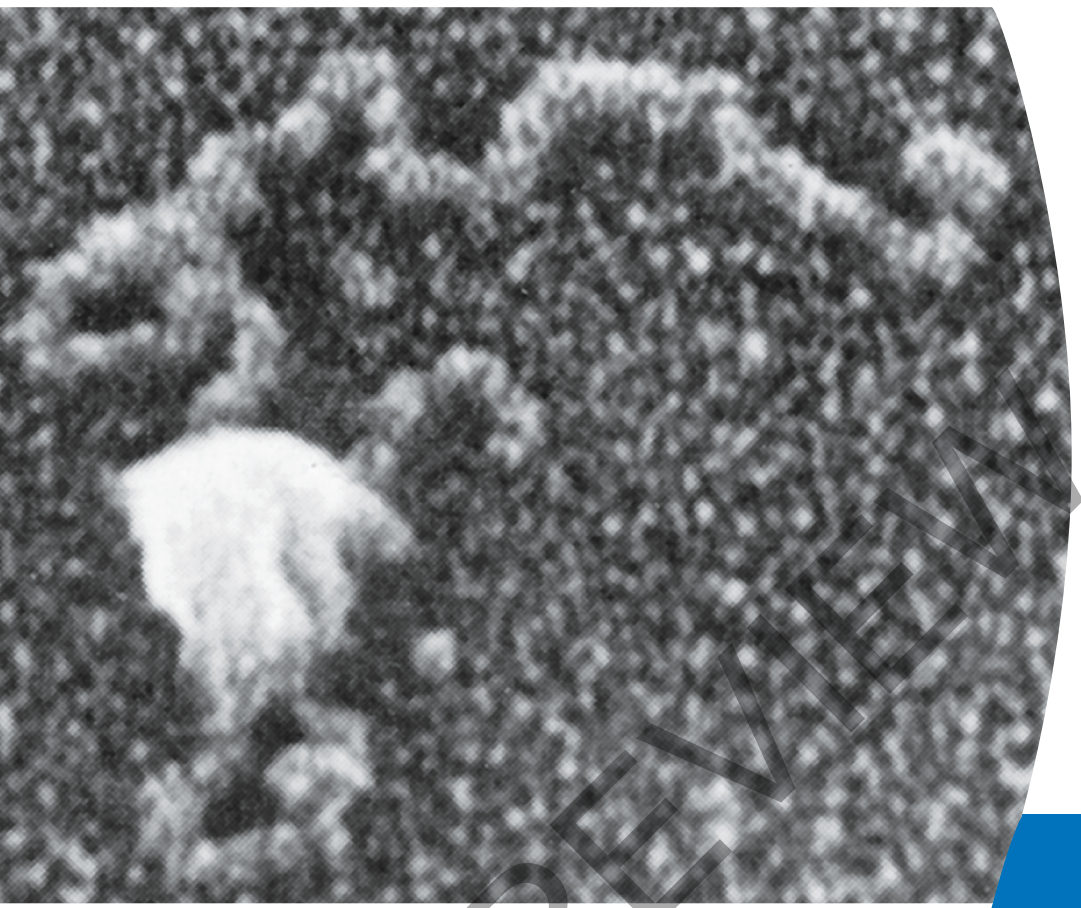
Molecular Biology of Transcription and RNA Processing

8



An electron micrograph of a spliceosome engaged in intron splicing.

Wilhelm Johansson introduced the term *gene* in 1909 to describe "the fundamental unit of inheritance." Johansson's definition encompasses the understanding that genes contain genetic information and are passed from one generation to the next and that genes are the basis of the fundamental structural, functional, developmental, reproductive, and evolutionary properties of organisms. This basic definition of the gene remains valid today, more than a century after being coined, but our knowledge of molecular genetics has expanded enormously, refining our understanding of the structure and function of genes and clarifying the roles genes play in producing traits.

CHAPTER OUTLINE

ESSENTIAL IDEAS

- Ribonucleic acid (RNA) molecules are transcribed from genes and are classified either as messenger RNA or as one of several types of functional RNA.
- Bacterial transcription is a four-step process that begins with promoter recognition by RNA polymerase and ends with the completion of transcript synthesis.
- Eukaryotes and archaea have homologous transcription proteins and processes. Eukaryotes use different RNA polymerases to transcribe different kinds of RNA. Each type of polymerase initiates transcription at a different type of promoter.
- Eukaryotic RNAs undergo three processing steps after transcription. Alternative events during and after transcription allow different transcripts and proteins to be produced from the same DNA sequence.

The central dogma of biology describes the flow of genetic information from DNA to RNA to protein (see Figure 1.8). It conveys that DNA is the repository of genetic information, which is converted through *transcription* into RNA, one type of which is then *translated* into protein. Transcription is the process by which RNA polymerase enzymes and other transcriptional proteins and enzymes use the template strand of DNA to synthesize a complementary RNA strand. Translation is the process by which *messenger RNA* is used to direct protein synthesis.

This chapter describes the mechanisms of RNA transcription in the three domains of life: bacteria, archaea, and eukaryotes. We will also examine the events that modify the precursor messenger RNA (mRNA) to yield the mature mRNA that subsequently undergoes translation to produce proteins. We will see that these transcriptional events are closely tied to the process of translation, the subject of the following chapter.

This chapter also discusses the shared evolutionary history and common ancestry of bacteria, archaea, and eukaryotes. We will see that, bacteria have a number of general features of transcription in common with archaea and eukaryotes. At the same time, we see that, differences among the members of these domains, including differences in cell structure, gene structure, and genome organization, lead to significant differences in how their genes are transcribed and translated.

Multiple types of RNA are introduced and described here, but the principal focus of discussion is mRNA. The discovery of mRNA and of its function raised numerous questions: How is a gene recognized by the transcription machinery? Where does transcription begin? Which strand of DNA is transcribed? Where does transcription end? How much transcript is made? How is RNA modified after transcription? We answer these questions in the chapter and set the answers in a context that compares and contrasts the process of transcription in bacterial, archaeal, and eukaryotic genomes.

8.1 RNA Transcripts Carry the Messages of Genes

In the late 1950s, with the structure of DNA in hand, molecular biology researchers focused on identifying and describing the molecules and mechanisms responsible for conveying the genetic message of DNA. RNA was known to be chemically similar to DNA and present in abundance in all cells, but its diversity and biological roles remained to be discovered. Some roles were strongly suggested by cell structure. For example, in eukaryotic cells, DNA is located in the nucleus, whereas protein synthesis takes place in the cytoplasm, suggesting that DNA could not code directly for proteins but RNA perhaps could. Bacteria, however, lack a nucleus, so an open research question was whether bacteria and eukaryotes used similar mechanisms and similar molecules to convey the genetic message for protein synthesis. The search was on to identify the types of RNA in cells and to identify the mechanisms by which the genetic message of DNA is conveyed for protein synthesis.

It is worth noting that the experimental evidence identifying archaea as occupying a separate domain from bacteria and eukaryotes was obtained after some of the fundamental information about transcription became known. We introduce transcription in archaea in a later section. These microbes, which like bacteria also lack a nucleus, reveal an intriguing blend of bacterial and eukaryotic features. The archaeal core transcriptional proteins are clearly homologous to the eukaryotic apparatus, while the regulation of these processes is more bacteria-like in nature.

RNA Nucleotides and Structure

Both DNA and RNA are polynucleotide molecules composed of nucleotide building blocks. One principal difference between the molecules is the single-stranded structure of RNA versus the double-stranded structure of DNA. Despite their single-stranded structure, however, RNA molecules can, and frequently do, adopt folded secondary structures by complementary base pairing of segments of the molecule. In certain instances, folded secondary structures are essential to RNA function, as we discuss in the following section.

The RNA nucleotides, like those of DNA, are composed of a five-carbon sugar, a nucleotide base, and one or more phosphate groups. Each RNA nucleotide carries one of four possible nucleotide bases. At the same time, RNA nucleotides have two critical chemical differences in comparison to DNA nucleotides. The first difference concerns the identity of the RNA nucleotide bases. The purines adenine and guanine in RNA are identical to the purines in DNA. Likewise, the pyrimidine cytosine is identical in RNA and DNA. In RNA, however, the second pyrimidine

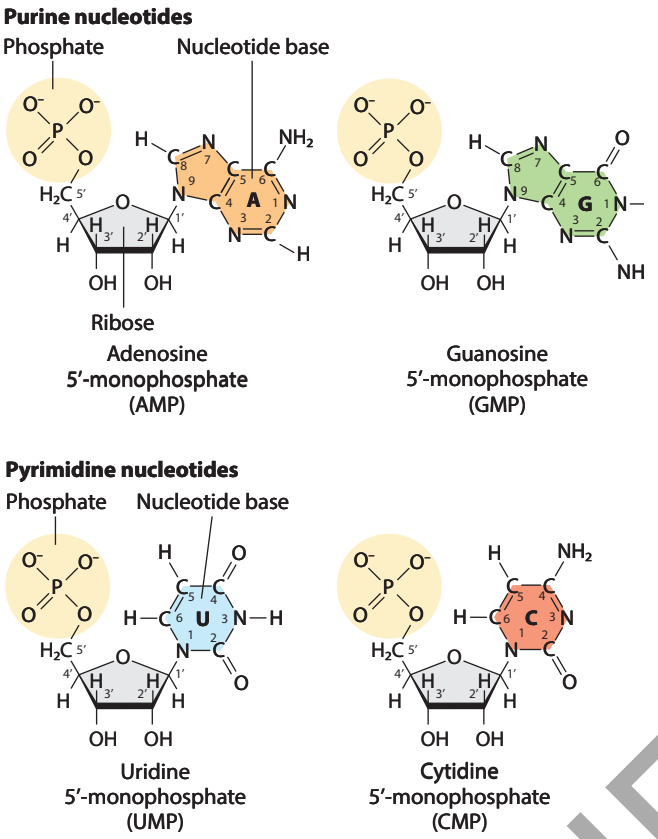


Figure 8.1 **The four RNA ribonucleotides.** Shown in their monophosphate forms, each ribonucleotide consists of the sugar ribose, one phosphate group, and one of the nucleotide bases adenine, guanine, cytosine, and uracil.

is **uracil** (U) rather than the thymine carried by DNA. The four RNA **ribonucleotides** (A, U, G, C) are shown in Figure 8.1. The structure of uracil is similar to that of thymine, but notice, by comparing the structure of uracil in Figure 8.1 with that of thymine in Figure 7.5, that thymine has a methyl group (CH_3) at the 5 carbon of the pyrimidine ring, whereas uracil does not. In all other respects, uracil is similar to thymine, and when uracil undergoes base pairing, its complementary partner is adenine.

The second chemical difference between RNA and DNA nucleotides is the presence of the sugar **ribose** in RNA rather than the deoxyribose occurring in DNA. The ribose gives RNA its name (ribonucleic acid). Compare the ribose molecules shown in Figure 8.1 to deoxyribose in Figure 7.5, and notice that ribose carries a hydroxyl group (OH) not found in deoxyribose at the 2′ carbon of the ring. Except for this difference, ribose and deoxyribose are identical, having a nucleotide base attached to the 1′ carbon and a hydroxyl group at the 3′ carbon.

The similarity of the sugars of RNA and DNA leads to the formation of essentially identical sugar-phosphate backbones in the molecules. RNA strands are assembled by formation of phosphodiester bonds, between the 5′ phosphate of one nucleotide and the 3′ hydroxyl of the adjacent nucleotide, that are identical to those found in DNA (Figure 8.2). RNA is synthesized from a DNA template strand using the same purine-pyrimidine complementary base pairing described for DNA except for the pairing between adenine of DNA with uracil of RNA. **RNA polymerase** enzymes catalyze the addition of each ribonucleotide to the 3′ end of the nascent strand and form phosphodiester bonds between a triphosphate group at the 5′ carbon of one nucleotide and the hydroxyl group at the 3′ carbon of the adjacent nucleotide, eliminating two phosphates (the pyrophosphate group), just as in DNA synthesis. Compare Figure 8.2 to Figure 7.6 to see the similarity of these nucleic acid synthesis processes.

Identification of Messenger RNA

In their search for the RNA molecule responsible for transmitting the genetic information content of DNA to the ribosome for protein production, researchers utilized many techniques. Among the methods used was the pulse-chase technique (see Section 7.3) to follow the trail of newly synthesized RNA in cells. The "pulse" step of this technique exposes cells to radioactive nucleotides that become incorporated into newly synthesized nucleic acids (see Chapter 7). After a short incubation period to incorporate the labeled nucleotides, a "chase" step replaces any remaining unincorporated radioactive nucleotides by introducing an excess of unlabeled nucleotides. An experimenter can then observe the location and movement of the labeled nucleic acid to determine the pattern of its movement and its ultimate destination and fate.

In 1957, microbiologist Elliot Volkin and geneticist Lazarus Astrachan used the pulse-chase method to examine transcription in bacteria immediately following infection by a bacteriophage. Exposing newly infected bacteria to radioactive uracil, they observed rapid incorporation of the label, indicating a burst of transcriptional activity. In the chase phase of the experiment, when radioactive uracil was removed, Volkin and Astrachan found that the radioactivity quickly dissipated, indicating that the newly synthesized RNA broke down rapidly. They concluded that the synthesis of a type of RNA with a very short life span is responsible for the production of phage proteins that drive progression of the infection.

Similar pulse-chase experiments were soon conducted with eukaryotic cells. In these experiments, cells were pulsed with radioactive uracil that was then chased with nonradioactive uracil. Immediately after the pulse, radioactivity was concentrated in the nucleus, indicating that newly synthesized RNA has a nuclear location. Over a short period, radioactivity migrated to the cytoplasm, where translation takes place. The radioactivity dissipated after lingering in the cytoplasm for a period of time. These experiments led researchers to conclude that the RNA synthesized in the nucleus was likely to act as an

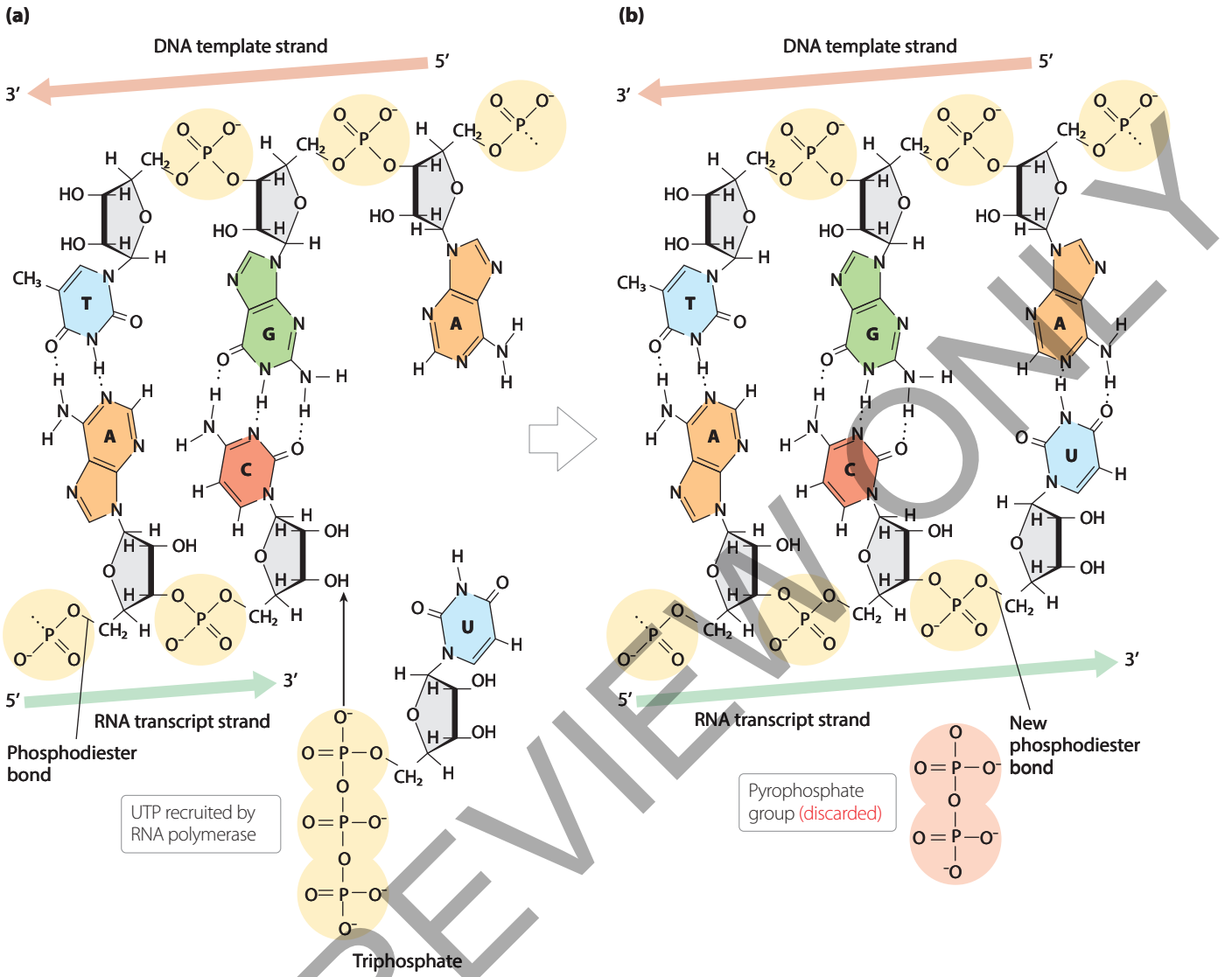


Figure 8.2 RNA synthesis.

intermediary carrying the genetic message of DNA to the cytoplasm for translation into proteins.

The discovery of mRNA was capped in 1961 when an experiment by the biologists Sydney Brenner, Francois Jacob, and Matthew Meselson identified an unstable form of RNA as the genetic messenger. Brenner and his colleagues designed an experiment using the bacteriophage T2 and *Escherichia coli* to investigate whether phage protein synthesis requires newly constructed ribosomes, or whether phage proteins could be produced using existing bacterial ribosomes and a messenger molecule to encode the proteins. The experiment found that newly synthesized phage RNA associates with bacterial ribosomes to produce phage proteins. The RNA that directed the protein synthesis formed and degraded quickly, leading the experimenters to conclude that a phage "messenger" RNA with a short half-life is responsible for protein synthesis during infection.

RNA Classification

A large variety of different RNA species exist within any cell. The most essential types of RNA are found in all cells in all three domains, but several others are specific to eukaryotic cells. **Table 8.1** identifies and briefly describes the most important types of RNA found in cells, although it is not an exhaustive list, as there are too many varieties of RNA to describe all of them here.

All RNAs are transcribed from RNA-encoding genes. The various types of RNA are constructed from the same building blocks but perform different roles in the cell. In light of these different roles, RNAs are divided into two general categories—*messenger RNA* and *functional RNA*.

Genes transcribing **messenger RNA (mRNA)** are protein-producing genes, and their transcripts direct protein synthesis by the process of translation. Messenger RNA is the short-lived intermediary form of RNA that

Table 8.1 Major RNA Molecules

Type of RNA	Function
Messenger RNA (mRNA)	Used to encode the sequence of amino acids in a polypeptide. May be polycistronic (encoding two or more polypeptides) in bacteria and archaea. Encodes single polypeptides in nearly all eukaryotes (see Section 8.2).
Ribosomal RNA (rRNA)	Along with numerous proteins, helps form the large and small ribosomal subunits that unite for translation of mRNA (see Sections 8.4 and 9.2).
Transfer RNA (tRNA)	Carries amino acids to ribosomes and binds there to mRNA by complementary base pairing in order to deposit the amino acids to elongate the polypeptide (see Sections 8.4 and 9.3).
Small nuclear RNA (snRNA)	Found in eukaryotic nuclei, where multiple snRNAs join with numerous proteins to form spliceosomes that remove introns from precursor mRNA (see Section 8.4).
MicroRNA (miRNA) and small interfering RNA (siRNA)	Eukaryotic regulatory RNAs that have different origins. Involved in eukaryotic regulation of gene expression (see Section 15.3).
Telomerase RNA	Along with several proteins, forms telomerase, the ribonucleoprotein complex essential for maintaining and elongating telomere length of eukaryotic chromosomes (see Section 7.4).

conveys the genetic message of DNA to ribosomes for translation. Messenger RNA is the only form of RNA that undergoes translation. Transcription of mRNA and post-transcriptional processing of mRNA are principal areas of focus in this chapter.

Functional RNAs perform a variety of specialized roles in the cell. The functional RNAs carry out their activities in nucleic acid form and are not translated. Two major categories of functional RNA are active in bacterial and eukaryotic translation. **Transfer RNA (tRNA)** is encoded in dozens of different forms in all genomes. Each tRNA is responsible for binding a particular amino acid that it carries to the ribosome. There the tRNA interacts with mRNA and deposits its amino acid for inclusion in the growing protein chain. **Ribosomal RNA (rRNA)** combines with numerous proteins to form the ribosome, the molecular machine responsible for translation. Certain bacterial rRNA molecules interact with mRNA to initiate translation.

Three additional types of functional RNA perform specialized functions in eukaryotic cells only. **Small nuclear RNA (snRNA)** of various types is found in the nucleus of eukaryotic cells, where it participates in mRNA processing. Certain snRNAs unite with nuclear proteins to form ribonucleoprotein complexes that are responsible for intron removal. We discuss these activities in later sections of this chapter. **Micro RNA (miRNA)** and **small interfering RNA (siRNA)** are recently recognized types of regulatory RNA that are particularly active in plant and animal cells. Micro RNAs and siRNAs have a widespread and important role in the post-transcriptional regulation of mRNA, regulating protein production through a process called *RNA interference*. Their transcription and activities are beyond the scope of this chapter, but they are central to the discussion of the regulation of gene expression in eukaryotes in Chapter 15.

Lastly, certain RNAs in eukaryotic cells have catalytic activity. In contrast to DNA, which is exclusively a repository of genetic information, catalytically active RNA molecules can catalyze biological reactions. Called **ribozymes,** catalytically active RNAs can activate cellular reactions, including the removal of introns in a process identified as *self-splicing,* described later in the chapter.

8.2 Bacterial Transcription Is a Four-Stage Process

Transcription is the synthesis of a single-stranded RNA molecule by RNA polymerase. It is most clearly understood and described in bacteria, and *E. coli* is the model experimental organism from which the majority of our knowledge of bacterial transcription has been derived. In this section, we examine the four stages of transcription in bacteria: (1) promoter recognition and identification, (2) the initiation of transcript synthesis, (3) transcript elongation, and (4) transcription termination.

Like all RNA polymerases, bacterial RNA polymerase uses one strand of DNA, the **template strand,** to assemble the transcript by complementary and antiparallel base pairing of RNA nucleotides with DNA nucleotides of the template strand (see Figure 1.9 for a review). The **coding strand** of DNA, also known as the **nontemplate strand,** is complementary to the template strand. The gene—that is, the stretch of DNA regions that produces an RNA transcript—contains several segments with distinct functions (Figure 8.3). The **promoter** of the gene is immediately **upstream**—that is, immediately 5′ to the start of transcription, which is identified as corresponding to the +1 nucleotide. The promoter is not transcribed. Instead, the promoter sequence is a transcription-regulating DNA

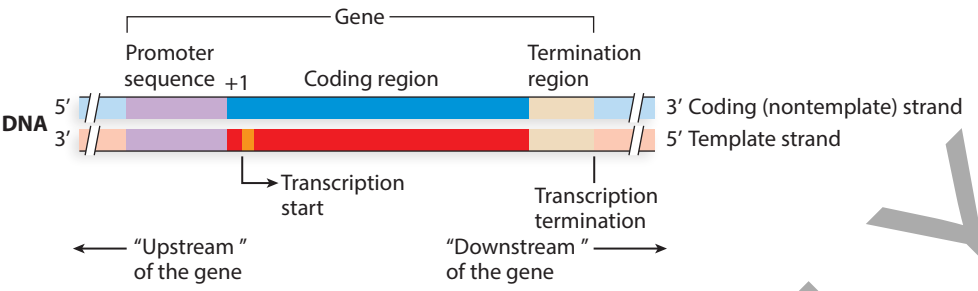


Figure 8.3 **A general diagram of gene structure and associated nomenclature.**

sequence that controls the access of RNA polymerase to the gene. The **coding region** is the portion of the gene that is transcribed into mRNA and contains the information needed to synthesize the protein product of the gene. The **termination region** is the portion of the gene that regulates the cessation of transcription. The termination region is located immediately **downstream**—that is, immediately 3′ to the coding segment of the gene.

Bacterial RNA Polymerase

A single type of *E. coli* RNA polymerase catalyzes transcription of all RNAs. The initial experimental evidence supporting this conclusion came from analysis of the effect of the antibiotic rifampicin on bacterial RNA synthesis. Rifampicin inhibits RNA synthesis by preventing RNA polymerase from catalyzing the formation of the first phosphodiester bond in the RNA chain. In rifampicin-sensitive (rif^S) bacterial strains, synthesis of all three major types of RNA (mRNA, tRNA, and rRNA) is inhibited in the presence of rifampicin. In contrast, rifampicin-resistant (rif^R) bacteria actively transcribe DNA into the three major RNAs when rifampicin is present. Molecular analysis identifies a single mutation of RNA polymerase in rif^R strains that allows it to remain catalytically active when exposed to rifampicin, and subsequent molecular studies have confirmed the presence of a single bacterial RNA polymerase.

Bacterial RNA polymerase is composed of a pentameric (five-polypeptide) **RNA polymerase core** that binds to a sixth polypeptide, called the **sigma subunit (σ),** which induces a conformational change in the core enzyme that switches it to its active form. In its active form, the RNA polymerase is described as a holoenzyme, a term meaning an intact complex of multiple subunits, with full enzymatic capacity. Figure 8.4 shows a common type of sigma subunit known as σ^{70}, but there are also other sigma subunits in *E. coli.*

The RNA polymerase core consists of two α subunits, designated αI and αII, two β subunits, and an ω (omega) subunit. The molecular weight of the five-subunit core RNA polymerase is approximately 390 kD (kiloDaltons), and with the sigma subunit added, the holoenzyme has a molecular weight of 430 kD. Each of these subunits have been evolutionarily conserved in archaea and in eukaryotes, as we discuss in the following section.

By itself, the core RNA polymerase can transcribe DNA template-strand sequence into RNA sequence, but the core is unable to efficiently bind to a promoter or initiate RNA synthesis without a sigma subunit. The joining of the sigma subunit to the core enzyme to form a holoenzyme induces a conformational shift in the core segment that enables it to bind specifically to particular promoter consensus sequences. The addition of the sigma subunit to the core RNA polymerase, with its five subunits and approximately 390-kD molecular weight, produces a holoenzyme having a molecular weight of approximately 430 kD. Each of the subunits has been evolutionarily conserved in archaea and in eukaryotes, as we discuss in a following section.

This single RNA polymerase is responsible for all bacterial transcription. Thus, the bacterial RNA polymerase must recognize promoters for protein-coding genes as well as for genes that produce functional RNAs, such as tRNA and rRNA. However, not all promoters of bacterial genes are identical. There is great diversity among bacterial promoter sequences, permitting certain genes to be expressed only under special circumstances. Bacteria manage the recognition of the promoters of these specialized genes by producing several different types of sigma subunits that can join the core polymerase. These so-called **alternative sigma subunits** alter the specificity of the holoenzymes for promoter regions by imparting

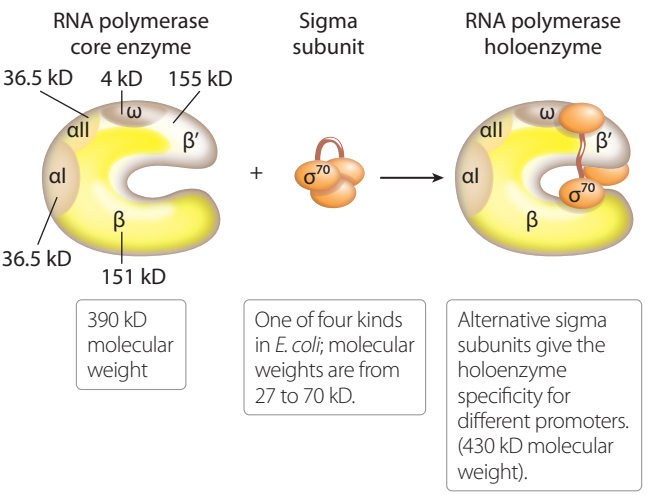


Figure 8.4 **Bacterial RNA polymerase core plus a sigma (σ) subunit forms the fully active holoenzyme.**

distinct conformational changes to the core. These differences enable transcription of specific genes under the appropriate conditions, or at the correct time.

Bacterial Promoters

A promoter is a double-stranded DNA sequence that is the binding site for RNA polymerase. Promoters are regulatory DNA sequences that bind transcription proteins, and their presence usually indicates that a gene is nearby. Bacterial promoters are located a short distance upstream of the coding sequence, typically within a few nucleotides of the start of transcription, represented by the +1 nucleotide. RNA polymerase is attracted to promoters by the presence of **consensus sequences,** short regions of DNA sequences that are highly similar, though not necessarily identical, to one another and are located in the same position relative to the start of transcription of different genes.

Although promoters are double stranded, promoter consensus sequences are usually written in a single-stranded shorthand form that gives the 5′-to-3′ sequence of the coding (non-template) strand of DNA (**Figure 8.5**). The most commonly occurring bacterial promoter contains two consensus sequence regions that each play an important functional role in recognition by RNA polymerase and the subsequent initiation of transcription. These consensus sequences are located upstream from the +1 nucleotide (the start of transcription) in a region flanking the gene where the nucleotides are denoted by negative numbers and are not transcribed. At the −10 position of the *E. coli* promoter is the **Pribnow box sequence,** or the **−10 consensus sequence,** consisting of 6 bp having the consensus sequence `5′-TATAAT-3′`. The Pribnow box is separated by about 25 bp from another 6-bp region, the **−35 consensus sequence,** identified by the nucleotides `5′-TTGACA-3′`. The nucleotide sequences that occur upstream, downstream, and between these consensus sequences are highly variable and contain no other consensus sequences. Thus, in a functional sense, the −10 (Pribnow) and −35 consensus sequences are important because of their nucleotide content, their location relative to one another, and their location relative to the start of transcription. In contrast to the consensus sequences themselves, the nucleotides between −10 and −35 are important as spacers between the consensus elements, but their specific sequences are not critical.

Natural selection has operated to retain strong sequence similarity in consensus regions and to retain the position of the consensus regions relative to the start of transcription. The effectiveness of evolution in maintaining promoter consensus sequences is illustrated by comparison with the sequences between and around −10 and −35, which are not conserved and which exhibit considerable variation. In addition, the spacing between the sequences and their placement relative to the +1 nucleotide is stable. RNA polymerase is a large molecule that binds to −10 and −35 consensus sequences and occupies the space between and immediately around the sites. Crystal structure models show that the enzyme spans enough DNA to allow it to contact promoter consensus regions and reach the +1 nucleotide. Once bound at a promoter in this fashion, RNA polymerase can initiate transcription. **Genetic Analysis 8.1** guides you through the identification of promoter consensus regions.

Transcription Initiation

RNA polymerase holoenzyme initiates transcription through a process involving two steps. In the first step, the holoenzyme makes an initial loose attachment to the double-stranded promoter sequence and then binds tightly to it to form the **closed promoter complex** (❶ in **Foundation Figure 8.6**). In the second step, the bound holoenzyme unwinds approximately 18 bp of DNA around the −10 consensus sequence to form the **open promoter complex** (❷). Following formation of the open promoter complex, the holoenzyme progresses downstream to initiate RNA synthesis at the +1 nucleotide on the template strand of DNA (❸).

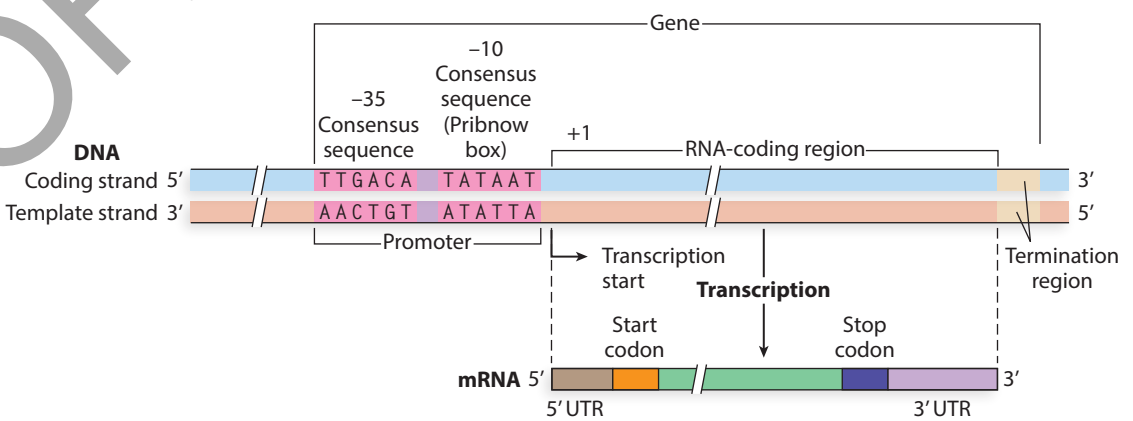
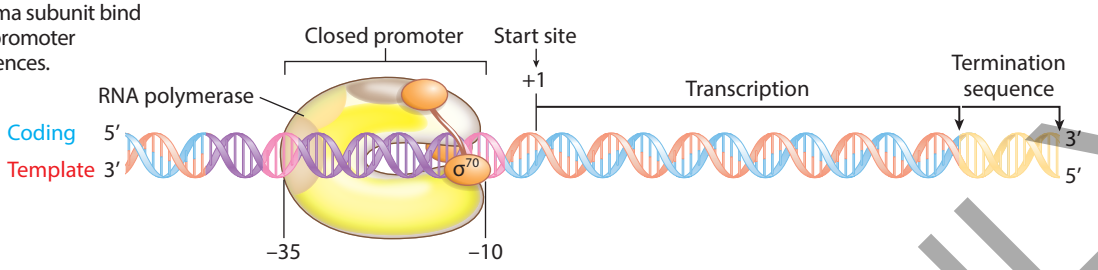


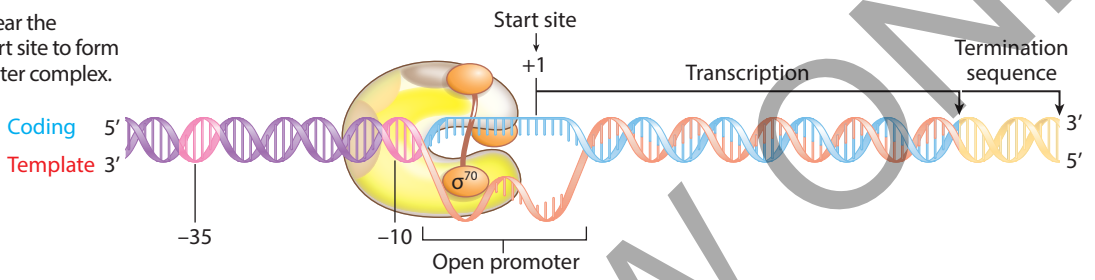
Figure 8.5 **Bacterial promoter structure.** Two promoter consensus sequences—the Pribnow box at −10 and the −35 sequence—are essential promoter regulatory elements.

Bacterial Transcription

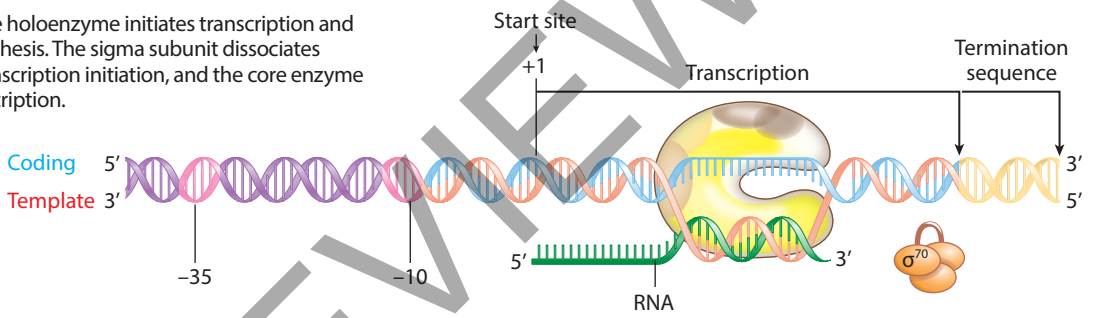
1. The RNA polymerase core enzyme and sigma subunit bind to –10 and –35 promoter consensus sequences.



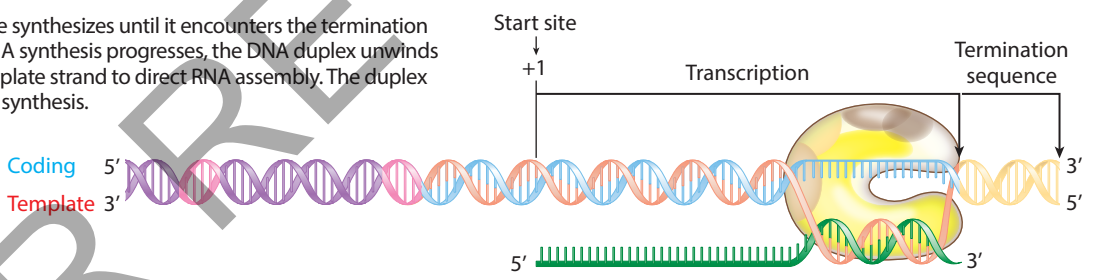
2. DNA unwinds near the transcription start site to form the open promoter complex.



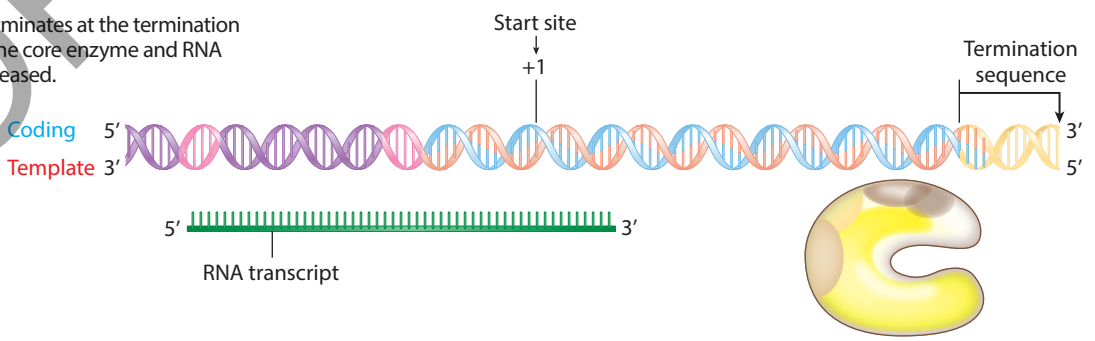
3. RNA polymerase holoenzyme initiates transcription and begins RNA synthesis. The sigma subunit dissociates shortly after transcription initiation, and the core enzyme continues transcription.



4. The core enzyme synthesizes until it encounters the termination sequence. As RNA synthesis progresses, the DNA duplex unwinds to allow the template strand to direct RNA assembly. The duplex closes following synthesis.



5. Transcription terminates at the termination sequence, and the core enzyme and RNA transcript are released.



PROBLEM DNA sequences in the promoter region of 10 *E. coli* genes are shown. Sequences at the −35 and −10 sites are boxed.

a. Use the sequence information provided to deduce the −35 and −10 consensus sequences.

b. Speculate on the relative effects on transcription of a mutation in a promoter consensus region versus a mutation in the sequence between consensus regions.

BREAK IT DOWN: Promoter consensus sequences are similar in different genes and bind transcriptionally active proteins (p. 273).

BREAK IT DOWN: Research methods directed at detecting promoters and assessing their functionality are described in Research Technique 8.1 (pp. 279–280) and Figure 8.11 (p. 282).

Gene	−35 region						−10 region						+1
A2	A	A	T	G	C	T	T	G	A	C	T	C	T	G	T	A	G	C	G	G	G	A	A	G	G	C	G	−	−	T	A	T	A	A	T	G	C	A	C	A	C	C	−	C	C	G	C
bio	A	A	A	A	C	G	T	G	T	T	T	T	T	G	T	T	G	T	T	A	A	T	T	C	G	G	T	G	T	A	G	A	C	T	T	G	T	−	−	−	A	A	A	C	C	T			
his	A	G	T	T	C	T	T	G	C	T	T	T	C	T	A	A	C	G	T	G	A	A	A	G	T	G	G	T	T	T	A	G	G	T	T	A	A	A	A	G	A	C	−	A	T	C	A			
lac	C	A	G	G	C	T	T	T	A	C	A	C	T	T	T	A	T	G	C	T	T	C	C	G	G	C	T	C	G	T	A	T	G	T	T	G	T	G	−	T	G	G	−	A	A	T	T			
lacI	G	A	A	T	G	G	C	G	C	A	A	A	A	C	T	T	T	T	C	G	C	G	G	T	A	T	G	G	−	C	A	T	G	A	T	A	G	C	G	C	C	C	−	G	G	A	A			
leu	A	A	A	A	G	T	T	G	A	C	A	T	C	C	G	T	T	T	T	T	G	T	A	T	C	C	A	G	−	T	A	A	C	T	C	T	A	A	A	A	G	C	−	A	T	A	T			
recA	A	A	C	A	C	T	T	G	A	T	A	C	T	G	T	A	T	G	A	G	C	A	T	A	C	A	G	−	−	T	A	T	A	A	T	T	G	C	T	T	C	−	−	A	A	C	A			
trp	A	G	C	T	G	T	T	G	A	C	A	A	T	T	A	A	T	C	A	T	C	G	A	A	C	T	A	G	−	T	T	A	A	C	T	A	G	T	A	C	G	C	−	A	A	G	T			
tRNA	A	A	C	A	C	T	T	T	A	C	A	G	C	G	G	G	C	C	G	T	C	A	T	T	T	G	A	−	−	T	A	T	G	A	T	G	C	G	C	C	C	C	−	G	C	T	T			
X1	T	C	C	G	C	T	T	G	T	C	T	T	C	C	T	A	G	G	C	C	G	A	C	T	C	C	C	−	−	T	A	T	A	A	T	G	C	G	C	C	T	C	C	A	T	C	G			

Solution Strategies

Evaluate

1. Identify the topic this problem addresses and the nature of the required answer.
2. Identify the critical information provided in the problem.

Deduce

3. Examine the −10 and −35 sequences of these promoters, and look for common patterns.

TIP: A consensus sequence identifies the most common nucleotide at each position in a DNA segment.

Solve

4. Determine the consensus sequence at the −10 and −35 regions.

TIP: Identify the most commonly occurring nucleotide in each position of the 6-nucleotide consensus region of these genes.

5. Compare and contrast the likely effects of consensus sequence mutations with those of mutations occurring between consensus regions.

Solution Steps

1. This question concerns bacterial promoters. The answer requires identification of consensus sequences for −35 and −10 regions of promoters and speculation about the consequences of promoter mutations.
2. The problem provides promoter sequence information for 10 *E. coli* genes and identifies the segment of each promoter containing the −10 and −35 regions.
3. The −10 and −35 sites are the location of RNA polymerase binding during transcription initiation. Count the numbers of A, T, C, and G in each position in the boxed regions.

Answer a

4. At the −10 site, and moving left to right (toward +1), the most common nucleotides in each position in the consensus region, and the number of times they occur in that position, are

T	A	T	A	A	T
(9)	(9)	(6)	(5)	(5)	(9)

At the −35 site, also moving left to right (toward the +1), the most common nucleotides in each position, and the number of times they occur in that position, are

T	T	G	A	C	A
(8)	(9)	(8)	(6)	(6)	(6)

Answer b

5. Mutation in a consensus sequence is likely to alter the efficiency with which a protein binds to the promoter and to decrease the amount of gene transcription. In contrast, mutations between consensus sequences are unlikely to alter gene transcription because the sequences in these intervening regions do not bind tightly to RNA polymerase.

For more practice, see Problems 4, 7, and 18.

Table 8.2 *Escherichia coli* RNA Polymerase Sigma Subunits

Subunit	Molecular Weight (Daltons)	Consensus Sequence		Function
		−35	−10	
σ^{28}	28	TAAA	GCCGATAA	Flagellar synthesis and chemotaxis
σ^{32}	32	CTTGAA	CCCCATTA	Heat shock genes
σ^{54}	54	CTGGPyAPyPu	TTGCA	Nitrogen metabolism
σ^{70}	70	TTGACA	TATAAT	Housekeeping genes

Bacterial promoters often differ from the consensus sequence by one or more nucleotides, and some are different at several nucleotides. Since considerable DNA-sequence variation occurs among promoters, it is reasonable to ask how RNA polymerase is able to recognize promoters and reliably initiate RNA synthesis. For an answer, we turn to the sigma subunits that confer promoter recognition and chain-initiation ability on RNA polymerase.

Four alternative sigma subunits identified in *E. coli* are named according to their molecular weight (Table 8.2). Each alternative sigma subunit leads to recognition of a different set of −10 and −35 consensus sequences by the holoenzyme. These different consensus sequence elements are found in promoters of different types of genes; thus, the sigma subunit that it becomes attached to determines the specific gene promoters a holoenzyme will recognize.

The sigma subunit σ^{70} is the most common in bacteria. It recognizes promoters of "housekeeping genes," the genes whose protein products are continuously needed by cells. Because of the constant need for their products, housekeeping genes are continuously expressed. Subunits σ^{54} and σ^{32} recognize promoters of genes involved in nitrogen metabolism and genes expressed in response to environmental stress such as heat shock and are utilized when the action of these genes is required. The fourth sigma subunit, σ^{28}, recognizes promoters for genes required for bacterial chemotaxis (chemical sensing and motility).

The specificity of each type of sigma subunit for different promoter consensus sequences produces RNA polymerase holoenzymes that have different DNA-binding specificities. Microbial geneticists estimate that each *E. coli* cell contains about 3000 RNA polymerase holoenzymes at any given time and that each of the four kinds of sigma subunits is represented to a differing degree among them. Because sigma subunits readily attach and detach from core enzymes in response to changes in environmental conditions, the organism is able to change its transcription patterns to adjust to different conditions.

Transcription Elongation and Termination

Upon reaching the +1 nucleotide, the holoenzyme begins RNA synthesis by using the template strand to direct RNA assembly. The holoenzyme remains intact until the first 8 to 10 RNA nucleotides have been joined. At that point, the sigma subunit dissociates from the core enzyme, which continues its downstream progression (❸ in Foundation Figure 8.6). The sigma subunit itself remains intact and can associate with another core enzyme to transcribe another gene.

Downstream progression of the RNA polymerase core is accompanied by DNA unwinding ahead of the enzyme to maintain approximately 18 bp of unwound DNA (❹). As the RNA polymerase passes, progressing at a rate of approximately 40 nucleotides per second, the DNA double helix reforms in its wake. When transcription of the gene is completed, the 5′ end of the RNA trails off the core enzyme (❺).

The end product of transcription is a single-stranded RNA that is complementary and antiparallel to the template DNA strand. The transcript has the same 5′-to-3′ polarity as the coding strand of DNA, the strand complementary to the template strand. The coding strand and the newly formed transcript also have identical nucleotide sequences, except for the presence of uracil in the transcript in place of thymine in the coding strand. For this reason, gene sequences are written in 5′-to-3′ orientation as single-stranded sequences based on the coding strand of DNA. This allows easy identification of the mRNA sequence of a gene by simply substituting U for T.

Gene transcription is not a one-time event, and shortly after one round of transcription is initiated, a second round begins with new RNA polymerase–promoter interaction. Following sigma subunit dissociation and core enzyme synthesis of 50 to 60 RNA nucleotides, a new holoenzyme can bind to the promoter and initiate a new round of transcription while the first core enzyme continues along the gene. In addition, if the transcript under construction is mRNA, the 5′ end is immediately available to begin translation. In contrast, transcripts that are functional RNAs, such as transfer and ribosomal RNA, must await the completion of transcription before undergoing the folding into secondary structures that readies them for cellular action.

Transcription Termination Mechanisms

Termination of transcription in bacterial cells is signaled by a DNA termination sequence that usually contains a repeating sequence producing distinctive 3′ RNA sequences.

Termination sequences are downstream of the stop codon; thus, they are transcribed after the coding region of the mRNA and so are not translated. Two transcription termination mechanisms occur in bacteria. The most common is **intrinsic termination,** a mechanism dependent only on the occurrence of specialized repeat sequences in DNA that induce the formation in RNA of a secondary structure leading to transcription termination. Less frequently, bacterial gene transcription terminates by **rho-dependent termination,** a mechanism characterized by a different terminator sequence and requiring the action of a specialized protein called the **rho protein.**

Intrinsic Termination Most bacterial transcription termination occurs exclusively as a consequence of termination sequences encoded in DNA—that is, by intrinsic termination. Intrinsic termination sequences have two features. First, they are encoded by a DNA sequence containing an **inverted repeat,** a DNA sequence repeated in opposite directions but with the same 5′-to-3′ polarity. Figure 8.7 shows the inverted repeats ("repeat 1" and "repeat 2") in a termination sequence, separated by a short spacer sequence that is not part of either repeat. The second feature of intrinsic termination sequences is a string of adenines on the template DNA strand that begins at the 5′ end of the repeat 2 region. Transcription of inverted repeats produces mRNA with complementary segments that are able to fold into a short double-stranded stem ending with a single-stranded loop. This secondary structure is a **stem-loop structure,** also known as a **hairpin.** A string of uracils complementary to the adenines on the template strand immediately follows the stem-loop structure at the 3′ end of the RNA.

The formation of a stem-loop structure followed immediately by a poly-U sequence near the 3′ end of RNA causes the RNA polymerase to slow down and destabilize. In addition, the 3′ U-A region of the RNA–DNA duplex contains the least stable of the complementary base pairs. Together, the instability created by RNA polymerase slowing and the U-A base pairs induces RNA polymerase to release the transcript and separate from the DNA. The behavior of RNA polymerase during intrinsic termination of transcription is like that of a bicycle rider at slow speed. Slow forward momentum creates instability and eventually the rider loses balance. In a similar way, RNA polymerase is destabilized as it slows while transcribing inverted repeat sequences, and it falls off DNA when the transcript is released where A-U base pairs form.

Rho-Dependent Termination In contrast to the more common intrinsic termination, certain bacterial genes require the action of rho protein to bind to nascent mRNA and catalyze separation of mRNA from RNA polymerase to terminate transcription. Genes whose transcription is rho-dependent have termination sequences that are

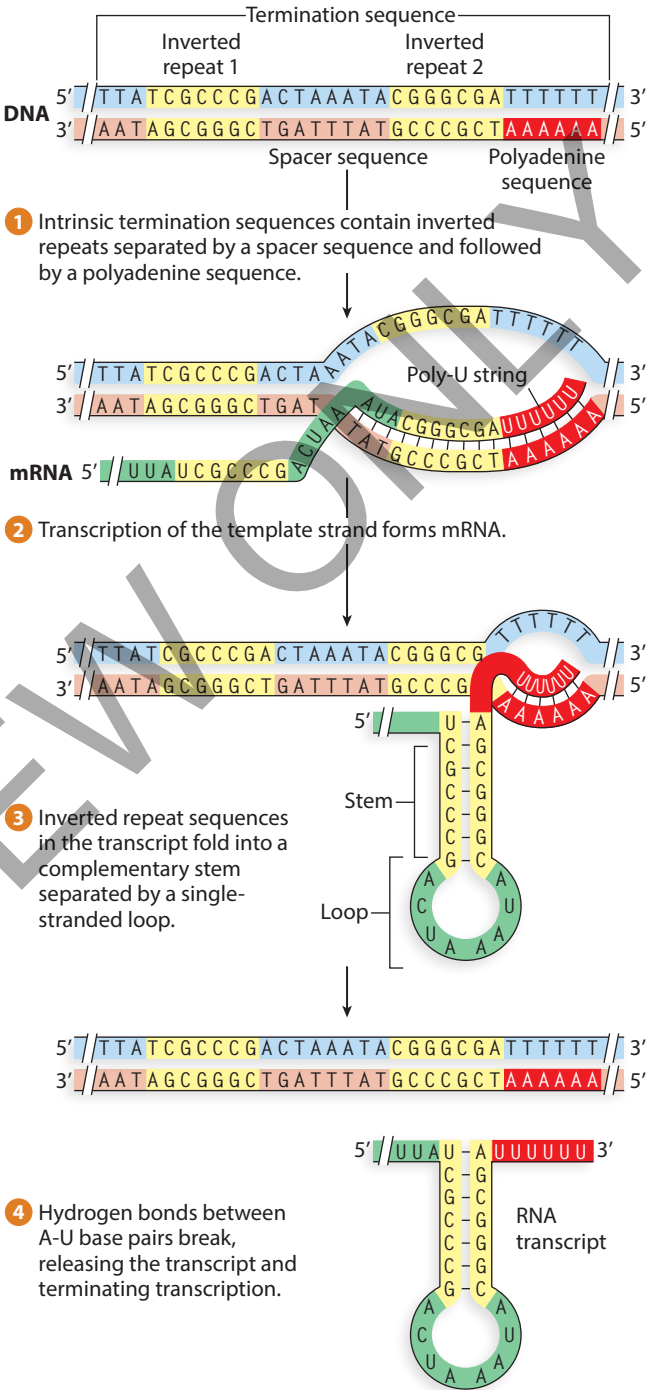


Figure 8.7 **Intrinsic termination of transcription is driven by the presence of inverted repeat DNA sequences.**

distinct from those in genes utilizing intrinsic termination. Stem-loop structures often form as part of rho-dependent termination, but rho-dependent terminator sequences do not have a string of uracil residues. Instead, the sequences contain a **rho utilization site,** or **rut site,** which is a stretch of approximately 50 nucleotides that is rich in cytosine and poor in guanine.

Rho protein is composed of six identical polypeptides and has two functional domains, both of which are

utilized during the two-step process of transcription termination. The first step is initiated when rho protein is activated by an ATP molecule that binds to one functional domain of rho. Activated rho protein utilizes its second domain to bind to the rut site of the RNA transcript. Using ATP-derived energy, rho then moves along the mRNA in the 3′ direction, eventually catching up to RNA polymerase that has slowed near a terminator sequence. As the rho travels, it catalyzes the breakage of hydrogen bonds between mRNA and the DNA template strand. The bond breakage releases the transcript from the RNA polymerase and induces the polymerase to release the DNA.

8.3 Archaeal and Eukaryotic Transcription Displays Structural Homology and Common Ancestry

Bacteria use a single RNA polymerase core enzyme and several alternative sigma subunits to transcribe all genes. Similarly, archaea have a single type of RNA polymerase. Eukaryotes, by contrast, each have multiple RNA polymerases that are specialized for the transcription of different genes. The archaeal and eukaryotic RNA polymerases responsible for the transcription of most polypeptide-producing genes share a common structure that is divergent from the bacterial RNA polymerase. Transcription in archaea and eukaryotes progresses through the same four stages we described for bacteria: promoter recognition, transcription initiation, transcript elongation, and transcription termination. Several structural and functional factors make transcription more complex in archaea and eukaryotes. First, eukaryotic promoters and consensus sequences are considerably more diverse than in *E. coli,* and eukaryotes have three different RNA polymerases that recognize different promoters, transcribe different genes, and produce different RNAs. Promoter consensus sequences in archaea are considerably less complicated than those in eukaryotes, but they appear to be more diverse than bacterial promoter sequences. Second, the molecular apparatus assembled at promoters to initiate and elongate transcription is more complex in eukaryotes and in archaea. Third, eukaryotic genes contain introns and exons, requiring extensive post-transcriptional processing of mRNA. Archaeal genes generally do not contain introns, although there is RNA splicing of archaeal pre-tRNAs in a similar manner to eukaryotic pre-tRNA splicing. We describe these details in a later section. Finally, eukaryotic DNA is permanently associated with a large amount of protein to form a compound known as *chromatin*.

Chromatin plays a central role in regulating eukaryotic transcription. Chromatin structure is a permanent feature and a dynamic feature of eukaryotic genomes. Its state controls the accessibility of DNA to transcription, either permitting or blocking RNA polymerase and transcription factor access to promoters. In later chapters, we discuss chromatin structure (Chapter 11) and explore the functional role of chromatin in the regulation of gene expression in eukaryotes (Chapter 15).

Eukaryotic and Archaeal RNA Polymerases

Three different RNA polymerases transcribe distinct classes of RNA coded by eukaryotic genomes: **RNA polymerase I (RNA pol I)** transcribes three ribosomal RNA genes, **RNA polymerase II (RNA pol II)** is responsible for transcribing messenger RNAs that encode polypeptides as well as for transcribing most small nuclear RNA genes, and **RNA polymerase III (RNA pol III)** transcribes all transfer RNA genes as well as one small nuclear RNA gene and one ribosomal RNA gene. RNA pol II and RNA pol III are responsible for miRNA and siRNA synthesis.

The RNA polymerases of members of all three domains of life share similarities of sequence and function. The *E. coli* RNA polymerase core enzyme has five units. Each of these subunits has a homolog in the 10 to 13 subunit (depending on the species) archaeal RNA polymerase and in the 10 to 12 subunit (depending on the species) eukaryotic RNA polymerase II (**Table 8.3**).

Despite differences in sizes and molecular complexity, the RNA polymerases have a similar overall structure, forming a characteristic shape one reminiscent of DNA polymerase (see Figure 7.23), with a "hand" composed of protein "fingers" to help RNA polymerase grasp DNA, and a "palm" in which polymerization takes place. These similarities of RNA polymerase structure and function are a direct result of the shared evolutionary history of bacteria, archaea, and eukaryotes.

Table 8.3 RNA Polymerase Composition

Bacteria	Archaea	Eukarya
Escherichia coli 5 subunits	*Sulfolobus solfataricus* 10 subunits	*Saccaromyces cerevisiae* (RNA pol II) 12 subunits
Homologous proteins:		
β′	RpoA′/A″	Rpb1
β	RpoB	Rpb2
αI	RpoD	Rpb3
ω	RpoK	Rpb6
αII	RpoL	Rpb11
Additional proteins:		
	RpoE, RpoF, RpoH,	Rpb4, Rpb5, Rpb7, Rpb8,
	RpoN and RpoP	Rpb9, Rpb10, Rpb12

Consensus Sequences for Eukaryotic RNA Polymerase II Transcription

RNA polymerase II transcribes eukaryotic polypeptide-coding genes into mRNA. The promoters for these genes are numerous and highly diverse, with different overall lengths and differences in the number and type of consensus sequences prominent among the sources of promoter variation. Given these characteristics, it is reasonable to ask how RNA polymerases locate promoter DNA for different genes.

Three lines of investigation help researchers to identify and characterize promoters of different polypeptide-coding genes: (1) promoters are identified by determining which DNA sequences are bound by proteins associated with RNA pol II during transcription, (2) putative promoter sequences from different genes are compared to evaluate their similarities, and (3) mutations that alter gene transcription are examined to identify how DNA base-pair changes affect transcription. **Research Technique 8.1** discusses the experimental identification and analysis of promoters.

Research Technique 8.1

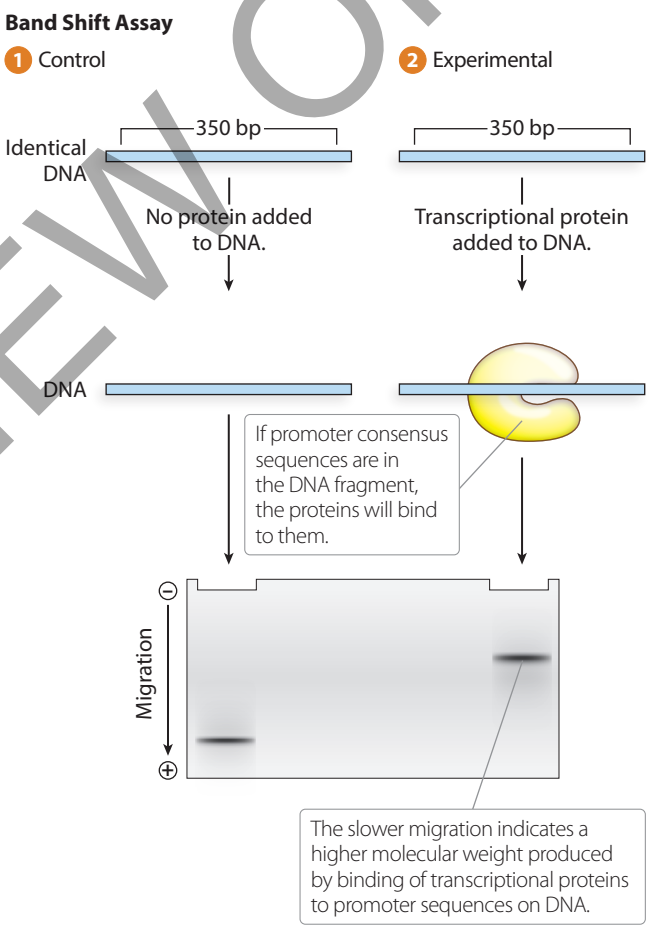
Band Shift Assay to Identify Promoters

PURPOSE The functional action of promoters in transcription depends on consensus DNA sequences that bind RNA polymerase and transcription factor proteins. To locate promoters, molecular biologists first scan DNA for potential promoter consensus sequences and then determine that the sequence binds transcriptionally active proteins.

MATERIALS AND PROCEDURES Fragments of DNA containing suspected promoter consensus sequence are examined by two experimental methods. The first, called *band shift assay,* verifies that the sequence of interest binds proteins. The second, called *DNA footprint protection assay,* identifies the exact location of the protein-binding sequence.

In band shift assay, two identical samples of DNA fragments that contain suspected consensus sequence are analyzed. One DNA sample is a control to which no transcriptional proteins are added. The experimental DNA sample, on the other hand, has transcriptional proteins added. Both the control and the experimental DNA samples are subjected to electrophoresis.

DNA footprint protection also begins with two identical samples of DNA fragments containing suspected consensus sequences. All fragments are end-labeled with ^{32}P. The experimental DNA is mixed with transcriptional proteins, but the control sample is not. Both samples are exposed to DNase I that randomly cuts DNA that is not protected by protein. The samples are run in separate lanes of an electrophoresis gel, and each end-labeled fragment produced is identified by autoradiography.

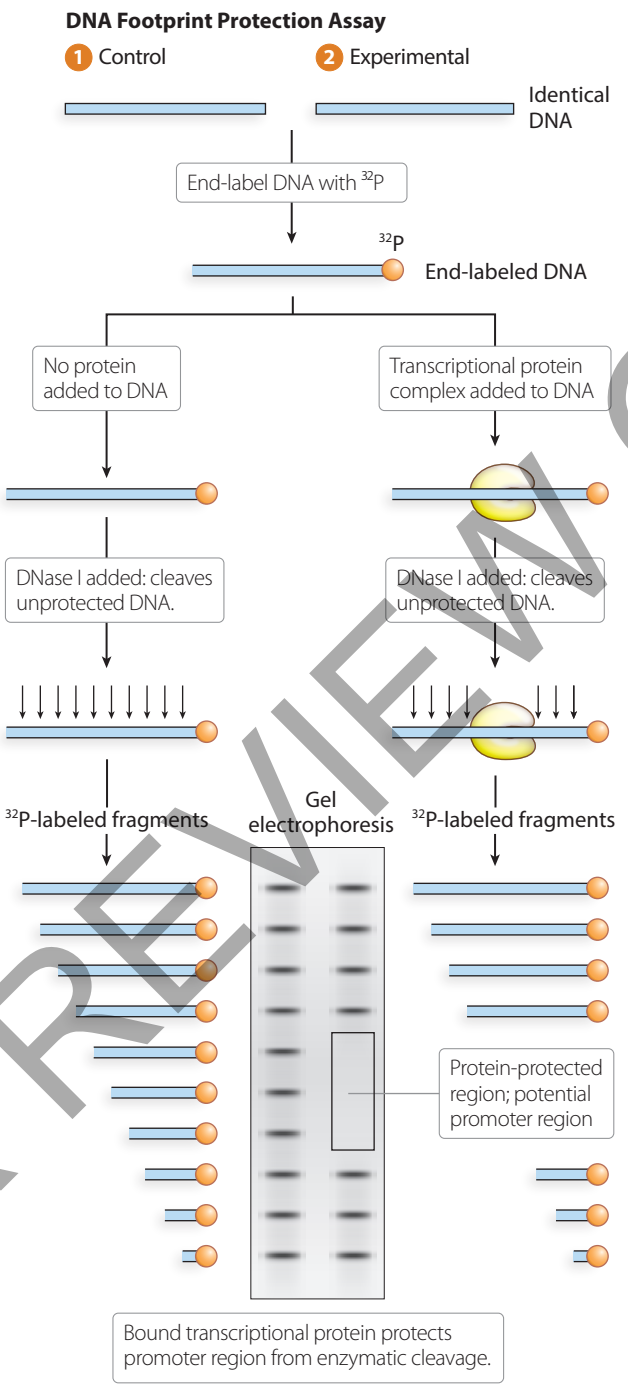


DESCRIPTION In the band shift assay result, notice that the electrophoretic mobility of experimental DNA is slower than that of control DNA. This is the anticipated result if the experimental sample contains consensus sequence that is bound by transcriptional proteins. The bound protein increases the molecular weight of the experimental sample and slows its migration relative to the same DNA without bound protein. In the DNA footprint protection assay, notice that the experimental DNA lane contains a gap in which no DNA fragments appear. The gap represents "footprint protection" for the portion of the fragment that is protected from DNase I digestion by bound transcriptional proteins. No such protection occurs for the control fragment that is randomly cleaved.

CONCLUSION Evidence from these two methods constitutes necessary but not sufficient evidence that the DNA fragment contains a promoter. The final piece of evidence that a DNA fragment contains a promoter rests on mutational analysis that identifies functional changes caused by mutations of specific nucleotides of promoter consensus sequences (see Figure 8.11).

(continued)

Research Technique 8.1 Continued



The most common eukaryotic promoter consensus sequence, the *TATA box*, is shown in Figure 8.8 as part of a set of three consensus segments that were the first eukaryotic promoter elements to be identified. A **TATA box,** also known as a **Goldberg-Hogness box,** is located approximately at position −25 relative to the beginning of the transcriptional start site. Consisting of 6 bp with the consensus sequence TATAAA, it is the most strongly conserved promoter element in eukaryotes. The figure shows two additional consensus sequence elements that are more variable in their frequency in promoters. A 4-bp consensus sequence identified as the **CAAT box** is most commonly located near −80 when it is present in the promoter. An upstream GC-rich region called the **GC-rich box,** with

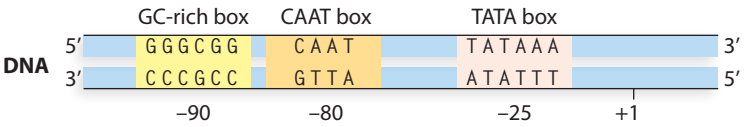


Figure 8.8 Three eukaryotic promoter consensus sequence elements. The TATA box and the CAAT box are common; the presence of the upstream GC-rich box is more variable.

a consensus sequence GGGCGG located –90 or more upstream of the transcription start, has a frequency that is less than that of CAAT box sequences.

Comparison of eukaryotic promoters reveals a high degree of variability in the type, number, and location of consensus sequence elements (Figure 8.9). Some promoters contain all three of the consensus sequences identified above, others contain one or two of these consensus elements, some contain none at all, and many contain other types of consensus sequence elements altogether. For example, the thymidine kinase gene contains TATA, CAAT, and GC-rich boxes along with an octamer (OCT) sequence, called an OCT box. The histone *H2B* gene contains two OCT boxes in addition to a TATA box and a pair of CAAT boxes. All of these consensus sequence elements play important roles in the binding of *transcription factors*, a group of transcriptional proteins described below.

Promoter Recognition

RNA polymerase II recognizes and binds to promoter consensus sequences in eukaryotes with the aid of proteins called **transcription factors (TF).** The TF proteins bind to promoter regulatory sequences and influence transcription initiation by interacting, directly or indirectly, with RNA polymerase. Transcription factors that influence mRNA transcription, and therefore interact with RNA pol II, are given the designation TFII. Individual TFII proteins also carry a letter designation, such as A, B, or C.

In most eukaryotic promoters, the TATA box is the principal binding site for transcription factors during promoter recognition. At the TATA box, a protein called TFIID, a multisubunit protein containing **TATA-binding protein (TBP)** and subunits of a protein called **TBP-associated factor (TAF),** binds the TATA box sequence. The assembled TFIID binds to the TATA box region to form the **initial committed complex** (Figure 8.10). Next, TFIIA, TFIIB, TFIIF, and RNA polymerase II join the

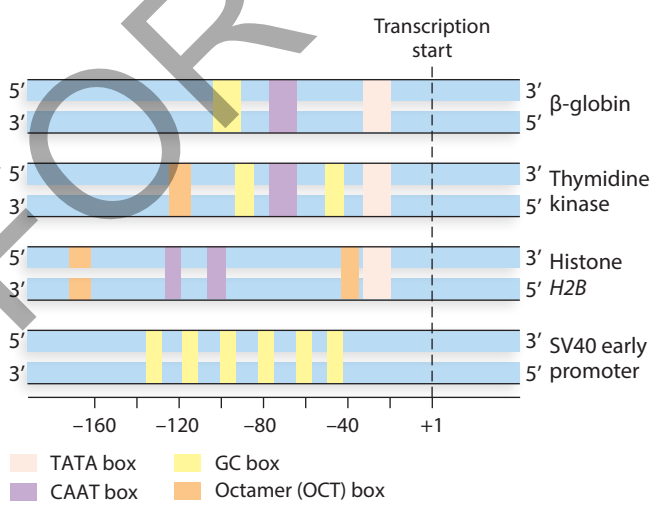


Figure 8.9 Examples of eukaryotic promoter variability.

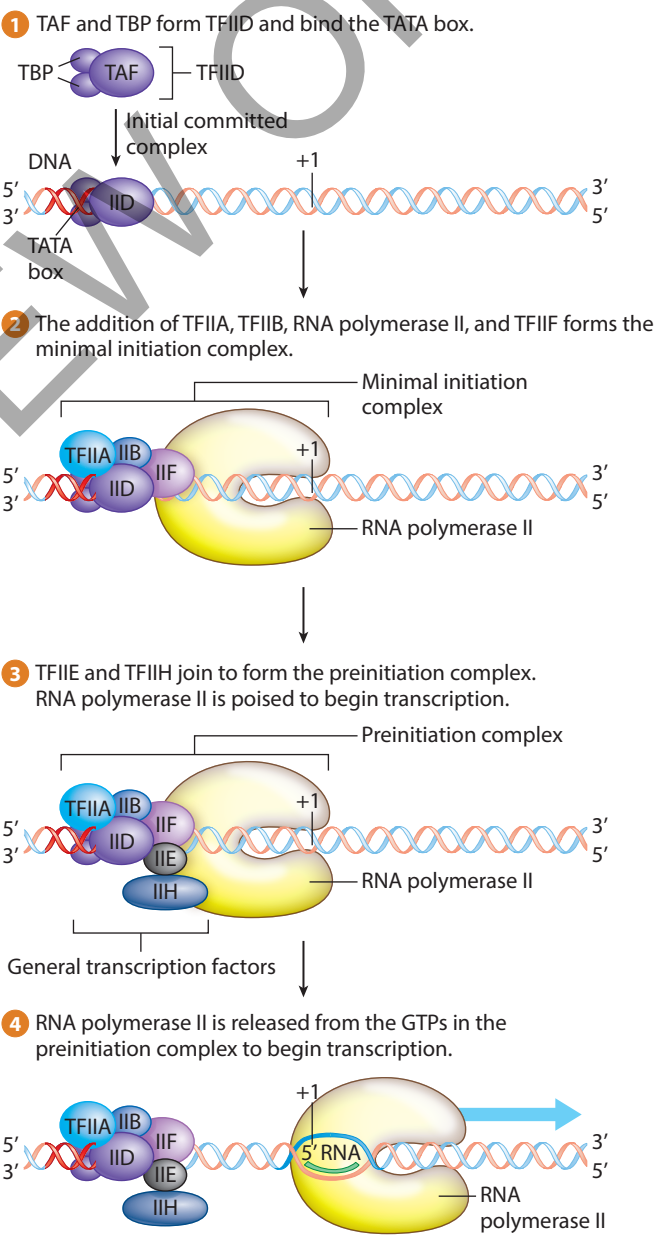


Figure 8.10 Six general transcription factor proteins bind the promoter region to set the stage for eukaryotic transcription by RNA polymerase II.

initial committed complex to form the **minimal initiation complex**, which in turn is joined by TFIIE and TFIIH to form the **preinitiation complex (PIC)**. The complete **initiation complex** contains six proteins that are commonly identified as **general transcription factors (GTFs)**. Once assembled, the complete initiation complex directs RNA polymerase II to the +1 nucleotide on the template strand, where it begins the assembly of messenger RNA.

While most of the eukaryotic genes that have been examined have a TATA box and undergo TBP binding, there is evidence that some metazoan genes may use a related factor called TLF (*T*BP-*l*ike *f*actor). The complexity of TBP, TLF, and associated proteins is analogous to the different sigma factors in prokaryotic systems, thus allowing differential recognition of promoters in eukaryotes.

Detecting Promoter Consensus Elements

The diversity of eukaryotic promoters begs an important question: How do researchers verify that a segment of DNA is a functionally important component of a promoter? The research has two components; the first, outlined in Research Technique 8.1, is discovering the presence and location of DNA sequences that transcription factor proteins will bind to. The second component involves mutational analysis to confirm the functionality of the sequence. Researchers produce many different point mutations in the DNA sequence under study and then compare the level of transcription generated by each mutant promoter sequence with transcription generated by the wild-type sequence.

Figure 8.11 shows a synopsis of promoter mutation analysis from an experiment performed by the molecular biologist Richard Myers and colleagues on a mammalian β-globin gene promoter. These researchers produced mutations of individual base pairs in TATA box, CAAT box, and GC-rich sequences, and of nucleotides between the consensus sequences, to identify the effect of each individual mutation on the relative transcription level of the gene. They found that most base-pair mutations in each of the three consensus regions significantly decreased the transcription level of the gene and found two base substitutions in the CAAT box region that significantly increased transcription. In contrast, mutations outside the consensus regions had nonsignificant effects on transcription level. Such results show the functional importance of specific DNA sequences in promoting transcription and confirm a functional role in transcription for TATA box, CAAT box, and GC-rich sequences.

Enhancers and Silencers

Promoters alone are often not sufficient to initiate transcription of eukaryotic genes, and other regulatory sequences are needed to drive transcription. This is particularly the case for multicellular eukaryotes that have different numbers and patterns of expressed genes in different cells and tissues, and that change their patterns of gene expression as the organisms grow and develop. These *tissue-specific* or *developmental* types of transcriptional regulation are fully discussed in later chapters (Chapters 15 and 20), but here we highlight two categories of DNA transcription-regulating sequences that lead to differential expression of genes.

Enhancer sequences are one important group of DNA regulatory sequences that increase the level of transcription of specific genes. Enhancer sequences bind specific proteins that interact with the proteins bound at gene promoters, and together promoters and enhancers drive transcription of certain genes. In many situations, enhancers are located upstream of the genes they regulate; but enhancers can be located downstream as well. Some enhancers are relatively close to the genes they regulate, but others are thousands to tens of thousands of base pairs away from their target genes. Thus, important questions for molecular biologists are: What proteins are

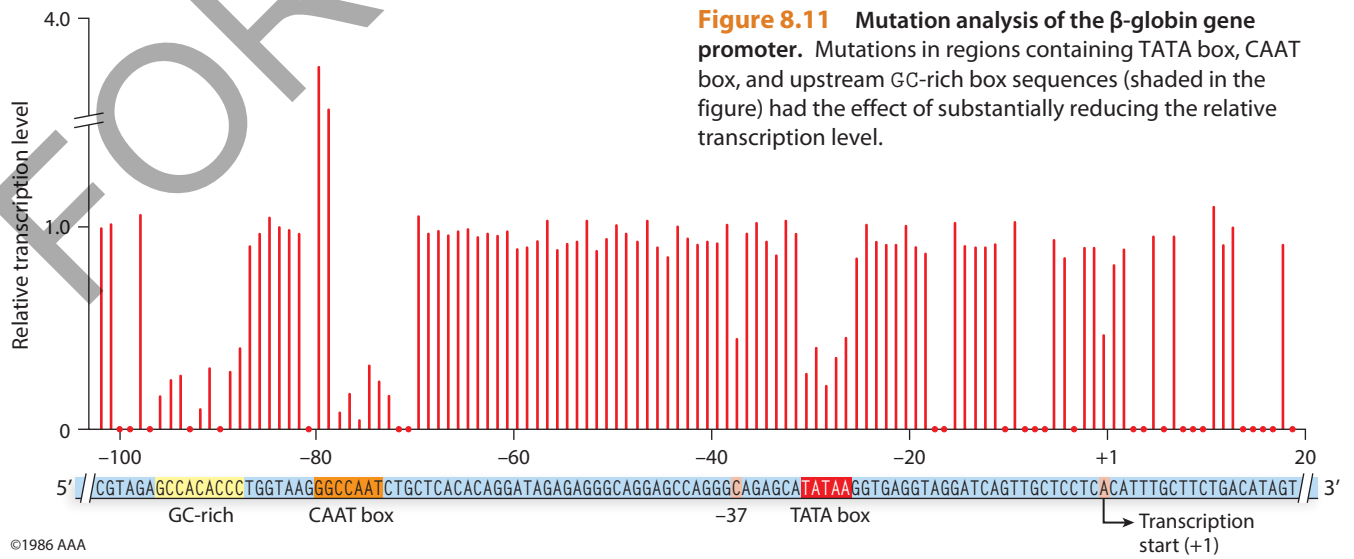


Figure 8.11 Mutation analysis of the β-globin gene promoter. Mutations in regions containing TATA box, CAAT box, and upstream GC-rich box sequences (shaded in the figure) had the effect of substantially reducing the relative transcription level.

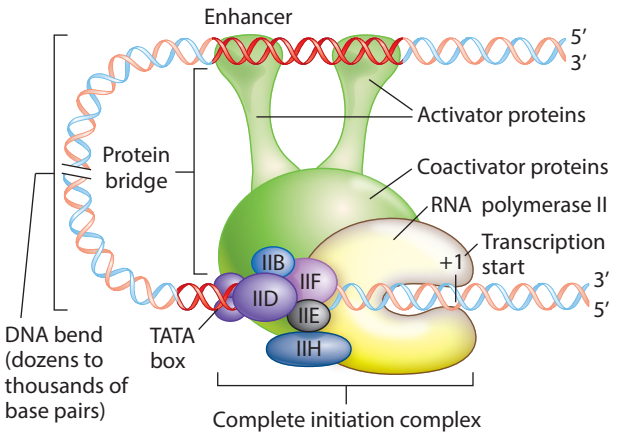


Figure 8.12 Enhancers activate transcription in cooperation with promoters. A protein bridge composed of transcriptional proteins forms between enhancer and promoter sequences, which may be separated by thousands of nucleotides.

bound to enhancers, and how do enhancer sequences regulate transcription of the gene given their different distances from the start of transcription?

The answers are that enhancers bind activator proteins and associated coactivator proteins to form a protein "bridge" that bends the DNA and links the complete initiation complex at the promoter to the activator–coactivator complex at the enhancer (Figure 8.12). The bend produced in the DNA may contain dozens to thousands of base pairs. The action of enhancers and the proteins they bind dramatically increases the efficiency of RNA pol II in initiating transcription, and as a result increases the level of transcription of genes regulated by enhancers.

At the other end of the transcription-regulating spectrum are **silencer sequences,** DNA elements that can act at a great distance to repress transcription of their target genes. Silencers bind transcription factors called repressor proteins, inducing bends in DNA that are similar to what is seen when activators and coactivators bind to enhancers—except with the consequence of reducing the transcription of targeted genes. Like enhancers, silencers can be located upstream or downstream of a target gene and can reside up to several thousand base pairs away from it. Thus enhancers and silencers may operate by similar general mechanisms but with opposite effects on transcription. We discuss these and other eukaryotic regulatory DNA sequences in more detail in Chapter 15.

RNA Polymerase I Promoters

The genes for rRNA are transcribed by RNA polymerase I, utilizing a transcription initiation mechanism similar to that used by RNA pol II. RNA polymerase I is the most specialized eukaryotic RNA polymerase, as it transcribes a limited number of genes. It is recruited to upstream promoter elements following the initial binding of transcription factors, and it transcribes ribosomal RNA genes found in the **nucleolus** (plural, **nucleoli**), a nuclear organelle containing rRNA and multiple tandem copies of the genes encoding rRNAs (tandem means "end to end"). In *Arabidopsis,* for example, each nucleolus contains about 700 copies of rRNA genes. Nucleoli play a key role in the manufacture of ribosomes. At nucleoli, transcribed ribosomal RNA genes are packaged with proteins to form the large and small ribosomal subunits.

Promoters recognized by RNA pol I contain two similar functional sequences near the start of transcription. The first is the **core element,** stretching from −45 to +20 and bridging the start of transcription, and the second is the **upstream control element,** spanning nucleotides −100 to −150 (Figure 8.13). The core element is essential for transcription initiation, and the upstream control element increases the level of gene transcription. Both of these elements are rich in guanine and cytosine; DNA sequence comparisons show that all upstream control elements have the same base pairs at approximately 85 percent of nucleotide positions, and the same is true of all core elements. Two upstream binding factor 1 (UBF1) proteins bind the upstream control element. A second protein complex, known as sigma-like factor 1 (SL1) protein, binds the core

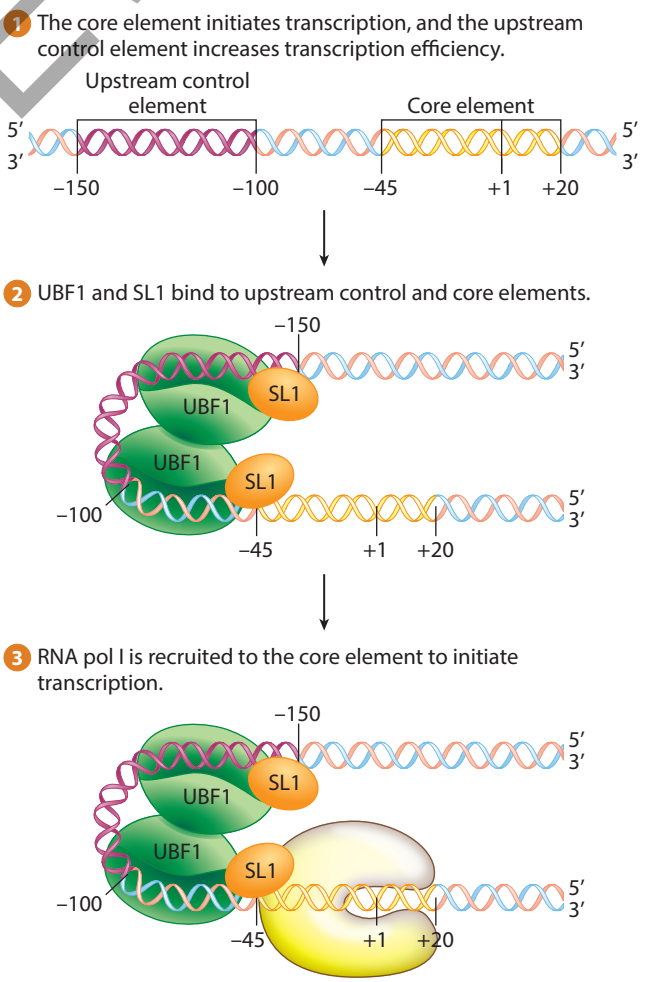


Figure 8.13 Promoter consensus sequences for transcription initiation by RNA polymerase I.

element. This complex recruits RNA pol I to the core element, to initiate transcription of rRNA genes.

RNA Polymerase III Promoters

The remaining eukaryotic RNA polymerase, RNA polymerase III, is primarily responsible for transcription of tRNA genes. However, it also transcribes one rRNA and other RNA-encoding genes. Each of these genes has a promoter structure that differs significantly from the structure of promoters recognized by RNA pol I or RNA pol II. Small nuclear RNA genes have three upstream elements, whereas the genes for 5S ribosomal RNA and transfer RNA each contain two **internal promoter elements** that are *downstream* of the start of transcription.

The upstream elements of small nuclear RNA genes are a TATA box, a **promoter-specific element (PSE),** and an octamer (OCT) (**Figure 8.14a**). A small number of transcription factors—TFIIIs, in this case—bind to these elements and recruit RNA polymerase III, which initiates transcription in a manner similar to that of the other polymerases.

The genes for 5S ribosomal RNA and transfer RNA have internal promoter elements called **internal control regions (ICRs)**; see **Figure 8.14b and c**. The ICRs are two short DNA sequences—designated box A and box B in some genes and box A and box C in other genes—located downstream of the start of transcription, between nucleotides +55 and +80 (**Figure 8.15**). To initiate transcription, box B or box C is bound by TFIIIA, which facilitates the subsequent binding of TFIIIC to box A. TFIIIB then binds to the other transcription factors. In the final initiation step, RNA polymerase III binds to the transcription factor complex and overlaps the +1 nucleotide. With RNA polymerase correctly positioned, transcription begins approximately 55 bp upstream of the beginning of box A, at the +1 nucleotide.

Termination in RNA Polymerase I or III Transcription

Each of the eukaryotic RNA polymerases utilizes a different mechanism to terminate transcription. Here we briefly describe termination in transcription by RNA pol I and RNA pol III, leaving termination of RNA pol II transcription for more extensive discussion in Section 8.4. Transcription by RNA polymerase III is terminated in a manner reminiscent of *E. coli* transcription termination. The RNA pol III transcribes a terminator sequence that creates a string of uracils in the transcript. The poly-U string is similar to the string that occurs in bacterial intrinsic termination (see Section 8.2). The RNA pol III

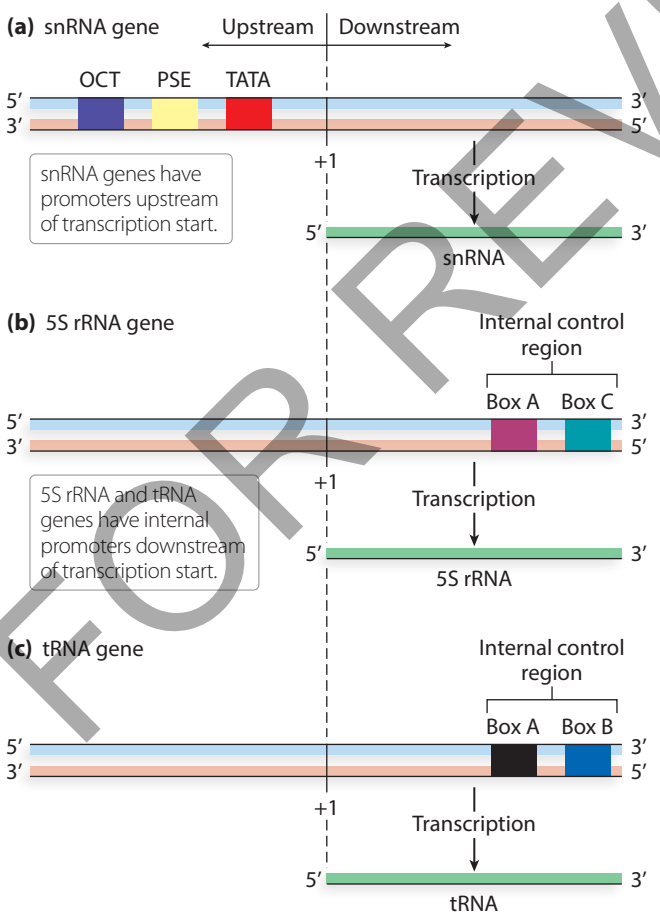


Figure 8.14 **Promoter variation in genes transcribed by RNA polymerase III.**

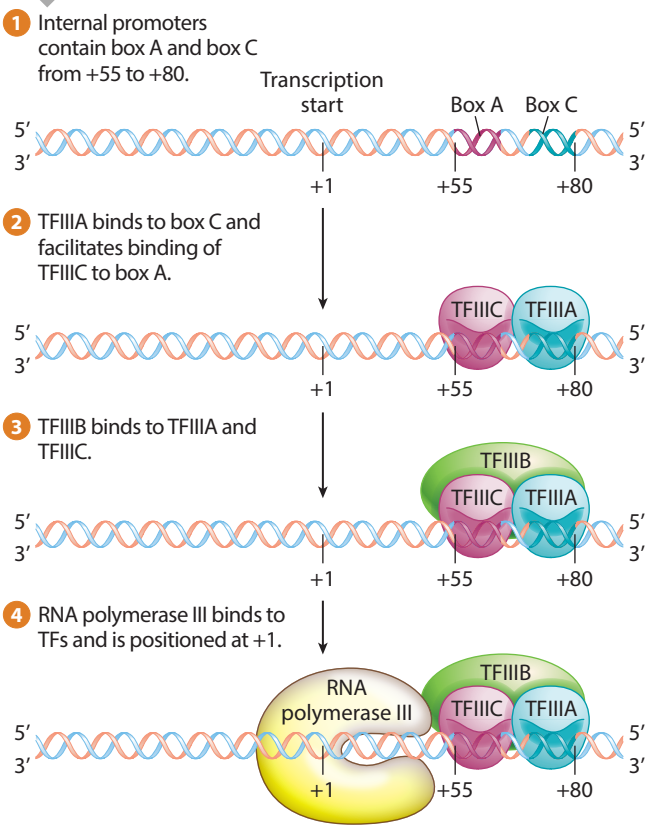


Figure 8.15 **Promoter internal control regions for transcription by RNA polymerase III.**

terminator sequence does not contain an inverted repeat, however, so no stem-loop structure forms near the 3′ end of RNA.

Transcription by RNA pol I is terminated at a 17-bp consensus sequence that binds **transcription-terminating factor I (TTFI).** The binding site for TTFI is the DNA consensus sequence

$$\mathrm{AGGTCGACCAG}^{\mathrm{A}}/_{\mathrm{T}}{}^{\mathrm{A}}/_{\mathrm{T}}\mathrm{NTCG}$$

In this sequence, adenine and thymine are equally likely to appear at two adjacent sites, as indicated by the diagonal lines; N signifies a location at which all four nucleotides are more or less equally frequent. A large rRNA precursor transcript is cleaved about 18 nucleotides upstream of the TTFI binding site, so the consensus sequence does not appear in the mature transcript.

Archaeal Transcription

The transcription machinery of archaea is distinct from that of bacteria and represents a simplified and ancestrally related version of the eukaryotic apparatus that is most similar to the RNA pol II holoenzyme. While bacterial transcription utilizes different sigma subunits to alter core polymerase specificity for distinct promoters, eukaryotes use a group of general transcription factors to facilitate the recognition of promoter consensus sequences. In the case of the eukaryotic RNA polymerase II holoenzyme, six general transcription factors are recruited to the promoter. Archaeal transcription follows the eukaryotic model, using three proteins homologous to eukaryotic transcription factors to identify two promoter consensus regions.

Studies examining archaeal promoters and transcription initiation in the thermophilic archaeal species *Sulfolobus shibatae* have identified a TATA-binding protein (TBP, a subunit of TFIID) and transcription factor B (TFB), a homolog of eukaryotic TFIIB, as the only proteins required for interaction with RNA polymerase in the initiation of archaeal transcription (Figure 8.16). TBP binds to a TATA box in the archaeal promoter, and TFB binds a BRE box (TFB-recognition element) that is immediately upstream of the TATA box. With TBP and TFB bound to their promoter elements, RNA polymerase is directed approximately 25 base pairs downstream to the transcription start site. A third component, TFIIE[α], a homolog of the eukaryotic GTP TFIIE, is not always required for transcription, but it enhances TATA box binding, thereby stimulating transcription.

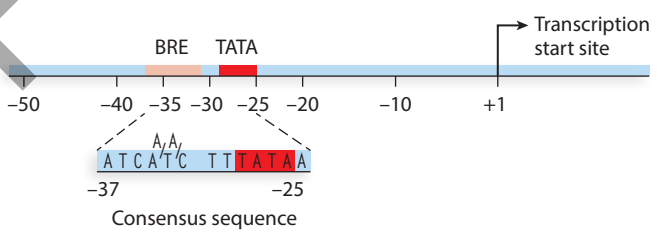


Figure 8.16 Archaea promoter consensus sequences. The TATA box and BRE box sequences bind TBP and TFB along with RNA polymerase to initiate transcription.

8.4 Post-Transcriptional Processing Modifies RNA Molecules

Bacterial, archaeal, and eukaryotic transcripts differ in several ways. For example, eukaryotic transcripts are more stable than bacterial and archaeal transcripts. The half-life of a typical eukaryotic mRNA is measured in hours to days, whereas bacterial mRNAs have an average half-life measured in seconds to minutes. A second difference is the separation, in time and in location, between transcription and translation. Recall that in bacteria the lack of a nucleus leads to coupling of transcription and translation. Similarly, archaea lack a nucleus, leading to the possibility of synchrony between transcription and translation. In eukaryotic cells, on the other hand, transcription takes place in the nucleus, and translation occurs later at free ribosomes or at those attached to the rough endoplasmic reticulum in the cytoplasm. A third difference is the presence of introns in eukaryotic genes that are absent from most bacterial and archaeal genes. Each of these differences comes into play as we consider post-transcriptional modifications of mRNA in eukaryotic cells, which is the focus of this section.

In discussing post-transcriptional processing, we highlight three processing steps that are coordinated during transcription to modify the initial eukaryotic gene mRNA transcript, called **pre-mRNA,** into **mature mRNA,** the fully processed mRNA that migrates out of the nucleus to the cytoplasm for translation. These modification steps are (1) **5′ capping,** the addition of a modified nucleotide to the 5′ end of mRNA; (2) **3′ polyadenylation,** cleavage at the 3′ end of mRNA and addition of a tail of multiple adenines to form the **poly-A tail**; and (3) **intron splicing,** RNA splicing to remove introns and ligate exons. We conclude the section with a discussion of the mechanisms directing alternative splicing and self-splicing RNAs.

Capping 5′ mRNA

After RNA pol II has synthesized the first 20 to 30 nucleotides of the mRNA transcript, a specialized enzyme, guanylyl transferase, adds a guanine to the 5′ end of the pre-mRNA, producing an unusual 5′-to-5′ bond that forms a triphosphate linkage. Additional enzymatic action then methylates the newly added guanine and may also methylate the next one or more nucleotides of the transcript. This addition of guanine to the transcript and the subsequent methylation is known as 5′ capping.

Guanylyl transferase initiates 5′ capping in three steps depicted in Figure 8.17. Before capping, the terminal 5′ nucleotide of mRNA contains three phosphate groups,

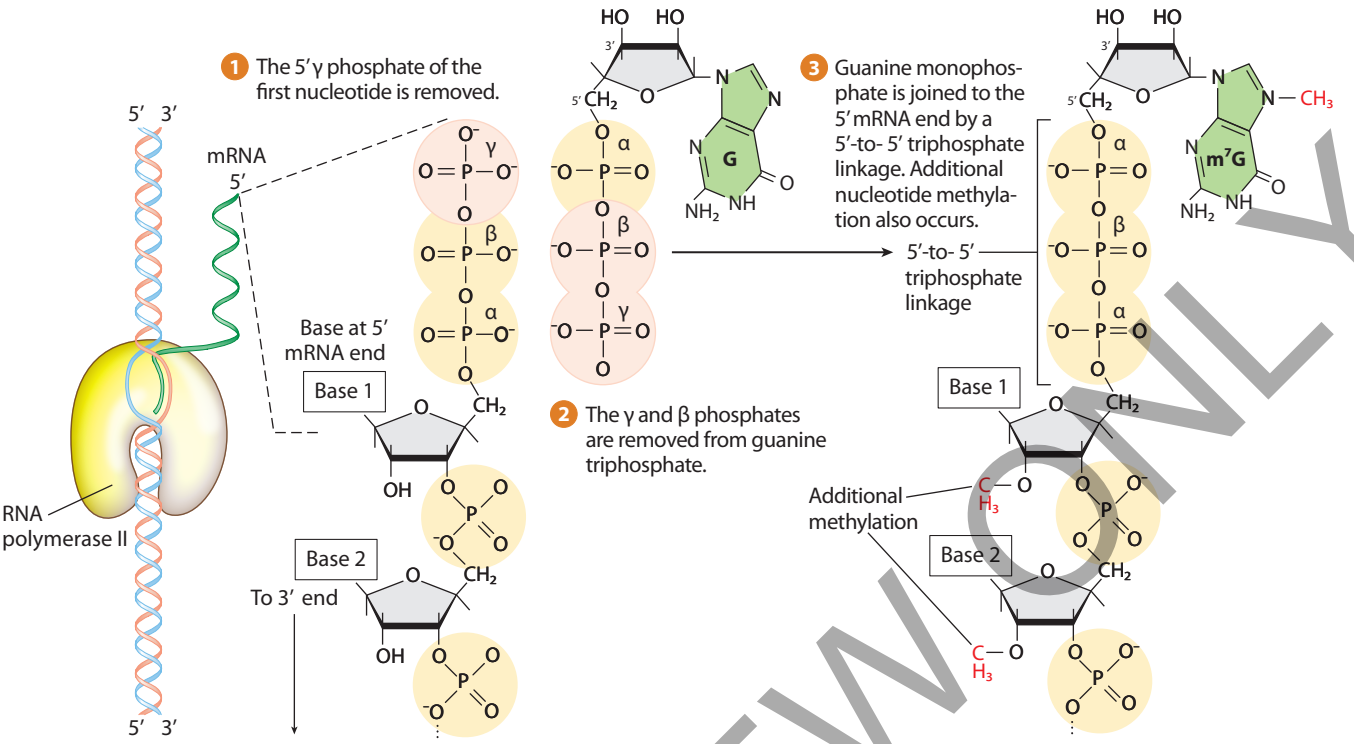


Figure 8.17 **Capping the 5′ end of eukaryotic pre-mRNA.**

labeled α, β, and γ in Figure 8.17. Guanylyl transferase first removes the γ phosphate, leaving two phosphates on the 5′ terminal nucleotide 1. The guanine triphosphate containing the guanine that is to be added loses two phosphates (γ and β) to form a guanine monophosphate 2. Then, guanylyl transferase joins the guanine monophosphate to the mRNA terminal nucleotide to form the 5′-to-5′ triphosphate linkage 3. Methyl transferase enzyme then adds a methyl (CH_3) group to the 7-nitrogen of the new guanine, forming 7-methylguanosine (m^7G). Methyl transferase may also add methyl groups to 2′–OH of nearby nucleotides of mRNA.

The 5′ cap has several functions, including (1) protecting mRNA from rapid degradation, (2) facilitating mRNA transport across the nuclear membrane, (3) facilitating subsequent intron splicing, and (4) enhancing translation efficiency by orienting the ribosome on mRNA.

Polyadenylation of 3′ Pre-mRNA

Termination of transcription by RNA pol II is not fully understood, but it appears likely to be tied to the processing and polyadenylation of the 3′ end of pre-mRNA. It is clear that the 3′ end of mRNA is not generated by transcriptional termination. Rather, the 3′ end of the pre-mRNA is created by enzymatic action that removes a segment from the 3′ end of the transcript and replaces it with a string of adenine nucleotides, the poly-A tail. This step of pre-mRNA processing is thought to be associated with subsequent termination of transcription.

Figure 8.18 illustrates these steps. Polyadenylation begins with the binding of a factor called cleavage and polyadenylation specificity factor (CPSF) near a six-nucleotide mRNA sequence, AAUAAA, that is downstream of the stop codon and thus not part of the coding sequence of the gene. This six-nucleotide sequence is known as the **polyadenylation signal sequence.** The binding of cleavage-stimulating factor (CStF) to a uracil-rich sequence several dozen nucleotides downstream of the polyadenylation signal sequence quickly follows, and the binding of two other cleavage factors, CFI and CFII, and polyadenylate polymerase (PAP) enlarges the complex 1. The pre-mRNA is then cleaved 15 to 30 nucleotides downstream of the polyadenylation signal sequence 2. The cleavage releases a transcript fragment bound by CFI, CFII, and CStF, which is later degraded 3. The 3′ end of the cut pre-mRNA then undergoes the enzymatic addition of 20 to 200 adenine nucleotides that form the 3′ poly-A tail through the action of CPSF and PAP 4. After addition of the first 10 adenines, molecules of poly-A-binding protein II (PABII) join the elongating poly-A tail and increase the rate of adenine addition 5. The 3′ poly-A tail has several functions, including (1) facilitating transport of mature mRNA across the nuclear membrane, (2) protecting mRNA from degradation, and (3) enhancing translation by enabling ribosomal recognition of messenger RNA.

Certain eukaryotic mRNA transcripts do not undergo polyadenylation. The most prominent of these are transcripts of genes producing *histone proteins,* which are key components of *chromatin,* the DNA–protein complex

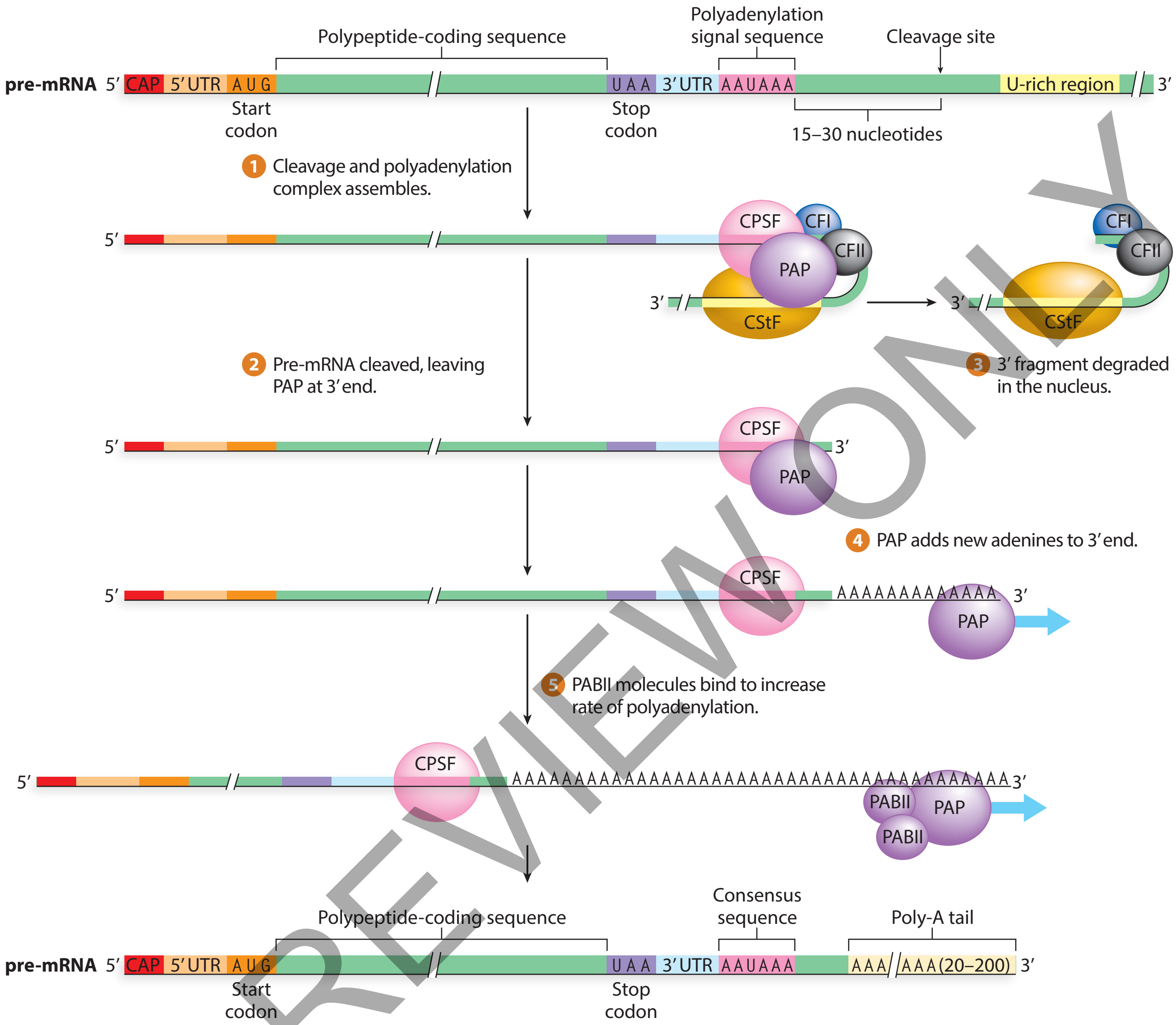


Figure 8.18 **Polyadenylation of the 3′ end of eukaryotic pre-mRNA.**

that makes up eukaryotic chromosomes (see Chapter 11). On these and other "tailless" mRNAs, the 3′ end contains a short stem-loop structure reminiscent of the ones seen in the intrinsic transcription termination mechanism of bacteria. There may be an evolutionary connection between bacterial transcription termination and stem-loop formation on "tailless" eukaryotic mRNAs.

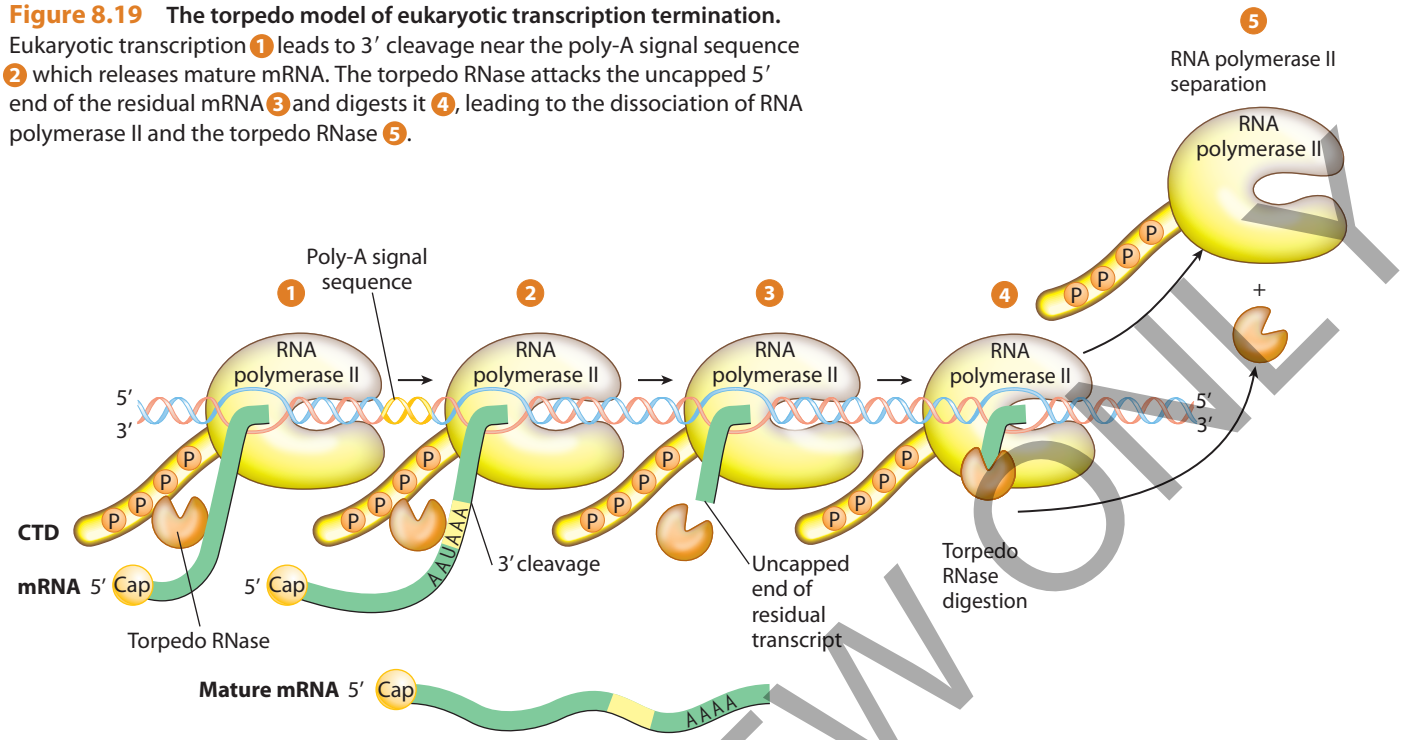
The Torpedo Model of Transcription Termination

The connection between polyadenylation and transcription termination lies in the activity of a specialized RNase (an RNA-destroying enzyme) that attacks and digests the residual RNA transcript attached to RNA pol II after 3′ transcript cleavage (**Figure 8.19**). Following polyadenylation and 3′ cleavage, the residual segment of the transcript still attached to RNA pol II is not capped at its 5′ end. This end is attacked by the specialized RNase that rapidly digests the remaining transcript. The RNase is thought of as a "torpedo" aimed at the residual mRNA attached to RNA pol II. Studies have shown that the torpedo RNase is a highly processive enzyme, meaning that it rapidly carries out its enzymatic action. Once the RNase destroys the residual mRNA and catches up to RNA pol II, it triggers dissociation of the polymerase from template strand DNA to terminate transcription.

Pre-mRNA Intron Splicing

The third step of pre-mRNA processing is intron splicing, which consists of removing intron segments from pre-mRNA and ligating the exons. Intron splicing requires

Figure 8.19 The torpedo model of eukaryotic transcription termination. Eukaryotic transcription ❶ leads to 3′ cleavage near the poly-A signal sequence ❷ which releases mature mRNA. The torpedo RNase attacks the uncapped 5′ end of the residual mRNA ❸ and digests it ❹, leading to the dissociation of RNA polymerase II and the torpedo RNase ❺.



exquisite precision to remove all intron nucleotides accurately without intruding on the exons, and without leaving behind additional nucleotides, so that the mRNA sequence encoded by the ligated exons will completely and faithfully direct synthesis of the correct polypeptide. As an example of the need for precision in intron removal, consider the following "precursor string" made up of exon-like blocks of letters forming three-letter words interrupted by unintelligible intron-like blocks of letters. If editing removes the "introns" accurately, the "edited string" can be divided into its three-letter words to form a "sentence." If an error in editing were to remove too many or too few letters, a nonsense sentence would result.

The finding that introns interrupt the genetically informative segments of eukaryotic genes was a stunning discovery reported independently by the molecular biologists Richard Roberts and Phillip Sharp in 1977. Nothing known about eukaryotic gene structure at the time suggested that most eukaryotic genes are subdivided into intron and exon elements. Roberts and Sharp shared the 1993 Nobel Prize in Physiology or Medicine for their codiscovery of "split genes" in the eukaryotic genome.

Sharp's research group discovered the split nature of eukaryotic genes by using a technique known as R-looping. In this method, DNA encoding a gene is isolated, denatured to single-stranded form, and then mixed with the mature mRNA transcript from the gene. Regions of the gene that encode sequences in mature mRNA will be complementary to those sequences in the mRNA and will hybridize with them to form a DNA–mRNA duplex. However, DNA segments encoding introns will not find complementary sequences in mature mRNA and will remain single-stranded, looping out from between the hybridized sequences.

Figure 8.20 shows a map of the *hexon* gene studied in R-looping experiments by Sharp and colleagues. The experimental results, photographed by electron microscopy, reveal four DNA–mRNA hybrid regions where exon DNA sequence pairs with mature mRNA sequence. Three single-stranded R-loop sequences are introns which do not pair with mRNA.

Splicing Signal Sequences

Eukaryotic pre-mRNA contains specific short sequences that define the 5′ and 3′ junctions between introns and their neighboring exons. In addition, there is a consensus sequence near each intron end to assist in its accurate identification. The **5′ splice site** is located at the 5′ intron end, where it abuts an exon (Figure 8.21). This site

	intron	intron	intron
Precursor string:	youmaynoxpghrcyeomtpwtipthepfxwubijrdlzmcolzotandsipthetea		
Edited string:	youmaynowtipthepotandsipthetea		
Sentence:	you may now tip the pot and sip the tea		

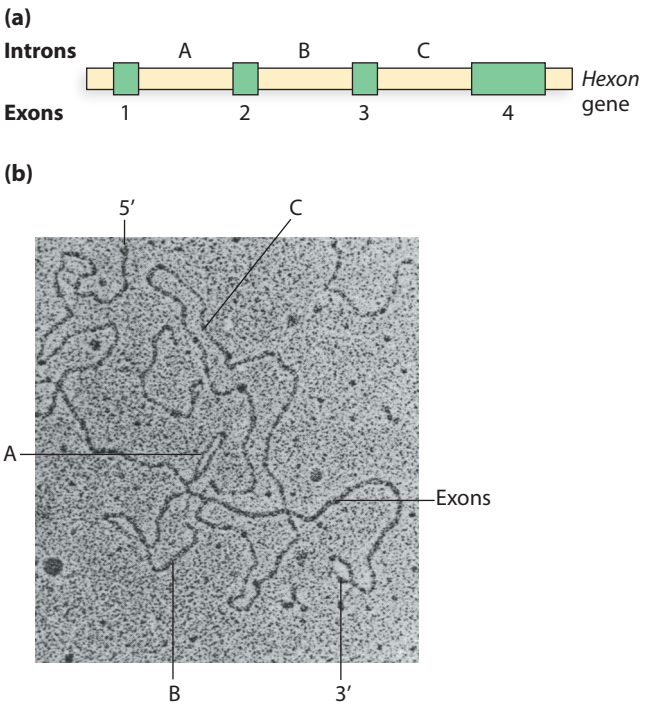


Figure 8.20 **R-loop experimental analysis.** **(a)** The *hexon* gene contains four exons (1 to 4) and three introns (A to C). **(b)** Electron micrographs show hybridization of mature mRNA with exon sequences of denatured *hexon* DNA. Intron sequences are not hybridized and remain single stranded.

contains a consensus sequence with a nearly invariant GU dinucleotide forming the 5′-most end of the intron. The consensus sequence includes the last three nucleotides of the adjoining exon, as well as the four or five nucleotides that follow the GU in the intron. At the **3′ splice site** on the opposite end of the intron, a consensus sequence of 11 nucleotides contains a pyrimidine-rich region and a nearly invariant AG dinucleotide at the 3′-most end of the intron. The third consensus sequence, called the branch site, is located 20 to 40 nucleotides upstream of the 3′ splice site. This consensus sequence is pyrimidine-rich and contains an invariant adenine, called the **branch point adenine,** near the 3′ end.

Mutation analysis shows that these consensus sequences are critical for accurate intron removal. Mutations altering nucleotides in any of the three consensus regions can produce abnormally spliced mature mRNA. The abnormal mRNAs—too short if exon sequence is mistakenly removed, too long if intron sequence is left behind, or altered in other ways that result in improper reading of mRNA sequence—produce proteins with incorrect sequences of amino acids (see Chapter 12).

Introns are removed from pre-mRNA by an snRNA–protein complex called the **spliceosome.** The spliceosome is something like a molecular workbench to which pre-mRNA is attached while spliceosome subunit components cut and splice it in a four-step process that, first, cleaves the 5′ splice site; second, forms a **lariat intron structure** that binds the 5′ intron end to the branch point adenine; third, cleaves the 3′ splice site; and finally, ligates exons and releases the lariat intron to be degraded to its nucleotide components. An electron micrograph of a spliceosome in action is seen in the opener photo for this chapter.

Figure 8.21 illustrates the steps of nuclear pre-mRNA splicing, beginning with the aggregation of five small nuclear ribonucleoproteins (snRNPs; pronounced "snurps") to form a spliceosome. The snRNPs are snRNA–protein subunits designated U1 to U6. The spliceosome is a large complex made up of multiple snRNPs, but its composition is dynamic; it changes throughout the different stages of splicing when individual snRNPs come and go as particular reaction steps are carried out.

Coupling of Pre-mRNA Processing Steps

Each intron–exon junction is subjected to the same spliceosome reactions, raising the question of whether there is a particular order in which introns are removed from pre-mRNA—or whether U1 and U2 search more or less randomly for 5′ splice-site and branch-site consensus sequences, inducing spliceosome formation when they happen to encounter an intron. The answer is that introns appear to be removed one by one, but not necessarily in order along the pre-mRNA. For example, a study of intron splicing of the mammalian *ovomucoid* gene demonstrates the successive steps of intron removal. The *ovomucoid* gene contains eight exons and seven introns. The pre-mRNA transcript is approximately 5.6 kb, and the mature mRNA is reduced to 1.1 kb. Northern blot analysis of *ovomucoid* pre-mRNAs at various stages of intron removal illustrates that each intron is removed separately, rather than all introns being removed at once. The order of intron removal does not precisely match their 5′-to-3′ order in pre-mRNA.

The three steps of pre-mRNA processing are tightly coupled. In comprehensive models developed over the last decade or so, the carboxyl terminal domain (CTD) of RNA polymerase II plays an important role in this coupling by functioning as an assembly platform and regulator of pre-mRNA processing machinery. The CTD is located at the site of emergence of mRNA from the polymerase and contains multiple heptad (seven-member) repeats of amino acids that can be phosphorylated. Binding of processing proteins to the CTD allows the mRNA to be modified as it is transcribed.

Current models propose that "gene expression machines" consisting of RNA polymerase II and an array of pre-mRNA-processing proteins are responsible for the coupling of transcription and pre-mRNA processing. **Foundation Figure 8.22** illustrates this gene expression machine model. The CTD of RNA polymerase II associates with multiple proteins that carry out capping (CAP), intron splicing (SF), and polyadenylation (pA) so that the processes of transcription and pre-mRNA processing occur simultaneously. At the initiation of transcription, phosphorylation (P) along

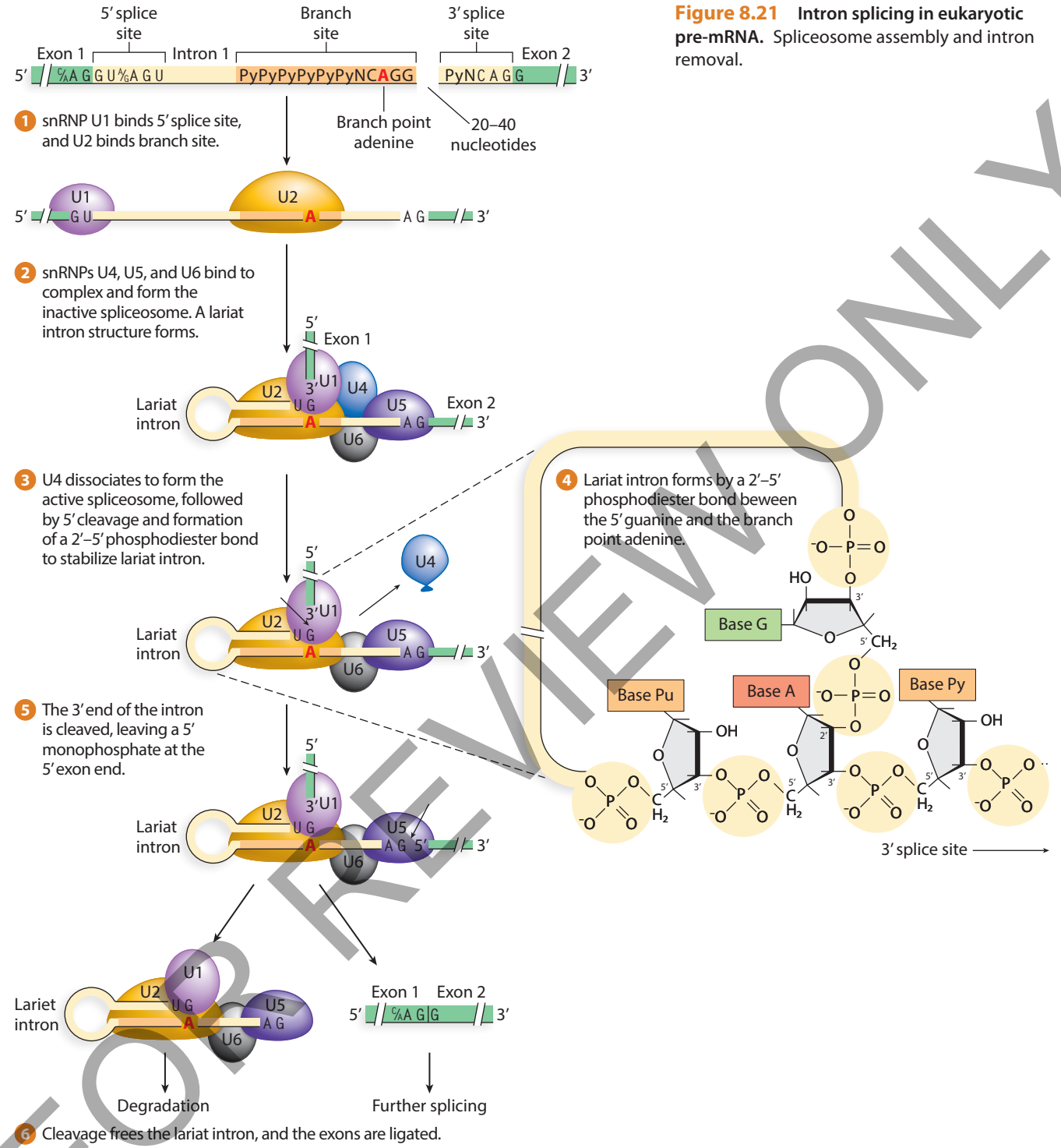


Figure 8.21 Intron splicing in eukaryotic pre-mRNA. Spliceosome assembly and intron removal.

the CTD assists the binding of 5′-capping enzymes, which carry out their capping function and then dissociate. During transcription elongation, specific transcription elongation factors bind the CTD and facilitate splicing-factor binding. The CTD also contains the torpedo RNase responsible for digestion of the residual transcript left attached to RNA pol II by 3′ cleavage linked to polyadenylation. The torpedo RNase is loaded onto the transcript from the CTD to quickly trigger transcription termination.

Alternative Transcripts of Single Genes

Before the complete sequencing of the human genome in the early 2000s, estimates of the number of human genes varied, having been as high as 80,000 to 100,000 genes 20 years or so earlier. A principal reason for this prediction was that human cells produce well over 100,000 distinct polypeptides. It came as something of a surprise, then, when gene annotation of the human genome revealed a total

FOUNDATION FIGURE **8.22**

The Gene Expression Machine Model for Coupling Transcription with pre-mRNA Processing

1. At the initiation of transcription the carboxyl terminal domain (CTD) of RNA polymerase II affiliates with capping (CAP), polyadenylation (pA), splicing factor (SF), and torpedo RNase (RNase).

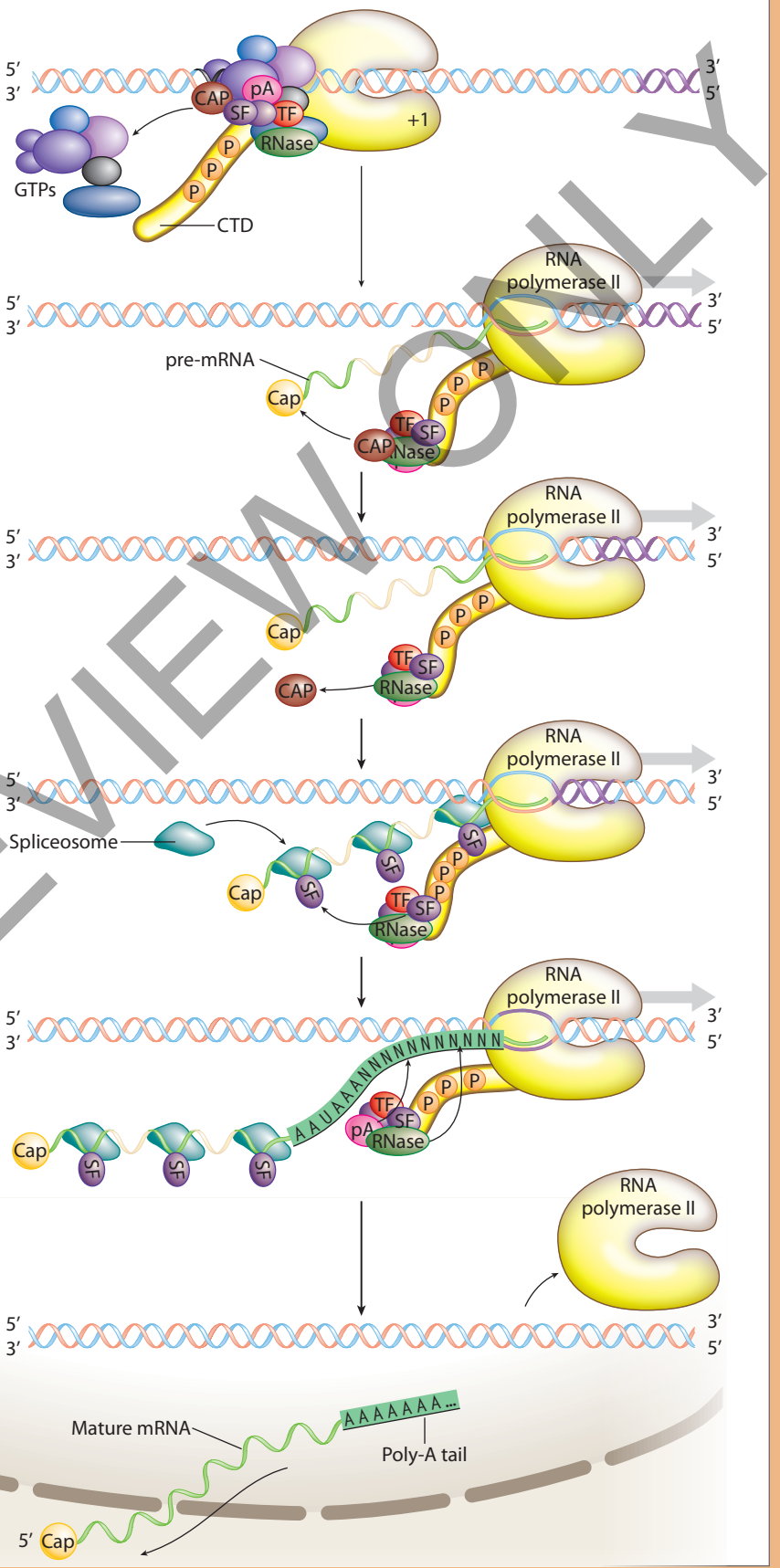
2. RNA pol II initiates transcription after dissociation of the general transcription factors (GTPs). The pre-mRNA processing proteins on the CTD begin their work, starting with the CAP proteins carrying out 5′ capping.

3. Capping proteins dissociate and pre-mRNA elongates.

4. Spliceosome complexes affiliate with pre-mRNA with the aid of SF proteins. Intron splicing takes place as RNA pol continues elongation of mRNA.

5. Polyadenylation proteins identify the pA signal sequence and carry out polyadenylation. Transcription terminates. Splicing continues to completion. Torpedo RNase digests the residual mRNA.

6. Fully processed mature mRNA dissociates from RNA pol II and is transported to cytoplasm for translation. The torpedo RNase digest residual transcript and triggers RNA pol II dissociation to terminate transcription.



content of approximately 22,800 genes. The difference between the number of genes and the number of polypeptides is mirrored by similar findings in other eukaryotic genomes, especially those of mammals. It is common for large eukaryotic genomes to express more proteins than there are genes in the genomes. Three transcription-associated mechanisms can account for the ability of single DNA sequences to produce more than one polypeptide: (1) pre-mRNA can be spliced in alternative patterns in different types of cells; (2) alternative promoters can initiate transcription at distinct +1 start points in different cell types; and (3) alternative locations of polyadenylation can produce different mature mRNAs. Collectively, these varied processes are identified as **alternative pre-mRNA processing.**

Alternative intron splicing is the mechanism by which post-transcriptional processing of identical pre-mRNAs in different cells can lead to mature mRNAs with different combinations of exons. These alternative mature mRNAs produce different polypeptides. In other words, alternative splicing is a mechanism by which a single DNA sequence can produce more than one specific protein. Alternative splicing is common in mammals—approximately 70 percent of human genes are thought to undergo alternative splicing—but it is less common in other animals, and it is rare in plants.

The products of the human *calcitonin/calcitonin gene-related peptide* (*CT/CGRP*) gene exemplify the process of alternative splicing (**Figure 8.23a**). The *CT/CGRP* gene produces the same pre-mRNA transcript in many cells, including thyroid cells and neuronal cells. The transcript contains six exons and five introns and includes two alternative polyadenylation sites, one in exon 4 and the other following exon 6. In thyroid cells, *CT/CGRP* pre-mRNA is spliced to form mature mRNA containing exons 1 through 4, using the first

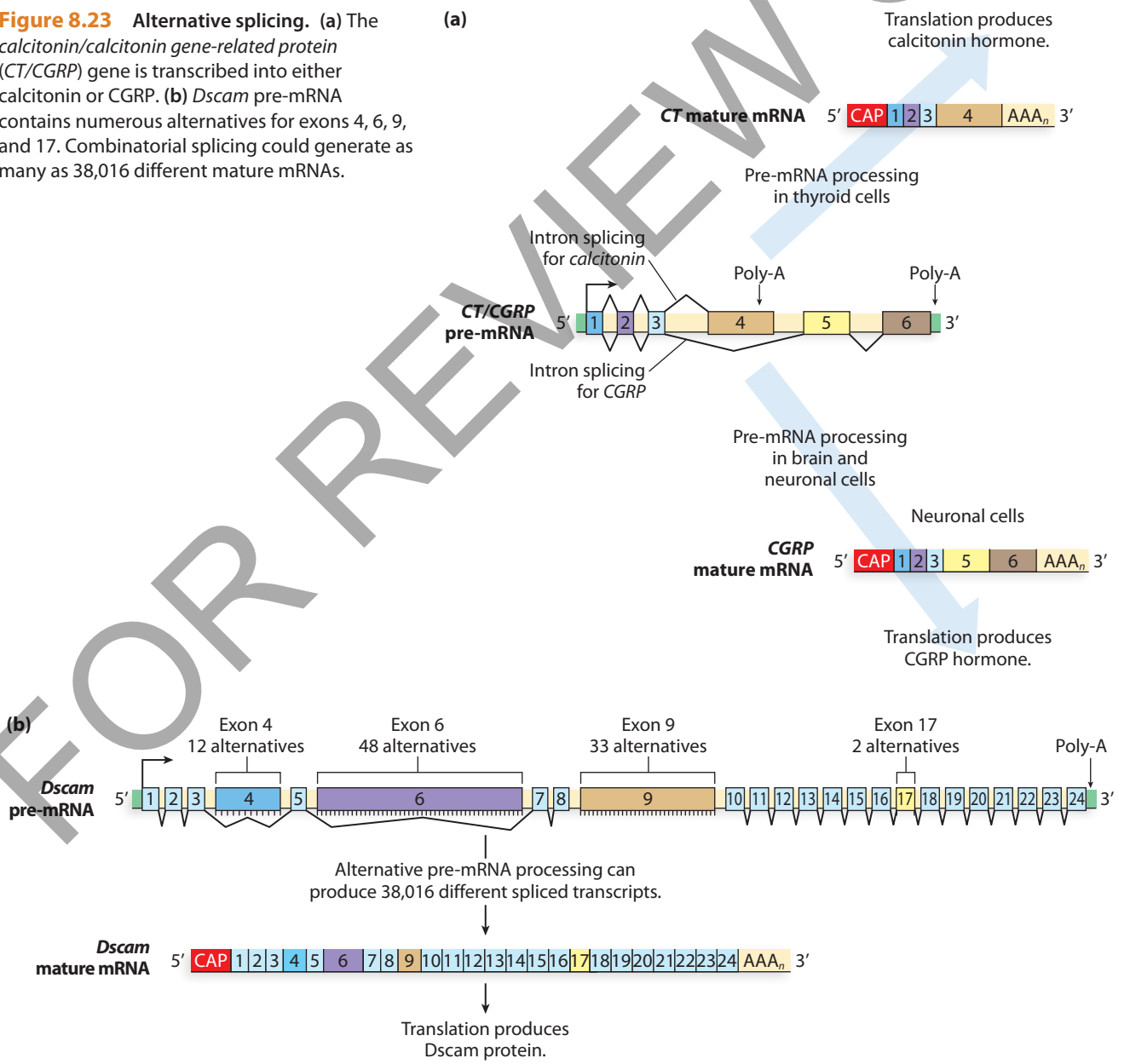


Figure 8.23 **Alternative splicing.** **(a)** The *calcitonin/calcitonin gene-related protein* (*CT/CGRP*) gene is transcribed into either calcitonin or CGRP. **(b)** *Dscam* pre-mRNA contains numerous alternatives for exons 4, 6, 9, and 17. Combinatorial splicing could generate as many as 38,016 different mature mRNAs.

poly-A site for polyadenylation. Translation produces calcitonin, a hormone that helps regulate calcium. In neuronal cells, the same pre-mRNA is spliced to form mature mRNA containing exons 1, 2, 3, 5, and 6. Polyadenylation takes place at the site that follows exon 6, since exon 4 is spliced out as though it were an intron. Translation in neuronal cells produces the hormone CGRP.

One of the most complex patterns of alternative splicing occurs in the *Drosophila Dscam* gene, which produces a protein directing axon growth in *Drosophila* larvae. Mature mRNA from *Dscam* contains 24 exons, but as shown in **Figure 8.23b**, numerous alternative sequences can be used as exons 4, 6, 9, and 17. In total, more than 38,000 different alternative splicing arrangements of *Dscam* are possible, although not all are observed in the organism.

The use of **alternative promoters** occurs when more than one sequence upstream of a gene can bind transcription factors and initiate transcription. Similarly, **alternative polyadenylation** is possible when genes contain more than one polyadenylation signal sequence that can activate 3′ pre-mRNA cleavage and polyadenylation. Alternative promoters and alternative polyadenylation are driven by the variable expression of transcriptional or polyadenylation proteins in a cell-type-specific manner. The variable expression of transcriptional and polyadenylation proteins generates characteristic mature mRNAs from specific genes in particular cells. The result is that transcription of a single gene may lead to the production of several different mature mRNAs in different types of cells, and to their translation into distinct proteins in each of those cell types.

A comprehensive example of a single gene for which all three alternative mechanisms operate to produce distinct polypeptides in different cells is that of the rat α-tropomyosin (α-*Tm*) gene that produces nine different mature mRNAs and, correspondingly, nine different tropomyosin proteins from a single gene. **Figure 8.24a** shows a map of α-*Tm*. The gene contains 14 exons, including alternatives for exons 1, 2, 6, and 9. The gene has two promoters (identified as P_1 and P_2) as well as five alternative polyadenylation sites (identified as A_1 to A_5). The nine distinct mature mRNAs from α-*Tm* are produced in muscle cells (two forms), brain cells (three forms), and fibroblast cells (four forms); see **Figure 8.24b**. Each different mature mRNA illustrates a unique pattern of promoter selection, intron splicing, and choice of polyadenylation site. All mature mRNAs, and their corresponding tropomyosin proteins, contain the genetic information of exons 3, 4, 5, 7, and 8; however, they may contain distinct information in the alternative exons that depends largely on the cell-type-specific selection of promoter and polyadenylation site.

In striated muscle cells, for example, promoter P_1 and polyadenylation site A_2 are used. The mature mRNA includes the alternative exons 1a, 2b, 6b, 9a, and 9b. In contrast,

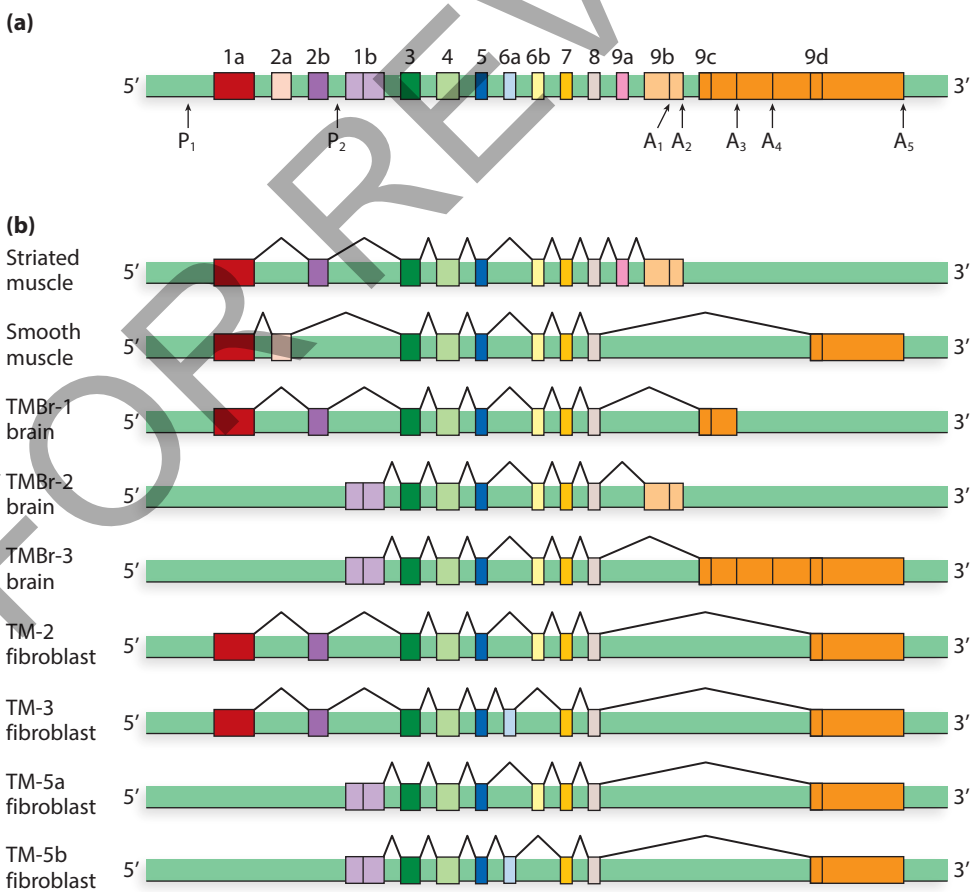


Figure 8.24 Alternative pre-mRNA processing of the rat α-tropomyosin gene. Alternative splicing patterns are indicated by the arched lines connecting exons. Nine distinct mature mRNAs produced by different types of muscle, brain, and fibroblast cells each produce a different tropomyosin protein.

tropomyosin in smooth muscle cells utilizes promoter P_1 and polyadenylation site A_5, and its mature mRNA contains exons 1a, 2a, 6b, and 9d. Brain cells produce three different tropomyosin proteins, each of which are translated from differentially spliced pre-mRNAs that also utilize different polyadenylation sites. In addition, two forms of the brain cell tropomyosin proteins are translated from mRNAs that utilize promoter P_2, and one from an mRNA utilizing P_1. Among the four different tropomyosin proteins produced in fibroblasts, the mRNAs all use polyadenylation site A_5, but they differ in selection of P_1 versus P_2, and alternative splicing occurs as well. **Genetic Analysis 8.2** guides you through analysis of the results of alternative mRNA processing.

Control of Alternative Splicing

We have seen that specific RNA sequences at 5′ and 3′ splice sites are crucial to accurate pre-mRNA splicing and that alternative splicing is widespread in many genomes, with some genes having a large number of alternative protein products from different splicing patterns of pre-mRNA.

Obviously, alternative splicing is carefully controlled in cells, but what mechanisms are involved in that control? The answer appears to be specific sequences in exons and in introns that bind splicing proteins to either enhance or suppress splicing at nearby splice sites. The sequences are identified as **exonic** or **intronic splicing enhancers (ESE** or **ISE)** and **exonic** or **intronic splicing silencers (ESS** or **ISS)**. ESE and ISE sequences, for example, attract protein rich in serine and arginine (one-letter abbreviations S and R, respectively) called SR proteins (**Figure 8.25**). SR proteins direct spliceosome activity to nearby splice sites. These proteins are the products of a large and diverse family of genes, and differential gene transcription in cells is key to SR-protein control of different splicing patterns.

ESS and ISS sequences seem to work in a manner similar to that of splice enhancer sequences, attracting splice repressor proteins that prevent splicing using nearby splice sites. Current evidence indicates that these splice repressor proteins are members of a diverse group of heterogeneous nuclear ribonucleoproteins (hnRNPs). Binding of hnRNPs to ESS or ISS sequences blocks the action of the spliceosome at nearby splice sites.

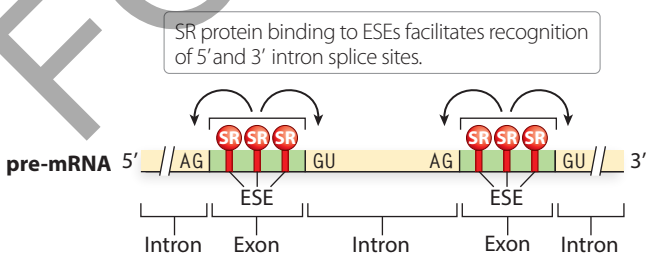


Figure 8.25 **SR-protein recruitment to ESEs, directing spliceosome components to nearby splice sites.** Binding of SR protein to ISEs has a similar result. In contrast, protein binding to ESS and ISS elements blocks nearby spliceosome binding.

Intron Self-Splicing

In addition to introns that are spliced by spliceosomes, certain other RNAs can contain introns that self-catalyze their own removal. Three categories of self-splicing introns, designated group I, group II, and group III introns, have been identified. The molecular biologist Thomas Cech and his colleagues discovered group I introns in 1981, when they observed that a 413-nucleotide precursor mRNA of an rRNA gene from the protozoan *Tetrahymena* could splice itself without the presence of any protein. Following up on this initial observation, Cech and others have shown that group I introns are large, self-splicing ribozymes (catalytically active RNAs) that catalyze their own excision from certain mRNAs and also from tRNA and rRNA precursors in bacteria, simple eukaryotes, and plants. **Intron self-splicing** takes place by way of two transesterification reactions (**Figure 8.26** 1, 2) that excise

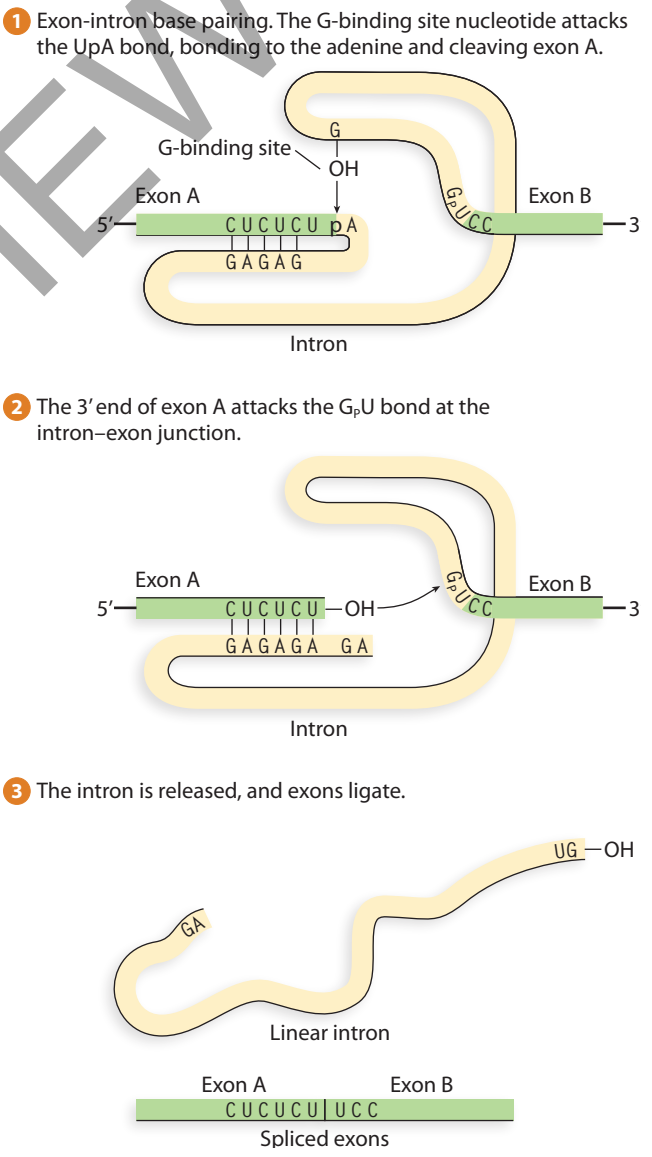
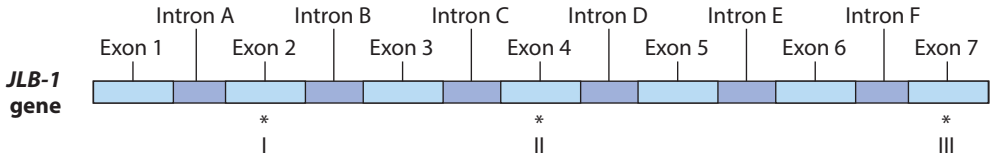


Figure 8.26 **Self-splicing of group I introns.**

GENETIC ANALYSIS 8.2

PROBLEM The *JLB-1* gene, expressed in several human organs, contains seven exons (1 to 7) and six introns (A to F). Three oligonucleotide probes (I to III), hybridizing to exons 2, 4, and 7, respectively, are indicated by asterisks below the gene map:



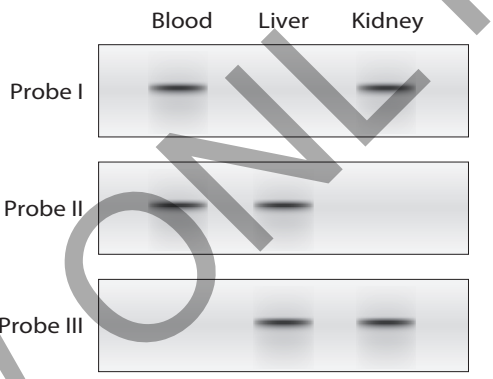
Mature mRNA is isolated from three tissues expressing the *JLB-1* gene and examined by northern blotting using the three oligonucleotide probes indicated above. The probes bind to complementary sequences in mRNA. Northern blot patterns of hybridization between each probe and mRNA isolated from blood, liver, and kidney cells are shown. For each northern blot:

a. Explain the meaning of the hybridization result.

b. Identify the biological process or processes accounting for the observed patterns of hybridization in the northern blots.

BREAK IT DOWN: Molecular probes bind only to their target sequences. A band appears in the northern blot only if the exon target of a probe is present in the mRNA (p. 289, See also p. 349).

BREAK IT DOWN: Differences in the results for different tissues indicate the presence of alternative transcripts of the gene (p. 289, see also p. 349).



Solution Strategies	Solution Steps
Evaluate	
1. Identify the topic this problem addresses and the nature of the required answer.	1. This problem concerns the production of mature mRNAs from a single human gene expressed in different organs. The answer requires identification of the specific mechanisms responsible for the data obtained from each organ.
2. Identify the critical information provided in the problem.	2. The problem gives gene structure, the binding location of each of three molecular probes hybridizing the gene, and the results of three northern blot analyses of mature mRNA from different organs.
Deduce	
3. Identify the regions of *JLB-1* that are anticipated to be part of the pre-mRNA.	3. Pre-mRNA from this gene is anticipated to include all intron and exon sequences.
4. Identify the regions expected to be found in mature mRNA.	4. Exon segments are expected in mature mRNA, along with modification at the 5′ mRNA end (capping) and the 3′ end (poly-A tailing).
Solve	
5. Determine the hybridization pattern of molecular probes in each tissue.	Answer a 5. Blood: Probes I and II hybridize, but probe III does not. This result indicates that exons 2 and 4 are present in the mature mRNA of blood, but exon 7 is not. Liver: Probe I fails to hybridize to mRNA from liver, indicating that exon 1 is missing from liver mRNA. Probes II and III hybridize liver mRNA, indicating that exons 4 and 7 are included in the mature transcript. Kidney: Probe II does not hybridize kidney mRNA, indicating that exon 4 is missing from it. Probes I and III find hybridization targets, indicating that exons 2 and 7 are present in the transcript.
6. Interpret the hybridization patterns in each tissue and identify the process or processes that reasonably account for the observed patterns.	Answer b 6. Blood: The absence of exon 7 is most likely due to either the use of an alternative polyadenylation site that generates 3′ cleavage of pre-mRNA ahead of exon 7 or to differential splicing that removes exon 7 from pre-mRNA during intron splicing. Liver: The absence of exon 2 is most likely due either to use of an alternative promoter that initiates transcription at a point past exon 2 or to differential splicing of liver pre-mRNA. Kidney: The absence of exon 4 is most likely the result of differential splicing of pre-mRNA.

TIP: Hybridization of a probe occurs when the probe finds its target sequence. The absence of hybridization indicates that the target sequence for a probe is not present.

TIP: Alternative promoters, alternative polyadenylation sites, and alternative splicing are three mechanisms that lead eukaryotic genomes to generate distinct proteins from the same gene.

For more practice, see Problems 2, 3, and 8.

the intron and allow exons to ligate 3. Cech and Sidney Altman shared the 1989 Nobel Prize in Physiology or Medicine for their contributions to the discovery and description of the catalytic properties of RNA.

Group II introns, which are also self-splicing ribozymes, are found in mRNA, tRNA, and rRNA of fungi, plants, protists, and bacteria. Group II introns form highly complex secondary structures containing many stem-loop arrangements. Their self-splicing takes place in a lariat-like manner utilizing a branch point nucleotide that in many cases is adenine. It is thought that nuclear pre-mRNA splicing may have evolved from group II self-splicing introns.

Group III introns and group II introns are similar in having elaborate secondary structures and lariat-like splicing structures that utilize a branch point nucleotide. Group III introns are much shorter than group II introns, however, and their secondary structures are different from those of group II introns.

Ribosomal RNA Processing

In bacteria, archaea, and eukaryotes, rRNAs are transcribed as large precursor molecules that are cleaved into smaller RNA molecules by removal and discarding of spacer sequences intervening between the sequences of the different RNAs. The *E. coli* genome, for example, contains seven copies of an rRNA gene. Each gene copy is transcribed into a single 30S precursor RNA that is processed by the removal of intervening sequences to yield 5S, 16S, and 23S rRNAs, along with several tRNA molecules (**Figure 8.27a**). All seven gene copies produce the same three rRNAs, but each gene generates a different

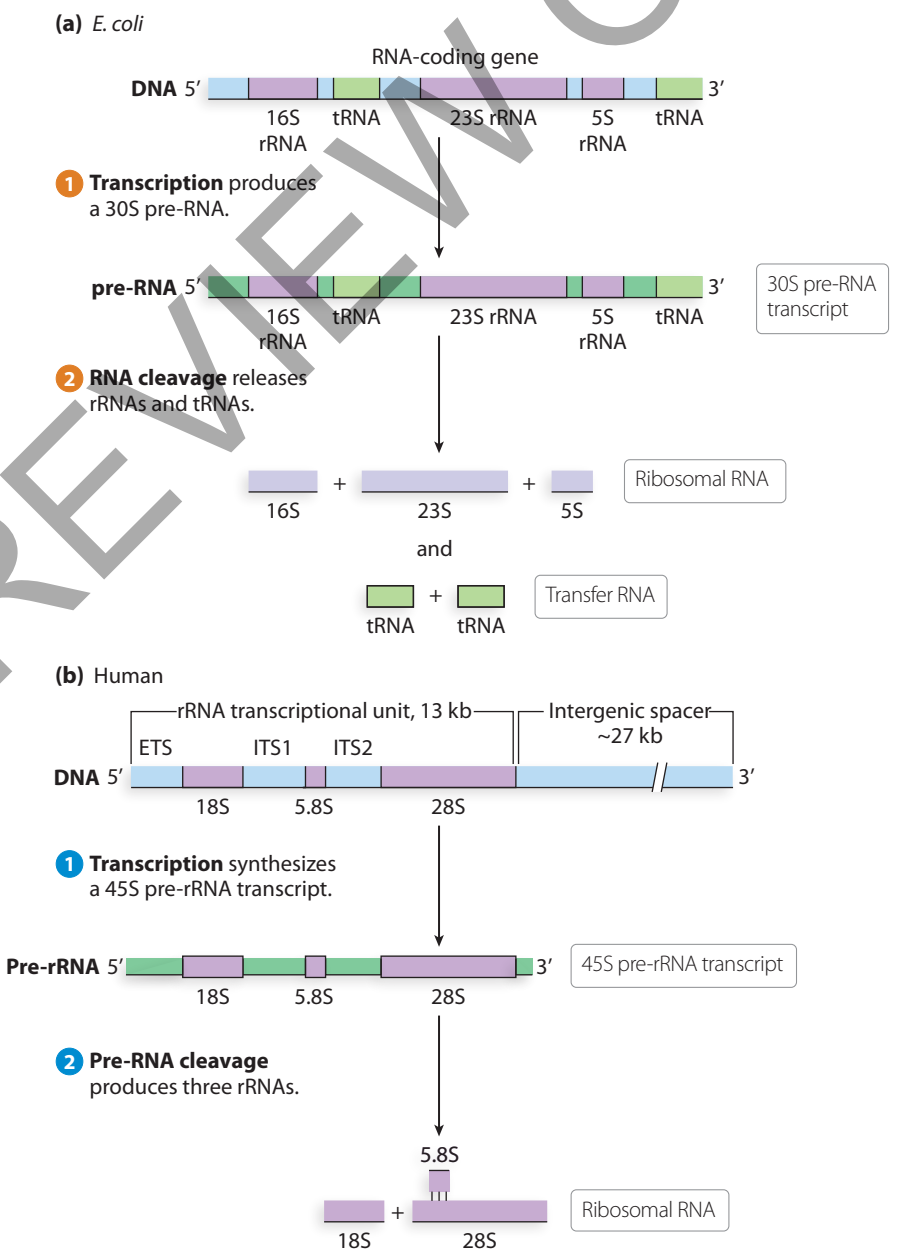


Figure 8.27 The processing of ribosomal and transfer RNA. **(a)** A large transcript is cleaved to produce rRNA and tRNA in *E. coli*. **(b)** Human rRNA genes are part of a 40-kb repeating sequence that produces three rRNAs.

set of tRNAs. There is evidence that archaea use a similar process to produce some rRNA molecules.

Eukaryotic genomes have hundreds of rRNA genes clustered in regions of repeated genes on various chromosomes. Each gene produces a 45S precursor rRNA that contains an external transcription sequence (ETS) and two internal transcription sequences (ITS1 and ITS2) that are removed by processing. The transcript is processed in multiple steps to yield three rRNA molecules weighing 5.8S, 18S, and 28S (Figure 8.27b). Eukaryotic genomes differ somewhat in the steps that process the 45S pre-rRNA transcript. In general, however, the 45S transcript is cleaved to a 41S intermediate from which the 18S transcript is then removed, followed by cleavage that produces the 28S and 5.8S transcripts. The 5.8S and 28S products pair with one another and become part of the same ribosomal subunit. After processing, the resulting rRNAs fold into complex secondary structures and are joined by proteins to form ribosomal subunits. Some chemical modifications of rRNA, particularly methylation of selected nucleotide bases, occur after completion of transcription.

Transfer RNA Processing

The production of tRNA, whether in bacteria, archaea, or eukaryotes, also requires post-transcriptional processing. Each type of tRNA has distinctive nucleotides and a specific pattern of folding, but all tRNAs have similar structures and functions (Figure 8.28). Some bacterial transfer RNA molecules are produced simultaneously with rRNAs, as described above (see Figure 8.27a). Other tRNAs are transcribed as part of a large pre-tRNA transcript that is then cleaved to yield multiple tRNA molecules. In eukaryotes, tRNA genes occur in clusters on specific chromosomes. Each eukaryotic tRNA gene is individually transcribed by RNA polymerase III, and a single pre-tRNA is produced from each gene.

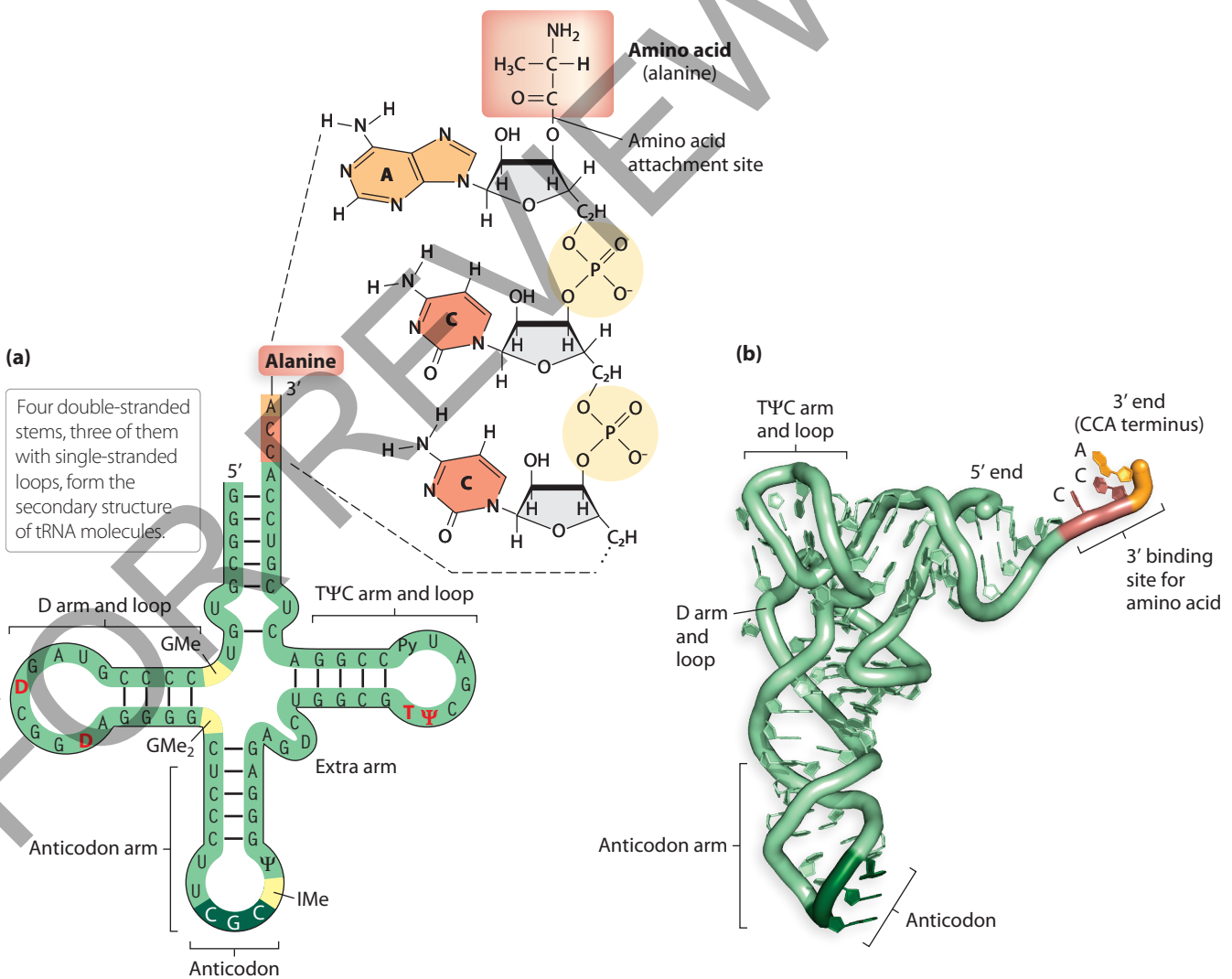


Figure 8.28 Transfer RNA structure. Each tRNA has a distinctive structure. The tRNA carrying alanine is illustrated in two-dimensions **(a)** and three-dimensions **(b)**.

The number of different tRNAs produced depends on the type of organism. In bacteria, the exact number of different tRNAs varies, but it is usually substantially *less than* 61, the number of codons found in mRNA. At a minimum, each species must have at least 20 different tRNAs, one for each amino acid, but most produce at least 30 to 40 different tRNAs. The low number of different tRNAs (compared to number of codons) results from a phenomenon called *third-base wobble*, a relaxation of the "rules" of complementary base pairing at the third base of codons (see Chapter 9). Although third-base wobble plays a role in reducing the number of distinct tRNA genes needed in eukaryotic genomes, eukaryotes nevertheless produce a larger number of different tRNAs than bacteria do. Some eukaryotic genomes contain a full complement of 61 different tRNA genes, one corresponding to each codon of the genetic code.

Bacterial tRNAs require processing before they are ready to assume their functional role of transporting amino acids to the ribosome. The precise processing events differ somewhat among tRNAs, but several features are common. First, many tRNAs are cleaved from large precursor tRNA transcripts to produce several individual tRNA molecules. Second, nucleotides are trimmed off the 5′ and 3′ ends of tRNA transcripts to prepare the mature molecule. Third, certain individual nucleotides in different tRNAs are chemically modified to produce a distinctive molecule. Fourth, tRNAs fold into a precise three-dimensional structure that includes four double-stranded stems, three of which are capped by single-stranded loops; each stem and loop constitutes an "arm" of the tRNA molecule. Fifth, tRNAs undergo post-transcriptional addition of bases. The most common addition is three nucleotides, `CCA`, at the 3′ end of the molecule. This region is the binding site for the amino acid the tRNA molecule transports to the ribosome. Figure 8.28 shows $tRNA_{Ala}$, which carries alanine. The `CCA` terminus is indicated, along with chemically modified nucleotides in each arm that are characteristic of this tRNA. Both a two-dimensional and a three-dimensional representation are shown.

Eukaryotic and archaeal tRNAs undergo processing modifications similar to those of bacterial tRNAs. In addition, however, eukaryotic pre-tRNAs may contain small introns that are removed during processing. For example, an intron 14 nucleotides in length is removed from the precursor molecule by a specialized nuclease enzyme that cleaves the 5′ and 3′ splice sites of tRNA introns. The cleaved tRNA then refolds to form the anticodon stem, and the enzyme RNA ligase joins the 5′ and 3′ ends of the tRNA.

Post-Transcriptional RNA Editing

A firmly established tenet in the central dogma of biology is the role of DNA as the repository and purveyor of genetic information. Notwithstanding the modifications made to precursor RNA transcripts after transcription, a fundamental principle of biology is that DNA dictates the sequence of mRNA nucleotides and controls the order of amino acids in proteins. And yet, in the mid-1980s, a phenomenon called **RNA editing** was uncovered that is responsible for post-transcriptional modifications that change the genetic information carried by mRNA.

Two kinds of RNA editing occur. In one kind of RNA editing, uracils are inserted into edited mRNA with the assistance of a specialized RNA called **guide RNA (gRNA).** A guide RNA, transcribed from a separate RNA-encoding gene, contains a sequence complementary to the region of mRNA that it edits. With the aid of a protein complex, a portion of guide RNA pairs with complementary nucleotides of pre-edited mRNA and acts as a template to direct the insertion (and occasionally the deletion) of uracil (Figure 8.29). Guide RNA releases edited mRNA after editing is complete. The protein translated from edited mRNA may differ from the protein produced from unedited transcript.

The second kind of RNA editing is by base substitution, and frequently consists of the replacement of cytosine with uracil (`C-to-U` editing) in mRNA by removal of the amino group from cytosines. We describe the details of this process, known as deamination, in Section 12.3 and here simply examine the consequences of the event. This type of RNA editing has been identified in mammals, most land plants, and several single-celled eukaryotes.

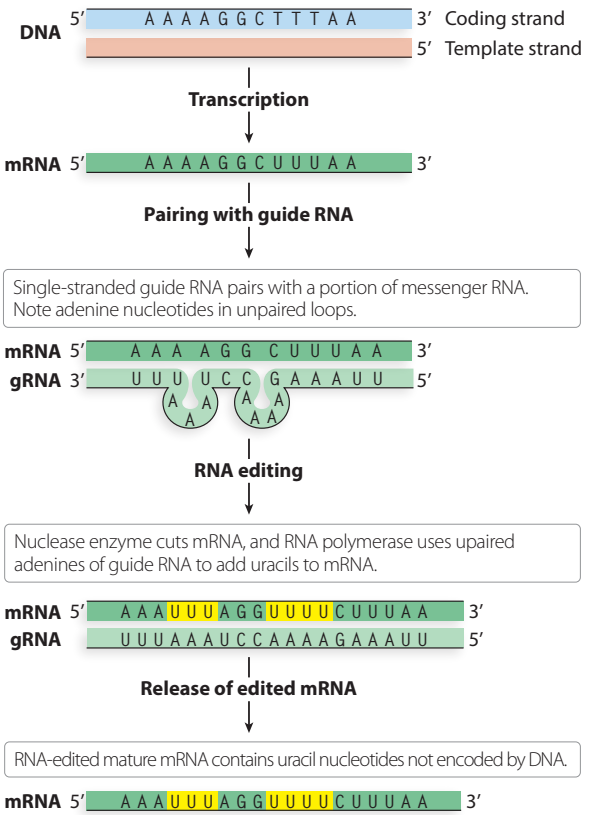


Figure 8.29 Guide RNA (gRNA) directs RNA editing.

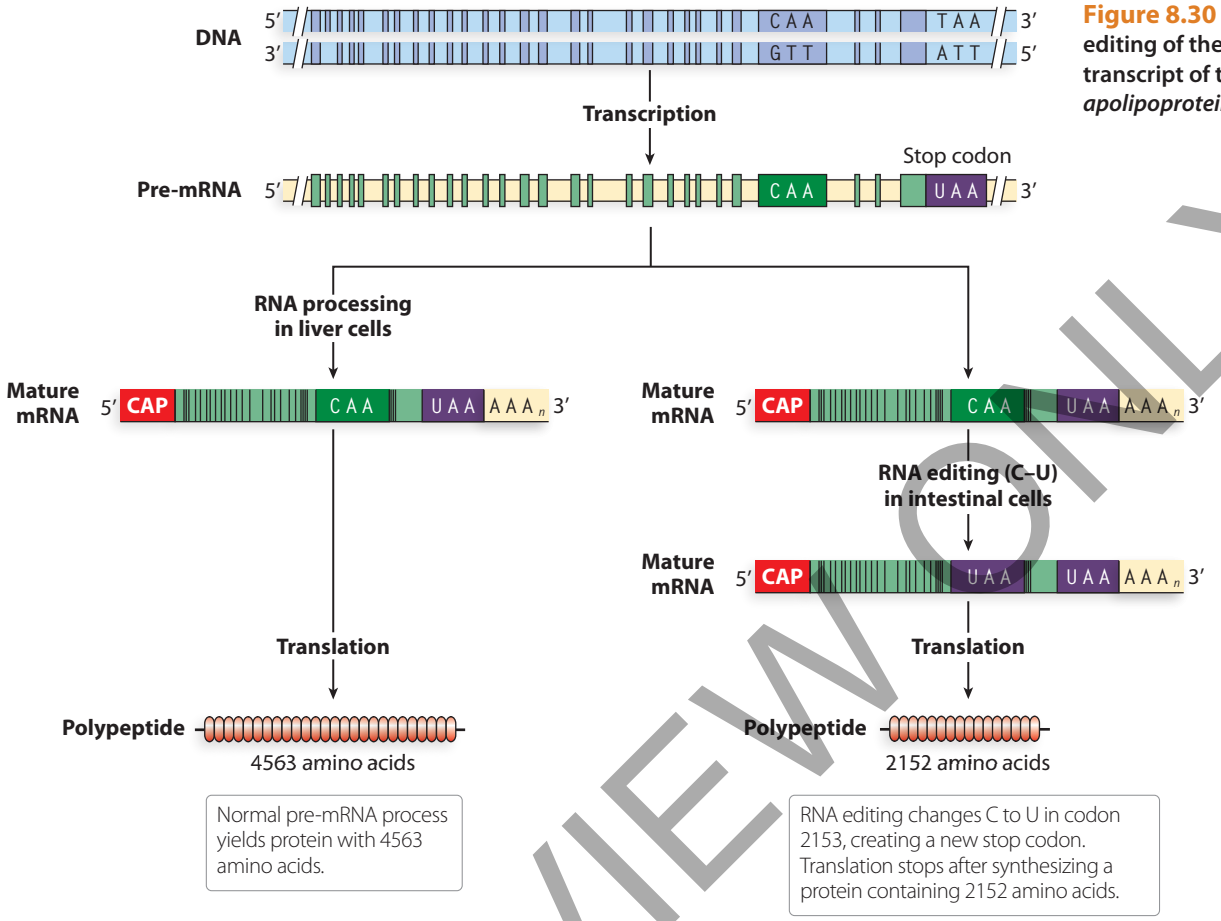


Figure 8.30 RNA editing of the mRNA transcript of the human *apolipoprotein B* gene.

The consequence of C-to-U RNA editing is demonstrated by the protein products of the mammalian *apolipoprotein B* gene (Figure 8.30). An identical gene containing 29 exons is found in all mammalian cells, and the same mRNA is transcribed in all tissues. Part of this messenger RNA sequence includes codon number 2153 that has the sequence CAA and is translated as glutamine in liver apolipoprotein B, a protein consisting of 4563 amino acids. In intestinal cells, however, RNA editing changes the cytosine in codon 2153 to a uracil, converting the codon to UAA. This C-to-U change produced by RNA editing creates a "stop" codon that halts translation after the assembly of the first 2152 amino acids of intestinal apolipoprotein B.

CASE STUDY

Sexy Splicing: Alternative mRNA Splicing and Sex Determination in *Drosophila*

The number of X chromosomes in the nuclei of *Drosophila* embryos is critical in sex determination, but the X/autosome (X/A) ratio proposed by Calvin Bridges (X/A = 1.0 in females and X/A = 0.5 in males) as the underlying cause is not the entire story (see Section 3.4). In fact, the process involves differential gene expression and pre-mRNA splicing. The molecular basis of *Drosophila* sex determination depends on a series of steps that begins with the transcription activation of the *sex-lethal* (*Sxl*) gene, includes alternative splicing of the pre-mRNA transcript of the *transformer* (*Tra*) gene, and culminates with one of two alternative splicing variants of the pre-mRNA transcripts of the *double-sex* (*Dsx*) gene. The Dsx protein directs further transcription activation and repression, leading to female or to male development.

The X/A ratio in fly embryos initially influences the transcription and translation of two X-linked activator proteins called SisA and SisB, and an autosomal gene producing a transcription repressor protein called Deadpan (Figure 8.31). Since the genes producing SisA and SisB are X-linked, early female embryos produce twice as much of each activator as do early male embryos, and the ratio of SisA + SisB to Deadpan differs between female and male embryos. In early female embryos, the ratio of SisA + SisB protein to Deadpan protein leads to transcription of the *Sxl* gene and to the production of Sxl protein. *Sxl* transcription is repressed in male embryos and no Sxl protein is produced.

Sxl protein is a splicing regulator that operates on the pre-mRNA transcript of the *Tra* gene. In female embryos,

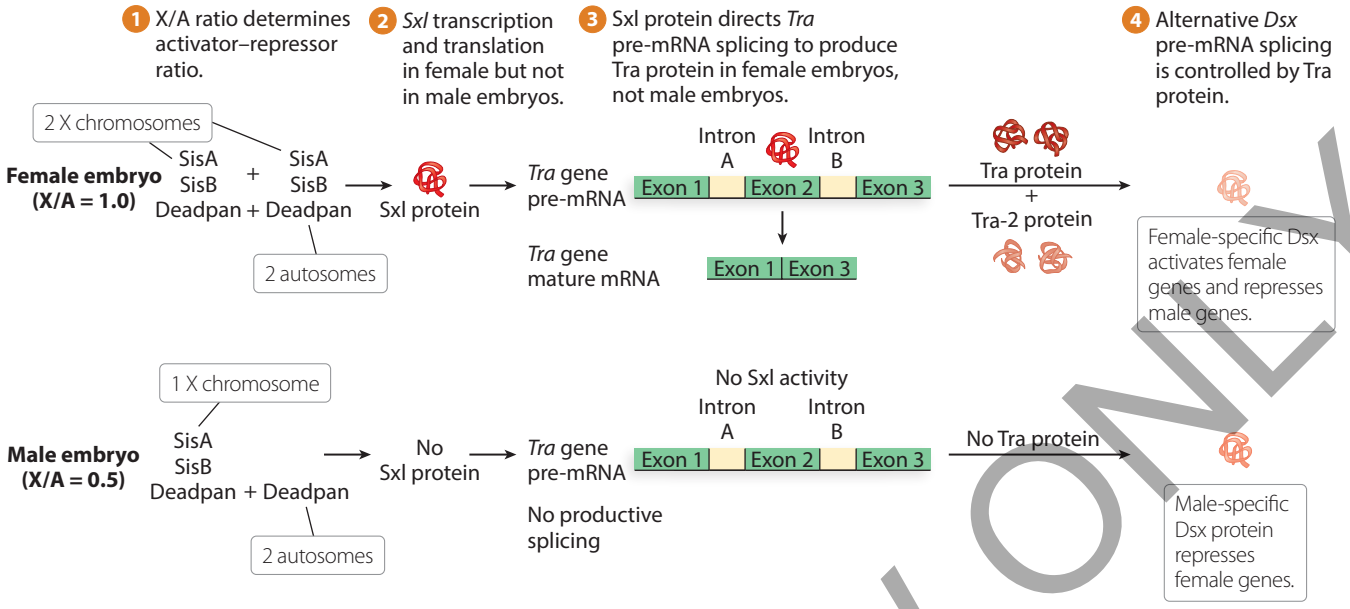


Figure 8.31 **The X/A ratio determines gene transcription and transcript splicing pattern to determine sex in fruit flies.**

Tra pre-mRNA is spliced to produce a functional Tra protein. In male embryos, the absence of Sxl protein leads to alternative *Tra* pre-mRNA splicing that does not produce functional Tra protein. The Tra protein is also a splicing regulator; it operates on the pre-mRNA of *Dsx* along with a second protein known as Tra-2. In female embryos, Tra protein and Tra-2 protein splice *Dsx* pre-mRNA in one alternative variant, which when translated produces female-specific Dsx protein. Female-specific Dsx activates transcription of female-specific genes and represses transcription of male-specific genes to produce female flies. Tra protein is absent in male embryos, and *Dsx* pre-mRNA is spliced in another alternative variant. Dsx protein in male embryos represses female-specific genes and allows transcription of unrepressed male-specific genes, leading to male sex development.

SUMMARY

MasteringGenetics™ For activities, animations, and review quizzes, go to the Study Area.

8.1 RNA Transcripts Carry the Messages of Genes

- RNA molecules are synthesized by RNA polymerases using as building blocks the RNA nucleotides `A`, `G`, `C`, and `U` to form single-stranded sequences complementary to DNA template strands.
- Messenger RNA is the transcript that undergoes translation to produce proteins. Five other major forms of functional RNA are transcribed, and may undergo modification, but are not translated.

8.2 Bacterial Transcription Is a Four-Stage Process

- Transcription has four stages: promoter recognition, chain initiation, chain elongation, and chain termination.
- A single RNA polymerase transcribes all bacterial genes. This polymerase is a holoenzyme composed of a five-subunit core enzyme and a sigma subunit that aids the recognition of different forms of bacterial promoters.
- Bacterial promoters have two consensus sequence regions located upstream of the transcription start at approximately −10 and −35.
- The core enzyme of bacterial RNA polymerase carries out RNA synthesis following chain initiation by the holoenzyme.
- Transcription of most bacterial genes terminates by an intrinsic mechanism that depends only on DNA terminator sequences. Certain bacterial genes have a rho-dependent mechanism of transcription termination.

8.3 Archaeal and Eukaryotic Transcription Displays Structural Homology and Common Ancestry

- Eukaryotic cells contain three types of RNA polymerases that transcribe mRNA and the various classes of functional RNA.
- RNA polymerase II transcribes mRNA by interaction with numerous transcription factors that lead the enzyme to recognize promoters controlling transcription of polypeptide-coding genes.

- Promoters recognized by RNA polymerase II have a TATA box and additional regulatory elements that bind transcription factors and RNA pol II during transcription initiation. Transcription shows similarities among all three domains of life due to the sharing of a common ancestor and the essential nature of transcription.
- Archaeal transcription is a simplified version of eukaryotic transcription and is dissimilar from bacterial transcription.
- Three archaeal transcription proteins, TBP, TFB, and less often TFIIE α, share homology with bacterial and eukaryotic proteins and initiate transcription by interacting with RNA polymerase.
- Eukaryotic promoter regulatory elements are recognized by their consensus sequences.
- Tissue-specific and developmental modifications in transcription are regulated by enhancer and silencer sequences.
- RNA polymerase I uses exclusive transcription factors to recognize upstream consensus sequences of ribosomal RNA genes.
- RNA polymerase III recognizes promoter consensus sequences that are upstream and downstream of the start of transcription.

8.4 Post-Transcriptional Processing Modifies RNA Molecules

- 5′ capping of eukaryotic messenger RNA adds a methylated guanine through the action of guanylyl transferase shortly after transcription is initiated.
- Polyadenylation at the 3′ end of eukaryotic messenger RNA is signaled by an AAUAAA sequence and is accomplished by a complex of enzymes.
- Intron splicing is controlled by cellular proteins that identify introns and exons and form spliceosome complexes that remove introns and ligate exons.
- Consensus sequences at the 5′ splice site, the 3′ splice site, and the branch point serve as guides during intron splicing.
- Alternative splicing is regulated by cell-type-specific variation of proteins that identify introns and exons.
- Some RNA molecules have catalytic activity and are able to self-splice introns without the aid of proteins.
- Ribosomal and transfer RNA molecules are generated by cleavage of large precursor molecules transcribed in bacterial, archaeal, and eukaryotic genomes.
- RNA editing is a post-transcriptional altering of nucleotide sequence, causing the transcripts to differ from the corresponding template DNA sequence.

KEYWORDS

3′ polyadenylation (3′ poly-A tailing) *(p. 285)*
3′ splice site *(p. 289)*
5′ capping *(p. 285)*
5′ splice site *(p. 288)*
–35 consensus sequence *(p. 273)*
alternative pre-mRNA processing (alternative intron splicing, promoter, polyadenylation) *(pp. 292, 293)*
branch point adenine *(p. 289)*
CAAT box *(p. 280)*
closed promoter complex *(p. 273)*
coding region *(p. 272)*
coding strand (nontemplate strand) *(p. 271)*
consensus sequence *(p. 273)*
core element *(p. 283)*
downstream *(p. 272)*
enhancer sequence *(p. 282)*
exonic and intronic splicing enhancers (ESEs and ISE) *(p. 294)*
exonic and intronic splicing silencers (ESS and ISS) *(p. 294)*
functional RNA (tRNA, rRNA, snRNA, miRNA, siRNA, ribozyme) *(p. 271)*
GC-rich box *(p. 280)*
general transcription factors (GTFs) *(p. 282)*
guide RNA (gRNA) *(p. 298)*
initial committed complex *(p. 281)*
initiation complex *(p. 282)*
internal control region (ICR) *(p. 284)*
internal promoter element *(p. 284)*
intrinsic termination *(p. 277)*
intron self-splicing *(p. 294)*
intron splicing *(p. 285)*
intronic splicing enhancer, suppressor (ISE, ISS) *(p. 294)*
inverted repeat *(p. 277)*
lariat intron structure *(p. 289)*
mature mRNA *(p. 285)*
messenger RNA (mRNA) *(p. 270)*
micro RNA (miRNA) *(p. 271)*
minimal initiation complex *(p. 282)*
nucleolus (nucleoli) *(p. 283)*
open promoter complex *(p. 273)*
polyadenylation signal sequence *(p. 286)*
precursor mRNA (pre-mRNA) *(p. 285)*
preinitiation complex (PIC) *(p. 282)*
Pribnow box (–10 consensus sequence) *(p. 273)*
promoter *(p. 271)*
promoter-specific element (PSE) *(p. 284)*
rho-dependent termination (rho protein) *(p. 277)*
rho utilization site (rut site) *(p. 277)*
ribonucleotide (A, U, G, C) *(p. 269)*
ribose *(p. 269)*
ribosomal RNA (rRNA) *(p. 271)*
ribozymes *(p. 271)*
RNA editing *(p. 298)*
RNA polymerase core *(p. 272)*
RNA polymerase *(p. 269)*
RNA polymerase I, II, III (RNA pol I, II, III) *(p. 278)*
sigma (σ) subunit (alternative sigma subunit) *(p. 272)*
silencer sequence *(p. 283)*
small interfering RNA (siRNA) *(p. 271)*
small nuclear RNA (snRNA) *(p. 271)*
spliceosome *(p. 289)*
stem-loop (hairpin structure) *(p. 277)*
TATA box (Goldberg-Hogness box) *(p. 280)*
TATA-binding protein (TBP) *(p. 281)*
TBP-associated factor (TAF) *(p. 281)*
template strand *(p. 271)*
termination region *(p. 272)*
transcription factors (TF) *(p. 281)*
transcription-terminating factor I (TTFI) *(p. 285)*
transfer RNA (tRNA) *(p. 271)*
upstream *(p. 271)*
upstream control element *(p. 283)*
uracil (U) *(p. 269)*

PROBLEMS

MasteringGenetics™ Visit for instructor-assigned tutorials and problems.

Chapter Concepts

For answers to selected even-numbered problems, see Appendix: Answers.

1. Based on discussion in this chapter,
   a. What is a gene?
   b. Why are genes for rRNA and tRNA considered to be genes even though they do not produce polypeptides?
2. In one to two sentences each, describe the three processes that commonly modify eukaryotic pre-mRNA.
3. Answer these questions concerning promoters.
   a. What role do promoters play in transcription?
   b. What is the common structure of a bacterial promoter with respect to consensus sequences?
   c. What consensus sequences are detected in the mammalian β-globin gene promoter?
   d. Eukaryotic promoters are more variable than bacterial promoters. Explain why.
   e. What is the meaning of the term *alternative promoter*? How does the use of alternative promoters affect transcription?
4. The diagram below shows a DNA duplex. The template strand is identified, as is the location of the +1 nucleotide.

   +1

   5′ _____|_____ 3′ template strand

   3′ _____ 5′ coding strand

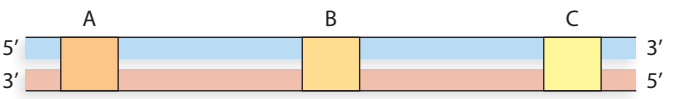
   a. Assume this region contains a gene transcribed in a bacterium. Identify the location of promoter consensus sequences and of the transcription termination sequence.
   b. Assume this region contains a gene transcribed to form mRNA in a eukaryote. Identify the location of the most common promoter consensus sequences.
   c. If this region is a eukaryotic gene transcribed by RNA polymerase III, where are the promoter consensus sequences located?
5. The following is a portion of an mRNA sequence:

   3′-AUCGUCAUGCAGA-5′

   a. During transcription, was the adenine at the left-hand side of the sequence the first or the last nucleotide added to the portion of mRNA shown? Explain how you know.
   b. Write out the sequence and polarity of the DNA duplex that encodes this mRNA segment. Label the template and coding DNA strands.
   c. Identify the direction in which the promoter region for this gene will be located.
6. Compare and contrast the properties of DNA polymerase and RNA polymerase, listing at least three similarities and at least three differences between the molecules.
7. The DNA sequences shown below are from the promoter regions of six bacterial genes. In each case, the last nucleotide in the sequence (highlighted in blue) is the +1 nucleotide that initiates transcription.
   a. Examine these sequences and identify the Pribnow box sequence at approximately −10 for each promoter.
   b. Determine the consensus sequence for the Pribnow box from these sequences.

Gene 1	. . .	TTCCGGCTCGTATGTTGTGTGG	A . . .
Gene 2	. . .	CGTCATTTGATATGATGCGCCCC	G . . .
Gene 3	. . .	CCACTGGCGGTGATACTGAGCAC	A . . .
Gene 4	. . .	TTTATTGCAGCTTATAATGGTTAC	A . . .
Gene 5	. . .	TGCTTCTGACTATAATAGACAGG	G . . .
Gene 6	. . .	AAGTAAACACGGTACGATGTACCAC	A . . .

8. Bacterial and eukaryotic gene transcripts can differ, in the transcripts themselves, in whether the transcripts are modified before translation, and in how the transcripts are modified. For each of these three areas of contrast, describe what the differences are and why the differences exist.
9. Describe the two types of transcription termination found in bacterial genes. How does transcription termination differ for eukaryotic genes?
10. What is the role of enhancer sequences in transcription of eukaryotic genes? Speculate about why enhancers are not part of transcription of bacterial genes.
11. Describe the difference between intron sequences and spacer sequences, such as the spacer sequence depicted in Figure 8.27b.
12. Draw a bacterial promoter and label its consensus sequences. How does this promoter differ from a eukaryotic promoter transcribed by RNA polymerase II? By RNA polymerase I? By RNA polymerase III?
13. How do SR proteins help guide pre-mRNA intron splicing? What is meant by the term *alternative splicing*, and how does variation in SR protein production play a role?
14. Three genes identified in the diagram as *A*, *B*, and *C* are transcribed from a region of DNA. The 5′-to-3′ transcription of genes *A* and *C* elongates mRNA in the right-to-left direction, and transcription of gene *B* elongates mRNA in the left-to-right direction. For each gene, identify the coding strand by designating it as an "upper strand" or "lower strand in the diagram."



Application and Integration

For answers to selected even-numbered problems, see Appendix: Answers.

15. The eukaryotic gene *Gen-100* contains four introns labeled A to D. Imagine that *Gen-100* has been isolated and its DNA has been denatured and mixed with polyadenylated mRNA from the gene.
    a. Illustrate the R-loop structure that would be seen with electron microscopy.
    b. Label the introns.
    c. Are intron regions single stranded or double stranded? Why?

16. The segment of the bacterial *TrpA* gene involved in intrinsic termination of transcription is shown below.

```
3′-TGGGTCGGGGCGGATTACTGCCCCGAAAAAAAACTTG-5′
5′-ACCCAGCCCCGCCTAATGACGGGGCTTTTTTTTGAAC-3′
```

    a. Draw the mRNA structure that forms during transcription of this segment of the *TrpA* gene.
    b. Label the template and coding DNA strands.
    c. Explain how a sequence of this type leads to intrinsic termination of transcription.

17. A 2-kb fragment of *E. coli* DNA contains the complete sequence of a gene for which transcription is terminated by the rho protein. The fragment contains the complete promoter sequence as well as the terminator region of the gene. The cloned fragment is examined by band shift assay (see Research Technique 8.1). Each lane of a single electrophoresis gel contains the 2-kb cloned fragment under the following conditions:

    Lane 1: 2-kb fragment alone

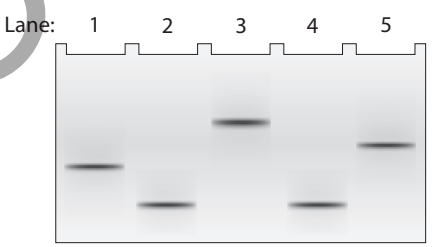
    Lane 2: 2-kb fragment plus the core enzyme

    Lane 3: 2-kb fragment plus the RNA polymerase holoenzyme

    Lane 4: 2-kb fragment plus rho protein

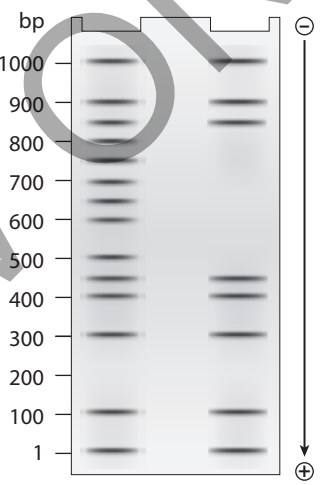
    a. Diagram the relative positions expected for the DNA fragments in this gel retardation analysis.
    b. Explain the relative positions of bands in lanes 1 and 3.
    c. Explain the relative positions of bands in lanes 1 and 4.

18. A 3.5-kb segment of DNA containing the complete sequence of a mouse gene is available. The DNA segment contains the promoter sequence and extends beyond the polyadenylation site of the gene. The DNA is studied by band shift assay (see Research Technique 8.1), and the following gel bands are observed.



    Match these conditions to a specific lane of the gel.
    a. 3.5-kb fragment plus TFIIB and TFIID
    b. 3.5-kb fragment plus TFIIB, TFIID, TFIIF, and RNA polymerase II
    c. 3.5-kb fragment alone
    d. 3.5-kb fragment plus RNA polymerase II
    e. 3.5-kb fragment plus TFIIB

19. A 1.0-kb DNA fragment from the 5′ end of the mouse gene described in the previous problem is examined by DNA footprint protection analysis (see Research Technique 8.1). Two samples are end-labeled with ^{32}P, and one of the two is mixed with TFIIB, TFIID, and RNA polymerase II. The DNA exposed to these proteins is run in the right-hand lane of the gel shown below and the control DNA is run in the left-hand. Both DNA samples are treated with DNase I before running the samples on the electrophoresis gel.



    a. What length of DNA is bound by the transcriptional proteins? Explain how the gel results support this interpretation.
    b. Draw a diagram of this DNA fragment bound by the transcriptional proteins, showing the approximate position of proteins along the fragment. Use the illustration style seen in Research Technique 8.1 as a model.
    c. Explain the role of DNase I.

20. Wild-type *E. coli* grow best at 37°C but can grow efficiently up to 42°C. An *E. coli* strain has a mutation of the sigma subunit that results in an RNA polymerase holoenzyme that is stable and transcribes at wild-type levels at 37°C. The mutant holoenzyme is progressively destabilized as the temperature is raised, and it completely denatures and ceases to carry out transcription at 42°C. Relative to wild-type growth, characterize the ability of the mutant strain to carry out transcription at
    a. 37°C b. 40°C c. 42°C
    d. What term best characterizes the type of mutation exhibited by the mutant bacterial strain? (*Hint*: The term was used in Chapter 4 to describe the Himalayan allele of the mammalian *C* gene.)

21. A mutant strain of *Salmonella* bacteria carries a mutation of the rho protein that has full activity at 37°C but is completely inactivated when the mutant strain is grown at 40°C.
    a. Speculate about the kind of differences you would expect to see if you compared a broad spectrum of mRNAs from

the mutant strain grown at 37°C and the same spectrum of mRNAs from the strain when grown at 40°C.

b. Are all mRNAs affected by the rho protein mutation in the same way? Why or why not?

22. The human β-globin wild-type allele and a certain mutant allele are identical in sequence except for a single base-pair substitution that changes one nucleotide at the end of intron 2. The wild-type and mutant sequences of the affected portion of pre-mRNA are

	Intron 2	Exon 3
wild type	5′-CCUCCCACAG	CUCCUG-3′
mutant	5′-CCUCCCACUG	CUCCUG-3′

a. Speculate about the way in which this base substitution causes mutation of β-globin protein.

b. This is one example of how DNA sequence change occurring somewhere other than in an exon can produce mutation. List other kinds of DNA sequence changes occurring outside exons that can produce mutation. In each case, characterize the kind of change you would expect to see in mutant mRNA or mutant protein.

23. Microbiologists describe the processes of transcription and translation as "coupled" in bacteria. This term indicates that a bacterial mRNA can be undergoing transcription at the same moment it is also undergoing translation.

a. How is coupling of transcription and translation possible in bacteria?

b. Is coupling of transcription and translation possible in single-celled eukaryotes such as yeast? Why or why not?

24. A full-length eukaryotic gene is inserted into a bacterial chromosome. The gene contains a complete promoter sequence and a functional polyadenylation sequence, and it has wild-type nucleotides throughout the transcribed region. However, the gene fails to produce a functional protein.

a. List at least three possible reasons why this eukaryotic gene is not expressed in bacteria.

b. What changes would you recommend to permit expression of this eukaryotic gene in a bacterial cell?

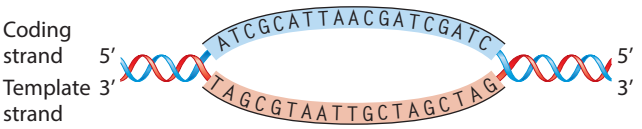
25. The accompanying illustration shows a portion of a gene undergoing transcription. The template and coding strands for the gene are labeled, and a segment of DNA sequence is given. For this gene segment:

a. Superimpose a drawing of RNA polymerase as it nears the end of transcription of the DNA sequence.

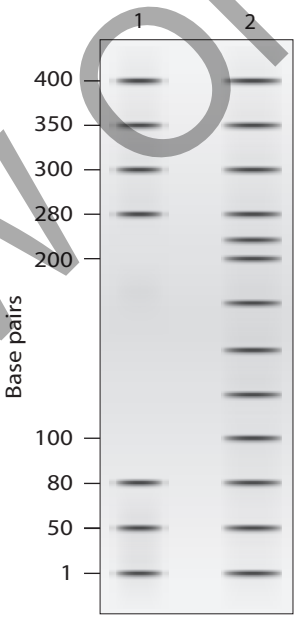
b. Indicate the direction in which RNA polymerase moves as it transcribes this gene.

c. Write the polarity and sequence of the RNA transcript from the DNA sequence given.

d. Identify the direction in which the promoter for this gene is located.



26. DNA footprint protection (described in Research Technique 8.1) is a method that determines whether proteins bind to a specific sample of DNA and thus protect part of the DNA from random enzymatic cleavage by DNase I. A 400-bp segment of cloned DNA is thought to contain a promoter. The cloned DNA is analyzed by DNA footprinting to determine if it has the capacity to act as a promoter sequence. The gel shown below has two lanes, each containing the cloned 400-bp DNA fragment treated with DNase I to randomly cleave unprotected DNA. Lane 1 is cloned DNA that was mixed with RNA polymerase II and several TFII transcription factors before exposure to DNase I. Lane 2 contains cloned DNA that was exposed only to DNase I. RNA pol II and TFIIs were not mixed with DNA before adding DNase I.



a. Explain why this gel provides evidence that the cloned DNA may act as a promoter sequence.

b. Approximately what length is the DNA region protected by RNA pol II and TFIIs?

c. What additional genetic experiments would you suggest to verify that this region of cloned DNA contains a functional promoter?

27. Suppose you have a 1-kb segment of cloned DNA that is suspected to contain a eukaryotic promoter including a TATA box, a CAAT box, and an upstream GC-rich sequence. The clone also contains a gene whose transcript is readily detectable. Your laboratory supervisor asks you to outline an experiment that will (1) determine if eukaryotic transcription factors (TF) bind to the fragment and, if so, (2) identify where on the fragment the transcription factors bind. All necessary reagents, equipment, and experimental know-how are available in the laboratory. Your assignment is to propose techniques to be used to address the three items your supervisor has listed and to describe the kind of results that would indicate binding of TF to the DNA, the location of the binding. (*Hint:* The techniques and general results are discussed in this chapter.)