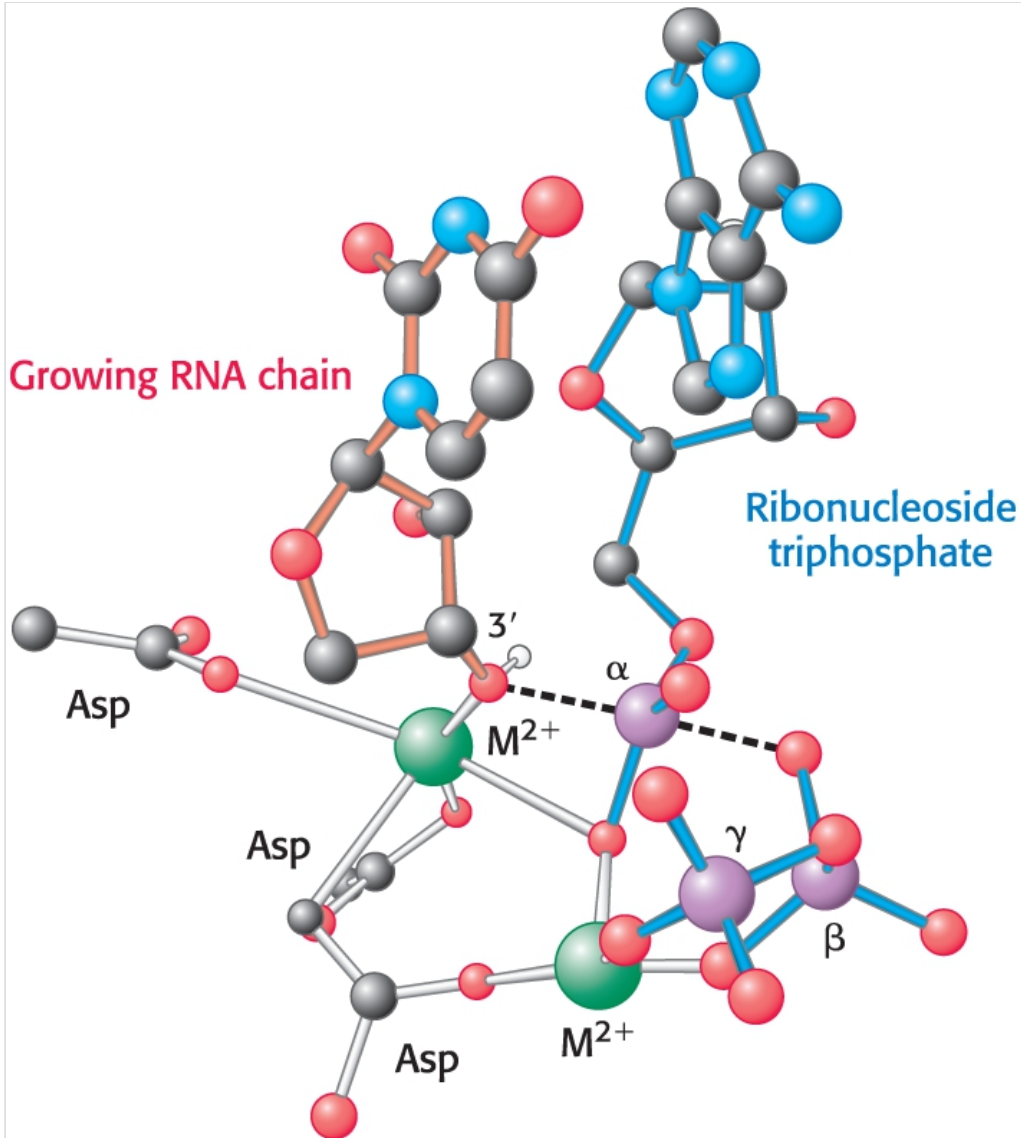
The chemistry of RNA synthesis is identical for all forms of RNA, including messenger RNAs, transfer RNAs, ribosomal RNAs, and small regulatory RNAs, so the basic steps just outlined apply to all forms. Their synthetic processes differ mainly in regulation, the specific RNA polymerase that creates them, and their posttranscriptional processing.

# RNA polymerases catalyze the formation of a phosphodiester bond

The fundamental reaction of RNA synthesis, like that of DNA synthesis, is the formation of a phosphodiester bond. The 3'-hydroxyl group of the last nucleotide in the chain makes a nucleophilic attack on the  $\alpha$ -phosphoryl group of the incoming nucleoside triphosphate, releasing a pyrophosphate.



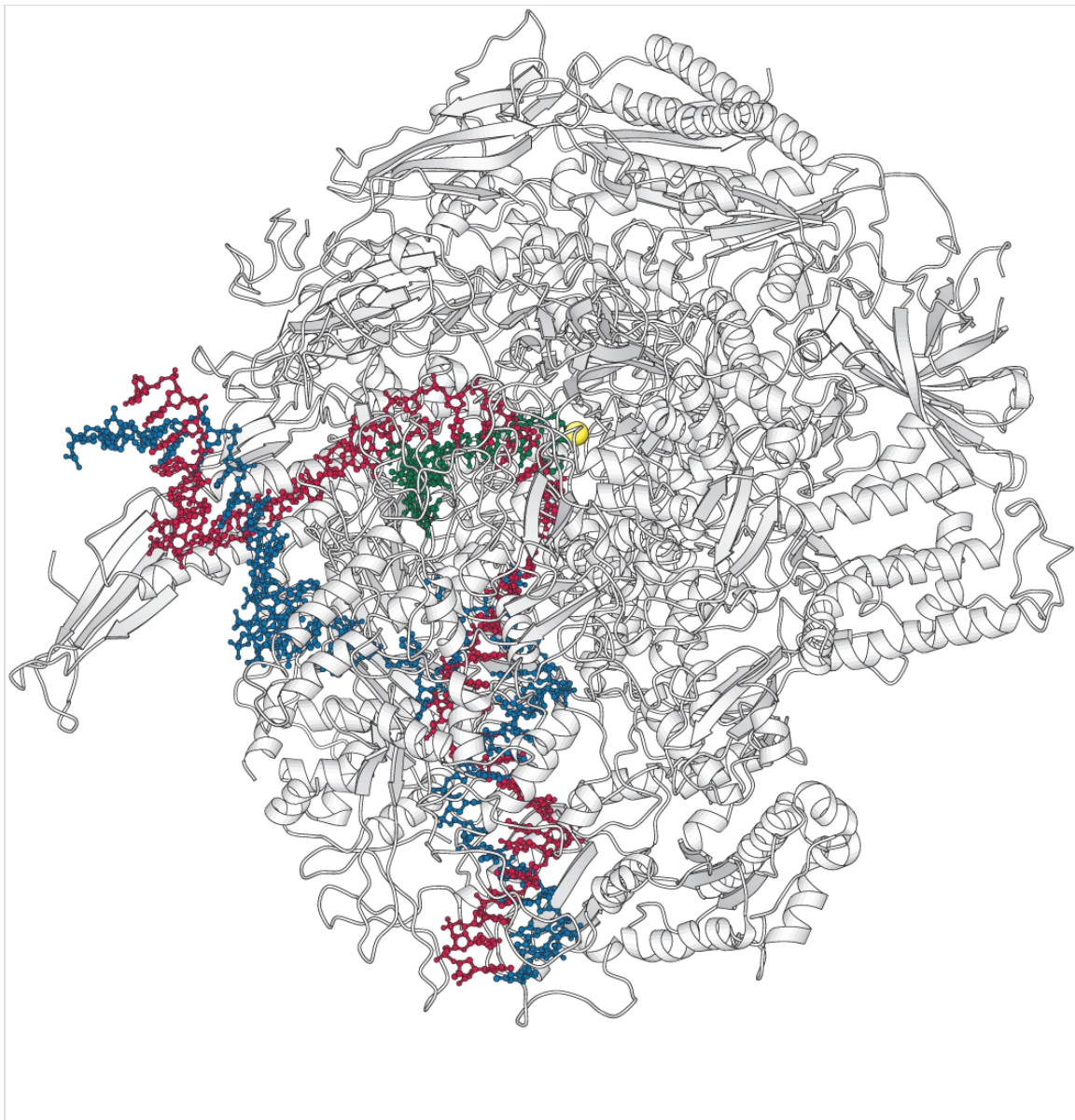
The catalytic sites of all RNA polymerases include two metal ions, normally magnesium ions ([Figure 29.2](#)). One ion remains tightly bound to the enzyme, whereas the other ion comes in with the nucleoside triphosphate and leaves with the pyrophosphate. Three conserved aspartate residues participate in binding these metal ions. Given the recent appreciation of the role of a third metal ion in the active site of DNA polymerases (see [Section 28.1](#)), it will be interesting to see if RNA polymerases show similarities to that model.



**FIGURE 29.2 The RNA polymerase active site contains metal ions.** In this model of the transition state for phosphodiester-bond formation in the active site of RNA polymerase, the 3'-hydroxyl group of the growing RNA chain attacks the  $\alpha$ -phosphoryl group of the incoming nucleoside triphosphate, resulting in the release of pyrophosphate. This transition state is structurally similar to that in the active site of DNA polymerase (see [Figure 28.2](#)).



The polymerization reactions that are catalyzed by both prokaryotic and eukaryotic RNA polymerases take place within a complex in DNA termed a **transcription bubble** (Figure 29.3). This complex consists of double-stranded DNA that has been locally unwound in a region of approximately 17 base pairs. The edges of the bases that normally take part in Watson–Crick base pairs are exposed in the unwound region. We will begin with a detailed examination of the elongation process, including the role of the DNA template read by RNA polymerase and the reactions catalyzed by the polymerase, before returning to the more complex processes of initiation and termination.



**FIGURE 29.3 RNA polymerase separates a region of the double helix to form the transcription bubble.** The red (template) and blue (nontemplate) strands of DNA are shown along with the RNA molecule being synthesized (shown in green). The position of the active site magnesium is indicated (yellow sphere). The DNA enters from the left and exits at the bottom.



## RNA chains are formed *de novo* and grow in the 5'-to-3' direction

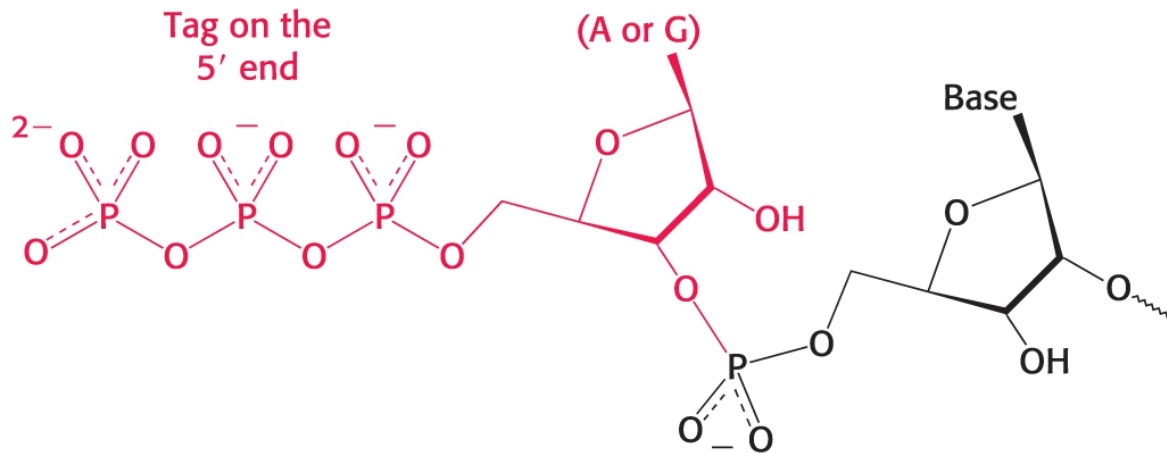
Let us begin our examination of transcription by considering the DNA template. The first nucleotide (the start site) of a DNA sequence to be transcribed is denoted as  $+1$  and the second one as  $+2$ ; the nucleotide preceding the start site is denoted as  $-1$ . These designations refer to the coding strand of DNA. Recall that the sequence of the *template strand* of DNA is the complement of that of the RNA transcript ([Figure 29.4](#)). In contrast, the *coding strand* of DNA has the same sequence as that of the RNA transcript except for thymine (T) in place of uracil (U). The coding strand is also known as the *sense*  $(+)$  strand, and the template strand as the *antisense*  $(-)$  strand.



**FIGURE 29.4** The RNA transcript is complementary to the template, or antisense  $(-)$ , strand

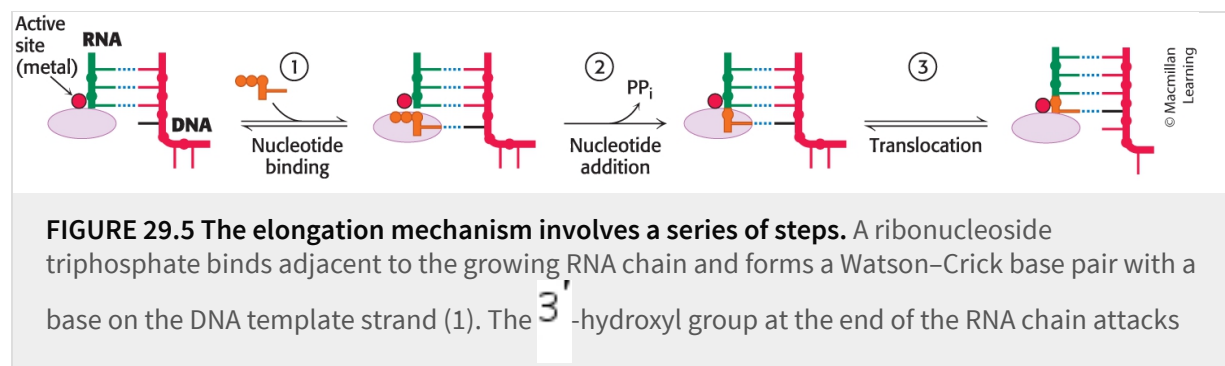


In contrast with DNA synthesis, RNA synthesis can start de novo, without the requirement for a primer. Most newly synthesized RNA chains carry a highly distinctive tag on the 5' end: the first base at that end is either pppG or pppA.



The presence of the triphosphate moiety confirms that RNA synthesis starts at the 5' end.

The dinucleotide shown above is synthesized by RNA polymerase as part of the complex process of initiation, which will be discussed later in the chapter. After initiation takes place, RNA polymerase elongates the nucleic acid chain as follows ([Figure 29.5](#)).



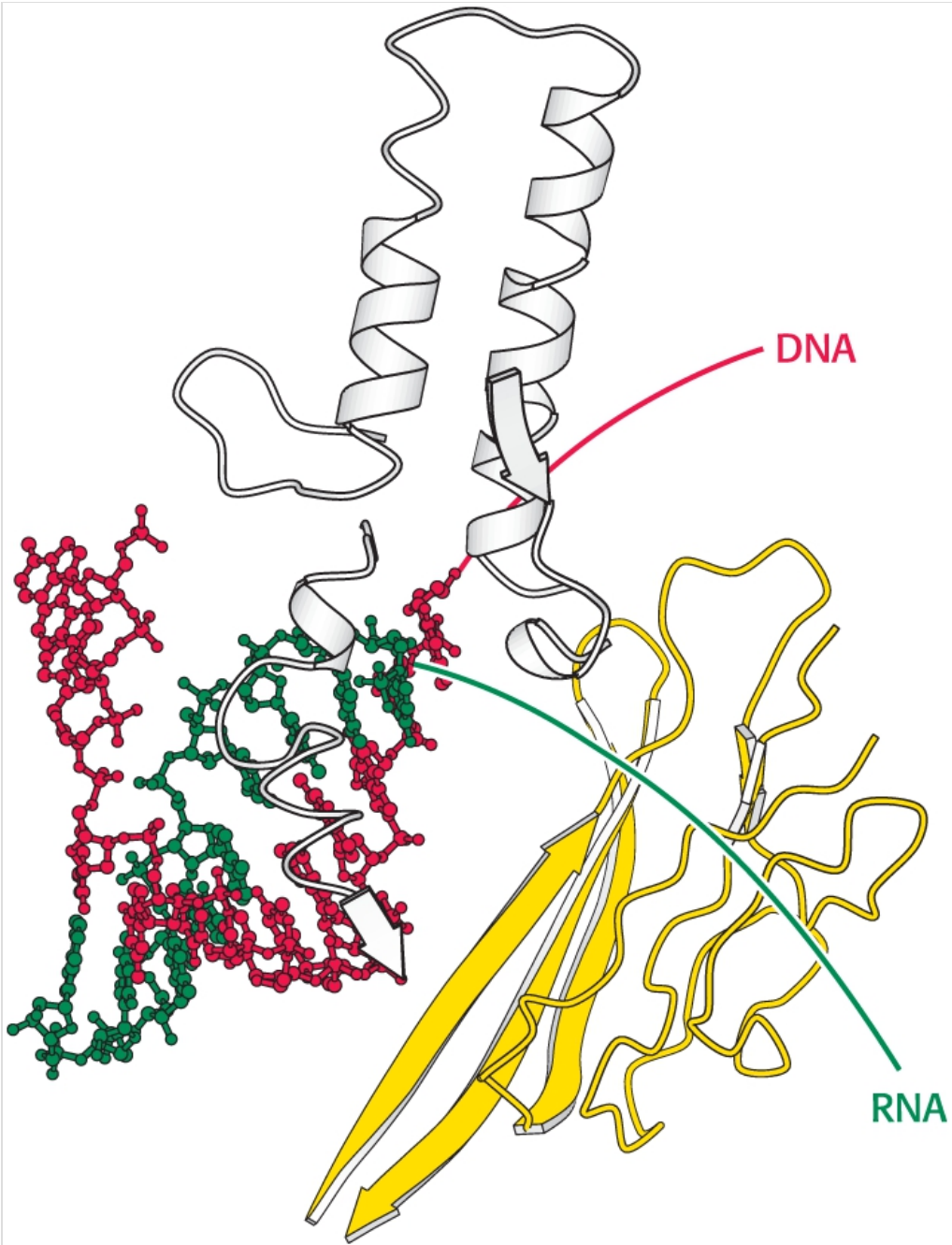
**FIGURE 29.5** The elongation mechanism involves a series of steps. A ribonucleoside triphosphate binds adjacent to the growing RNA chain and forms a Watson–Crick base pair with a base on the DNA template strand (1). The 3'-hydroxyl group at the end of the RNA chain attacks

the newly bound nucleotide and forms a new phosphodiester bond, releasing pyrophosphate (2). After nucleotide addition, the RNA–DNA hybrid can translocate through the RNA polymerase (3), bringing a new DNA base into position to base-pair with an incoming nucleoside triphosphate.




1. A ribonucleoside triphosphate binds in the active site of the RNA polymerase directly adjacent to the growing RNA chain, and it forms a Watson–Crick base pair with the template strand.
2. The 3'-hydroxyl group of the growing RNA chain, which is oriented and activated by the tightly bound metal ion, attacks the  $\alpha$ -phosphoryl group to form a new phosphodiester bond, displacing pyrophosphate.
3. Next, the RNA–DNA hybrid must move relative to the polymerase to bring the 3' end of the newly added nucleotide into proper position for the next nucleotide to be added. This translocation step does not include breaking any bonds between base pairs and is reversible; but, once it has taken place, the addition of the next nucleotide, favored by the triphosphate cleavage and pyrophosphate release and cleavage, drives the polymerization reaction forward.

The lengths of the RNA–DNA hybrid and of the unwound region of DNA stay rather constant as RNA polymerase moves along the DNA template. The length of the RNA–DNA hybrid is determined by a structure within the enzyme that forces the RNA–DNA hybrid to separate, allowing the RNA chain to exit from the enzyme and the DNA chain to rejoin its DNA partner ([Figure 29.6](#)).



**FIGURE 29.6** A structure within RNA polymerase forces the separation of the RNA-DNA hybrid. Notice that the DNA strand exits in one direction and the RNA product exits in another.

[Drawn from 1I6H.pdb.]

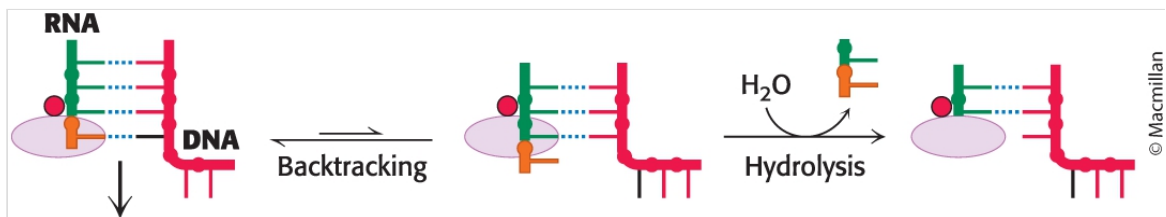
INTERACT with this model in  Achieve





## RNA polymerases backtrack and correct errors

The RNA–DNA hybrid can also move in the direction opposite that of elongation ([Figure 29.7](#)). This backtracking is energetically less favorable than moving forward because it breaks the bonds between a base pair. However, backtracking is very important for *proofreading*. The incorporation of an incorrect nucleotide introduces a non-Watson–Crick base pair. In this case, breaking the bonds between this base pair and backtracking is energetically less costly.



**FIGURE 29.7** The RNA–DNA hybrid can occasionally backtrack within the RNA polymerase. In the backtracked position, hydrolysis can take place, producing a configuration equivalent to that after translocation. Backtracking is more likely if a mismatched base is added, facilitating proofreading.



After the polymerase has backtracked, the phosphodiester bond one base pair before the one that has just formed is adjacent to the metal ion in the active site. In this position, a hydrolysis reaction in which a water molecule attacks the phosphate can result in the cleavage of the

phosphodiester bond and the release of a dinucleotide that includes the incorrect nucleotide.

Studies of single molecules of RNA polymerase have confirmed that the enzymes pause and backtrack to correct errors. Furthermore, these proofreading activities are often enhanced by accessory proteins.

The final error rate of the order of one mistake per  $10^4$  or  $10^5$  nucleotides is higher than that for DNA replication, including all error-correcting mechanisms. The lower fidelity of RNA synthesis can be tolerated because mistakes are not transmitted to progeny. For most genes, many RNA transcripts are synthesized; a few defective transcripts are unlikely to be harmful.

---

## SELF-CHECK QUESTION



List some ways in which RNA polymerases are similar to DNA polymerases, and how they are different.

---

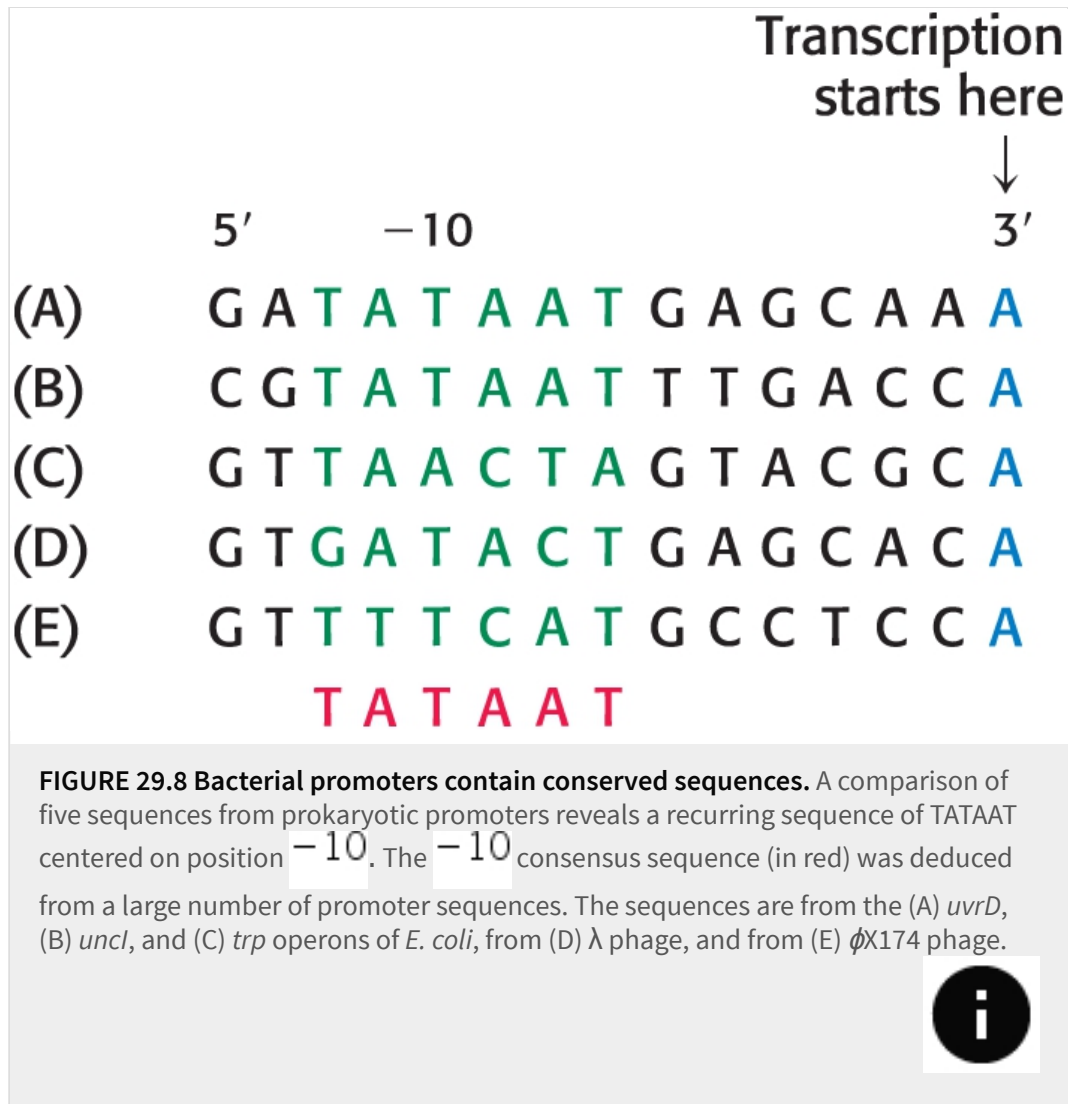
## RNA polymerase binds to promoter sites on the DNA template in bacteria to initiate transcription

While the elongation process is common to all organisms, the processes of initiation and termination differ substantially in bacteria and eukaryotes. We begin with a discussion of these processes in bacteria, starting with initiation of transcription.

The bacterial RNA polymerase discussed earlier with the composition  $\alpha_2\beta\beta'\omega$  is referred to as the *core enzyme*. The inclusion of an additional subunit ( $\sigma$ ) produces the *holoenzyme* with composition  $\alpha_2\beta\beta'\omega\sigma$ . The  **$\sigma$  subunit** helps find the promoters, sites on DNA where transcription begins. At these sites, the  $\sigma$  subunit participates in the initiation of RNA synthesis and then dissociates from the rest of the enzyme.

Sequences upstream of the promoter site are important in determining where transcription begins. A striking pattern is evident when the sequences of bacterial promoters are compared: Two common motifs are present on the upstream side of the transcription start site. They are known as the  $-10$  sequence and the  $-35$  sequence because they are centered at about 10 and 35 nucleotides upstream of the start site. The region containing these sequences is called the *core promoter*. The  $-10$  and  $-35$  sequences are each 6 bp long. Their **consensus sequences**, deduced from analyses of many promoters (**Figure 29.8**), are:





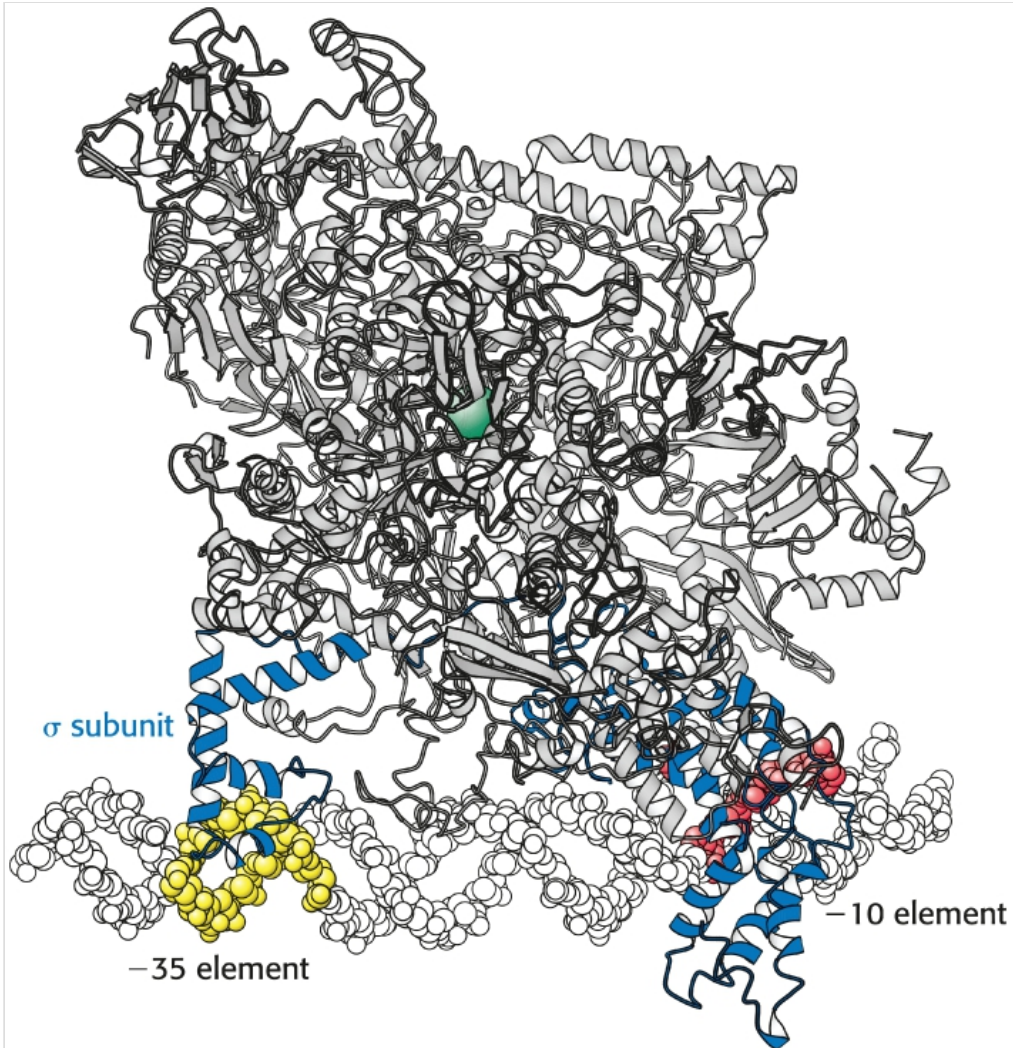
Promoters differ markedly in their efficacy. Some genes are transcribed frequently — as often as every 2 seconds in *E. coli*. The promoters for these genes are referred to as *strong promoters*. In contrast, other genes are transcribed much less frequently, about once in 10 minutes; the promoters for these genes are *weak promoters*. The  $-10$  and  $-35$  regions of most strong promoters have sequences that correspond closely to the consensus sequences, whereas weak promoters tend to have multiple substitutions at these sites. Indeed, mutation of a single base in either the  $-10$  sequence or the  $-35$  sequence can diminish promoter activity.

The distance between these conserved sequences also is important; a separation of 17 nucleotides is optimal. Thus, the efficiency or strength of a promoter sequence serves to regulate transcription. Regulatory proteins that bind to specific sequences near promoter sites and interact with RNA polymerase ([Chapter 31](#)) also markedly influence the frequency of transcription of many genes.

Outside the core promoter in a subset of highly expressed genes is the *upstream element* (also called the UP element). This sequence is present from 40 to 60 nucleotides upstream of the transcription start site. The UP element is bound by the  $\alpha$  subunit of RNA polymerase and serves to increase the efficiency of transcription by creating an additional interaction site for the polymerase.


## Sigma subunits of RNA polymerase in bacteria recognize promoter sites

To initiate transcription, the  $\alpha_2\beta\beta'\omega$  core of RNA polymerase must bind the promoter. However, it is the  $\sigma$  subunit that makes this binding possible by enabling RNA polymerase to recognize promoter sites. In the presence of the  $\sigma$  subunit, the RNA polymerase binds weakly to the DNA and slides along the double helix until it dissociates or encounters a promoter. The  $\sigma$  subunit recognizes the promoter through several interactions with the nucleotide bases of the promoter DNA. The structure of a bacterial RNA polymerase holoenzyme bound to a promoter site shows the  $\sigma$  subunit interacting with DNA at the  $-10$  and  $-35$  regions essential to promoter recognition ([Figure 29.9](#)). Therefore, the  $\sigma$  subunit is responsible for the specific binding of the RNA polymerase to a promoter site on the template DNA. The  $\sigma$  subunit is generally released when the nascent RNA chain reaches 9 or 10 nucleotides in length. After its release, it can associate with another core enzyme and assist in a new round of initiation.



**FIGURE 29.9** *E. coli* RNA polymerase holoenzyme complex binds to conserved promoter sequences. The  $\sigma$  subunit (blue) of the bacterial RNA polymerase holoenzyme makes sequence-specific contacts with the  $-10$  (red) and  $-35$  (yellow) promoter elements. The active site of the polymerase is revealed by a bound metal ion (purple).

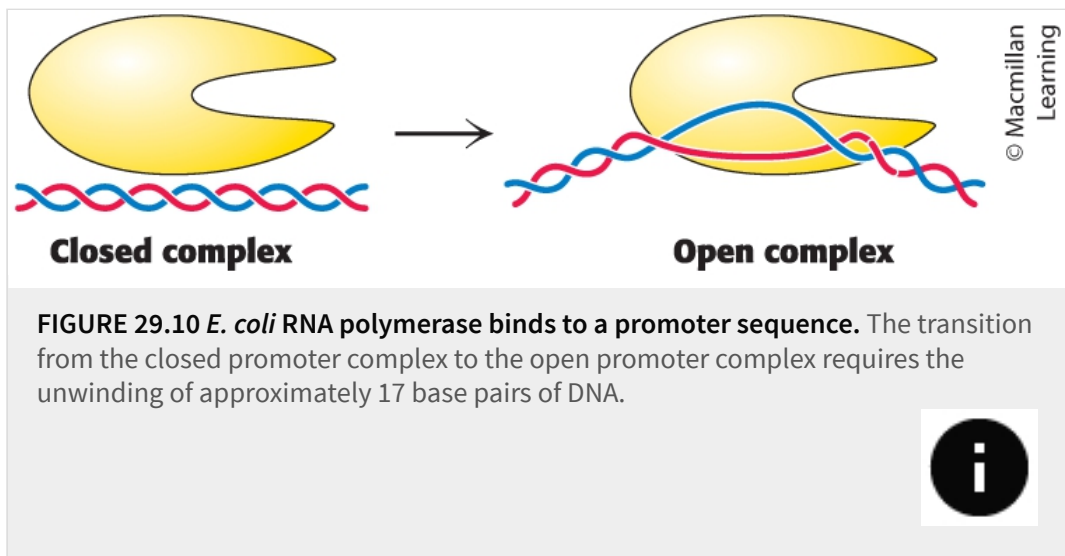
[Drawn from 1L9Z.pdb].

INTERACT with this model in  Achieve



# The template double helix must be unwound for transcription to take place

Although the bacterial RNA polymerase can search for promoter sites when bound to double-helical DNA, a segment of the DNA double helix must be unwound before synthesis can begin. The transition from the closed promoter complex (in which DNA is double helical) to the open promoter complex (in which a DNA segment is unwound) is an essential event in both bacterial and eukaryotic transcription. In bacteria it is the RNA polymerase itself that accomplishes this ([Figure 29.10](#)), while in eukaryotes additional proteins are required to unwind the DNA template.



We know that, in bacteria, the free energy necessary to break the bonds between approximately 17 base pairs in the double helix is derived from additional interactions between the template and the bacterial RNA polymerase. These interactions become possible when the DNA distorts to wrap around the RNA polymerase; they also occur between the single-stranded DNA regions and other parts of the enzyme. These interactions stabilize the open promoter complex and help pull the template strand into the active site. The  $-35$  element remains in a double-helical state, whereas the  $-10$  element is unwound. The stage is

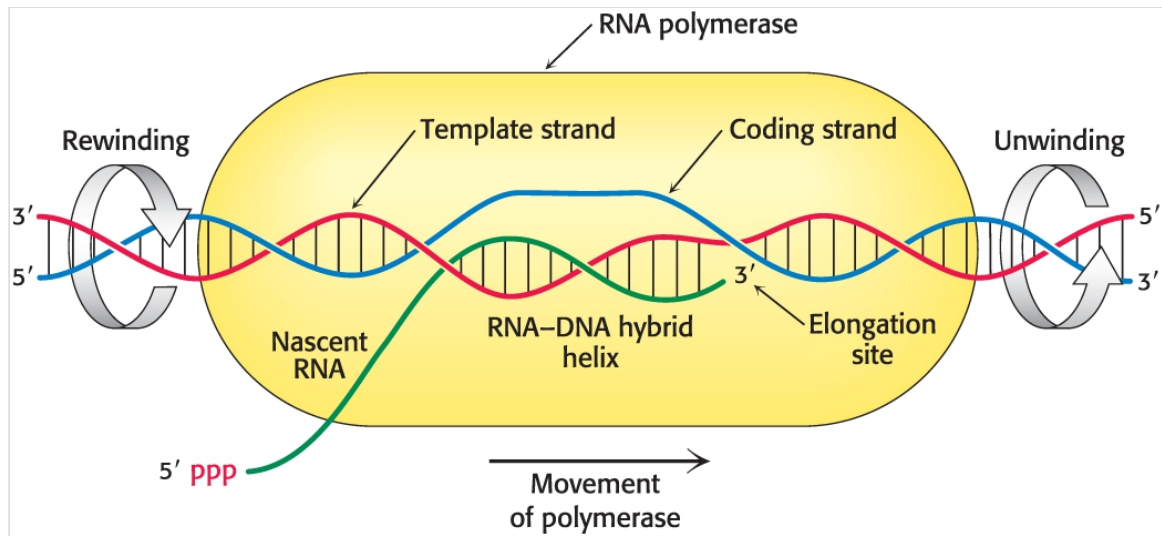
now set for the formation of the first phosphodiester bond of the new RNA chain.

## Elongation takes place at transcription bubbles that move along the DNA template

The elongation phase of RNA synthesis begins with the formation of the first phosphodiester bond, after which repeated cycles of nucleotide addition can take place. However, until about 10 nucleotides have been added, RNA polymerase sometimes releases the short RNA, which dissociates from the DNA and gets degraded. Once RNA polymerase passes this point, the enzyme stays bound to its template until a termination signal is reached.

The region containing the unwound DNA template and nascent RNA corresponds to the transcription bubble ([Figure 29.11](#)). The newly synthesized RNA forms a hybrid helix with the template DNA strand. This RNA–DNA helix is about 8 bp long, which corresponds to nearly one turn of a double helix. The 3'-hydroxyl group of the RNA in this hybrid helix is positioned so that it can attack the  $\alpha$ -phosphorus atom of an incoming ribonucleoside triphosphate. The core bacterial RNA polymerase also contains a binding site for the coding strand of DNA.





**FIGURE 29.11** The transcription bubble moves with the RNA polymerase during elongation. In this schematic representation of a transcription bubble in the elongation of an RNA transcript, duplex DNA is unwound at the forward end of RNA polymerase and rewound at its rear end. The RNA-DNA hybrid rotates during elongation.



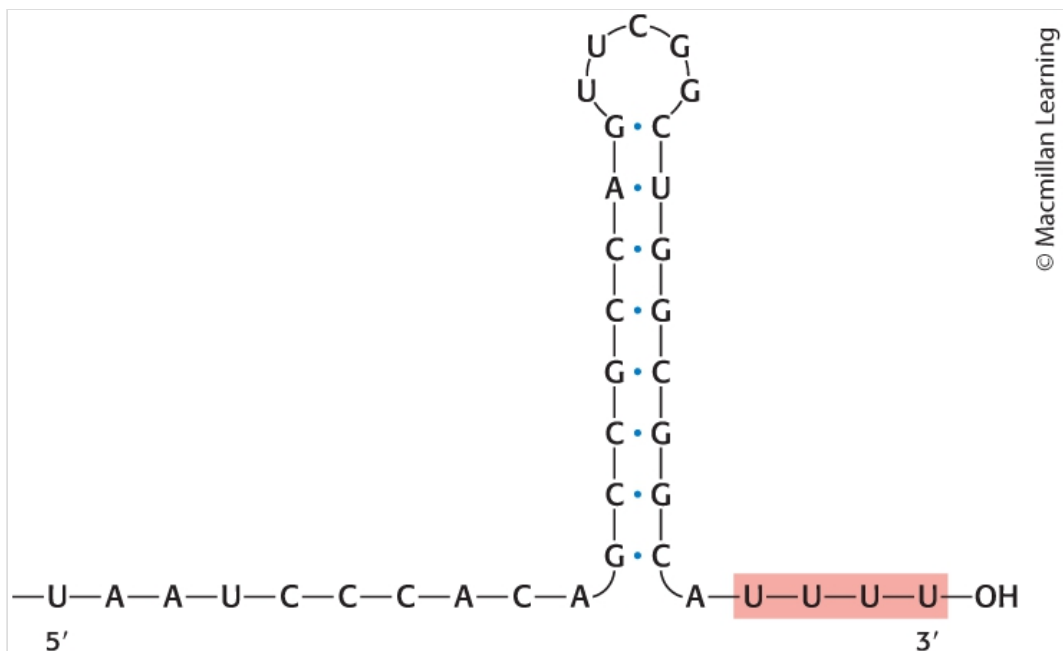
As in the initiation phase, about 17 bp of DNA are unwound throughout the elongation phase. The transcription bubble moves a distance of 170 Å (17 nm) in a second, which corresponds to a rate of elongation of about 50 nucleotides per second. Although rapid, it is much slower than the rate of DNA synthesis, which is 800 nucleotides per second.

## Sequences within the newly transcribed RNA signal termination

How does the RNA polymerase know where to stop transcription? In the termination phase of transcription, the formation of phosphodiester bonds ceases, the RNA-DNA hybrid dissociates, the unwound region of DNA rewinds, and RNA polymerase releases the DNA. This process is as precisely controlled as initiation. So what determines where

transcription is terminated? In both eukaryotes and bacteria, the transcribed regions of DNA templates contain so-called *intrinsic* termination signals.

In bacteria, the simplest intrinsic termination signal is a palindromic GC-rich region followed by an AT-rich region. The RNA transcript of this DNA palindrome is self-complementary ([Figure 29.12](#)). Hence, its bases can pair to form a hairpin structure with a stem and loop, a structure favored by its high content of G and C residues. Guanine–cytosine base pairs are more stable than adenine–thymine pairs, primarily because of the preferred base-stacking interactions in G–C base pairs ([Section 1.3](#)). This stable hairpin is followed by a sequence of four or more uracil residues, which also are crucial for termination. The RNA transcript ends within or just after them.



**FIGURE 29.12** Some genes contain intrinsic termination signals. A typical intrinsic termination signal found at the 3' end of an mRNA transcript consists of a series of bases that form a stable stem-loop structure followed by a string of U residues.

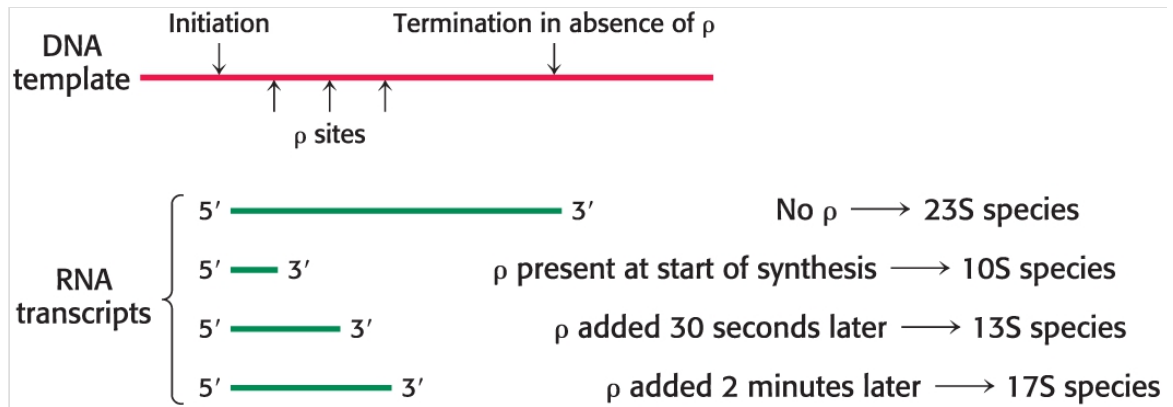


How does this combination hairpin–oligo(U) structure terminate transcription? First, RNA polymerase likely pauses immediately after it has synthesized a stretch of RNA that folds into a hairpin. Furthermore, the RNA–DNA hybrid helix produced after the hairpin is unstable because its rU–dA base pairs are the weakest of the four kinds. Hence, the pause in transcription caused by the hairpin permits the weakly bound nascent RNA to dissociate from the DNA template and then from the enzyme. The solitary DNA template strand rejoins its partner to reform the DNA duplex, and the transcription bubble closes.

## In bacteria, the rho protein helps to terminate the transcription of some genes

Bacterial RNA polymerase needs no help to terminate transcription at the intrinsic sites described above. At other sites, however, termination requires the participation of an additional factor. This discovery was prompted by the observation that some RNA molecules synthesized in vitro by RNA polymerase acting alone are longer than those made in vivo. The missing factor, a protein that caused the correct termination, was isolated and named **rho** ( $\rho$ ).

Additional information about the action of  $\rho$  was obtained by adding this termination factor to an incubation mixture at various times after the initiation of RNA synthesis (**Figure 29.13**). RNAs with sedimentation coefficients of 10S, 13S, and 17S were obtained when  $\rho$  was added at initiation, a few seconds after initiation, and 2 minutes after initiation, respectively. If no  $\rho$  was added, transcription yielded a 23S RNA product. It is evident that the template contains at least three termination sites that respond to  $\rho$  (yielding 10S, 13S, and 17S RNA) and one termination site that does not (yielding 23S RNA). Thus, specific termination at a site producing 23S RNA can take place in the absence of  $\rho$ . However,  $\rho$  detects additional termination signals that are not recognized by RNA polymerase alone.



**FIGURE 29.13  $\rho$  protein promotes the termination of transcription.** The DNA template shown here contains three  $\rho$ -dependent termination sites and one site that does not require the action of the  $\rho$  protein.

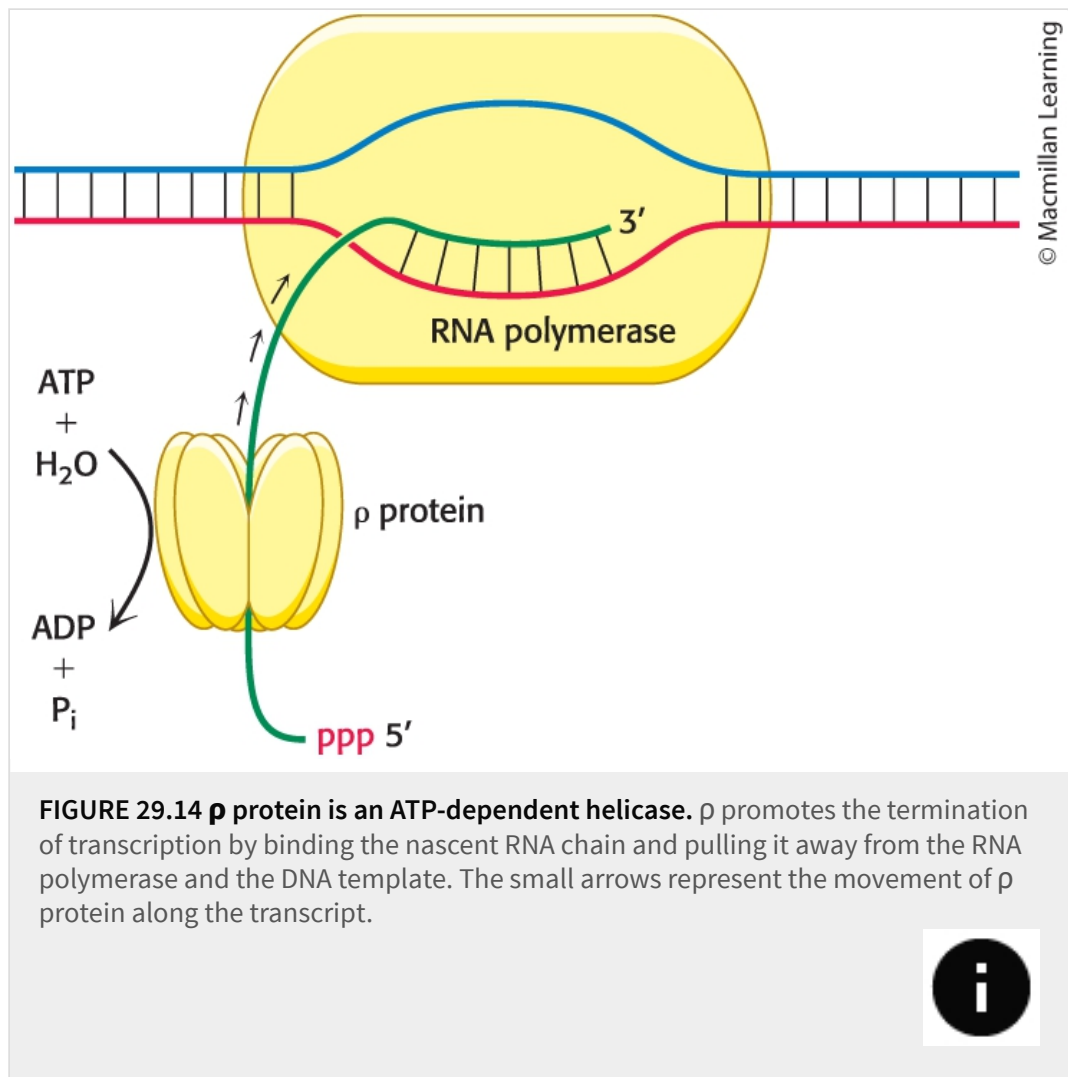


The  $\rho$  protein promotes about 20% of termination events in bacteria, but exactly how it selects its target termination signals is not clear. Unlike the hairpin–oligo(U) sequence of intrinsic termination sites, the identification of conserved patterns in  $\rho$ -dependent terminators has proven more difficult.

How does  $\rho$  promote the termination of RNA synthesis? A key clue is the finding that  $\rho$  is hexameric and hydrolyzes ATP in the presence of single-stranded RNA but not in the presence of DNA or duplex RNA. Thus  $\rho$  is a helicase, homologous to the hexameric helicases that we encountered in our discussion of DNA replication ([Section 28.1](#)). The role of  $\rho$  in the termination of transcription in bacteria is as follows ([Figure 29.14](#)):

- The  $\rho$  protein is brought into action by sequences located in the nascent RNA that are rich in cytosine and poor in guanine.
- A stretch of nucleotides is bound in such a way that the RNA passes through the center of the structure.

- The helicase activity of  $\rho$  enables the protein to pull the nascent RNA while pursuing RNA polymerase.
- When  $\rho$  catches RNA polymerase at the transcription bubble, it breaks the RNA–DNA hybrid by functioning as an RNA–DNA helicase.



Proteins in addition to  $\rho$  may promote termination. For example, the NusA protein enables RNA polymerase in *E. coli* to recognize a characteristic class of termination sites. A common feature of transcription termination, whether it relies on a protein or not, is that

the functioning signals lie in newly synthesized RNA rather than in the DNA template.

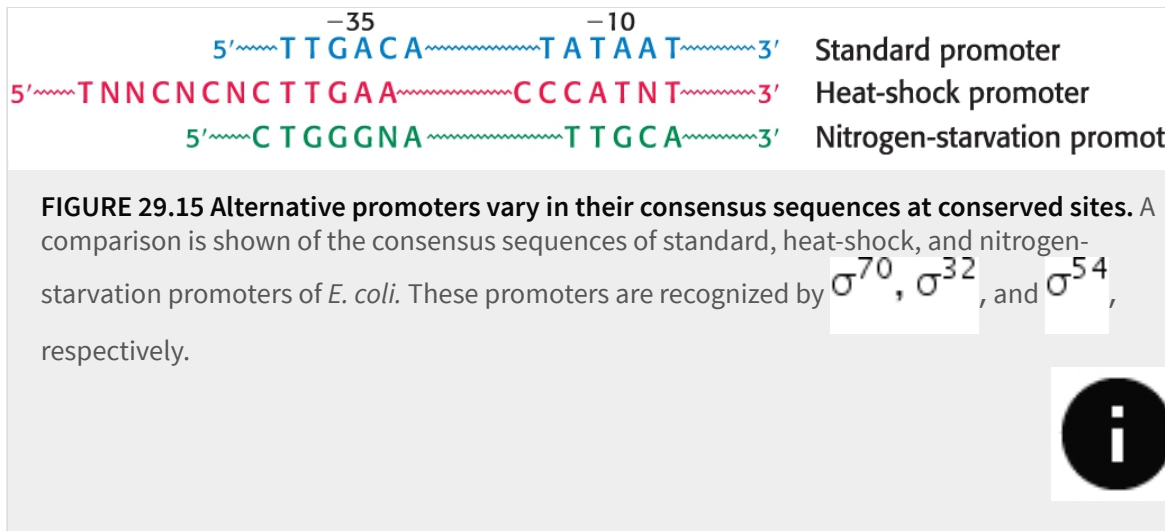
## 29.3 Transcription Is Highly Regulated

As we will see in [Chapter 31](#), the level at which different genes are transcribed is highly regulated. Regulated gene expression is critical for the development of multicellular organisms, the differentiation of various cell types, and the response of bacteria to changes in their environment. Here, we will discuss a few examples of how transcription can be controlled.

### Alternative sigma subunits in bacteria control transcription in response to changes in conditions

As noted above, the  $\sigma$  factor allows for the specific binding of the bacterial RNA polymerase to a promoter site on the template DNA. *E. coli* has seven distinct  $\sigma$  factors for recognizing several types of promoter sequences in *E. coli* DNA. The type that recognizes the consensus sequences described earlier is called  $\sigma^{70}$  because it has a mass of 70 kDa. A different  $\sigma$  factor comes into play when the temperature is raised abruptly. *E. coli* responds by synthesizing  $\sigma^{32}$ , which recognizes the promoters of so-called *heat-shock genes*. These promoters exhibit  $-10$  sequences that are somewhat different from the  $-10$  sequence for standard promoters ([Figure 29.15](#)). The increased transcription of heat-shock genes leads to the coordinated synthesis of a series of protective proteins. Other  $\sigma$  factors respond to environmental conditions, such as nitrogen starvation. These findings demonstrate

that  $\sigma$  plays the key role in determining when and where RNA polymerase initiates transcription.



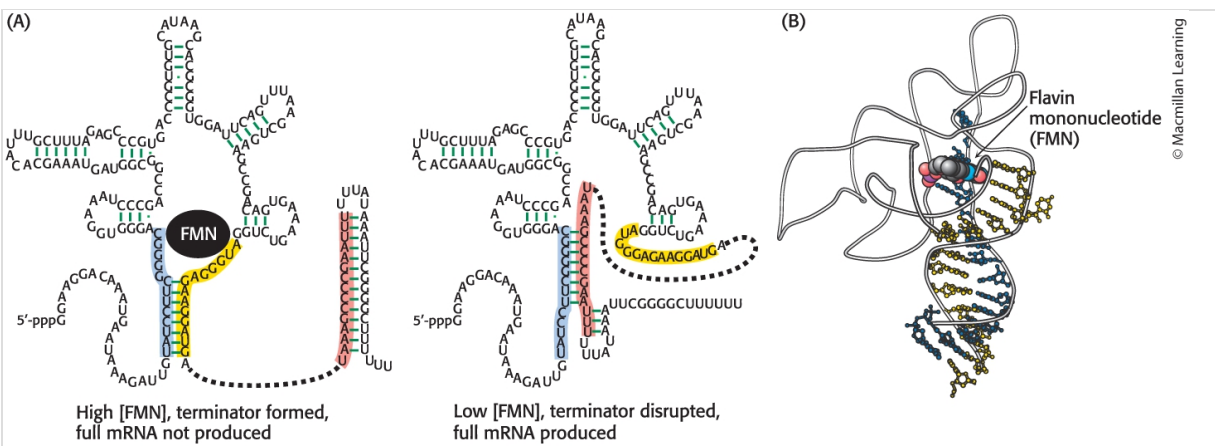
Some other bacteria contain a much larger number of  $\sigma$  factors. For example, the genome of the soil bacterium *Streptomyces coelicolor* encodes more than 60  $\sigma$  factors recognized on the basis of their amino acid sequences. This repertoire allows these cells to adjust their gene-expression programs to the wide range of conditions, with regard to nutrients and competing organisms, that they may experience.

## Some messenger RNAs directly sense metabolite concentrations

As we shall explore in [Chapter 31](#), the expression of many genes is controlled in response to the concentrations of metabolites and signaling molecules within cells. One set of control mechanisms found in both prokaryotes and eukaryotes depends on the remarkable ability of some mRNA molecules to form secondary structures that are capable of directly binding small molecules. These structures are termed [riboswitches](#).




Consider a riboswitch that controls the synthesis of genes that participate in the biosynthesis of riboflavin in the bacterium *Bacillus subtilis* (**Figure 29.16**). When flavin mononucleotide (FMN), a key intermediate in riboflavin biosynthesis, is present at high concentration, it binds to the RNA transcript. Binding of FMN to the transcript induces a hairpin structure that favors premature termination. By trapping the RNA transcript in this termination-favoring conformation, FMN prevents the production of functional full-length mRNA. However, when FMN is present at low concentration, it does not readily bind to the mRNA. Without FMN bound, the transcript adopts an alternative conformation without the terminator hairpin, allowing the production of the full-length mRNA. The occurrence of riboswitches serves as a vivid illustration of how RNAs are capable of forming elaborate, functional structures, though in the absence of specific information we tend to depict them as simple lines.



**FIGURE 29.16 Riboswitches use metabolite concentrations to regulate transcription.** (A) The 5' end of an mRNA that encodes proteins engaged in the production of flavin mononucleotide (FMN) folds to form a structure that is stabilized by binding FMN. This structure includes a terminator that leads to premature termination of the mRNA. At lower concentrations of FMN, an alternative structure that lacks the terminator is formed, leading to the production of full-length mRNA. (B) In the three-dimensional structure of a related FMN-binding riboswitch bound to FMN, the blue and yellow stretches correspond to regions highlighted in the same colors in part A. The yellow strand contacts the bound FMN, stabilizing the structure.

[Drawn from 3F2Q.pdb].

INTERACT with this model in  Achieve

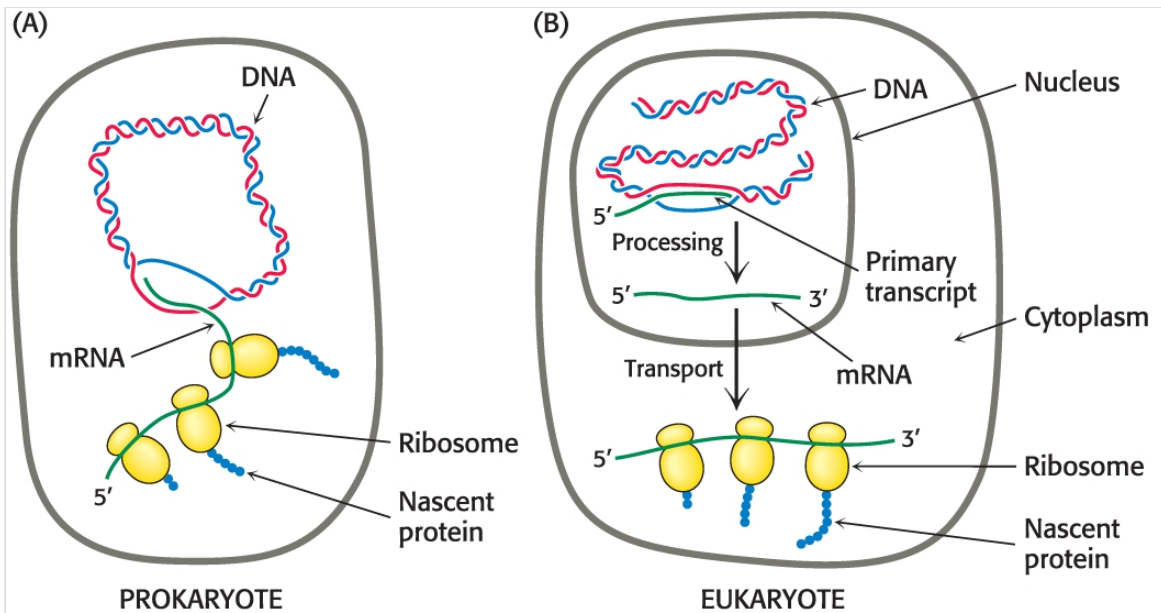


## Control of transcription in eukaryotes is highly complex

We turn now to transcription in eukaryotes, a much more complex process than in bacteria. Eukaryotic cells have a remarkable ability to regulate precisely the time at which each gene is transcribed and how much RNA is produced. This ability led to the evolution of multicellular eukaryotes with distinct tissues. That is, multicellular eukaryotes use differential transcriptional regulation to create different cell types.

Gene expression is influenced by three important characteristics unique to eukaryotes: the nuclear membrane, complex transcriptional regulation, and RNA processing.

1. *The nuclear membrane allows transcription and translation to take place in different cellular compartments.* Transcription takes place in the membrane-bound nucleus, whereas translation takes place outside the nucleus in the cytoplasm. In bacteria, the two processes are closely coupled ([Figure 29.17](#)). Indeed, the translation of bacterial mRNA begins while the transcript is still being synthesized. The spatial and temporal separation of transcription and translation enables eukaryotes to regulate gene expression in much more intricate ways, contributing to the richness of eukaryotic form and function.



**FIGURE 29.17** Transcription and translation are closely coupled in prokaryotes, whereas they are spatially and temporally separate in eukaryotes. (A) In prokaryotes, the primary transcript serves as mRNA and is used immediately as the template for protein synthesis. (B) In eukaryotes mRNA precursors are processed and spliced in the nucleus before being transported to the cytoplasm for translation into protein.

[Information from J. Darnell, H. Lodish, and D. Baltimore. *Molecular Cell Biology*, 2nd ed. (Scientific American Book 1990), p. 230.]



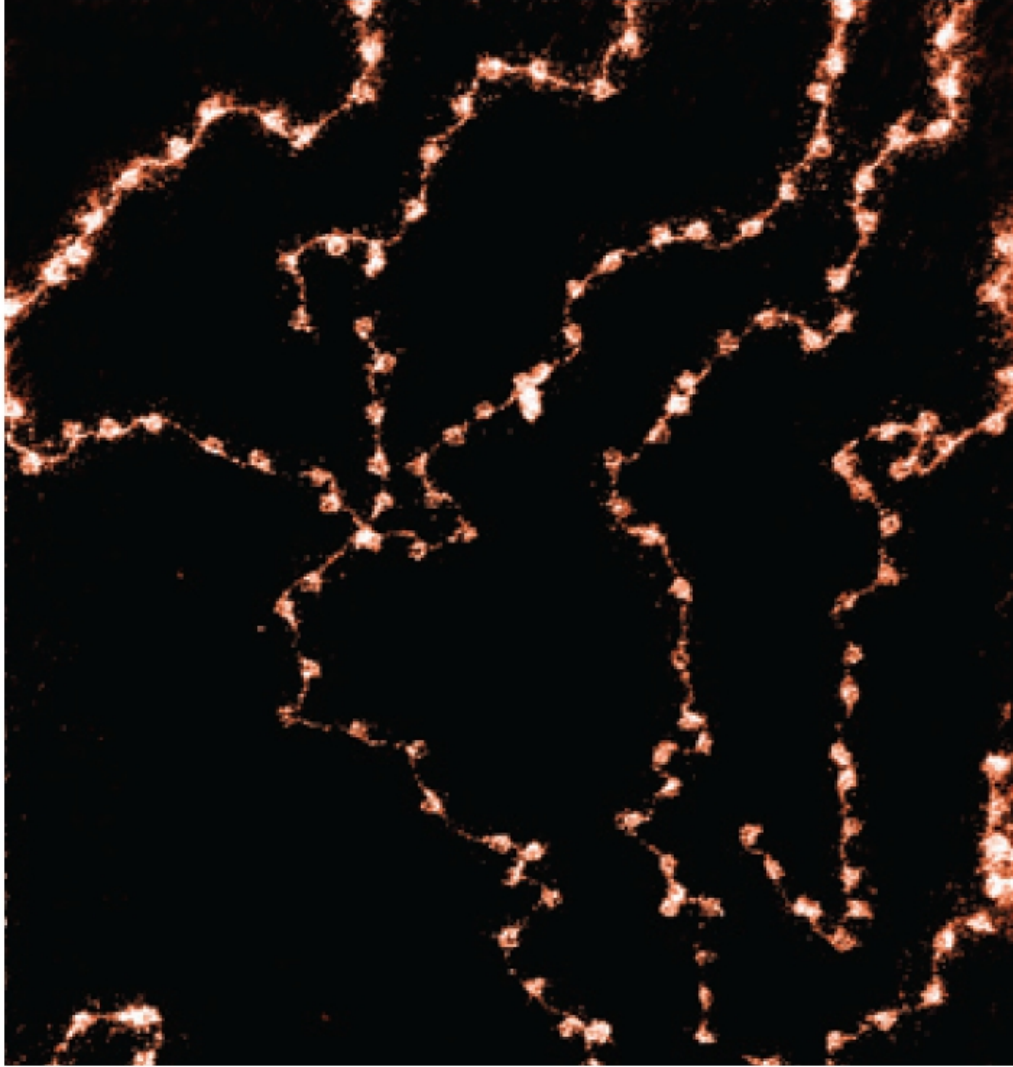
2. *A variety of types of promoter elements enables complex transcriptional regulation.* Like bacteria, eukaryotes rely on conserved sequences in DNA to regulate the initiation of transcription. But bacteria have only three promoter elements (the  $-10$ ,  $-35$ , and UP elements), whereas eukaryotes use a variety of types of promoter elements, each identified by its own conserved sequence. Not all possible types will be present together in the same promoter. In eukaryotes, elements that regulate transcription can be found upstream or downstream of the start site and sometimes at distances much farther from the start site than in prokaryotes. For example, enhancer elements located on DNA far from the start site increase the promoter activity of specific genes.

3. *The degree of RNA processing is much greater in eukaryotes than in bacteria.* Although both bacteria and eukaryotes modify RNA, eukaryotes very extensively process nascent RNA destined to become mRNA. This processing includes modifications to both ends and, most significantly, splicing out segments of the primary transcript. RNA processing is described in [Section 29.4](#).

## Eukaryotic DNA is organized into chromatin

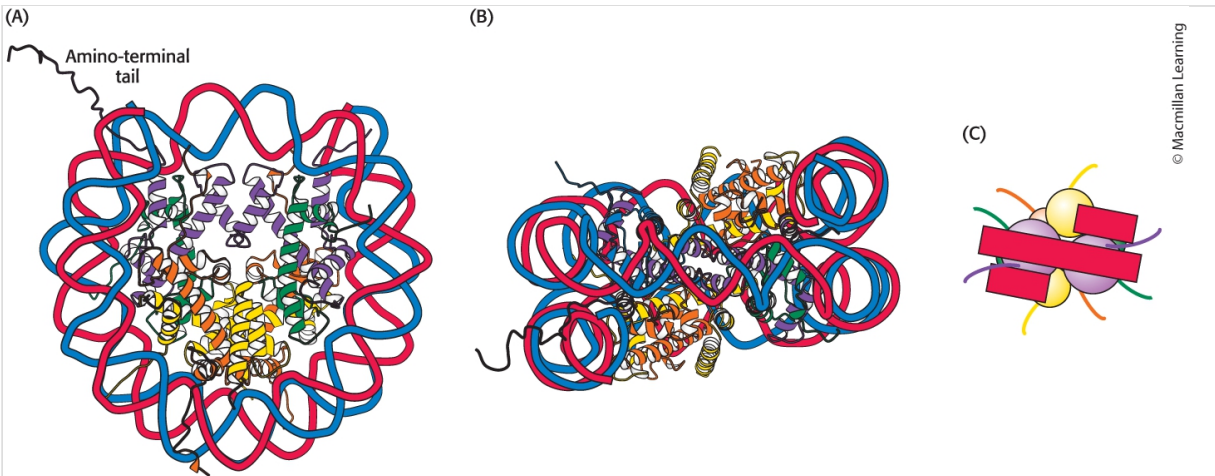
Whereas bacterial genomic DNA is relatively accessible to the proteins involved in transcription, eukaryotic DNA is packaged into **chromatin**, a complex formed between the DNA and a particular set of proteins. Chromatin compacts and organizes eukaryotic DNA, and its presence has dramatic consequences for gene regulation. Although the principles for the construction of chromatin are relatively simple, the chromatin structure for a complete genome is quite complicated. Importantly, in any given eukaryotic cell, some genes and their associated regulatory regions are relatively accessible for transcription and regulation, whereas other genes are tightly packaged, less accessible, and therefore inactive. Eukaryotic gene regulation frequently requires the manipulation of chromatin structure.

Chromatin viewed with the electron microscope has the appearance of beads on a string ([Figure 29.18](#)). Partial digestion of chromatin with DNase exposes these particles, which consist of fragments of DNA (the “string”) wrapped around octamers of proteins called *histones* (the “beads”). The complex formed by a histone octamer and a 145-bp DNA fragment is called the **nucleosome** ([Figure 29.19](#)).



Don. W. Fawcett/Science Source.

**FIGURE 29.18 Eukaryotic chromatin structure resembles beads on a string.** In this electron micrograph of chromatin, the “beads” correspond to DNA complexed with specific proteins into nucleosomes. Each bead has a diameter of approximately 100 Å.



**FIGURE 29.19** The structure of the nucleosome consists of eight histone proteins surrounded by DNA. (A) A view showing the DNA wrapping around the histone core. (B) A 90-degree rotation of the view in part A. The DNA forms a left-handed superhelix as it wraps around the core. (C) A schematic view.

[Drawn from 1A0I.pdb.]



The overall structure of the nucleosome was revealed through electron microscopic and x-ray crystallographic studies pioneered by Aaron Klug and his colleagues. More recently, the three-dimensional structures of reconstituted nucleosomes have been determined to higher resolution by x-ray diffraction methods. The histone octamer is a complex of four different types of histones (H2A, H2B, H3, and H4) that are homologous and similar in structure.

The eight histones in the core are arranged into a  $(H3)_2(H4)_2$  tetramer and a pair of H2A–H2B dimers. The tetramer and dimers come together to form a left-handed superhelical ramp around which the DNA wraps. In addition, each histone has an amino-terminal tail that extends out from the core structure. These tails are flexible and contain many lysine and arginine residues. As we shall see in [Chapter 31](#), covalent modifications of these tails play an essential role in regulating gene expression.

# Three types of RNA polymerase synthesize RNA in eukaryotic cells

In bacteria, RNA is synthesized by a single kind of polymerase. In contrast, the nucleus of a typical eukaryotic cell contains three types of RNA polymerase differing in template specificity and location in the nucleus ([Table 29.2](#)). The three polymerases are named for the order in which they were discovered, which has no bearing on the relative importance of their function. We will discuss them in an order that reflects their similarities in localization, function, and regulation. We will emphasize RNA polymerase II, since it transcribes all of the protein-coding genes and has therefore been the focus of much research investigating transcriptional mechanisms.

**TABLE 29.2 Eukaryotic RNA polymerases**

Type	Location	Cellular transcripts	Effects of $\alpha$ -amanitin
I	Nucleolus	18S, 5.8S, and 28S rRNA	Insensitive
II	Nucleoplasm	mRNA precursors and snRNA	Strongly inhibited
III	Nucleoplasm	tRNA and 5S rRNA	Inhibited by high concentrations

*RNA polymerase I* is located in specialized structures within the nucleus called nucleoli, where it transcribes the tandem array of genes for 18S, 5.8S, and 28S rRNA. The other rRNA molecule (5S rRNA) and all the tRNA molecules are synthesized by *RNA polymerase III*, which is located in the nucleoplasm rather than in nucleoli. *RNA polymerase II*, which

also is located in the nucleoplasm, synthesizes the precursors of mRNA as well as several small RNA molecules, such as those of the splicing apparatus and many of the precursors to small regulatory RNAs. All three of the polymerases are large proteins, containing from 8 to 14 subunits and having total molecular masses greater than 500 kDa (or 0.5 MDa), and it is likely that they evolved from a single enzyme that was present in a common ancestor of eukaryotes, bacteria, and archaea. In fact, many components of the eukaryotic transcriptional machinery evolved from those in a common ancestor.

Although all eukaryotic RNA polymerases are homologous to one another and to prokaryotic RNA polymerases, RNA polymerase II contains a unique **carboxyl-terminal domain** (CTD) on the 220-kDa subunit; this domain is unusual because it contains multiple repeats of a YSPTSPS consensus sequence. The activity of RNA polymerase II is regulated by phosphorylation, mainly on the serine residues of the CTD.

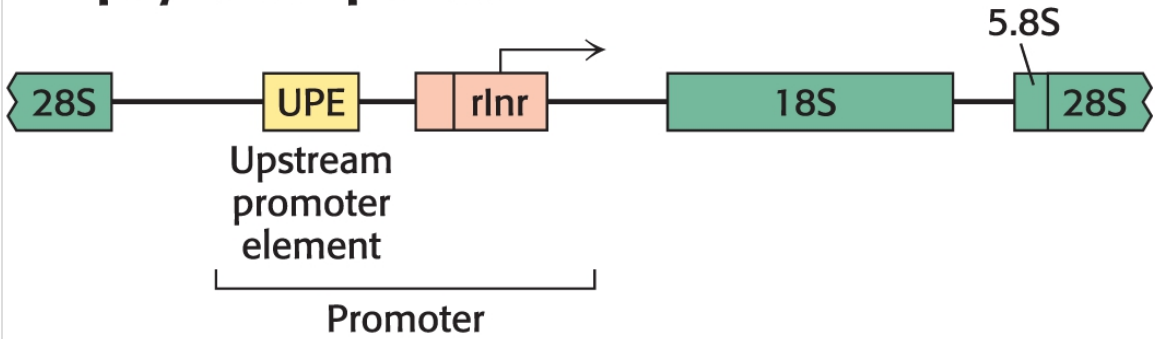
The different polymerases were originally distinguished through their variable responses to the toxin  $\alpha$ -amanitin, a cyclic octapeptide that contains several modified amino acids and is produced by a genus of poisonous mushroom (**Figure 29.20**).  $\alpha$ -Amanitin binds very tightly ( $K_d = 10$  nM) to RNA polymerase II and thereby blocks the elongation phase of RNA synthesis. Higher concentrations of  $\alpha$ -amanitin (1  $\mu$ M) inhibit RNA polymerase III, whereas RNA polymerase I is insensitive to this toxin. This pattern of sensitivity is highly conserved throughout the animal and plant kingdoms.



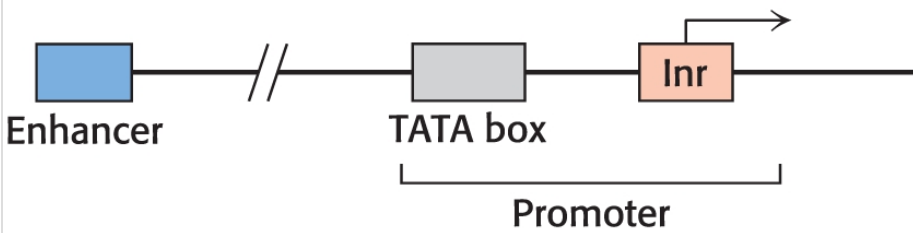


Finally, eukaryotic polymerases differ from each other in the promoters to which they bind. Eukaryotic genes, like prokaryotic genes, require promoters for transcription initiation. Like prokaryotic promoters, eukaryotic promoters consist of conserved sequences that attract the polymerase to the start site. However, eukaryotic promoters differ distinctly in sequence and position, depending on the type of RNA polymerase that binds to them ([Figure 29.21](#)).

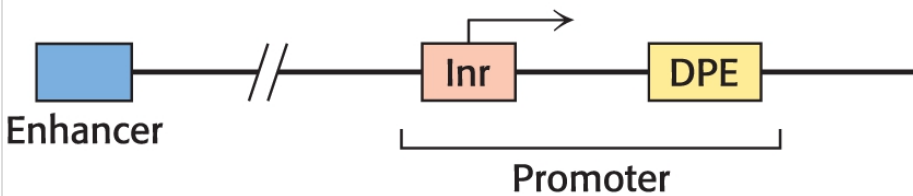
### RNA polymerase I promoter



### RNA polymerase II promoter

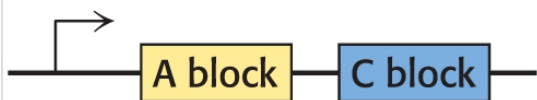


or

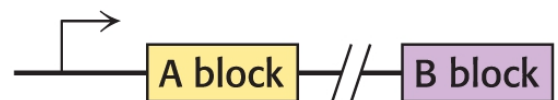


### RNA polymerase III promoter

Type I: 5S rRNA



Type II: tRNA



**FIGURE 29.21 Eukaryotic promoters share common elements.** Each eukaryotic RNA polymerase recognizes a set of promoter elements — sequences in DNA that promote transcription. The RNA polymerase I promoter consists of a ribosomal initiator (rInr) and an upstream promoter element (UPE). The RNA polymerase II promoter likewise includes an initiator element (Inr) and may also include either a TATA box or a downstream promoter element (DPE). Separate from the promoter region, enhancer elements bind specific transcription factors. RNA polymerase III promoters consist of conserved sequences that lie within the transcribed genes.

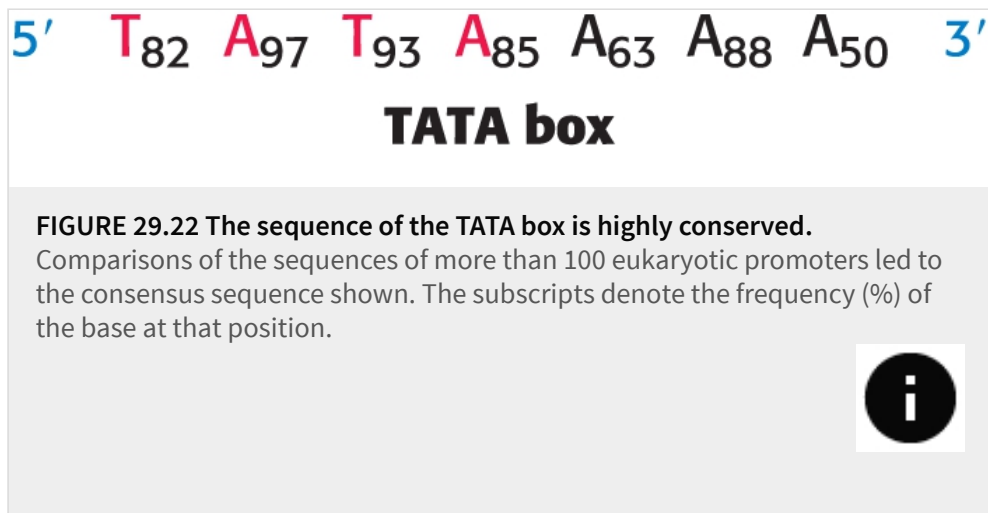


- *The promoter sequences for RNA polymerase I are located in stretches of DNA separating the ribosomal DNA (rDNA) it transcribes. These rRNA genes are arranged in several hundred tandem repeats, each containing a copy of each of three rRNA genes. At the transcriptional start site lies a TATA-like sequence called the ribosomal initiator element (rInr). Farther upstream, 150 to 200 bp from the start site, is the upstream promoter element (UPE). Both elements aid transcription by binding proteins that recruit RNA polymerase I.*
- *Promoters for RNA polymerase II, like prokaryotic promoters, include a set of consensus sequences that define the start site and recruit the polymerase. However, the promoter can contain any combination of a number of possible consensus sequences. Unique to eukaryotes, they also include enhancer elements that can be more than 1 kb from the start site.*
- *Promoters for RNA polymerase III are within the transcribed sequence, downstream of the start site. This is contrast to promoters for RNA polymerase I and II, which are upstream of the transcription start site. There are two types of intergenic promoters for RNA polymerase III. Type I promoters, found in the 5S rRNA gene, contain two short, conserved sequences, the A block and the C block. Type II promoters, found in tRNA genes, consist of two 11-bp sequences, the A block and the B block, situated about 15 bp from either end of the gene.*

## Three common elements can be found in the RNA polymerase II promoter region

RNA polymerase II transcribes all of the protein-coding genes in eukaryotic cells. Promoters for RNA polymerase II, like those for bacterial polymerases, are generally located upstream of the start site for transcription. Because these sequences are on the *same* molecule of DNA as the genes being transcribed, they are called *cis-acting elements*.

1. The most commonly recognized *cis-acting* element for genes transcribed by RNA polymerase II is called the **TATA box** on the basis of its consensus sequence ([Figure 29.22](#)). The TATA box is usually found between positions  $-30$  and  $-100$ . Note that the eukaryotic TATA box closely resembles the prokaryotic  $-10$  sequence (TATAAT) but is farther from the start site. The mutation of a single base in the TATA box markedly impairs promoter activity. Thus, the precise sequence, not just a high content of AT pairs, is essential.



2. The TATA box is often paired with an *initiator element* (Inr), a sequence found at the transcriptional start site, between positions  $-3$  and  $+5$ . This sequence defines the start site because the other

promoter elements are at variable distances from that site. Its presence increases transcriptional activity.

3. A third element, the *downstream core promoter element* (DPE), is commonly found in conjunction with the Inr in transcripts that lack the TATA box. In contrast with the TATA box, the DPE is found downstream of the start site, between positions  $+28$  and  $+32$ .

## Regulatory cis-acting elements are recognized by different mechanisms

Additional regulatory sequences are located between  $-40$  and  $-150$ .

Many promoters contain a *CAAT box*, and some contain a *GC box* ([Figure 29.23](#)). Constitutive genes (genes that are continuously expressed rather than regulated) tend to have GC boxes in their promoters. The positions of these upstream sequences vary from one promoter to another, in contrast with the quite constant location of the  $-35$  region in prokaryotes. Another difference is that the CAAT box and the GC box can be effective when present on the template (antisense) strand, unlike the  $-35$  region, which must be present on the coding (sense) strand.

5' GGNCAATCT 3'

**CAAT box**

5' GGGCGG 3'

**GC box**

**FIGURE 29.23** The sequences of the CAAT box and GC box are also highly conserved. Consensus sequences are shown for the CAAT and GC boxes of eukaryotic promoters for mRNA precursors.



These differences between prokaryotes and eukaryotes correspond to fundamentally different mechanisms for the recognition of cis-acting elements. The  $-10$  and  $-35$  sequences in prokaryotic promoters are binding sites for RNA polymerase and its associated  $\sigma$  factor. In contrast, the TATA, CAAT, and GC boxes and other cis-acting elements in eukaryotic promoters are recognized by proteins other than RNA polymerase itself.

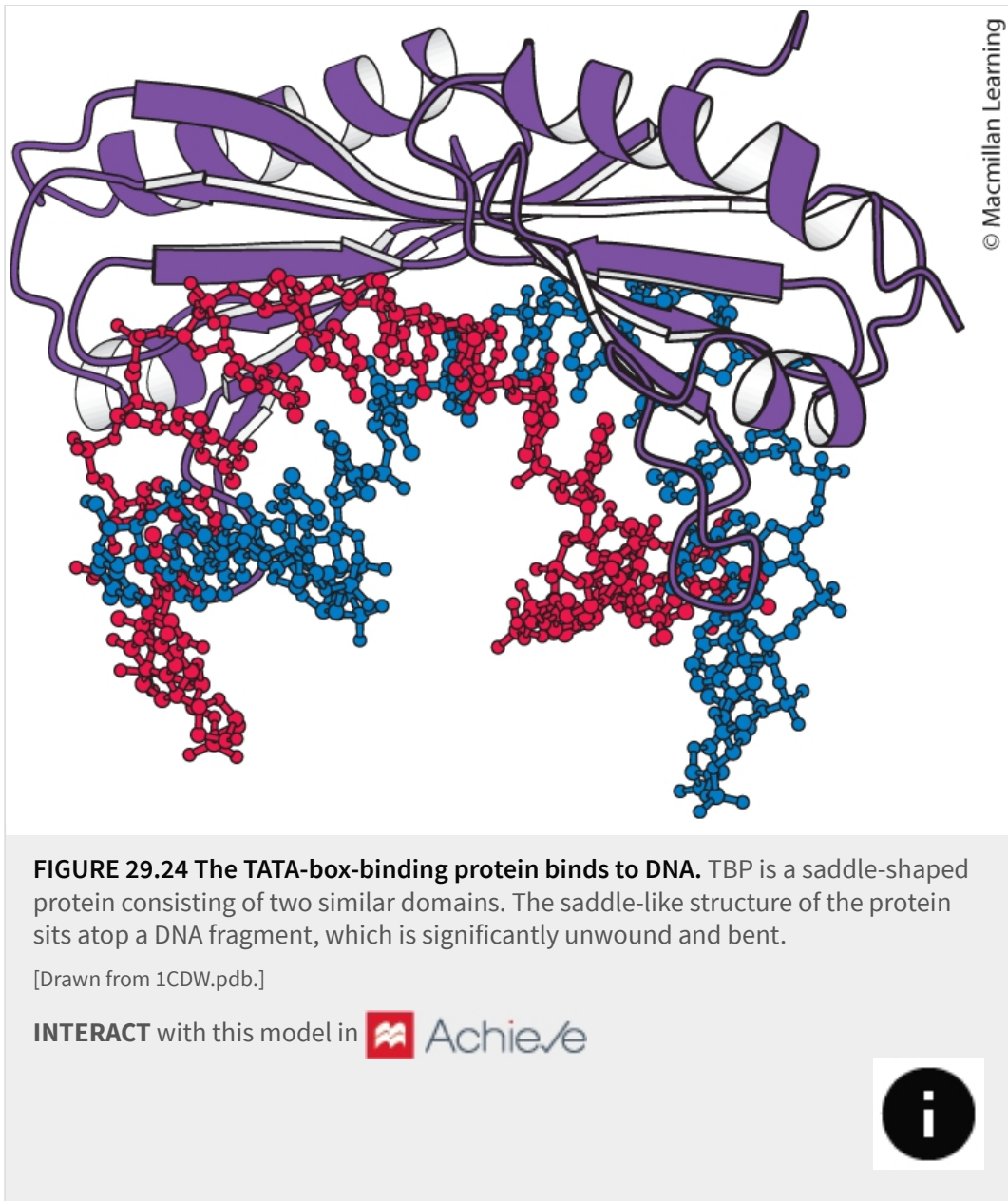
# The TFIID protein complex initiates the assembly of the active transcription complex in eukaryotes

Cis-acting elements constitute only part of the puzzle of eukaryotic gene expression. **Transcription factors** that bind to these elements also are required. For example, RNA polymerase II is guided to the start site by a set of transcription factors known collectively as *TFII* (*TF* stands for transcription factor, and *II* refers to RNA polymerase II). Individual TFII factors are called TFIIA, TFIIB, and so on.

In TATA-box promoters, the key initial event is the recognition of the TATA box by the TATA-box-binding protein (TBP), a 30-kDa component of the 700-kDa TFIID complex. In TATA-less promoters, other proteins in the TFIID complex bind the core promoter elements; however, because less is known about these interactions, we will consider only the TATA-box–TBP binding interaction. TBP binds  $10^5$  times as tightly to the TATA box as to nonconsensus sequences; the dissociation constant of the TBP–TATA-box complex is approximately 1 nM.

The TATA box of DNA binds to the concave surface of TBP, inducing large conformational changes in the bound DNA (**Figure 29.24**). The double helix is substantially unwound to widen its minor groove, enabling it to make extensive contact with the antiparallel  $\beta$  strands on the concave side of TBP. Hydrophobic interactions are prominent at this interface. Four phenylalanine residues, for example, are intercalated between base pairs of the TATA box. The flexibility of AT-rich sequences is generally exploited here in bending the DNA. Immediately outside the TATA box, classical B-DNA resumes. The TBP–TATA-box complex is distinctly asymmetric, a property that is crucial for specifying a unique start site and ensuring that transcription proceeds unidirectionally.





TBP bound to the TATA box is the heart of the initiation complex ([Figure 29.25](#)). The surface of the TBP saddle provides docking sites for the binding of other components, with additional transcription factors assembling on this nucleus in a defined sequence. TFIIA is recruited, followed by TFIIB; then TFIIF, RNA polymerase II, TFIIIE, and TFIIF join the other factors to form a complex called the *pre-initiation complex* (PIC).



**FIGURE 29.25 General transcription factors are essential in initiating transcription by RNA polymerase II.** The step-by-step assembly of these general transcription factors begins with the binding of TFIID (purple) to the TATA box. [The TATA-box-binding protein (TBP), a component of TFIID, recognizes the TATA box.] After assembly, TFIIH opens the DNA double helix and phosphorylates the carboxyl-terminal domain (CTD), allowing the polymerase to leave the promoter and begin transcription. The red arrow marks the transcription start site.



These additional transcription factors play specific roles in this complex. As we saw above, TFIID recognizes core promoter elements and is central to the assembly process. While TFIIA is not essential for the assembly or function of the PIC *in vitro*, it may aid in the binding of TFIID to the DNA. TFIIB is a DNA-binding protein that recognizes specific cis-acting promoter elements called *B recognition elements*, which are often found near the TATA box. TFIIF aids in the recruitment of polymerase II, while TFIIIE brings TFIIH to the complex. TFIIH is a multisubunit complex with helicase and protein kinase activities, both of which are critical in the initiation of transcription. The helicase activity unwinds the DNA template, and the kinase activity phosphorylates specific amino acids in the CTD of polymerase II.

During the formation of the PIC, the carboxyl-terminal domain (CTD) is unphosphorylated and plays a role in transcription regulation through its binding to an enhancer-associated complex called *mediator* ([Section 31.4](#)). Phosphorylation of the CTD by TFIIH marks the transition from initiation to elongation. The phosphorylated CTD stabilizes transcription elongation by RNA polymerase II and recruits RNA-processing enzymes that act during the course of elongation. The importance of the carboxyl-terminal domain is highlighted by the finding that yeast cells containing mutant polymerase II with fewer than 10 repeats in the CTD are not viable.

The PIC described above initiates transcription at a low (basal) frequency, and the transcription factors associated with it are referred

to as *basal* or *general* transcription factors. Additional transcription factors that bind to other sites are required to achieve a high rate of mRNA synthesis. Their role is to selectively stimulate *specific* genes. In summary, transcription factors and other proteins that bind to regulatory sites on DNA can be regarded as passwords that cooperatively open multiple locks, giving RNA polymerase access to specific genes.

---

## SELF-CHECK QUESTION



The function of the  $\sigma$  subunit of *E. coli* RNA polymerase is analogous to the function of general transcription factors in eukaryotes. Briefly describe their common function.

---

## Enhancer sequences can stimulate transcription at start sites thousands of bases away

The activities of many promoters in higher eukaryotes are greatly increased by another type of cis-acting element called an **enhancer**.

- Enhancer sequences have no promoter activity of their own yet can exert their stimulatory actions over distances of several thousand base pairs.
- Enhancers can be upstream, downstream, or even in the middle of a transcribed gene.
- Enhancers are effective when present on either the coding or noncoding DNA strand.

- A particular enhancer is effective only in certain cells; for example, the immunoglobulin enhancer functions in B lymphocytes but not elsewhere.



Cancer can result if the relation between genes and enhancers is disrupted. In Burkitt lymphoma and B-cell leukemia, a chromosomal translocation brings the proto-oncogene *myc* (a transcription factor itself) under the control of a powerful immunoglobulin enhancer. The consequent dysregulation of the *myc* gene is hypothesized to play a role in the progression of the cancer.

The discovery of promoters and enhancers has allowed us to gain a better understanding of how genes are selectively expressed in eukaryotic cells. The regulation of eukaryotic gene transcription, discussed in [Chapter 31](#), is the fundamental means of controlling gene expression.

## 29.4 Some RNA Transcription Products Are Processed

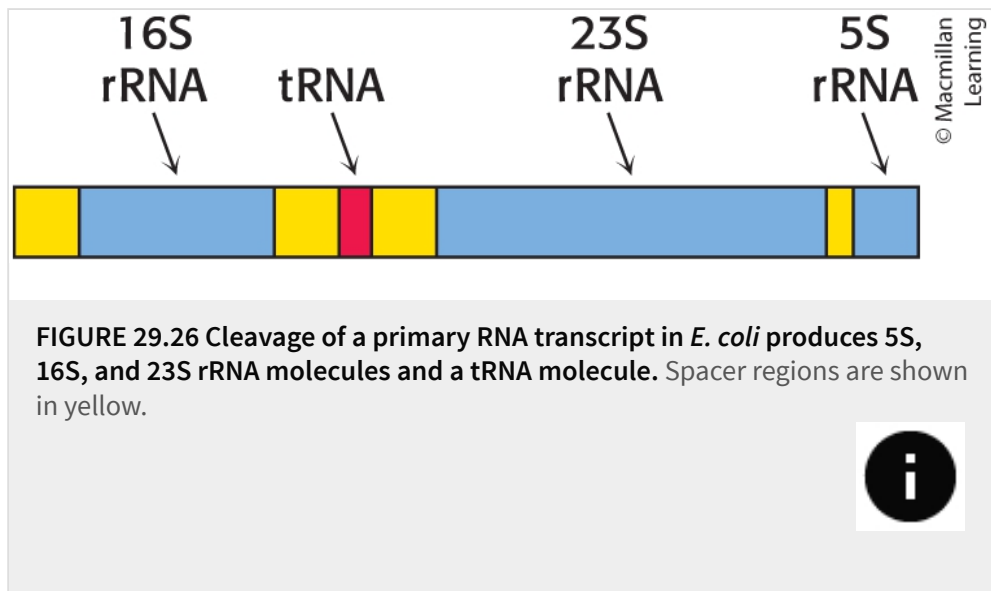
Virtually all the initial products of eukaryotic transcription are further processed, and even some prokaryotic transcripts are modified. As we will see next, the particular processing steps and the factors taking part vary according to the type of RNA precursor and the type of RNA polymerase that produced it.

### Precursors of transfer and ribosomal RNA are cleaved and chemically modified after transcription

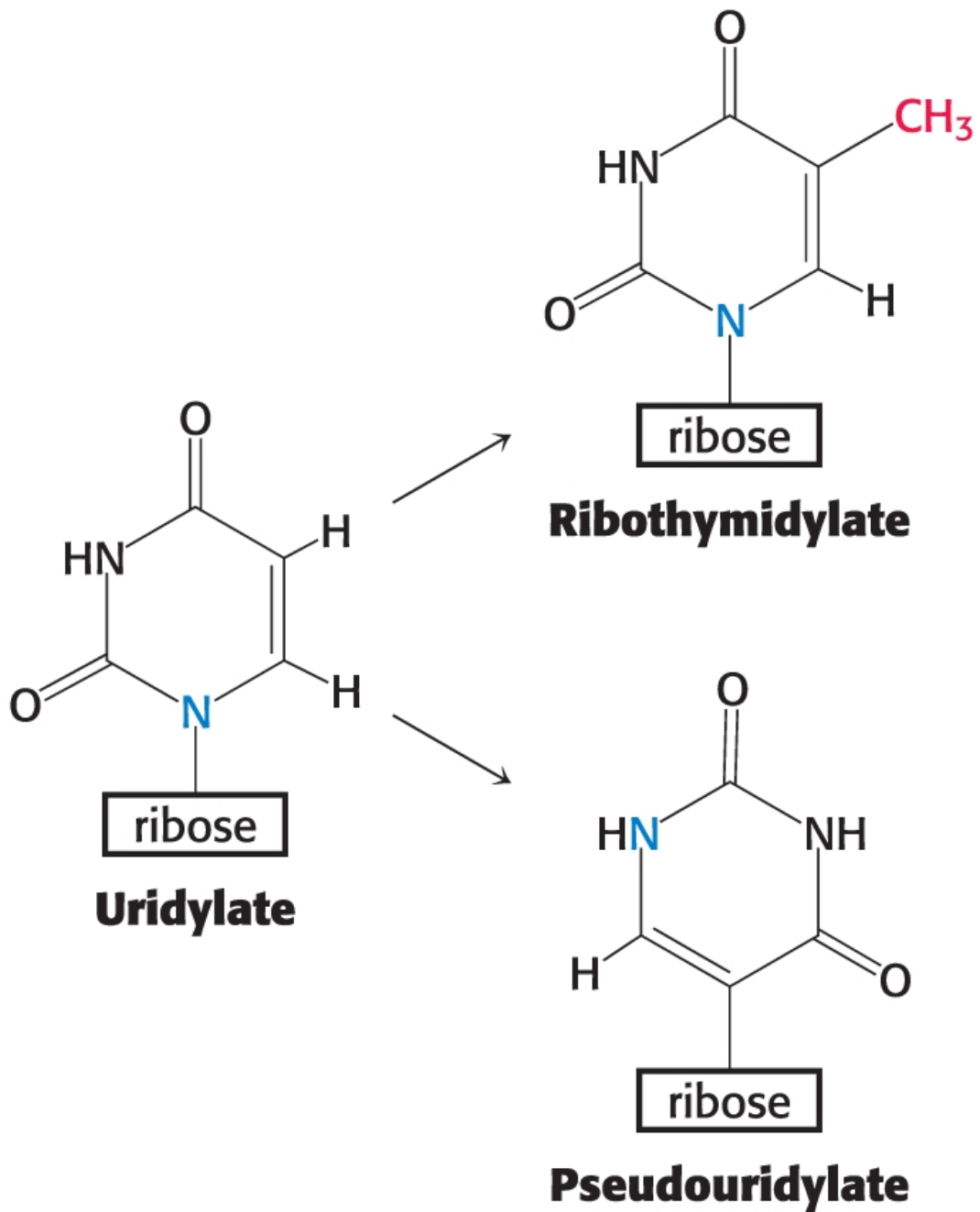
In bacteria, messenger RNA molecules undergo little or no modification after synthesis by RNA polymerase. Indeed, many mRNA molecules are translated while they are being transcribed. In contrast, transfer RNA (tRNA) and ribosomal RNA (rRNA) molecules are generated by modifications of nascent RNA chains.

- *The transcript can be cleaved at specific sites along its sequence.* For example, in *E. coli*, the three rRNAs and a tRNA are excised from a single primary RNA transcript that also contains spacer regions ([Figure 29.26](#)). Other transcripts contain arrays of several kinds of tRNA or several copies of the same tRNA. The nucleases that cleave and trim these precursors of rRNA and tRNA are highly precise. *Ribonuclease P* (RNase P), for example, generates the correct 5' terminus of all tRNA molecules in *E. coli*. Sidney Altman and his coworkers showed that this interesting enzyme contains a catalytically active RNA molecule. *Ribonuclease III* (Rnase III) excises 5S, 16S, and 23S rRNA precursors from the primary

transcript by cleaving double-helical hairpin regions at specific sites.



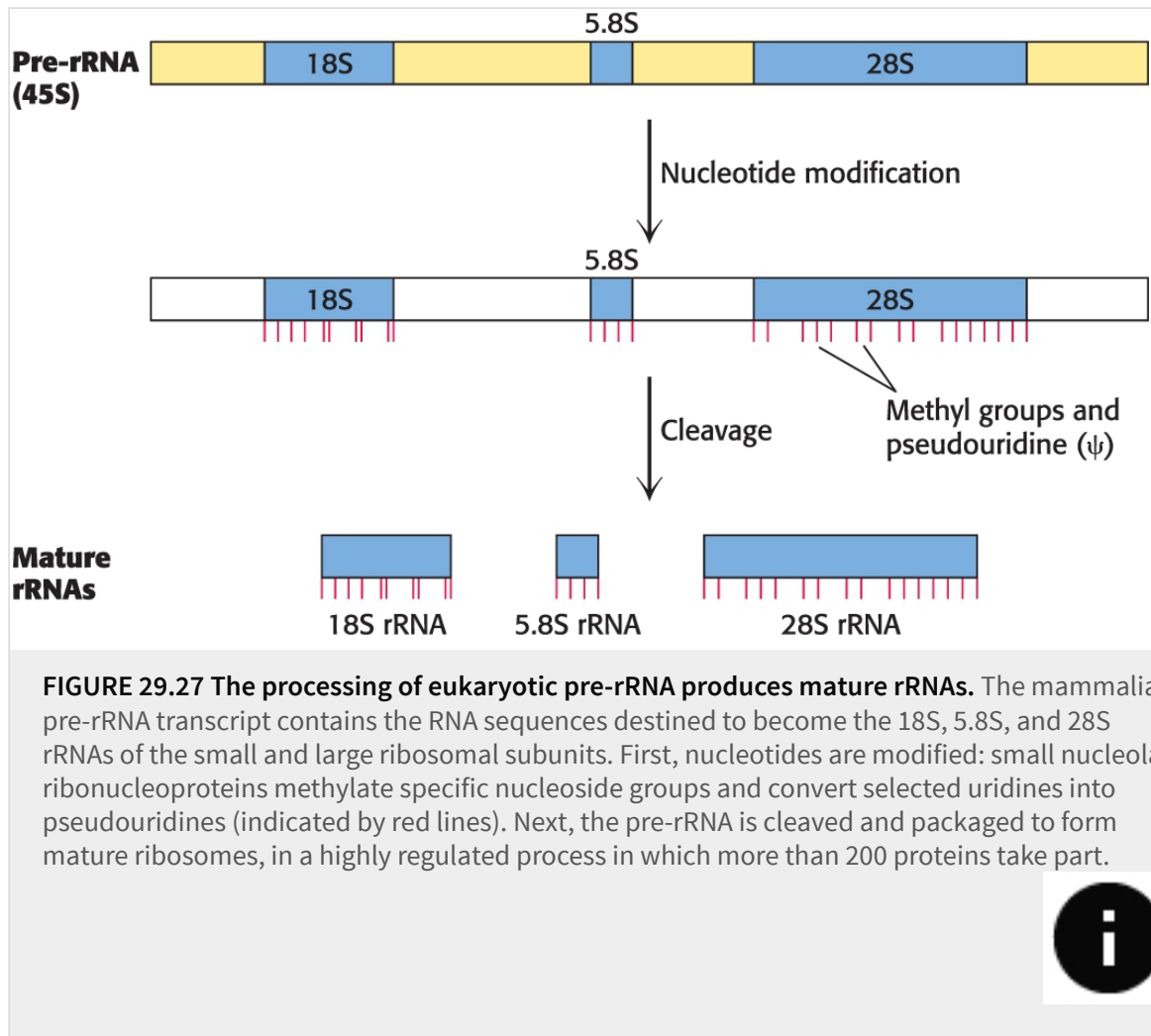
- *Nucleotides can be added to the termini of some RNA chains.* For example, CCA, a terminal sequence required for the function of all tRNAs, is added to the 3' ends of tRNA molecules for which this terminal sequence is not encoded in the DNA. The enzyme that catalyzes the addition of CCA is atypical for an RNA polymerase in that it does not use a DNA template.
- *Bases and ribose units of RNAs can be modified.* For example, some bases of rRNA are methylated. Furthermore, all tRNA molecules contain unusual bases formed by the enzymatic modification of a standard ribonucleotide in a tRNA precursor. For example, uridylylate residues are modified after transcription to form ribothymidylate and pseudouridylylate. These modifications generate diversity, allowing greater structural and functional versatility.





# RNA polymerase I produces three ribosomal RNAs

Several RNA molecules are key components of ribosomes. In eukaryotes, RNA polymerase I transcription produces a single precursor (45S in mammals) that encodes three RNA components of the ribosome: the 18S rRNA, the 28S rRNA, and the 5.8S rRNA ([Figure 29.27](#)).

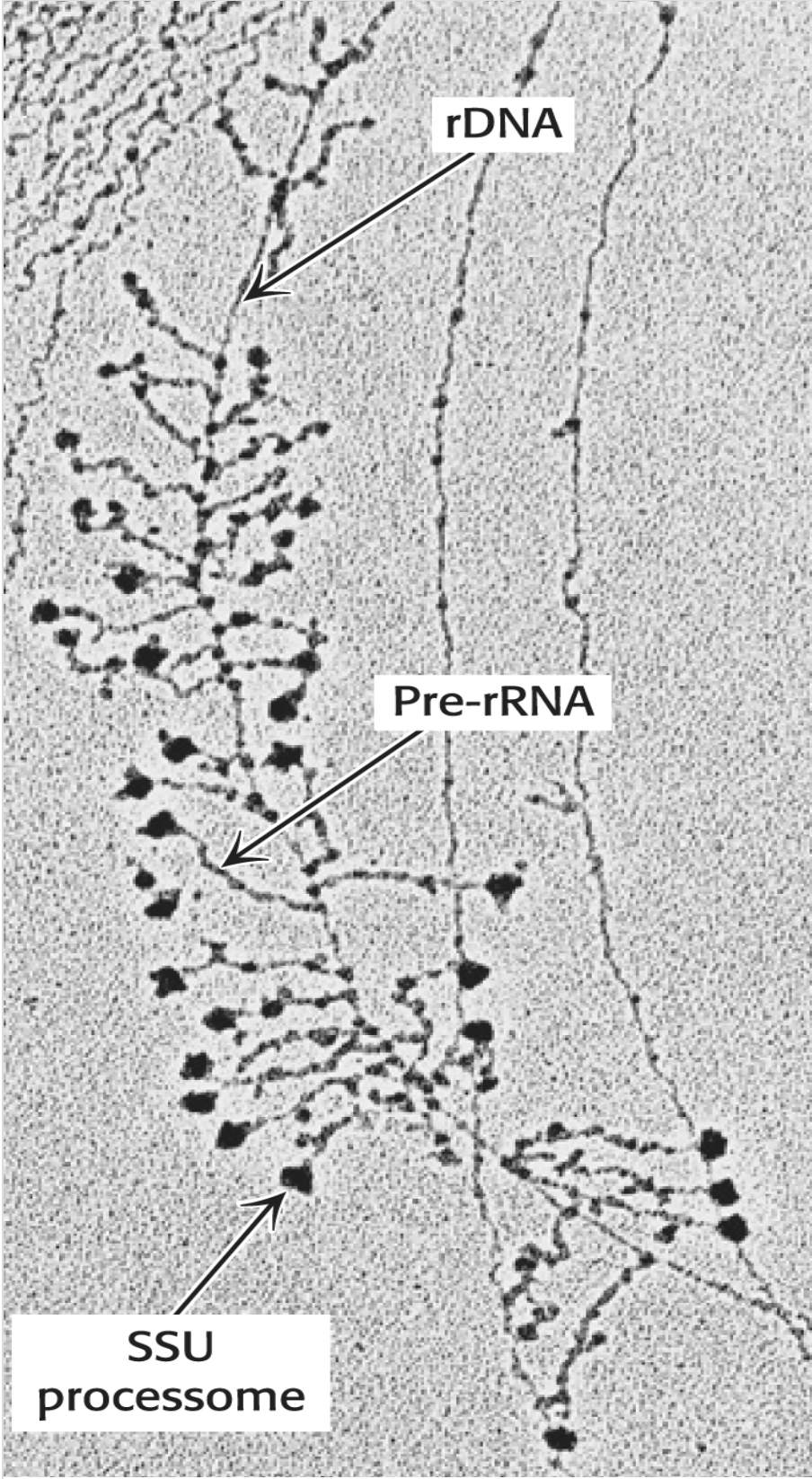


**FIGURE 29.27 The processing of eukaryotic pre-rRNA produces mature rRNAs.** The mammalian pre-rRNA transcript contains the RNA sequences destined to become the 18S, 5.8S, and 28S rRNAs of the small and large ribosomal subunits. First, nucleotides are modified: small nucleolar ribonucleoproteins methylate specific nucleoside groups and convert selected uridines into pseudouridines (indicated by red lines). Next, the pre-rRNA is cleaved and packaged to form mature ribosomes, in a highly regulated process in which more than 200 proteins take part.

The 18S rRNA is the RNA component of the small ribosomal subunit (40S), and the 28S and 5.8S rRNAs are two RNA components of the large ribosomal subunit (60S). The other RNA component of the large

ribosomal subunit, the 5S rRNA, is transcribed by RNA polymerase III as a separate transcript. Processing of the precursor proceeds as follows:

- First, the nucleotides of the pre-rRNA sequences destined for the ribosome undergo extensive modification, on both ribose and base components, directed by many **small nucleolar ribonucleoproteins (snoRNPs)**, each of which consists of one snoRNA and several proteins.
- The pre-rRNA is then assembled with ribosomal proteins, as guided by processing factors, to form a large ribonucleoprotein. For instance, the small-subunit (SSU) processome is required for 18S rRNA synthesis and can be visualized in electron micrographs as a terminal knob at the 5' ends of the nascent rRNAs (**Figure 29.28**).



Reprinted by permission from Macmillan Publishers Ltd, from Dragon, F., Gallagher, J., Compagnone-Post, P. et al. A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* 417, 967-970 (2002); permission conveyed through Copyright Clearance Center, Ltd.

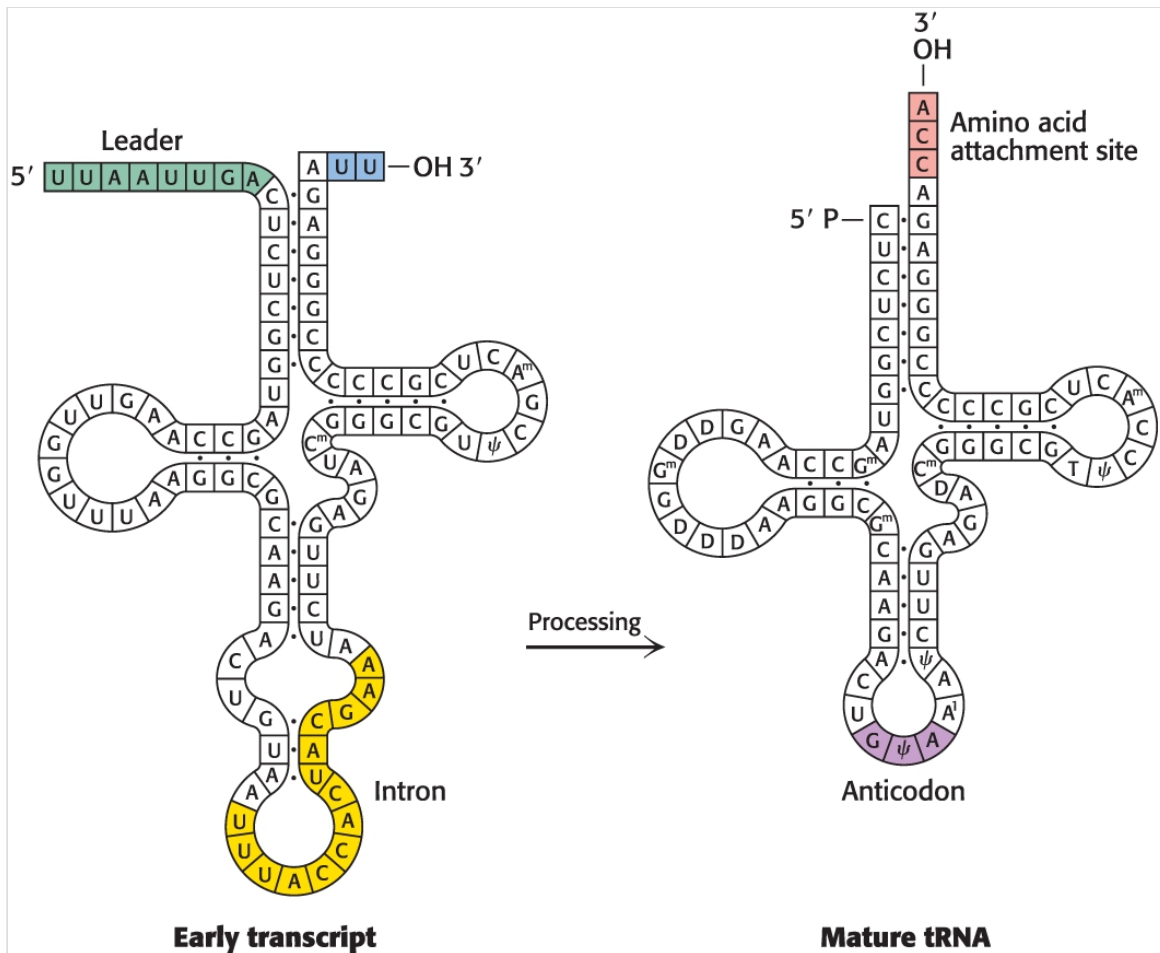
**FIGURE 29.28 Electron microscopy gives a window into rRNA transcription and processing in eukaryotes.** The assemblies of transcribed rRNA into precursor ribosomes resemble fir trees: the trunk is the rDNA, and each branch is a pre-rRNA transcript. Transcription starts at the top of the tree, where the shortest transcripts can be seen, and progresses down the rDNA to the end of the gene. The terminal knobs visible at the end of some pre-rRNA transcripts likely correspond to the SSU processome, a large ribonucleoprotein required for processing the pre-rRNA.



- Finally, rRNA cleavage (sometimes coupled with additional processing steps) releases the mature rRNAs assembled with ribosomal proteins as ribosomes. Like those of RNA polymerase I transcription itself, most of these processing steps take place in the cell's nucleolus.

## RNA polymerase III produces transfer RNAs

Eukaryotic tRNA transcripts are among the most processed of all RNA polymerase III transcripts. Like those of prokaryotic tRNAs, the 5' leader is cleaved by RNase P, the 3' trailer is removed, and CCA is added by the CCA-adding enzyme ([Figure 29.29](#)). Eukaryotic tRNAs are also heavily modified on base and ribose moieties; these modifications are important for function. In contrast with prokaryotic tRNAs, many eukaryotic pre-tRNAs are also spliced by an endonuclease and a ligase to remove an intron.



**FIGURE 29.29 Transfer RNA precursors are processed.** The conversion of a yeast tRNA precursor into a mature tRNA requires the removal of a 14-nucleotide intron (yellow), the cleavage of a 5' leader (green), and the removal of UU and the attachment of CCA at the 3' end (red). In addition, several bases are modified.



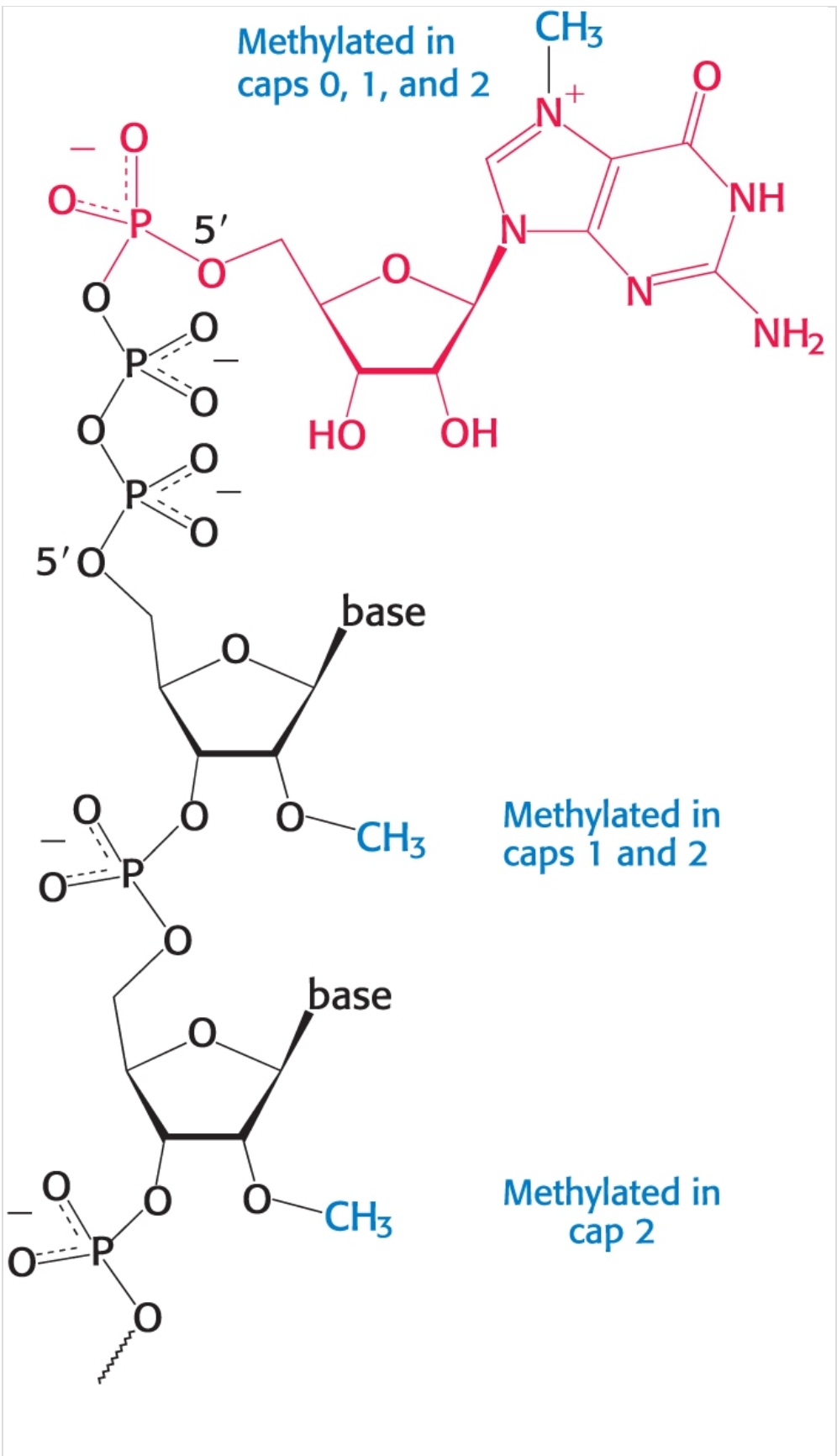
The product of RNA polymerase II, the pre-mRNA transcript, acquires a 5' cap and a 3'

## poly(A) tail

Perhaps the most extensively studied transcription product is the product of RNA polymerase II: most of this RNA will be processed to mRNA. The immediate product of RNA polymerase II is sometimes referred to as precursor-to-messenger RNA, or **pre-mRNA**. Most pre-mRNA molecules are spliced to remove the introns, which we will discuss in greater detail below. In addition, both the 5' and the 3' ends are modified, and both modifications are retained as the pre-mRNA is converted into mRNA.

As in prokaryotes, eukaryotic transcription usually begins with A or G. However, the 5' triphosphate end of the nascent RNA chain is immediately modified:

- First, a phosphoryl group is released by hydrolysis.
- The diphosphate 5' end then attacks the  $\alpha$ -phosphorus atom of GTP to form a very unusual 5' - 5' triphosphate linkage. This distinctive terminus is called a **5' cap** ([Figure 29.30](#)).



**FIGURE 29.30** The 5' ends of eukaryotic mRNAs are capped. Caps at the 5' end of eukaryotic mRNA include 7-methylguanylate (red) attached by a triphosphate linkage to the ribose at the 5' end. None of the riboses are methylated in cap 0, one is methylated in cap 1, and both are methylated in cap 2.



- The N-7 nitrogen of the terminal guanine is then methylated by S-adenosylmethionine to form cap 0. The adjacent riboses may be methylated to form cap 1 or cap 2.

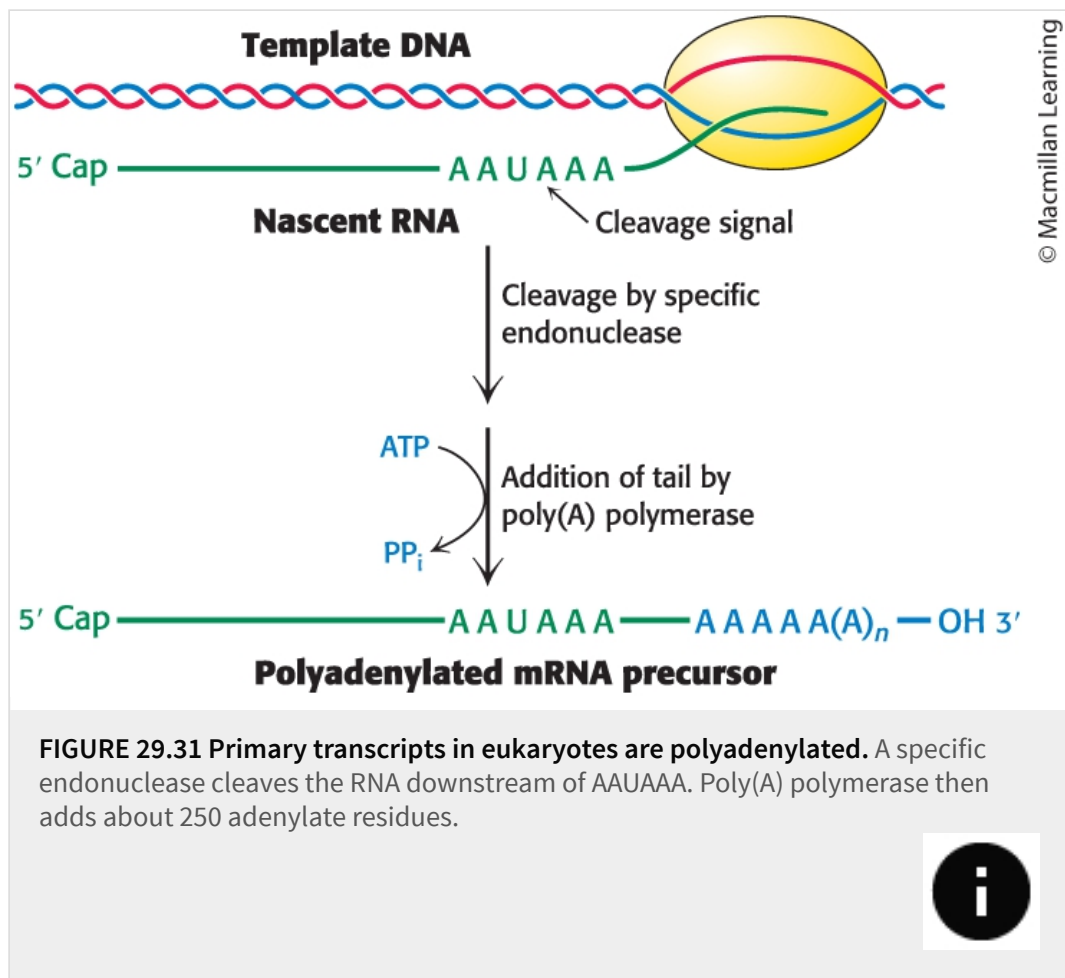
Caps contribute to the stability of mRNAs by protecting their 5' ends from phosphatases and nucleases. In addition, caps enhance the translation of mRNA by eukaryotic protein-synthesizing systems. Transfer RNA and ribosomal RNA molecules, in contrast with messenger RNAs and with small RNAs that participate in splicing, do not have caps.

As mentioned earlier, pre-mRNA is also modified at the 3' end. Most eukaryotic mRNAs contain a string of adenine nucleotides — a **poly(A) tail** — at that end. This poly(A) tail is added *after* transcription has ended, since the DNA template does not encode this sequence. Indeed, the nucleotide preceding poly(A) is not the last nucleotide to be transcribed. Some primary transcripts contain hundreds of nucleotides beyond the 3' end of the mature mRNA.

How is the 3' end of the pre-mRNA given its final form? Eukaryotic primary transcripts are cleaved by a specific endonuclease that recognizes the sequence AAUAAA (**Figure 29.31**). Cleavage does not take



place if this sequence or a segment of some 20 nucleotides on its 3' side is deleted. The presence of internal AAUAAA sequences in some mature mRNAs indicates that AAUAAA is only part of the cleavage signal; its context also is important. After cleavage of the pre-RNA by the endonuclease, a *poly(A) polymerase* adds about 250 adenylate residues to the 3' end of the transcript; ATP is the donor in this reaction.



The role of the poly(A) tail is still not firmly established despite much effort. However, evidence is accumulating that it enhances translation efficiency and the stability of mRNA. Blocking the synthesis of the poly(A) tail by exposure to 3'-deoxyadenosine (cordycepin) does not interfere with the synthesis of the primary transcript. Messenger RNA

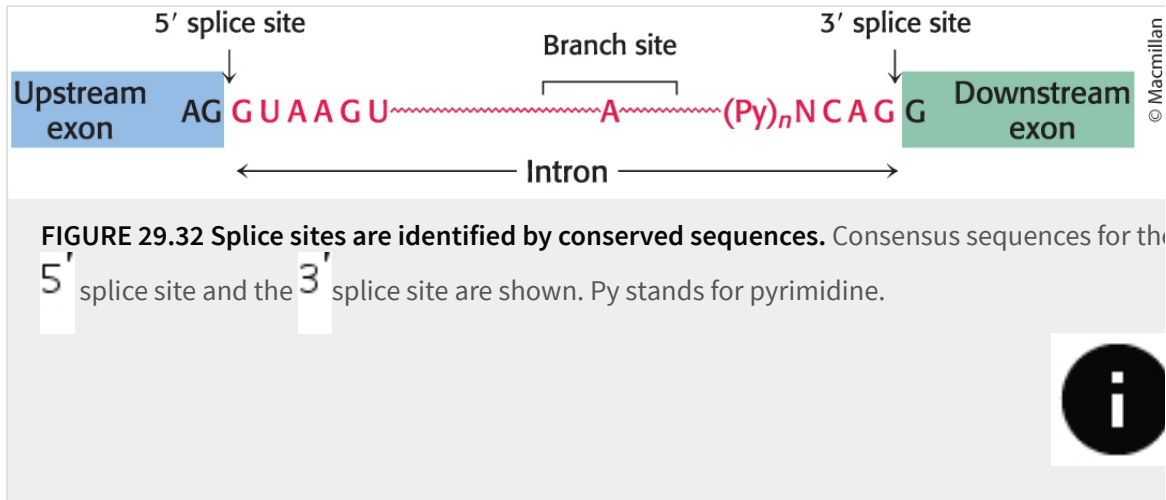
without a poly(A) tail can be transported out of the nucleus. However, an mRNA molecule without a poly(A) tail is usually much less effective as a template for protein synthesis than one with a poly(A) tail. Indeed, some mRNAs are stored in an unadenylated form and receive the poly(A) tail only when translation is imminent. The half-life of an mRNA molecule may be determined in part by the rate of degradation of its poly(A) tail.

## Sequences at the ends of introns specify splice sites in mRNA precursors

Most genes in higher eukaryotes are composed of exons and introns ([Section 8.7](#)). The introns must be excised and the exons linked to form the final mRNA in a process called [RNA splicing](#). This splicing must be exquisitely sensitive; splicing just one nucleotide upstream or downstream of the intended site would create a one-nucleotide shift, which would alter the reading frame on the 3' side of the splice to give an entirely different amino acid sequence, likely including a premature stop codon. Thus, the correct splice site must be clearly marked.

Does a particular sequence denote the splice site? The sequences of thousands of intron–exon junctions within RNA transcripts are known. In eukaryotes from yeast to mammals, these sequences have a common structural motif: the intron begins with GU and ends with AG. The consensus sequence at the 5' splice in vertebrates is AGGUAAGU, where the GU is invariant ([Figure 29.32](#)). At the 3' end of an intron, the consensus sequence is a stretch of 10 pyrimidines (U or C; termed the *polypyrimidine tract*), followed by any base, then by C, and ending with the invariant AG. Introns also have an important internal site located between 20 and 50 nucleotides upstream of the 3' splice site; it is called

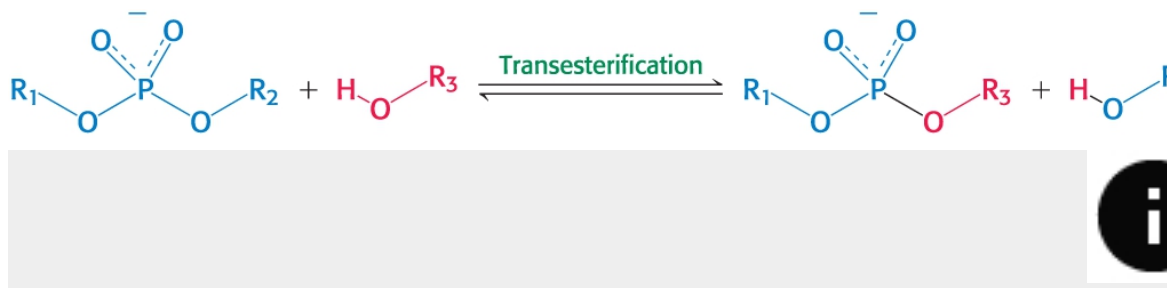
the *branch site* for reasons that will be evident shortly. In yeast, the branch-site sequence is nearly always UACUAAC, whereas in mammals a variety of sequences are found.

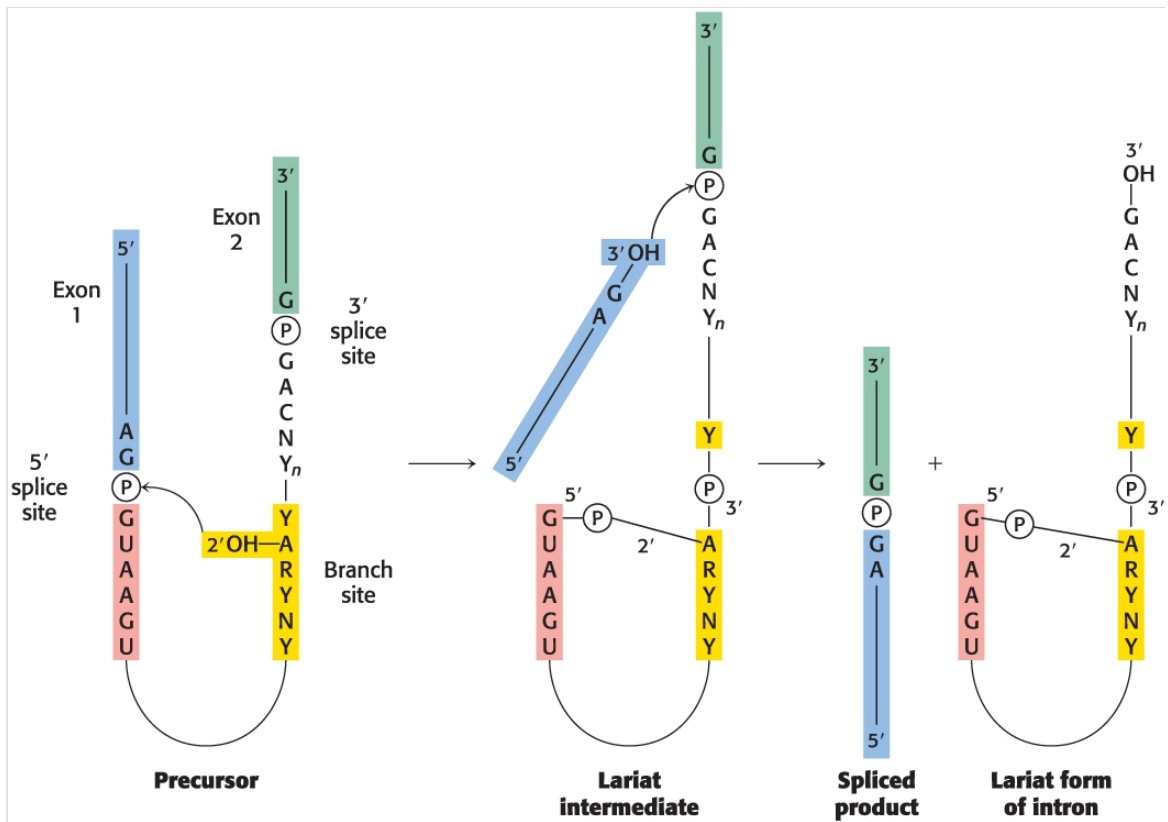


The 5' and 3' splice sites and the branch site are essential for determining where splicing takes place. Mutations in each of these three critical regions lead to aberrant splicing. Introns vary in length from 50 to 10,000 nucleotides, and so the splicing machinery may have to find the 3' site several thousand nucleotides away. Specific sequences near the splice sites (in both the introns and the exons) play an important role in splicing regulation, particularly in designating splice sites when there are many alternatives. Researchers are currently attempting to determine the factors that contribute to splice-site selection for individual mRNAs. Despite our knowledge of splice-site sequences, predicting pre-mRNAs and their protein products from genomic DNA sequence information remains a challenge.

## Splicing consists of two sequential transesterification reactions

The splicing of nascent mRNA molecules is a complicated process. It requires the cooperation of several small RNAs and proteins that form a large complex called a **spliceosome**. However, the chemistry of the splicing process is simple. Splicing begins with the cleavage of the phosphodiester bond between the upstream exon (exon 1) and the 5' end of the intron (**Figure 29.33**). The attacking group in this reaction is the 2'-OH group of an adenylate residue in the branch site. A 2'-5' phosphodiester bond is formed between this A residue and the 5' terminal phosphate of the intron in a transesterification reaction.



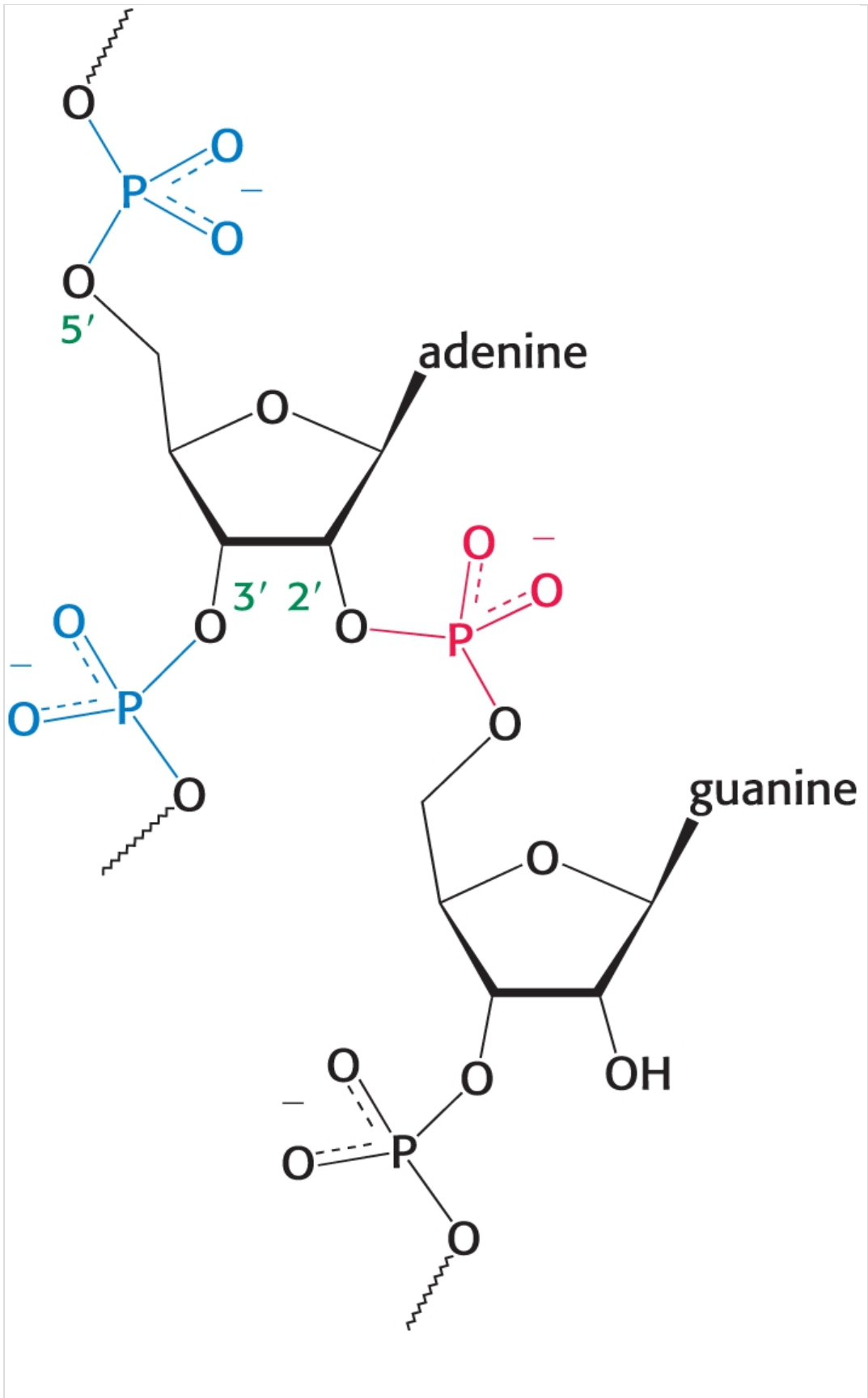


**FIGURE 29.33** The splicing mechanism for mRNA precursors involves a lariat intermediate. The upstream (5') exon is shown in blue, the downstream (3') exon in green, and the branch site in yellow. Y stands for a pyrimidine nucleotide, R for a purine nucleotide, and N for any nucleotide. The 5' splice site is attacked by the 2'-OH group of the branch-site adenosine residue. The 3' splice site is attacked by the newly formed 3'-OH group of the upstream exon. The exons are joined, and the intron is released in the form of a lariat.

[Information from P. A. Sharp, *Cell* 42:397-408, 1985.]



Note that this adenylate residue is also joined to two other nucleotides by normal 3'-5' phosphodiester bonds ([Figure 29.34](#)). Hence, a branch is generated at this site, and a lariat (loop) intermediate is formed.



**FIGURE 29.34 The splicing branch point has a unique structure.** The structure of the branch point in the lariat intermediate in which the adenylate residue is joined to three nucleotides by phosphodiester bonds. The new 2'-to-5' linkage is shown in red, and the usual 3'-to-5' linkages are shown in blue.



The 3'-OH terminus of exon 1 then attacks the phosphodiester bond between the intron and exon 2. In another transesterification reaction, exons 1 and 2 become joined, and the intron is released in lariat form. Splicing is thus accomplished by two transesterification reactions rather than by hydrolysis followed by ligation.

Both transesterification reactions are promoted by the pair of bound magnesium ions, in reactions reminiscent of those for DNA and RNA polymerases. The first reaction generates a free 3'-OH group at the 3' end of exon 1, and the second reaction links this group to the 5'-phosphate of exon 2. The number of phosphodiester bonds stays the same during these steps, which is crucial because it allows the splicing reaction itself to proceed without an energy source such as ATP or GTP.

## Small nuclear RNAs in spliceosomes catalyze the splicing of mRNA precursors

The nucleus contains many types of small RNA molecules with fewer than 300 nucleotides, referred to as **small nuclear RNAs (snRNAs)**. A few of them — designated U1, U2, U4, U5, and U6 — are essential for splicing mRNA precursors. The secondary structures of these RNAs are highly conserved in organisms ranging from yeast to human beings.

snRNA molecules are associated with specific proteins to form complexes termed **small nuclear ribonucleoproteins (snRNPs)**; investigators often speak of them as “snurps” (**Table 29.3**). SnRNPs and their role in RNA splicing were discovered by Joan Steitz and Michael Lerner in 1980. One major piece of evidence suggesting a role for snRNPs in splicing was the base complementarity between portions of the U1 snRNA and the splice sites found in the unprocessed mRNAs.

**TABLE 29.3 Roles of small nuclear ribonucleoproteins (snRNPs) in the splicing of mRNA precursors**

snR NP	Size of snRNA (nucleotides)	Role
U1	165	Binds the 5' splice site
U2	185	Binds the branch site
U5	116	Binds the 5' splice site and then the 3' splice site
U4	145	Masks the catalytic activity of U6
U6	106	Catalyzes splicing

SnRNPs associate with hundreds of other proteins (called *splicing factors*) and the mRNA precursors to form the large (60S) spliceosomes. The large and dynamic nature of the spliceosome made the determination of the detailed three-dimensional structure a great challenge. However, with the maturation of cryo-electron microscopy (**Section 4.5**), the structures of spliceosomes from several species in a number of different stages of their function have been determined (**Figure 29.35**). These structures have added to our understanding of the splicing process (**Figure 29.36**).

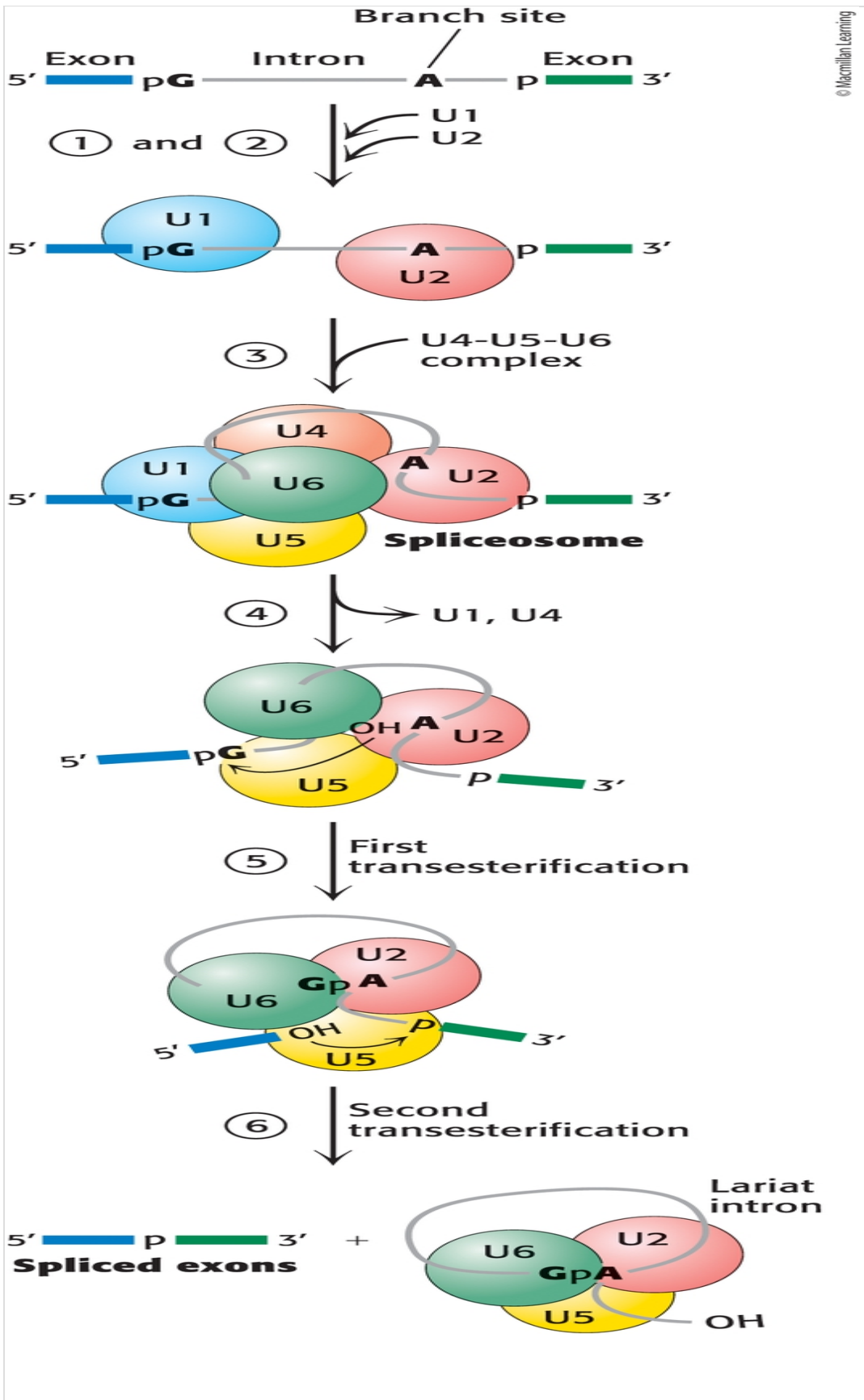




**FIGURE 29.35 The spliceosome consists of proteins and RNAs.** This three-dimensional structure displays one form of the spliceosome from a yeast determined by high-resolution cryo-electron microscopy. The key RNA components U2, U5, and U6 are shown in red, yellow, and blue, respectively.

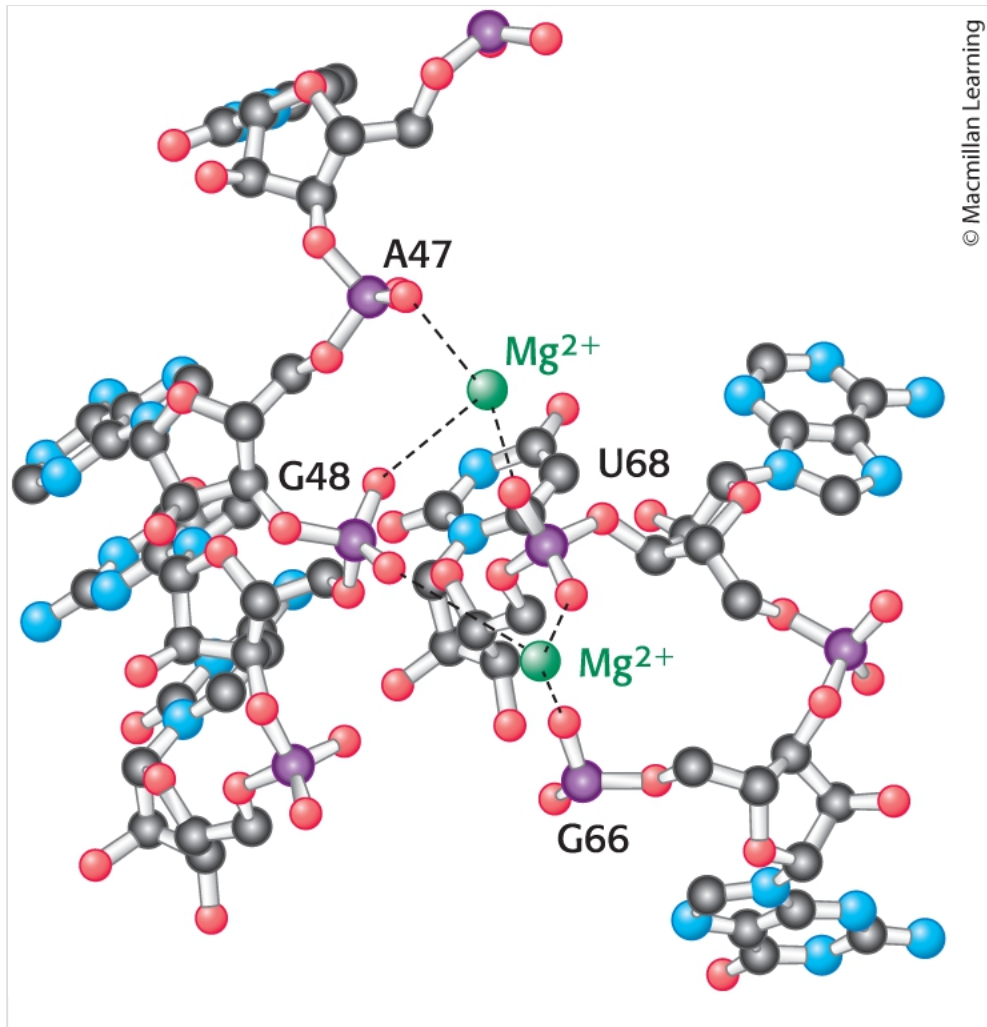
[Drawn from 3JB9.pdb.]







2. U2 snRNP then binds the branch site in the intron by base-pairing between a highly conserved sequence in U2 snRNA and the pre-mRNA. U2 snRNP binding requires ATP hydrolysis.
3. A preassembled U4-U5-U6 tri-snRNP joins this complex of U1, U2, and the mRNA precursor to form the spliceosome. This association also requires ATP hydrolysis. Experiments with a reagent that cross-links neighboring pyrimidines in base-paired regions revealed that in this assembly U5 interacts with exon sequences in the 5' splice site and subsequently with the 3' exon.
4. Next, U6 disengages from U4 and undergoes an intramolecular rearrangement that permits base-pairing with U2 as well as interaction with the 5' end of the intron, displacing U1 and U4 from the spliceosome. U4 serves as an inhibitor that masks U6 until the specific splice sites are aligned. The catalytic center includes two bound magnesium ions bound primarily by phosphate groups from the U6 RNA ([Figure 29.37](#)).



© Macmillan Learning

**FIGURE 29.37** The spliceosome's catalytic center includes two key magnesium ions. These ions bind to the U6 RNA and promote the two transesterification reactions that are central to RNA splicing.



5. These rearrangements result in the first transesterification reaction, cleaving the 5' exon and generating the lariat intermediate.
6. Further rearrangements of RNA in the spliceosome facilitate the second transesterification. In these rearrangements, U5 aligns the

free 5' exon with the 3' exon such that the 3'-hydroxyl group of the 5' exon is positioned to make a nucleophilic attack on the 3' splice site to generate the spliced product. U2, U5, and U6 bound to the excised lariat intron are released, completing the splicing reaction.

Many of the steps in the splicing process require ATP hydrolysis. How is the free energy associated with ATP hydrolysis used to power splicing? To achieve the well-ordered rearrangements necessary for splicing, ATP-powered RNA helicases must unwind RNA helices and allow alternative base-pairing arrangements to form. Thus, two features of the splicing process are noteworthy. First, RNA molecules play key roles in directing the alignment of splice sites and in carrying out catalysis. Second, ATP-powered helicases unwind RNA duplex intermediates that facilitate catalysis and induce the release of snRNPs from the mRNA.

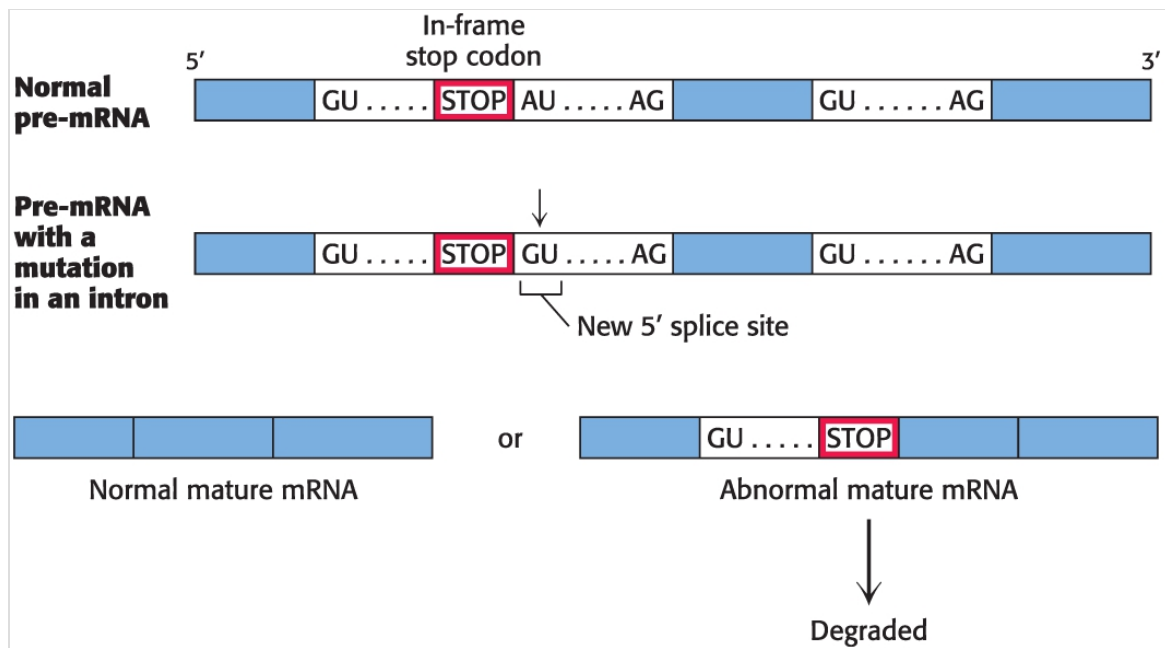
## Mutations that affect pre-mRNA splicing cause disease



Mutations in either the pre-mRNA (cis-acting) or the splicing factors (trans-acting) can cause defective pre-mRNA splicing that manifests in disease. In fact, mutations affecting splicing have been estimated to cause at least 15% of all genetic diseases. We will look at two examples here.

First, we will consider the possible effects of cis-acting mutations on hemoglobin function. Mutations in the pre-mRNA cause some forms of thalassemia, a group of hereditary anemias characterized by the defective synthesis of hemoglobin ([Section 3.3](#)). Cis-acting mutations that cause aberrant splicing can occur at the 5' or 3' splice sites in

either of the two introns of the hemoglobin  $\beta$  chain or in its exons. Typically, mutations in the 5' splice site alter that site such that the splicing machinery cannot recognize it, forcing the machinery to find another 5' splice site in the intron and introducing the potential for a premature stop codon. The defective mRNA is normally degraded rather than translated. Alternatively, mutations in the intron itself may create a new 5' splice site; in this case, either one of the two splice sites may be recognized ([Figure 29.38](#)). Consequently, some normal protein can be made, and so the disease is less severe.



**FIGURE 29.38** Mutations in splice sites can cause diseases such as thalassemia. An A-to-G mutation within the first intron of the gene for the human hemoglobin  $\beta$  chain creates a new 5' splice site (GU). Both 5' splice sites are recognized by the U1 snRNP; thus, splicing may sometimes create a normal mature mRNA and an abnormal mature mRNA that contains intron sequences. The normal mature mRNA is translated into a hemoglobin  $\beta$  chain. Because it includes intron sequences, the abnormal mature mRNA now has a premature stop codon and is degraded.



Second, we will consider the possible effects of trans-acting mutations on eyesight. Disease-causing mutations may also appear in splicing factors. Retinitis pigmentosa is a disease of acquired blindness, first described in 1857, with an incidence of 1/3500. About 5% of the autosomal dominant form of retinitis pigmentosa is likely due to mutations in the hPrp8 protein, a pre-mRNA splicing factor that is a component of the U4-U5-U6 tri-snRNP. How a mutation in a splicing factor that is present in all cells causes disease only in the retina is not clear; nevertheless, retinitis pigmentosa is a good example of how mutations that disrupt spliceosome function can cause disease.

## Most human pre-mRNAs can be spliced in alternative ways to yield different proteins

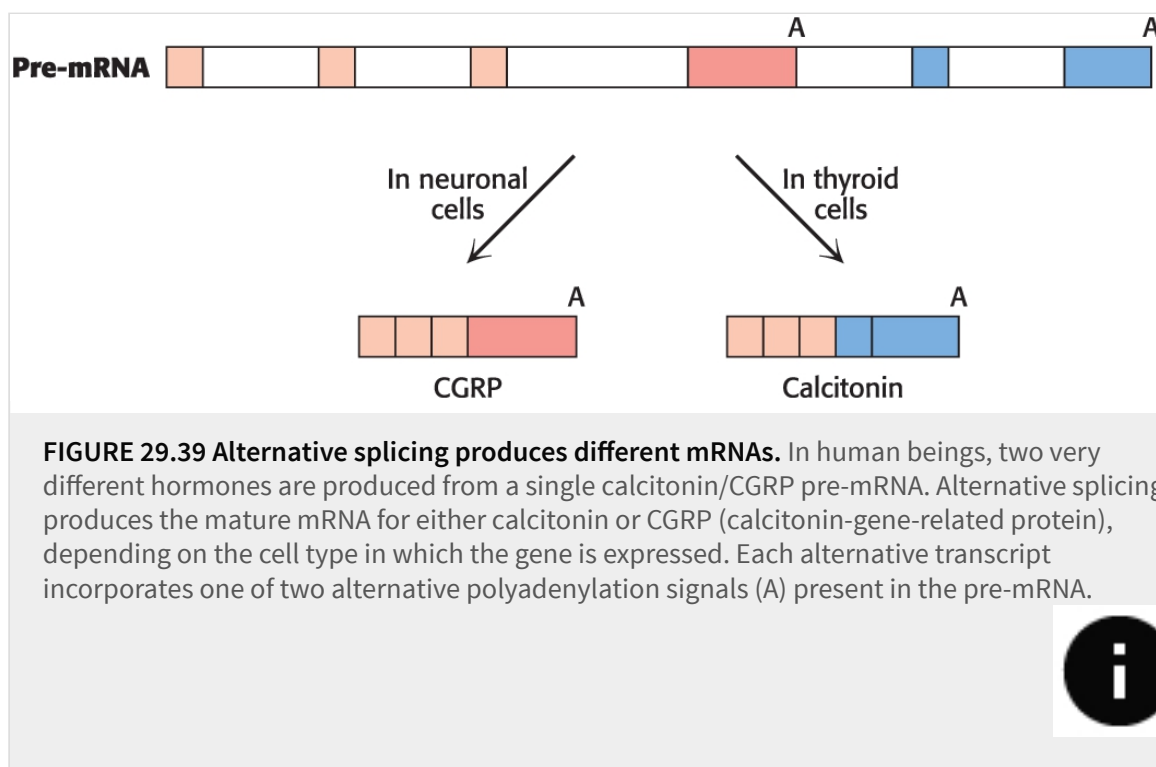
As a result of [alternative splicing](#), different combinations of exons from the same gene may be spliced into a mature RNA, producing distinct forms of a protein for specific tissues, developmental stages, or signaling pathways. What controls which splicing sites are selected? The selection is determined by the binding of trans-acting splicing factors to cis-acting sequences in the pre-mRNA. Most alternative splicing leads to changes in the coding sequence, resulting in proteins with different functions.

Alternative splicing provides a powerful mechanism for generating protein diversity. It expands the versatility of genomic sequences through combinatorial control. Consider a gene with five positions at which splicing can take place. With the assumption that these



alternative splicing pathways can be regulated independently, a total of  $2^5 = 32$  different mRNAs can be generated.

Sequencing of the human genome has revealed that most pre-mRNAs are alternatively spliced, leading to a much greater number of proteins than would be predicted from the number of genes. An example of alternative splicing leading to the expression of two different proteins, each in a different tissue, is provided by the gene encoding both calcitonin and calcitonin-gene-related peptide (CGRP; [Figure 29.39](#)). In the thyroid gland, the inclusion of exon 4 in one splicing pathway produces calcitonin, a peptide hormone that regulates calcium and phosphorus metabolism. In neuronal cells, the exclusion of exon 4 in another splicing pathway produces CGRP, a peptide hormone that acts as a vasodilator. A single pre-mRNA thus yields two very different peptide hormones, depending on cell type.



In the above example, only two proteins result from alternative splicing; however, in other cases, many more can be produced. An extreme

example is the *Drosophila* pre-mRNA that encodes DSCAM, a neuronal protein affecting axon connectivity. Alternative splicing of this pre-mRNA has the potential to produce 38,016 different combinations of exons, a greater number than the total number of genes in the *Drosophila* genome. However, only a fraction of these potential mRNAs appear to be produced, owing to regulatory mechanisms that are not yet well understood.

Several human diseases that can be attributed to defects in alternative splicing are listed in [Table 29.4](#). Further understanding of alternative splicing and the mechanisms of splice-site selection will be crucial to understanding how the proteome represented by the human genome is expressed.

**TABLE 29.4 Selected human disorders attributed to defects in alternative splicing**

Disorder	Gene or its product
Acute intermittent porphyria	Porphobilinogen deaminase
Breast and ovarian cancer	<i>BRCA1</i>
Cystic fibrosis	<i>CFTR</i>
Frontotemporal dementia	T protein
Hemophilia A	Factor VIII
HGPRT deficiency (Lesch–Nyhan syndrome)	Hypoxanthine-guanine phosphoribosyltransferase
Leigh encephalomyelopathy	Pyruvate dehydrogenase E1 $\alpha$

---

Severe combined immunodeficiency

Adenosine deaminase

---

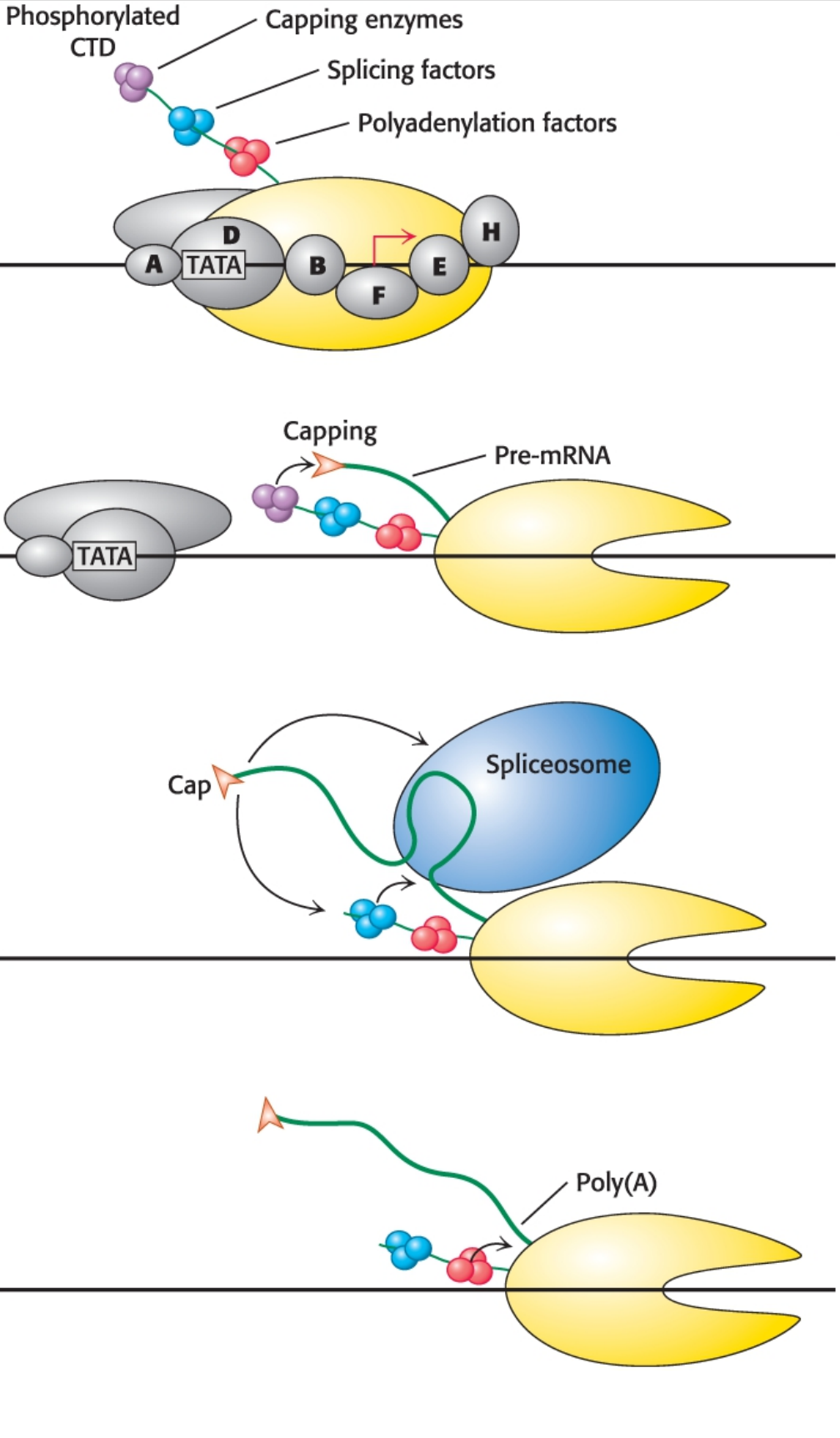
Spinal muscle atrophy

*SMN1* or *SMN2*

---

## Transcription and mRNA processing are coupled

Although we have described the transcription and processing of mRNAs as separate events in gene expression, experimental evidence suggests that the two steps are coordinated by the carboxyl-terminal domain of RNA polymerase II. We have seen that the CTD consists of a unique repeated seven-amino-acid sequence, YSPTSPS. Either S<sub>2</sub>, S<sub>5</sub>, or both may be phosphorylated in the various repeats. The phosphorylation state of the CTD is controlled by a number of kinases and phosphatases and leads the CTD to bind many of the proteins having roles in RNA transcription and processing. The CTD contributes to efficient transcription by recruiting certain proteins to the pre-mRNA ([Figure 29.40](#)). These proteins include:



**FIGURE 29.40 The CTD couples transcription to pre-mRNA processing.** The transcription factor TFIIF phosphorylates the carboxyl-terminal domain (CTD) of RNA polymerase II, signaling the transition from transcription initiation to elongation. The phosphorylated CTD binds factors required for pre-mRNA capping, splicing, and polyadenylation. These proteins are brought in close proximity to their sites of action on the nascent pre-mRNA as it is transcribed during elongation.

[Information from P. A. Sharp, *Trends in Biochemical Sciences* 30:279–281, 2005.]



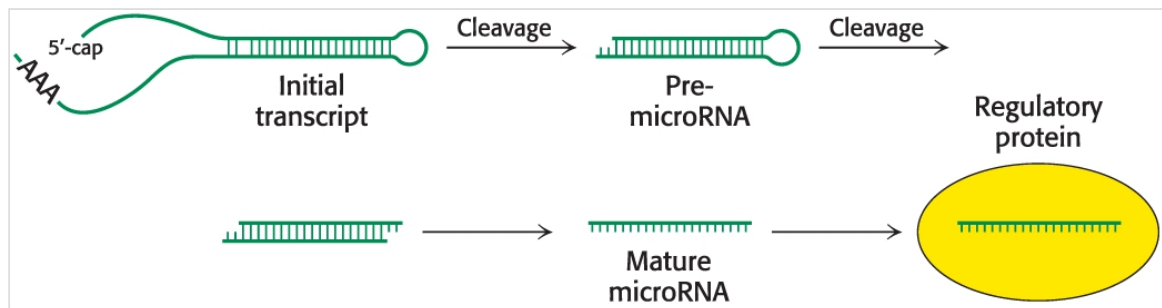
1. Capping enzymes, which methylate the 5' guanine on the pre-mRNA immediately after transcription begins
2. Components of the splicing machinery, which initiate the excision of each intron as it is synthesized
3. An endonuclease that cleaves the transcript at the poly(A) addition site, creating a free 3'-OH group that is the target for 3' adenylation

These events take place sequentially, directed by the phosphorylation state of the CTD.

## Small regulatory RNAs are cleaved from larger precursors

Cleavage plays a role in the processing of small single-stranded RNAs (approximately 20–23 nucleotides) called [microRNAs](#). MicroRNAs play key roles in gene regulation in eukaryotes, as we shall see in [Chapter 31](#). They are generated from initial transcripts produced by RNA polymerase II and, in some cases, RNA polymerase III. These transcripts fold into hairpin structures that are cleaved by specific nucleases at various stages ([Figure 29.41](#)). The final single-stranded

RNAs are bound by regulatory proteins, where the RNAs help target the regulation of specific genes.



**FIGURE 29.41 Small regulatory RNAs are involved in gene regulation.** The initial transcription product is first cleaved to a small double-stranded RNA called a pre-microRNA. One of the strands of the pre-microRNA, the mature microRNA, is then bound by a regulatory protein.



## RNA editing can lead to specific changes in mRNA

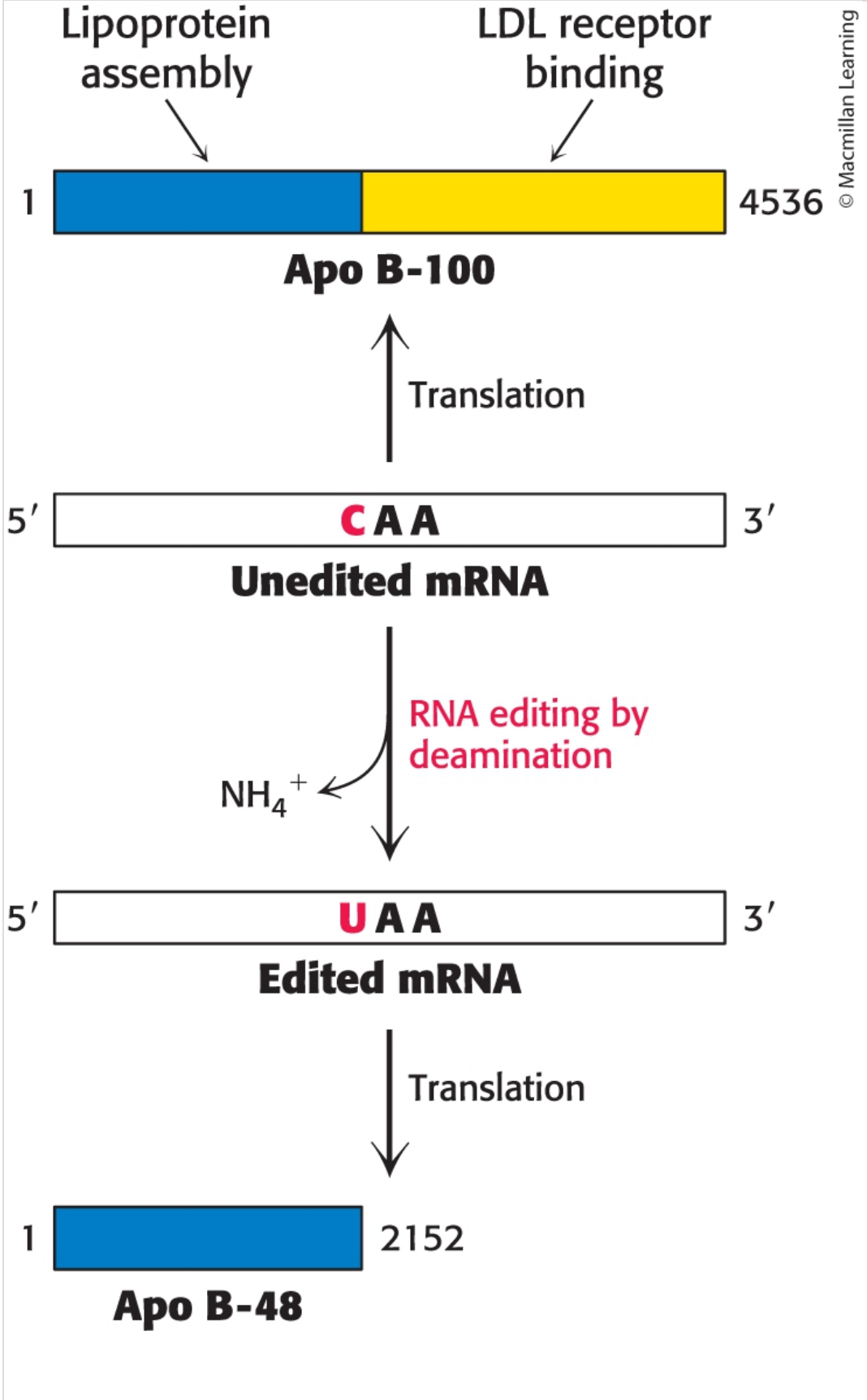
Remarkably, the amino acid sequence information encoded by some mRNAs is altered after transcription. This phenomenon is referred to as [RNA editing](#), a posttranscriptional change in the nucleotide sequence of RNA that is caused by processes other than RNA splicing. RNA editing is prominent in some systems; next, we will consider three examples.

RNA editing is key to the process of lipid transport by apolipoprotein B (apo B). Apo B plays an important role in the transport of triacylglycerols and cholesterol by forming an amphipathic spherical shell around the lipids carried in lipoprotein particles ([Section 27.3](#)). Apo B exists in two forms, a 512-kDa *apo B-100* and a 240-kDa *apo B-48*. The larger form, synthesized by the liver, participates in the transport of lipids synthesized in the cell. The smaller form, synthesized by the

small intestine, carries dietary fat in the form of chylomicrons. Apo B-48 contains the 2152 N-terminal residues of the 4536-residue apo B-100. This truncated molecule can form lipoprotein particles but cannot bind to the low-density-lipoprotein receptor on cell surfaces.

What is the relationship between these two forms of apo B?

Experiments revealed that a totally unexpected mechanism for generating diversity is at work: the changing of the nucleotide sequence of mRNA *after* its synthesis ([Figure 29.42](#)). A specific cytidine residue of mRNA is deaminated to uridine, which changes the codon at residue 2153 from CAA (Gln) to UAA (stop). The deaminase that catalyzes this reaction is present in the small intestine, but not in the liver, and is expressed only at certain developmental stages.





**FIGURE 29.42 RNA editing alters the coding sequence of mRNAs.** Enzyme-catalyzed deamination of a specific cytidine residue in the mRNA for apolipoprotein B-100 changes a codon for glutamine (CAA) to a stop codon (UAA). Apolipoprotein B-48, a truncated version of the protein lacking the LDL receptor-binding domain, is generated by this posttranscriptional change in the mRNA sequence.

[Information from P. Hodges and J. Scott, *Trends in Biochemical Sciences* 17:77, 1992.]



RNA editing also plays a role in the regulation of postsynaptic receptors. Glutamate opens cation-specific channels in the vertebrate central nervous system by binding to receptors in postsynaptic membranes. RNA editing changes a single glutamine codon (CAG) in the mRNA for the glutamate receptor to the codon for arginine (CGG). The substitution of Arg for Gln in the receptor prevents  $\text{Ca}^{2+}$ , but not  $\text{Na}^{+}$ , from flowing through the channel.

In trypanosomes (parasitic protozoans), a different kind of RNA editing markedly changes several mitochondrial mRNAs. Nearly half the uridine residues in these mRNAs are inserted by RNA editing. A guide RNA molecule identifies the sequences to be modified, and a poly(U) tail on the guide RNA donates uridine residues to the mRNAs undergoing editing.

DNA sequences evidently do not always faithfully represent the sequence of encoded proteins; crucial functional changes to mRNA can take place. RNA editing is likely much more common than was formerly thought. The chemical reactivity of nucleotide bases — including the susceptibility to deamination that necessitates complex DNA-repair mechanisms — has been harnessed as an engine for generating molecular diversity at the RNA and, hence, protein levels.

## 29.5 The Discovery of Catalytic RNA Revealed a Unique Splicing Mechanism

RNAs form a surprisingly versatile class of molecules. As we have seen, splicing is catalyzed largely by RNA molecules, with proteins playing a secondary role. RNA is also a key component of ribonuclease P, which catalyzes the maturation of tRNA by endonucleolytic cleavage of nucleotides from the 5' end of the precursor molecule. Finally, as we shall see in [Chapter 30](#), the RNA component of ribosomes is the catalyst that carries out protein synthesis.

### Some RNAs can promote their own splicing

The versatility of RNA first became clear from observations of the processing of ribosomal RNA in a single-cell eukaryote, a ciliated protozoan in the genus *Tetrahymena*. In *Tetrahymena*, a 414-nucleotide intron is removed from a 6.4-kb precursor to yield the mature 26S rRNA molecule.

In an elegant series of studies of this splicing reaction, Thomas Cech and his coworkers established that, in the absence of protein, the RNA spliced itself to precisely excise the intron. Indeed, the RNA alone is catalytic and, under certain conditions, is thus a [ribozyme](#). More than 1500 similar introns have since been found in species as widely dispersed as bacteria and eukaryotes, though not in vertebrates. Collectively, they are referred to as group I [self-splicing introns](#).



Glenn Asakawa, University of Colorado Boulder

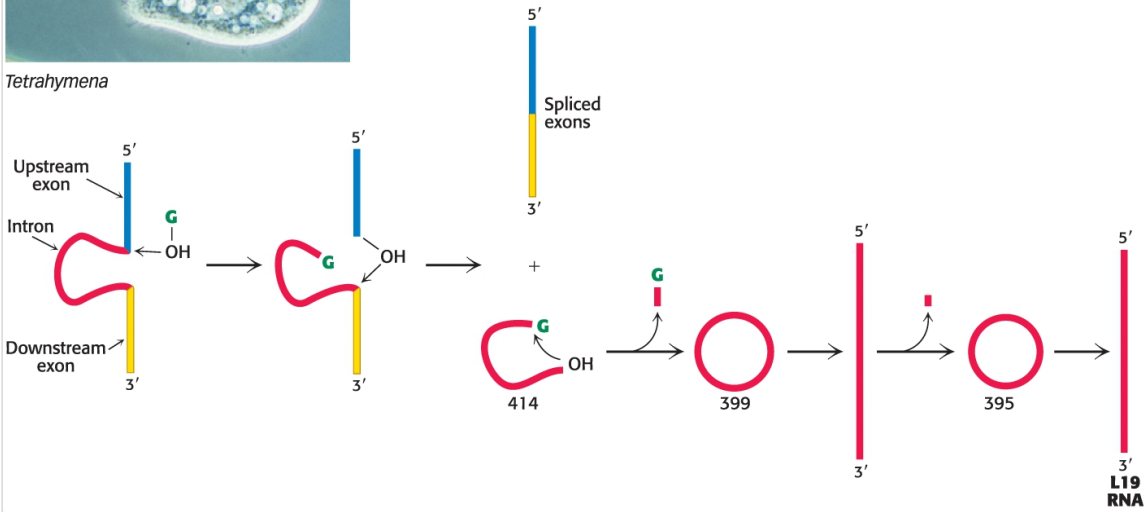
**THOMAS CECH** Fascinated by the structure of chromosomes, Thomas Cech pursued this topic as a graduate student and postdoctoral fellow. He then accompanied his wife, Dr. Carol Cech, to the University of Colorado in Boulder (UCB), where they had both been offered faculty positions. They had met in chemistry class while they were undergraduate students at Grinnell College. At UCB, Dr. Cech set out to purify enzymes involved in the splicing of a particular RNA molecule, but he and his

coworkers soon discovered that the RNA molecule spliced itself. This fundamental discovery of RNA catalysis, which changed our view of both modern biochemistry and our evolutionary past, was recognized with the Nobel Prize in Chemistry in 1989. Dr. Cech has remained at UCB, teaching with infectious enthusiasm throughout his career. He also served as an investigator (and president from 2000 to 2009) of the Howard Hughes Medical Institute (HHMI), a leading private funder of biomedical research.

The self-splicing reaction in the group I intron requires an added guanosine nucleotide ([Figure 29.43](#)). Nucleotides were originally included in the reaction mixture because it was thought that ATP or GTP might be needed as an energy source. Instead, the nucleotides were found to be necessary as cofactors. The required cofactor proved to be a guanosine unit, in the form of guanosine, GMP, GDP, or GTP. G (denoting any one of these species) serves not as an energy source but as an attacking group that becomes transiently incorporated into the RNA. G binds to the RNA and then attacks the 5' splice site to form a phosphodiester bond with the 5' end of the intron. This transesterification reaction generates a 3'-OH group at the end of the upstream exon. This 3'-OH group then attacks the 3' splice site in a second transesterification reaction that joins the two exons and leads to the release of the 414-nucleotide intron.



*Tetrahymena*

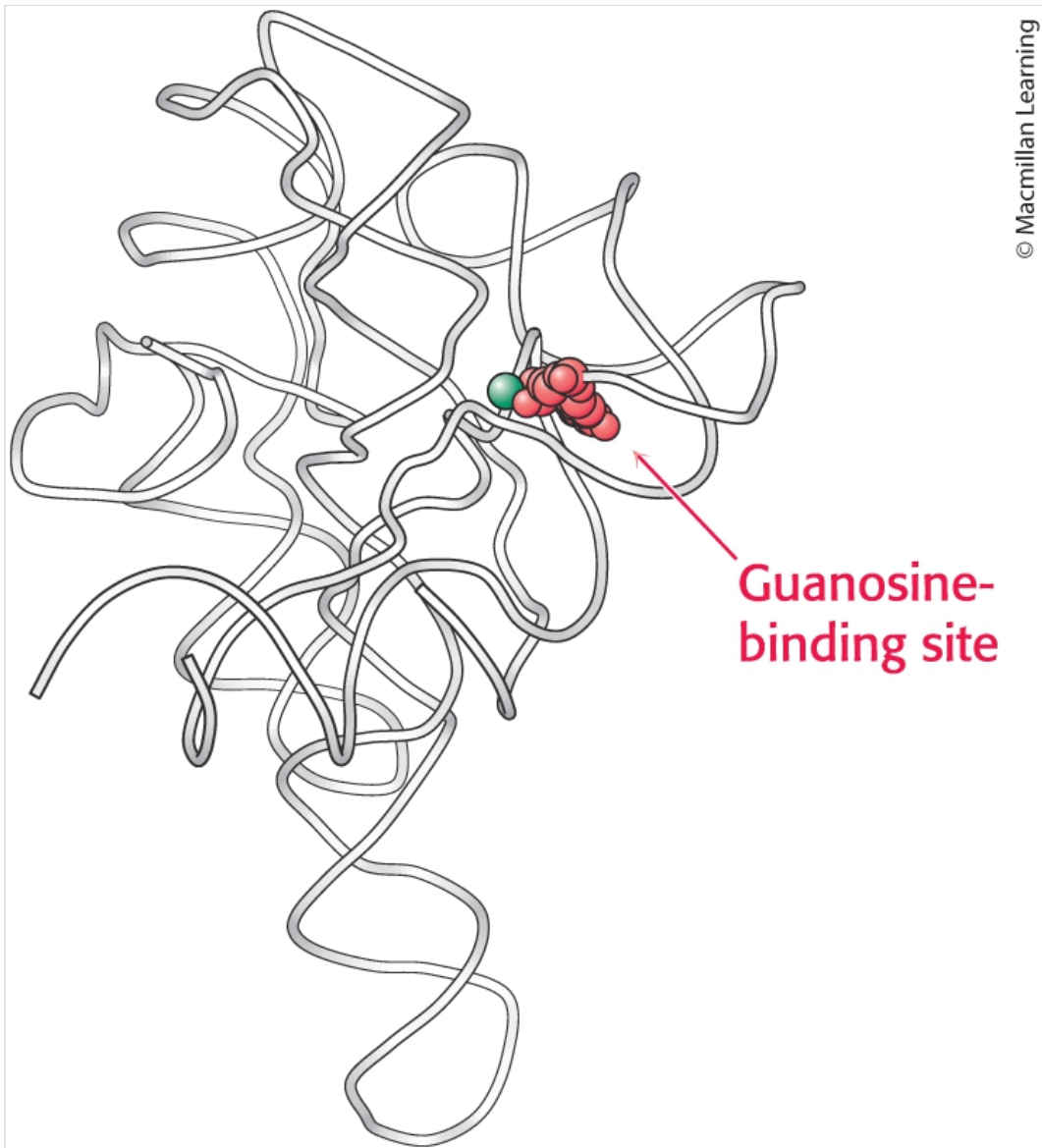


Dr. David J. Patterson/Science Source  
© Macmillan Learning

**FIGURE 29.43 Some introns are capable of self-splicing.** A ribosomal RNA precursor representative of the group I introns, from the protozoan *Tetrahymena*, splices itself in the presence of a guanosine cofactor (G, shown in green). A 414-nucleotide intron (red) is released in the first splicing reaction. This intron then splices itself twice again to produce a linear RNA that has lost a total of 19 nucleotides. This L19 RNA is catalytically active.


[Information from T. Cech, RNA as an enzyme. Copyright © 1986 by Scientific American, Inc. All rights reserved.]

Self-splicing depends on the structural integrity of the RNA precursor. Much of the group I intron is needed for self-splicing. This molecule, like many RNAs, has a folded structure formed by many double-helical stems and loops ([Figure 29.44](#)), with a well-defined pocket for binding the guanosine. Examination of the three-dimensional structure of a catalytically active group I intron determined by x-ray crystallography reveals the coordination of magnesium ions in the active site analogous to that observed in protein enzymes such as DNA polymerase.



**FIGURE 29.44 Self-splicing introns have a complex structure.** The structure of a large fragment of the self-splicing intron from *Tetrahymena* reveals a complex folding pattern of helices and loops. The guanosine binding site is highlighted in red.

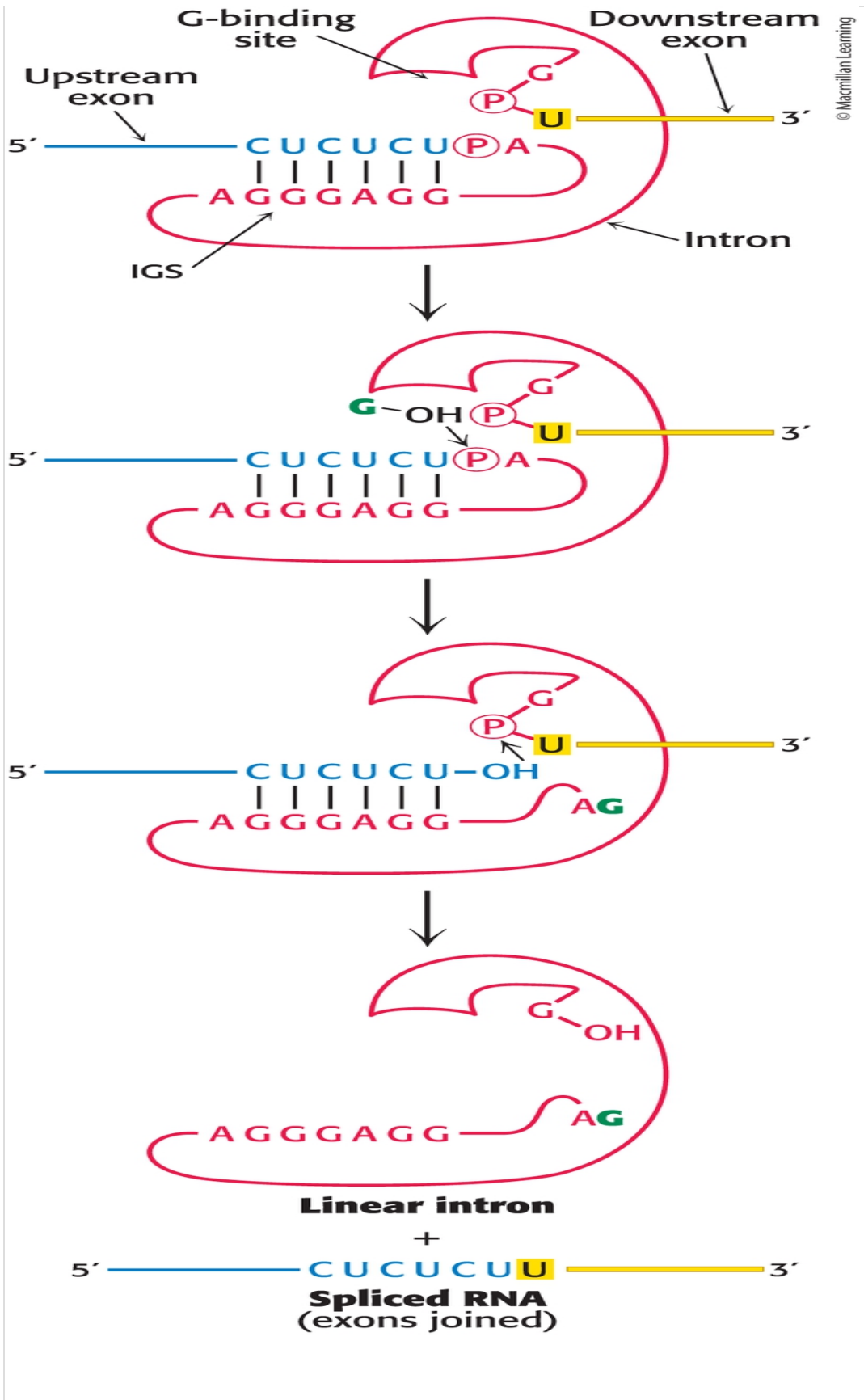
[Drawn from 1GRZ.pdb].

**INTERACT** with this model in  Achieve



Analysis of the base sequence of the rRNA precursor suggested that the splice sites are aligned with the catalytic residues by base-pairing

between the *internal guide sequence* (IGS) in the intron and the 5' and 3' exons ([Figure 29.45](#)). The IGS first brings together the guanosine cofactor and the 5' splice site so that the 3'-OH group of G can make a nucleophilic attack on the phosphorus atom at this splice site. The IGS then holds the downstream exon in position for attack by the newly formed 3'-OH group of the upstream exon. A phosphodiester bond is formed between the two exons, and the intron is released as a linear molecule. Like catalysis by protein enzymes, self-catalysis of bond formation and breakage in this rRNA precursor is highly specific.





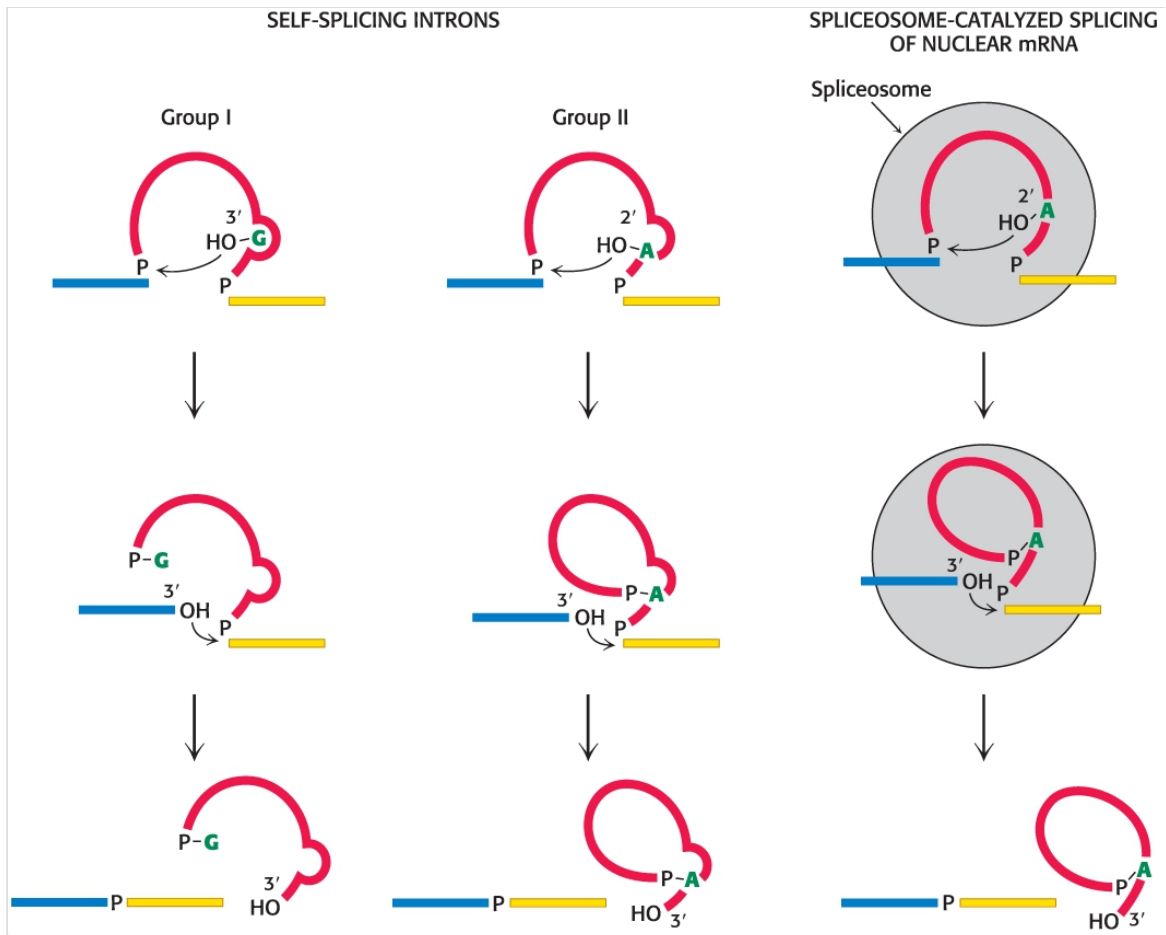
**FIGURE 29.45 The catalytic mechanism of the group I self-splicing intron includes a series of transesterification reactions.**

[Information from T. Cech, RNA as an enzyme. Copyright © 1986 by Scientific American, Inc. All rights reserved.]



The finding of enzymatic activity in the self-splicing intron and in the RNA component of RNase P has opened new areas of inquiry and changed the way in which we think about molecular evolution. As mentioned in an earlier chapter, the discovery that RNA can be a catalyst as well as an information carrier suggests that an RNA world may have existed early in the evolution of life, before the appearance of DNA and protein.

Messenger RNA precursors in the mitochondria of yeast and fungi also undergo self-splicing, as do some RNA precursors in the chloroplasts of unicellular organisms such as *Chlamydomonas*. Self-splicing reactions can be classified according to the nature of the unit that attacks the upstream splice site. Group I self-splicing is mediated by a guanosine cofactor, as in *Tetrahymena*. The attacking moiety in group II splicing is the  $2' - \text{OH}$  group of a specific adenylate of the intron ([Figure 29.46](#)).



**FIGURE 29.46** In self-splicing mechanisms, the catalytic site is formed by the intron itself. The exons being joined are shown in blue and yellow, and the attacking unit is shown in green. The catalytic site is formed by the intron itself (red) in group I and group II splicing. In contrast, the splicing of nuclear mRNA precursors is catalyzed by snRNAs and their associated proteins in the spliceosome.

[Information from P. A. Sharp, *Science* 235:766–771, 1987.]



Group I and group II self-splicing resembles spliceosome-catalyzed splicing in two respects. First, in the initial step, a ribose hydroxyl group attacks the 5' splice site. The newly formed 3'-OH terminus of the upstream exon then attacks the 3' splice site to form a phosphodiester bond with the downstream exon. Second, both reactions are

transesterifications in which the phosphate moieties at each splice site are retained in the products. The number of phosphodiester bonds stays constant.

Group II splicing is like the spliceosome-catalyzed splicing of mRNA precursors in several additional ways. First, the attack at the 5' splice site is carried out by a part of the intron itself (the 2'-OH group of adenosine) rather than by an external cofactor (G). Second, the intron is released in the form of a lariat. Third, in some instances, the group II intron is transcribed in pieces that assemble through hydrogen bonding to the catalytic intron, in a manner analogous to the assembly of the snRNAs in the spliceosome.

The similarities in mechanism have led to the suggestion that the spliceosome-catalyzed splicing of mRNA precursors evolved from RNA-catalyzed self-splicing. Group II splicing may well be an intermediate between group I splicing and the splicing in the nuclei of higher eukaryotes. A major step in this transition was the transfer of catalytic power from the intron itself to other molecules. The formation of spliceosomes gave genes a new freedom because introns were no longer constrained to provide the catalytic center for splicing. Another advantage of external catalysts for splicing is that they can be more readily regulated.

However, it is important to note that similarities do not establish ancestry. The similarities between group II introns and mRNA splicing may be a result of convergent evolution. Perhaps there are only a limited number of ways to carry out efficient, specific intron excision. The determination of whether these similarities stem from ancestry or from chemistry will require expanding our understanding of RNA biochemistry.

---

## SELF-CHECK QUESTION



Compare and contrast the three different splicing mechanisms that have been identified.

---

## RNA enzymes can promote many reactions, including RNA polymerization

In this chapter, we have discussed the role of many different RNA molecules, including those that promote reactions such as splicing. In [Chapter 30](#), we will discuss the key roles that RNAs play in protein biosynthesis, and we will see that ribosomal RNAs have a significant role in the catalysis of peptide bonds.

While the ribosomal RNAs have a role in the polymerization of amino acids into proteins, some RNAs are also capable of catalyzing the polymerization of other RNA molecules or even themselves. In vitro experiments have identified the possibility of self-ligating ribozymes: RNA molecules capable of joining other, short RNAs to their own end. The ability of RNAs to direct the synthesis of other catalytic molecules gives us a glimpse into how, in a primordial RNA world, ribozymes may have accelerated chemical reactions critical for life.

# Chapter 29 Summary

## 29.1 RNA Molecules Play Different Roles, Primarily in Gene Expression

- RNA molecules are integral to many different cellular functions; for example, protein biosynthesis.
- Some RNAs can direct their own modification.

## 29.2 RNA Polymerases Catalyze Transcription

- RNA polymerases synthesize all cellular RNA molecules according to instructions given by DNA templates.
- The direction of RNA synthesis is  $5' \rightarrow 3'$ , as in DNA synthesis.

RNA polymerases, unlike DNA polymerases, do not need a primer.

- RNA polymerase in *E. coli* is a multisubunit enzyme. The subunit composition of the  $\sim 500\text{-kDa}$  holoenzyme is  $\alpha_2\beta\beta'\omega\sigma$  and that of the core enzyme is  $\alpha_2\beta\beta'\omega$ .
- Transcription is initiated at promoter sites.
- The  $\sigma$  subunit enables the holoenzyme to recognize promoter sites. The  $\sigma$  subunit usually dissociates from the holoenzyme after the initiation of the new chain.
- Elongation takes place at transcription bubbles that move along the DNA template at a rate of about 50 nucleotides per second.
- The nascent RNA chain contains stop signals that end transcription. One stop signal is an RNA hairpin, which is followed by several U residues. A different stop signal is read by the *rho* protein, an ATPase.
- In *E. coli*, precursors of transfer RNA and ribosomal RNA are cleaved and chemically modified after transcription, whereas

messenger RNA is used unchanged as a template for protein synthesis.

### 29.3 Transcription Is Highly Regulated

- Bacteria use alternate  $\sigma$  factors to adjust the transcription levels of various genes in response to external stimuli.
- Some genes are regulated by riboswitches, structures that form in RNA transcripts and bind specific metabolites.
- Eukaryotic DNA is tightly bound to basic proteins called histones; the combination is called chromatin. DNA wraps around an octamer of core histones to form a nucleosome.
- There are three types of eukaryotic RNA polymerases in the nucleus, where transcription takes place: RNA polymerase I makes ribosomal RNA precursors, II makes messenger RNA precursors, and III makes transfer RNA precursors.
- Eukaryotic promoters are composed of several different elements. The activity of many promoters is greatly increased by enhancer sequences that have no promoter activity of their own.

### 29.4 Some RNA Transcription Products Are Processed

- The 5' ends of mRNA precursors become capped and methylated during transcription.
- A 3' poly(A) tail is added to most mRNA precursors after the nascent chain has been cleaved by an endonuclease.
- The splicing of mRNA precursors is carried out by spliceosomes, which consist of small nuclear ribonucleoproteins. Splice sites in mRNA precursors are specified by sequences at ends of introns and by branch sites near the 3' ends of introns.

- RNA editing alters the nucleotide sequence of some mRNAs, such as the one for apolipoprotein B.

## 29.5 The Discovery of Catalytic RNA Revealed a Unique Splicing Mechanism

- Some RNA molecules undergo self-splicing in the absence of protein.
  - Spliceosome-catalyzed splicing may have evolved from self-splicing.
  - The discovery of catalytic RNA has opened new vistas in our exploration of early stages of molecular evolution and the origins of life.
- 

## Key Terms

[transcription](#)

[RNA polymerase](#)

[promoter](#)

[transcription bubble](#)

[sigma \( \$\sigma\$ \) subunit](#)

[consensus sequence](#)

[rho \( \$\rho\$ \) protein](#)

[riboswitch](#)

[chromatin](#)

[nucleosome](#)

[carboxyl-terminal domain \(CTD\)](#)

[TATA box](#)

[transcription factor](#)

[enhancer](#)

[small nucleolar ribonucleoprotein \(snoRNP\)](#)

[pre-mRNA](#)

5' cap

poly(A) tail

RNA splicing

spliceosome

small nuclear RNA (snRNA)

small nuclear ribonucleoprotein (snRNP)

alternative splicing

microRNA



RNA editing

ribozyme

self-splicing introns


---

## Problems

1. Why is RNA synthesis not as carefully monitored for errors as is DNA synthesis?  1,  2

2. What are the functions of RNA polymerases? Select all that apply.  1,  3

- a. Polymerization of polypeptides from RNA transcripts
- b. Elongation of ribosomal RNA (rRNA)
- c. Initiation of transcription at promoter sites
- d. Elongation of messenger RNA (mRNA) transcripts
- e. Initiation of translation from RNA transcripts

3. The overall structures of RNA polymerase and DNA polymerase are very different, yet their active sites show considerable similarities. What do the similarities suggest about the evolutionary relationship between these two important enzymes?  1



4. The sequence of part of an mRNA transcript is

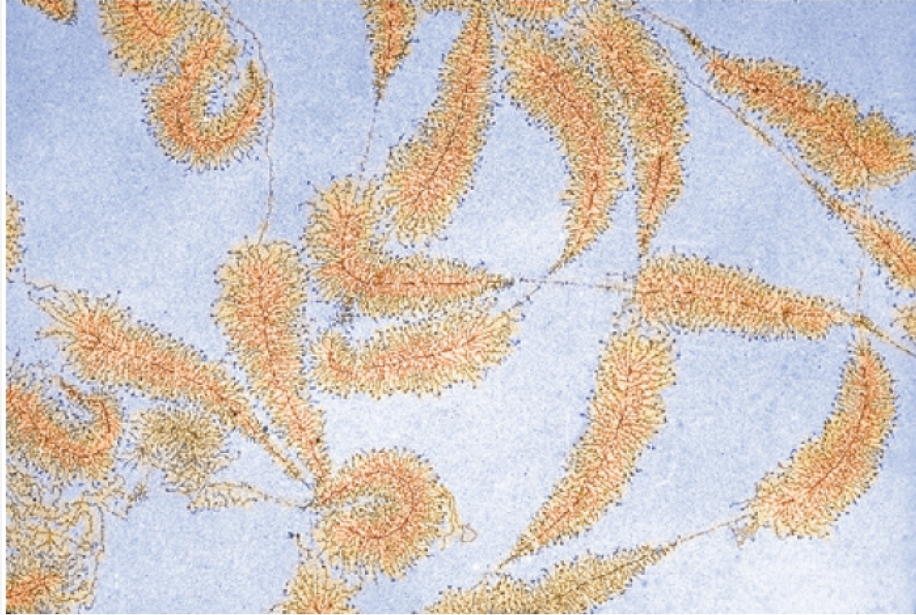
5' - AUGGGGAACAGCAAGAGUGGGGCCUGUCCAAGGAG - 3'

What is the sequence of the DNA coding strand? Of the DNA template strand? 1

5. Sigma protein by itself does not bind to promoter sites. Predict the effect of a mutation enabling  $\sigma$  to bind to the <sup>-10</sup> region in the absence of other subunits of RNA polymerase. 2

6. The molecular weight of an amino acid is approximately 110 Da, and *E. coli* RNA polymerase has a transcription rate of approximately 5050 nucleotides per second. What is the minimum length of time required by *E. coli* polymerase for the synthesis of an mRNA encoding a 100-kDa protein? Round your answer to the nearest whole number. 1

7. The autoradiograph below depicts several bacterial genes undergoing transcription. Identify the DNA. What are the strands of increasing length? Where is the beginning of transcription? The end of transcription? What can you conclude about the number of enzymes participating in RNA synthesis on a given gene? 2



Science Source

8. Fill in the blanks to complete the statements about the capping process: Capping occurs at the **(a)** \_\_\_\_\_ end of the mRNA. The nucleoside that is added is **(b)** \_\_\_\_\_. The nucleoside is connected to the mRNA through a **(c)** \_\_\_\_\_ linkage. The added nucleoside is modified by methylation at **(d)** \_\_\_\_\_ atom. The mRNA is further modified by methylation of ribose residues at the **(e)** \_\_\_\_\_-hydroxyl position. ❖ 3

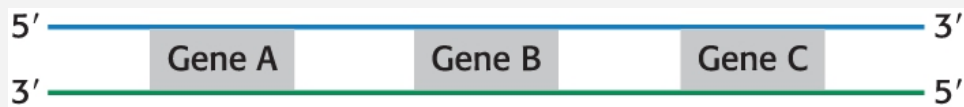
9. Suppose that RNA polymerase was transcribing a eukaryotic gene with several introns all contained within the coding region. In what order would the RNA polymerase encounter the elements in the DNA sequence of the gene, from earliest to latest? ❖ 2

- a. Splice donor site
- b. 3' untranslated region (UTR)
- c. Transcription start site
- d. Translation start codon

e. 5' untranslated region (UTR)

f. Translation termination codon

10. For the three genes shown below, transcription occurs from right to left for gene A and gene B, whereas transcription occurs from left to right for gene C. Identify which strand (top or bottom) is the template strand for each gene. 🎯 1



11. Which would you expect to have more types of  $\sigma$  factors: a free-living prokaryote in a fluctuating environment, or a prokaryote in a relatively stable environment with adequate nutrients? Why? 🎯 2

12. The three different eukaryotic RNA polymerases perform different functions and can be distinguished experimentally by their sensitivity to the toxin  $\alpha$ -amanitin. Match the appropriate function and  $\alpha$ -amanitin sensitivity to each RNA polymerase. Each function and sensitivity will be used only once. 🎯 3

Type of polymerase	Function and sensitivity
--------------------	--------------------------

1. RNA polymerase I	1. Inhibited by high concentrations of $\alpha$ -amanitin
---------------------	---

2. RNA polymerase II

2. Not inhibited by  $\alpha$ -amanitin

3. RNA polymerase III

3. Synthesizes pre-ribosomal RNA

4. Synthesizes primarily tRNA and small RNAs

5. Synthesizes mostly pre-messenger RNA

6. Inhibited by low concentrations of  $\alpha$ -amanitin

**13.** Fill in the blanks to complete the statements about ribosomal RNA processing in eukaryotes: RNA polymerase I generates the **(a)** \_\_\_\_\_ kb rRNA transcript, which has a sedimentation coefficient of **(b)** \_\_\_\_\_. The RNA transcript is processed in a series of steps to yield three smaller rRNA molecules: **(c)** \_\_\_\_\_. The **(d)** \_\_\_\_\_ rRNA assembles with the **(e)** \_\_\_\_\_ rRNA, the **(f)** \_\_\_\_\_ rRNA, and proteins into **(g)** \_\_\_\_\_ ribosomal subunits. The 5S rRNA genes are transcribed by **(h)** \_\_\_\_\_. The 18S rRNA is associated with **(i)** \_\_\_\_\_ ribosomal subunits. ❖ 3

**14.** Which of the following statements regarding promoters is correct? Select all that apply. ❖ 1, ❖ 4, ❖ 6

- In prokaryotes, the promoter is recognized by general transcription factors, which recruit the RNA polymerase holoenzyme.
- In both prokaryotes and eukaryotes, the promoter is located in the 5' direction, upstream from the transcription start site.
- In eukaryotes, the promoter attracts the small and large ribosomal subunits with the help of initiation factors.

- d. In prokaryotes, the promoter contains a  $-35$  and  $-10$  region upstream of the transcription start site.
- e. In eukaryotes, the promoter recruits the preinitiation complex, which includes the TATA-binding protein.

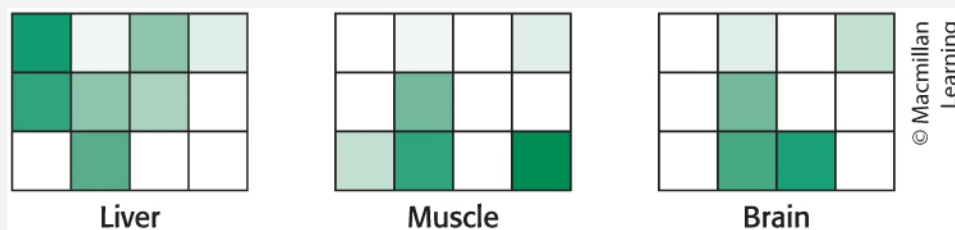
**15.** Match each RNA type with its appropriate function. Each function will be used only once. 🎯 3

Type of RNA	Function
1. small nuclear RNA (snRNA)	1. Helps to regulate the expression of specific genes
2. ribosomal RNA (rRNA)	2. Carries genetic information to ribosomes for protein synthesis
3. transfer RNA (tRNA)	3. Combines with proteins to form ribosomes
4. messenger RNA (mRNA)	4. Brings amino acids to sites of mRNA during protein synthesis
5. microRNA (miRNA)	5. Involved in the splicing process when mRNA is formed

**16.** A gene contains eight sites where alternative splicing is possible. Assuming that the splicing pattern at each site is independent of that at all other sites, how many splicing products are possible? 🎯 5

**17.** What processes considered in this chapter make the proteome more complex than the genome? What processes might further enhance this complexity? 🎯 5

18. Nuclei were isolated from brain, liver, and muscle. The nuclei were then incubated with  $\alpha$ - $[^{32}\text{P}]$  UTP under conditions that allow RNA synthesis, except that an inhibitor of RNA initiation was present. The radioactive RNA was isolated and annealed to various DNA sequences that had been attached to a gene chip. In the adjoining graphs, the intensity of the shading indicates roughly how much mRNA was attached to each DNA sequence. 4



- Why does the intensity of hybridization differ between genes?
- What is the significance of the fact that some of the RNA molecules display different hybridization patterns in different tissues?
- Some genes are expressed in all three tissues. What is the nature of these genes?
- Suggest a reason why an initiation inhibitor was included in the reaction mixture.

19. Eukaryotic messenger RNA can undergo postsynthetic processing after transcription and before translation. One of the processing steps is splicing, where portions of the RNA are removed and the remaining RNA are joined together. Classify the statements regarding mRNA splicing as true or false. 5

- a. Splicing occurs while the mRNA is attached to the nucleosome.
- b. One mRNA can sometimes code for more than one protein by splicing at alternative sites.
- c. Splicing occurs while the mRNA is still in the nucleus.
- d. In splicing, intron sequences are removed from the mRNA in the form of lariats (loops) and are degraded.
- e. Splicing of mRNA does not involve any proteins.