

7 Transcription and Post-Transcriptional Modification

7.1 Transcription of RNA from DNA

All cellular RNAs are synthesized from a DNA template through the process of transcription (**Figure 7.1**). Transcription is in many ways similar to the process of replication, but one fundamental difference relates to the length of the template used. During replication, all the nucleotides in the DNA template are copied, but, during transcription, only small parts of the DNA molecule—usually a single gene or, at most, a few genes—are transcribed into RNA. Because not all gene products are needed at the same time or in the same cell, it would be highly inefficient for a cell to constantly transcribe all of its genes. Furthermore, much of the DNA does not code for a functional product, and transcription of such sequences would be pointless. Transcription is, in fact, a highly selective process—individual genes are transcribed only as their products are needed. But this selectivity imposes a fundamental problem on the cell—the problem of how to recognize individual genes and transcribe them at the proper time and place.

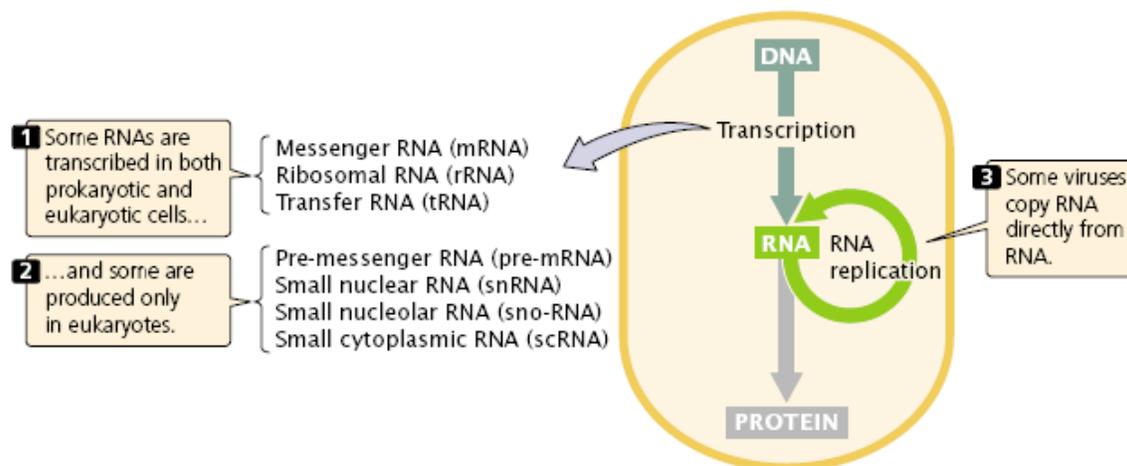


Figure 7.1: All cellular types of RNA are transcribed from DNA.

Like replication, transcription requires three major components:

- (1) A DNA template;
- (2) The raw materials (substrates) needed to build a new RNA molecule; and
- (3) The transcription apparatus, consisting of the proteins necessary to catalyze the synthesis of RNA.

7.1.1 The Template

7.1.1.1 The Transcribed Strand

The template for RNA synthesis, as for DNA synthesis, is a single strand of the DNA double helix. Unlike replication, however, transcription typically takes place on only one of the two nucleotide strands of DNA (**Figure 7.2**). The nucleotide strand used for transcription is termed the **template strand**. The other strand, called the **nontemplate strand**, is not ordinarily transcribed. Thus, in any one section of DNA, only one of the nucleotide strands normally carries the genetic information that is transcribed into RNA (there *are* some exceptions to this rule).

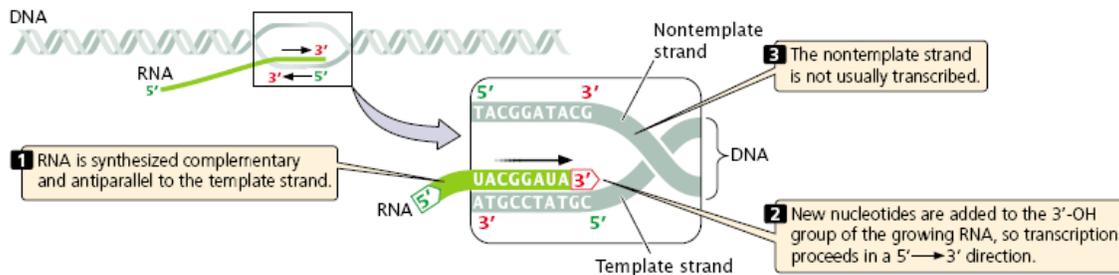


Figure 7.2: RNA molecules are synthesized that are complementary and antiparallel to one of the two nucleotide strands of DNA, the template strand.

In most organisms, each gene is transcribed from a single strand, but different genes may be transcribed from different strands (**Figure 7.3**).

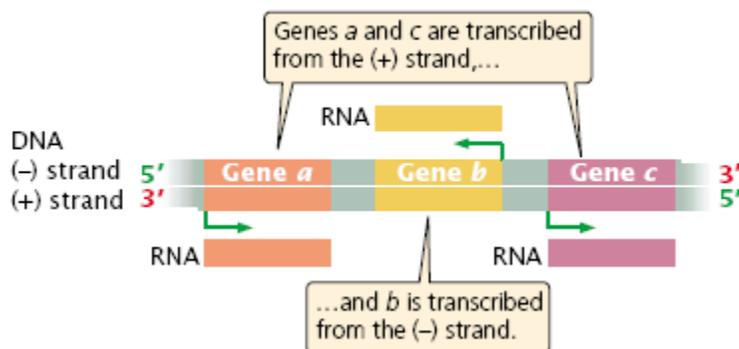


Figure 7.3: RNA is transcribed from one DNA strand. In most organisms, each gene is transcribed from a single DNA strand, but different genes may be transcribed from one or the other of the two DNA strands.

Concepts: Within a single gene, only one of the two DNA strands, the template strand, is generally transcribed into RNA.

7.1.1.2 The Transcription Unit

A **transcription unit** is a stretch of DNA that codes for an RNA molecule and the sequences necessary for its transcription. In eukaryotes, alternative RNA molecules can be produced from each transcription unit. How does the complex of enzymes and proteins that performs transcription—the transcription apparatus—recognize a transcription unit? How does it know which DNA strand to read, and where to start and stop? This information is encoded by the DNA sequence. Included within a transcription unit are three critical regions: a promoter, an RNA coding sequence, and a terminator (**Figure 7.4**). The **promoter** is a DNA sequence that the transcription apparatus recognizes and binds. It indicates which of the two DNA strands is to be read as the template and the direction of transcription. The promoter also determines the transcription start site, the first nucleotide that will be transcribed into RNA. In most transcription units, the promoter is located next to the transcription start site but is not, itself, transcribed.

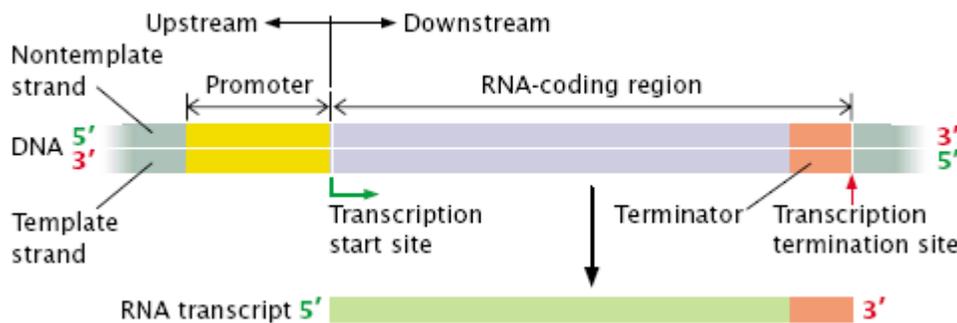


Figure 7.4: A transcription unit includes a promoter, an RNA-coding region, and a terminator.

The second critical region of the transcription unit is the **RNA-coding region**, a sequence of DNA nucleotides that is copied into an RNA molecule. A third component of the transcription unit is the **terminator**, a sequence of nucleotides that signals where transcription is to end. Terminators are usually part of the coding sequence; that is, transcription stops only after the terminator has been copied into RNA.

7.1.2 The Substrate for Transcription

RNA is synthesized from **ribonucleoside triphosphates (rNTPs)** (**Figure 7.5**). In synthesis, nucleotides are added one at a time to the 3'-OH group of the growing RNA molecule. Two phosphates are cleaved from the incoming

ribonucleoside triphosphate; the remaining phosphate participates in a phosphodiester bond that connects the nucleotide to the growing RNA molecule.

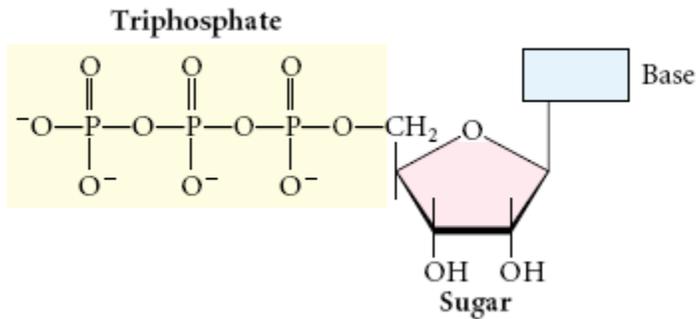


Figure 7.5: Ribonucleoside triphosphates are substrates used in RNA synthesis.

The overall chemical reaction for the addition of each nucleotide is:



where PP_i represents two atoms of inorganic phosphorus. Nucleotides are always added to the 3' end of the RNA molecule, and the direction of transcription is therefore 5'→3' (**Figure 7.6**), the same as the direction of DNA synthesis during replication. RNA is made complementary and antiparallel to one of the DNA strands (the template strand).

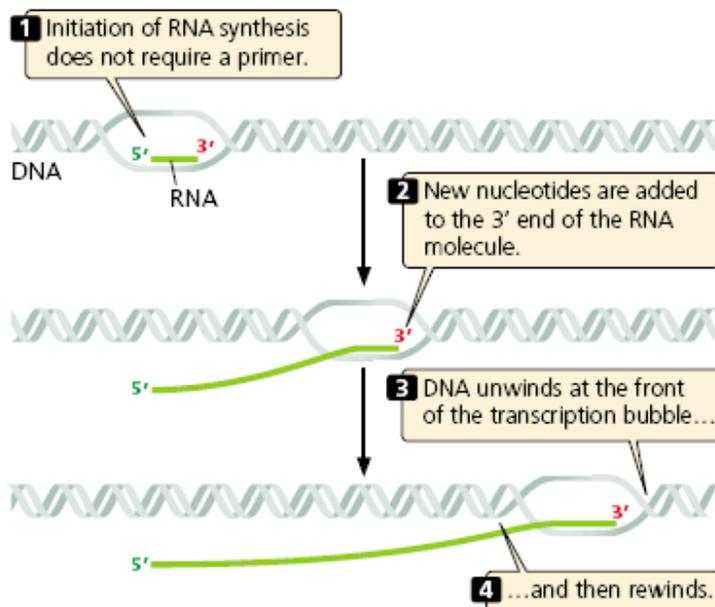


Figure 7.6: In transcription, nucleotides are always added to the 3' end of the RNA molecule.

Concepts: RNA is synthesized from ribonucleoside triphosphates. Transcription is 5'→3': each new nucleotide is joined to the 3'-OH group of the last nucleotide added to the growing RNA molecule. RNA synthesis does not require a primer.

7.1.3 The Transcription Apparatus

Recall that, in replication, a number of different enzymes and proteins are required to bring about DNA synthesis. Although transcription might initially appear to be quite different, because a single enzyme—**RNA polymerase**—carries out all the required steps of transcription, on closer inspection, the processes are actually similar. The action of RNA polymerase is enhanced by a number of accessory proteins that join and leave the polymerase at different stages of the process. Each accessory protein is responsible for providing or regulating a special function. Thus, transcription, like replication, requires an array of proteins.

7.1.3.1 Bacterial RNA Polymerase

Bacterial cells typically possess only one type of RNA polymerase, which catalyzes the synthesis of all classes of bacterial RNA: mRNA, tRNA, and rRNA. Bacterial RNA polymerase is a large, multimeric enzyme (meaning that it consists of several polypeptide chains). At the heart of bacterial RNA polymerase are four subunits (individual polypeptide chains) that make up the **core enzyme**: two copies of a subunit called alpha (α), a single copy of beta (β), and single copy of beta prime (β') (**Figure 7.7**). The core enzyme catalyzes the elongation of the RNA molecule by the addition of RNA nucleotides.

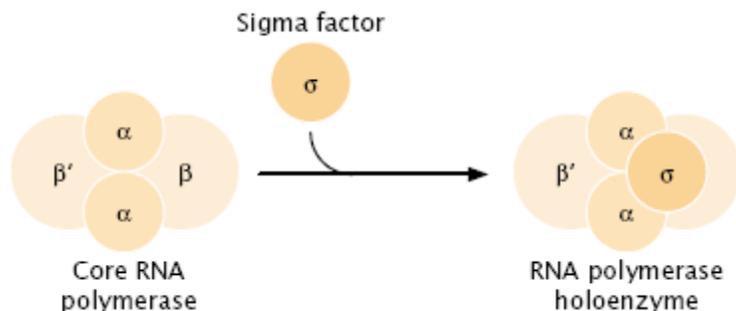


Figure 7.7: In bacterial RNA polymerase, the core enzyme consists of four subunits: two copies of alpha (α), a single copy of beta (β), and single copy of beta prime (β'). The core enzyme catalyzes the elongation of the RNA molecule by the addition of RNA nucleotides. The sigma factor (σ) joins the core to form the holoenzyme, which is capable of binding to a promoter and initiating transcription.

Other functional subunits join and leave the core enzyme at particular stages of the transcription process. The **sigma** (σ) **factor** controls the binding of the RNA polymerase to the promoter. Without sigma, RNA polymerase will initiate transcription at a random point along the DNA. After sigma has associated with the

core enzyme (forming a **holoenzyme**), RNA polymerase binds stably only to the promoter region and initiates transcription at the proper start site. Sigma is required only for promoter binding and initiation; when a few RNA nucleotides have been joined together, sigma detaches from the core enzyme.

Many bacteria possess multiple types of sigma. *E. coli*, for example, possesses sigma 28 (σ^{28}), sigma 32 (σ^{32}), sigma 54 (σ^{54}), and sigma 70 (σ^{70}), named on the basis of their molecular weights. Each type of sigma initiates the binding of RNA polymerase to a particular set of promoters. For example, σ^{32} binds to promoters of genes that protect against environmental stress, σ^{54} binds to promoters of genes used during nitrogen starvation, and σ^{70} binds to many different promoters. Other subunits provide the core RNA polymerase with additional functions. Rho (ρ) and NusA, for example, facilitate the termination of transcription.

7.1.3.2 Eukaryotic RNA Polymerases

Eukaryotic cells possess three distinct types of RNA polymerase, each of which is responsible for transcribing a different class of RNA: **RNA polymerase I** transcribes rRNA; **RNA polymerase II** transcribes pre-mRNAs, snoRNAs, and some snRNAs; and **RNA polymerase III** transcribes small RNA molecules—specifically tRNAs, small rRNA, and some snRNAs (**Table 7.1**). All three eukaryotic polymerases are large, multimeric enzymes, typically consisting of more than a dozen subunits. Some subunits are common to all three RNA polymerases, whereas others are limited to one of the polymerases. As in bacterial cells, a number of accessory proteins bind to the core enzyme and affect its function.

Table 7.1: Eukaryotic RNA polymerases

Type	Location	Transcribes
RNA polymerase I	Nucleolus	Large rRNAs
RNA polymerase II	Nucleoplasm	Pre-mRNA, some snRNAs, snoRNAs
RNA polymerase III	Nucleoplasm	tRNAs, small rRNA, snRNAs
Mitochondrial	Mitochondria	Mitochondrial RNA
Chloroplast	Chloroplasts	Chloroplast RNA

Concepts: Bacterial cells possess a single type of RNA polymerase, consisting of a core enzyme and other subunits that participate in various stages of transcription. Eukaryotic cells possess three distinct types of RNA polymerase: RNA polymerase I transcribes rRNA; RNA polymerase II transcribes pre-mRNA, snoRNAs, and some snRNAs; and RNA polymerase III transcribes tRNAs, small rRNAs, and some snRNAs.

7.2 The Process of Bacterial Transcription

Now that we've considered some of the major components of transcription, we're ready to take a detailed look at the process. Transcription can be conveniently divided into three stages:

- (1) **Initiation**, in which the transcription apparatus assembles on the promoter and begins the synthesis of RNA;
- (2) **Elongation**, in which RNA polymerase moves along the DNA, unwinding it and adding new nucleotides, one at a time, to the 3' end of the growing RNA strand; and
- (3) **Termination**, the recognition of the end of the transcription unit and the separation of the RNA molecule from the DNA template.

7.2.1 Initiation

Initiation includes all the steps necessary to begin RNA synthesis, including (i) promoter recognition, (ii) formation of the transcription bubble, (iii) creation of the first bonds between rNTPs, and (iv) escape of the transcription apparatus from the promoter.

Transcription initiation requires that the transcription apparatus recognize and bind to the promoter. At this step, the selectivity of transcription is enforced; the binding of RNA polymerase to the promoter determines which parts of the DNA template are to be transcribed and how often. Different genes are transcribed with different frequencies, and promoter binding is primarily responsible for determining the frequency of transcription for a particular gene. Promoters also have different affinities for RNA polymerase. Even within a single promoter, the affinity can vary over time, depending on its interaction with RNA polymerase and a number of other factors.

7.2.1.1 Bacterial Promoters

Essential information for the transcription unit—where it will start transcribing, which strand is to be read, and in what direction the RNA polymerase will move—is imbedded in the nucleotide sequence of the **promoter**. Promoters are sequences in the DNA that are recognized by the transcription apparatus and are required for transcription to take place. In bacterial cells, promoters are usually adjacent to an RNA coding sequence.

The examination of many promoters in *E. coli* and other bacteria reveals a general feature: although most of the nucleotides within the promoters vary in sequence, short stretches of nucleotides are common to many. Furthermore, the spacing and location of these nucleotides relative to the transcription start site are similar in most promoters. These short stretches of common nucleotides are called **consensus sequences**.

The term “consensus sequence” refers to sequences that possess considerable similarity or consensus. By definition, the consensus sequence comprises the most commonly encountered nucleotides found at a specific location. For example, consider the following nucleotides found near the transcription start site of four prokaryotic genes.

5' - A A T A A A - 3'
 5' - T A T T T T - 3'
 5' - T T T A A T - 3'
 5' - T A A A A T - 3'

Consensus sequence = 5' - T A T A A T - 3'

If two bases are equally frequent, they are designated by listing both bases separated by a line or a slash, as in 5' - T A T A A A/T - 3'. Purines can be indicated by the abbreviation R, pyrimidines by Y, and any nucleotide by N. For example, the consensus sequence 5' - T A **Y** A **R** **N** A - 3' means that the third nucleotide in the consensus sequence (**Y**) is usually a pyrimidine, but either pyrimidine is equally likely. Similarly, the fifth nucleotide in the sequence (**R**) is most likely one of the purines, but both are equally frequent. In the sixth position (**N**), no particular base is more common than any other. The presence of consensus in a set of nucleotides usually implies that the sequence is associated with an important function.

Consensus exists in a sequence because natural selection has favored a restricted set of nucleotides in that position. The most commonly encountered consensus sequence, found in almost all bacterial promoters, is located just upstream of the start site, centered on position -10. Called the **-10 consensus sequence** or, sometimes, the Pribnow box, its sequence is

5' TATAAT 3'
 3' ATATTA 5'

often written simply as **TATAAT (Figure 7.8)**.

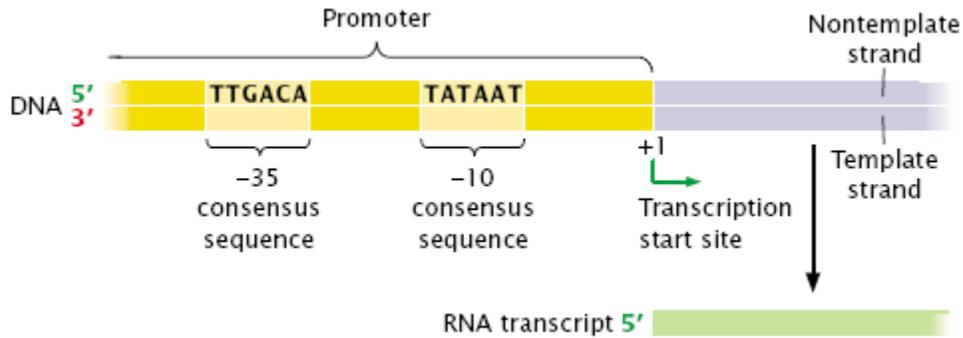


Figure 7.8: In bacterial promoters, consensus sequences are found upstream of the start site, approximately at positions -10 and -35 .

Remember that TATAAT is just the *consensus* sequence—representing the most commonly encountered nucleotides at each of these positions. In most prokaryotic promoters, the actual sequence is not TATAAT (**Figure 7.9**).

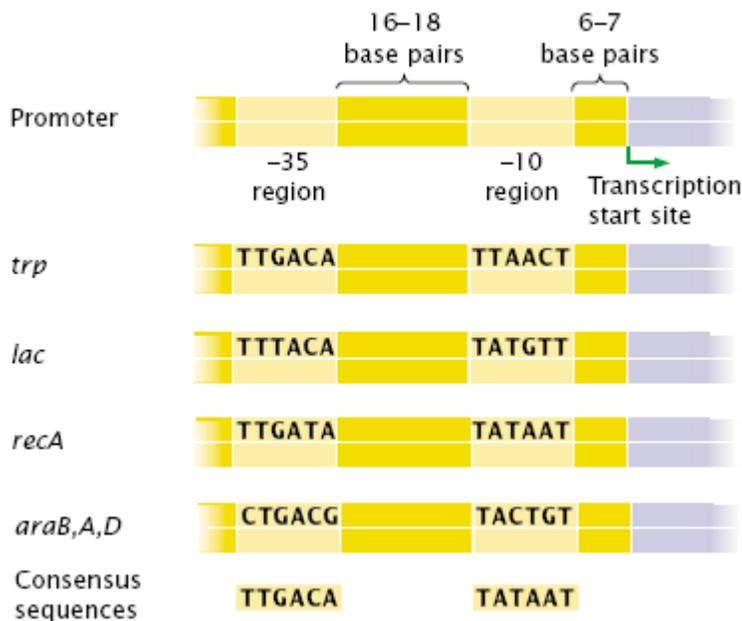


Figure 7.9: In most prokaryotic promoters, the actual sequence is not TATAAT. The sequences shown are found in five *E. coli* promoters, including those of genes for tryptophan biosynthesis (*trp*), tyrosine tRNA (*tRNA^{Tyr}*), lactose metabolism (*lac*), a recombination protein (*recA*), and arabinose metabolism (*araB, A, D*). These sequences are on the nontemplate strand and read 5'→3', left to right.

Another consensus sequence common to most bacterial promoters is **TTGACA**, which lies approximately 35 nucleotides upstream of the start site and is termed the **-35 consensus sequence** (**Figure 7.8**). The nucleotides on either side of the -10 and

-35 consensus sequences and those between them vary greatly from promoter to promoter, suggesting that they are relatively unimportant in promoter recognition.

The function of these consensus sequences in bacterial promoters has been studied by inducing mutations at various positions within the consensus sequences and observing the effect of the changes on transcription. The results of these studies reveal that most base substitutions within the -10 and -35 consensus sequences reduce the rate of transcription; these substitutions are termed *down mutations* because they slow down the rate of transcription. Occasionally, a particular change in a consensus sequence increases the rate of transcription; such a change is called an *up mutation*.

The sigma factor associates with the core enzyme (**Figure 7.10a**) to form a holoenzyme, which binds to the -35 and -10 consensus sequences in the DNA promoter (**Figure 7.10b**). Although it binds only the nucleotides of consensus sequences, the enzyme extends from -50 to -20 when bound to the promoter. The holoenzyme initially binds weakly to the promoter but then undergoes a change in structure that allows it to bind more tightly and unwind the double-stranded DNA (**Figure 7.10c**). Unwinding begins within the -10 consensus sequence and extends downstream for about 17 nucleotides, including the start site.

Some bacterial promoters contain a third consensus sequence that also takes part in the initiation of transcription. Called the **upstream element**, this sequence contains a number of A-T pairs and is found at about -40 to -60.

The alpha subunit of the RNA polymerase interacts directly with this upstream element, greatly enhancing the rate of transcription in those bacterial promoters that possess it. A number of other proteins may bind to sequences in and near the promoter; some stimulate the rate of transcription and others repress it.

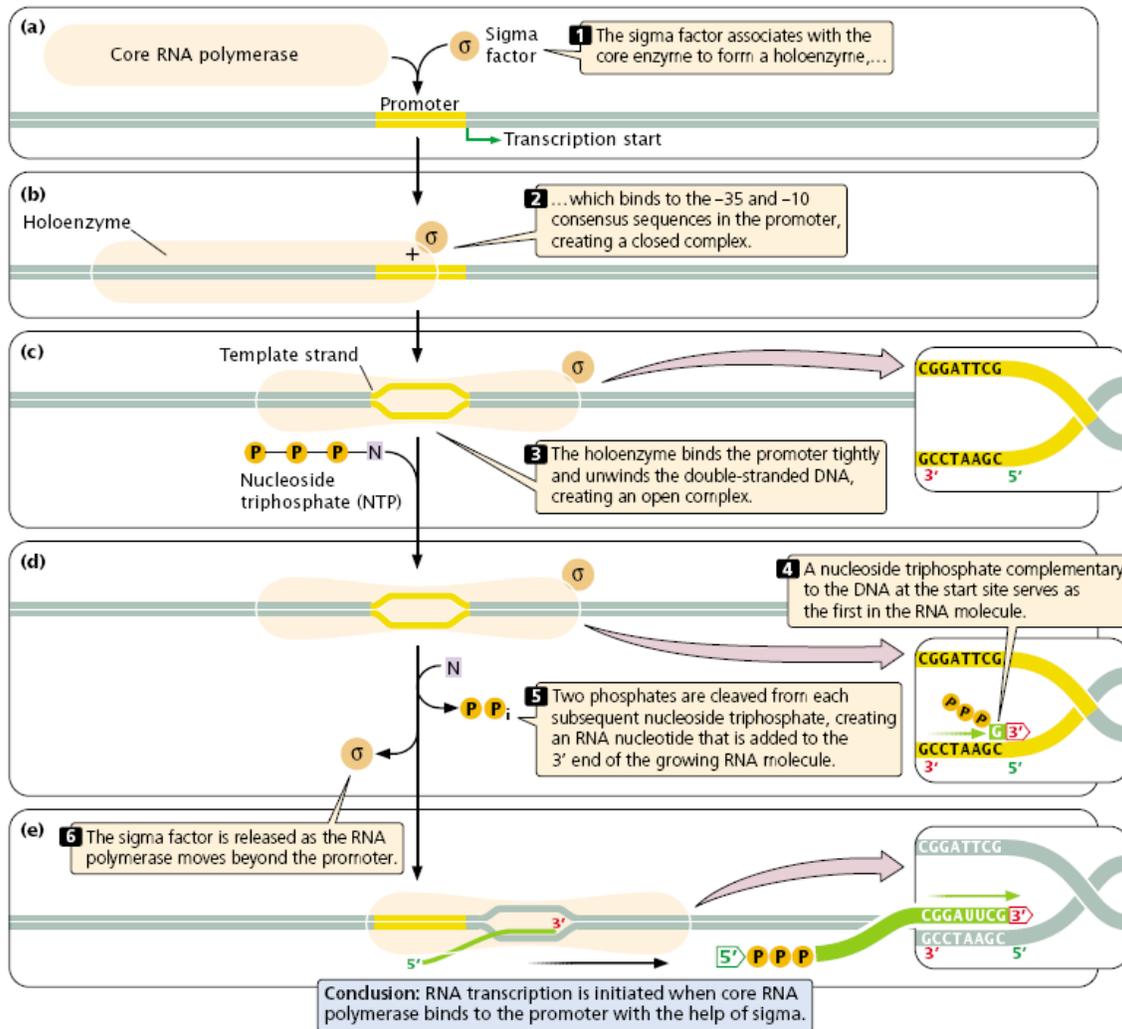


Figure 7.10: Transcription in bacteria is carried out by RNA polymerase, which must bind to the sigma factor to initiate transcription.

7.2.1.2 Promoter Strength: Interaction of RNA Polymerase with Promoter

Promoter strength depends on several factors.

- (1) The three important regions of the promoter that affect how it functions and the efficiency of transcription are (i) the -35 consensus region, (ii) the -10 consensus Pribnow sequence, and (iii) the initiation site. The -35 consensus sequence is the site of the initial recognition between RNA polymerase and the DNA, and the -10 consensus Pribnow sequence is the center of the region of the DNA that is unwound. The nucleotide sequence immediately surrounding the starting point of initiation of transcription may also influence initiation and how quickly the RNA polymerase leaves the promoter site. All of these features

have an influence on the efficiency of RNA polymerase to carry out transcription. Promoters with **nucleotide sequences** that favour efficient transcription by RNA polymerase are referred to as **strong promoters**. Promoters with nucleotide sequences that favour inefficient transcription by RNA polymerase are **weak promoters**.

- (2) In addition to nucleotide sequences at the three critical regions, promoter strength is also influenced by the ability to separate the DNA strands. **Strand separation** is directly influenced by the nucleotide composition (AT-rich regions separate more easily than GC-rich regions).
- (3) Promoter strength is also influenced by the **degree of supercoiling** of the double helix. Bacterial and eukaryotic RNA polymerases can initiate transcription more easily at most promoters when the DNA is supercoiled rather than relaxed.
- (4) Interference with the activity at the **topoisomerases**, which directly affects the degree of supercoiling or unwinding of the DNA, also affects transcription.
- (5) The conformation of the DNA also influences the strength of the promoter. In bacterial cells, RNA polymerase binding to the DNA results in **bending of the DNA**, changing the strength of the promoter. Several transcriptional activators also bend DNA and alter the localized conformation of the promoter region. The bending of DNA plays an important role in gene expression because it influences the interaction of DNA with the sigma factor of RNA polymerase that establishes the strength of the promoter.

Concepts: A promoter is a DNA sequence that is adjacent to a gene and required for transcription. Promoters contain short consensus sequences that are important in the initiation of transcription.

7.2.1.3 Initial RNA Synthesis

After the holoenzyme has attached to the promoter, RNA polymerase is positioned over the start site for transcription (at position +1) and has unwound the DNA to produce a single-stranded template. The orientation and spacing of consensus sequences on a DNA strand determine which strand will be the template for transcription, and thereby determine the direction of transcription. The start site itself is not marked by a consensus sequence but often has the sequence CAT, with the start site at the A. The position of the start site is determined not by the sequences located there but by the location of the consensus sequences, which positions RNA polymerase so that the enzyme's active site is aligned for initiation of

transcription at +1. If the consensus sequences are artificially moved upstream or downstream, the location of the starting point of transcription correspondingly changes.

To begin the synthesis of an RNA molecule, RNA polymerase pairs the base on a ribonucleoside triphosphate with its complementary base at the start site on the DNA template strand (**Figure 7.10d**). No primer is required to initiate the synthesis of the 5' end of the RNA molecule. Two of the three phosphates are cleaved from the ribonucleoside triphosphate as the nucleotide is added to the 3' end of the growing RNA molecule. However, because the 5' end of the first ribonucleoside triphosphate does not take part in the formation of a phosphodiester bond, all three of its phosphates remain. An RNA molecule therefore possesses, at least initially, three phosphates at its 5' end (**Figure 7.10e**).

7.2.2 Elongation

After initiation, RNA polymerase moves downstream along the template, progressively unwinding the DNA at the leading (downstream) edge of the transcription bubble, joining nucleotides to the RNA molecule according to the sequence on the template, and rewinding the DNA at the trailing (upstream) edge of the bubble. In bacterial cells at 37°C, about 40 nucleotides are added per second. This rate of RNA synthesis is much lower than that of DNA synthesis, which is more than 1,500 nucleotides per second in bacterial cells.

Transcription takes place within a short stretch of about 18 nucleotides of unwound DNA—the transcription bubble. Within this region, RNA is continuously synthesized, with single-stranded DNA used as a template. About 8 nucleotides of newly synthesized RNA are paired with the DNA-template nucleotides at any one time. As the transcription apparatus moves down the DNA template, it generates positive supercoiling ahead of the transcription bubble and negative supercoiling behind it. Topoisomerase enzymes probably relieve the stress associated with the unwinding and rewinding of DNA in transcription, as they do in DNA replication.

Concepts: Transcription is initiated at the start site, which, in bacterial cells, is set by the binding of RNA polymerase to the consensus sequences of the promoter. Transcription takes place within the transcription bubble. DNA is unwound ahead of the bubble and rewound behind it.

7.2.3 Termination

RNA polymerase moves along the template, adding nucleotides to the 3' end of the growing RNA molecule until it transcribes a terminator. Most terminators are found upstream of the point of termination. Transcription therefore does not suddenly end when polymerase reaches a terminator, like a car stopping in front of a stop sign. Rather, transcription ends after the terminator has been transcribed, like a car that stops only after running over a speed bump. At the terminator, several overlapping events are needed to bring an end to transcription: (i) RNA polymerase must stop synthesizing RNA, (ii) the RNA molecule must be released from RNA polymerase, (iii) the newly made RNA molecule must dissociate fully from the DNA, and (iv) RNA polymerase must detach from the DNA template. Bacterial cells possess two major types of terminators.

7.2.3.1 Rho-Independent Terminators

Rho-independent terminators are able to cause the end of transcription in the absence of rho. Rho-independent terminators have two common features. First, they contain inverted repeats (sequences of nucleotides on one strand that are inverted and complementary). When inverted repeats have been transcribed into RNA, a hairpin secondary structure forms (**Figure 7.11**). Second, in rho-independent terminators, a string of approximately six adenine nucleotides follows the second inverted repeat in the template DNA. Their transcription produces a string of uracil nucleotides after the hairpin in the transcribed RNA.

The presence of a hairpin in an RNA transcript causes RNA polymerase to slow down or pause, which creates an opportunity for termination. The adenine–uracil base pairings downstream of the hairpin are relatively unstable compared with other base pairings, and the formation of the hairpin may itself destabilize the DNA–RNA pairing, causing the RNA molecule to separate from its DNA template. When the RNA transcript has separated from the template, RNA synthesis can no longer continue (see **Figure 7.11**).

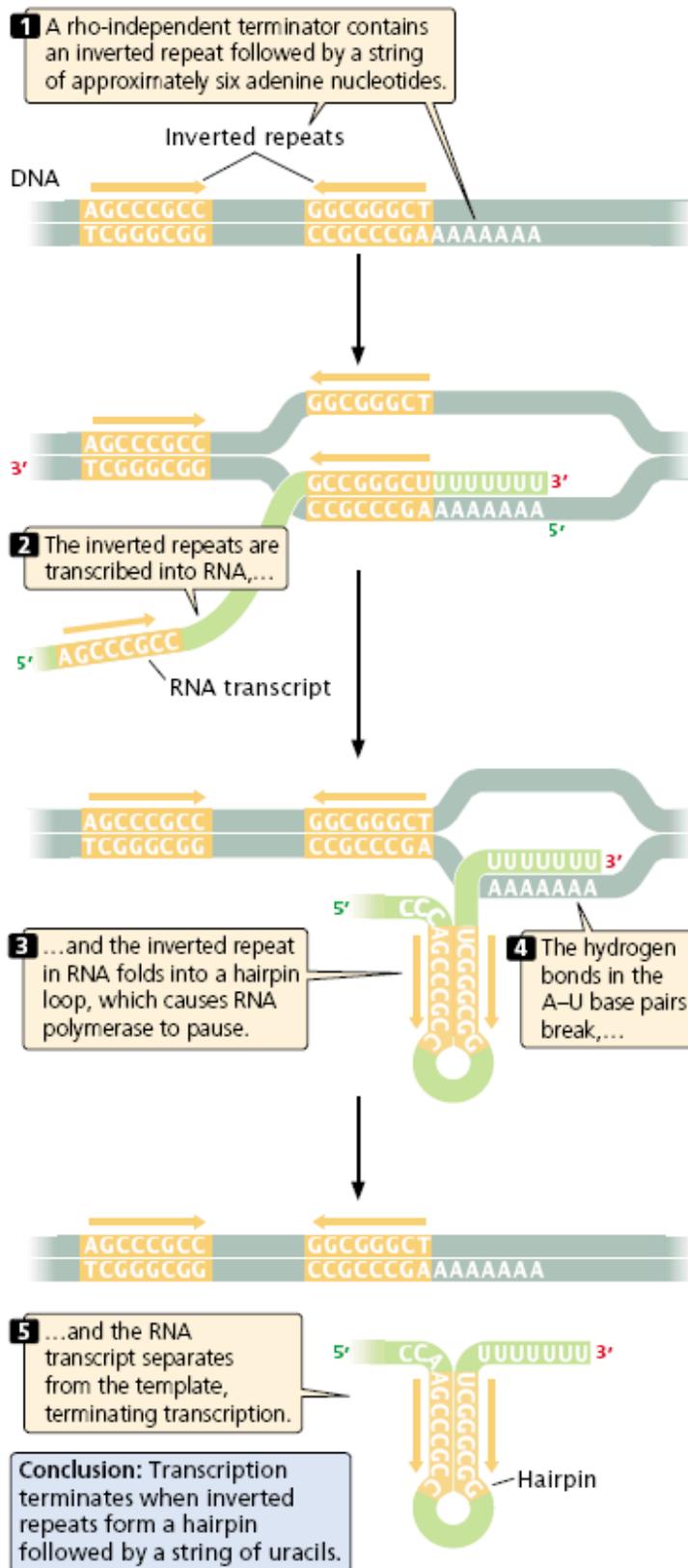


Figure 7.11: Termination by bacterial rho-independent terminators is a multistep process.

7.2.3.2 Rho-Dependent Terminators

Rho-dependent terminators are able to cause the termination of transcription only in the presence of an ancillary protein called the **rho factor**. Rho-dependent terminators have two features: (1) DNA sequences that produce a pause in transcription; and (2) a DNA sequence that encodes a stretch of RNA upstream of the terminator that is devoid of any secondary structures. This unstructured RNA serves as binding site for the rho protein, which binds the RNA and moves toward its 3' end, following the RNA polymerase (**Figure 7.12**). When RNA polymerase encounters the terminator, it pauses, allowing rho to catch up. The rho protein has helicase activity, which it uses to unwind the RNA–DNA hybrid in the transcription bubble, bringing an end to transcription.

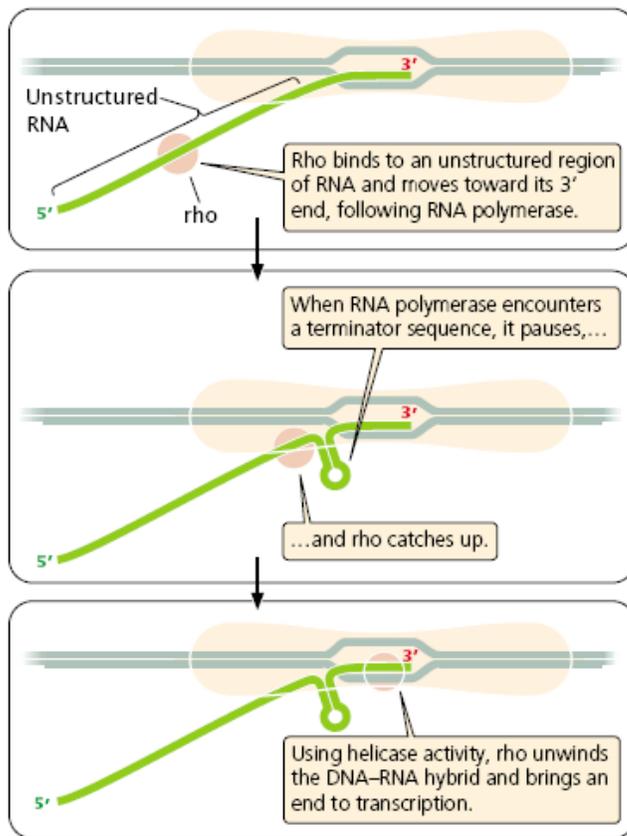


Figure 7.12: The termination of transcription in some bacterial genes requires the presence of the rho protein.

7.3 The Basic Rules of Transcription

- (1) Transcription is a selective process; only certain parts of the DNA are transcribed.
- (2) RNA is transcribed from single-stranded DNA. Normally, only one of the two DNA strands—the template strand—is copied into RNA.
- (3) Ribonucleoside triphosphates are used as the substrates in RNA synthesis. Two phosphates are cleaved from a ribonucleoside triphosphate, and the resulting nucleotide is joined to the 3'-OH group of the growing RNA strand.
- (4) RNA molecules are antiparallel and complementary to the DNA template strand. Transcription is always in the 5'→3' direction, meaning that the RNA molecule grows at the 3' end.
- (5) Transcription depends on RNA polymerase—a complex, multimeric enzyme. RNA polymerase consists of a core enzyme, which is capable of synthesizing RNA, and other subunits that may join transiently to perform additional functions.
- (6) The core enzyme of RNA polymerase requires a sigma factor in order to bind to a promoter and initiate transcription.
- (7) Promoters contain short sequences crucial in the binding of RNA polymerase to DNA; these consensus sequences are interspersed with nucleotides that play no known role in transcription.
- (8) RNA polymerase binds to DNA at a promoter, begins transcribing at the start site of the gene, and ends transcription after a terminator has been transcribed.

Concepts: Transcription ends after RNA polymerase transcribes a terminator. Bacterial cells possess two types of terminator: a rho-independent terminator, which RNA polymerase can recognize by itself; and a rho-dependent terminator, which RNA polymerase can recognize only with the help of the rho protein.

7.4 The Process of Eukaryotic Transcription

The process of eukaryotic transcription is similar to that of bacterial transcription. Eukaryotic transcription also includes initiation, elongation, and termination, and the basic principles of transcription already outlined apply to eukaryotic transcription. However, there are some important differences. (i) Eukaryotic cells possess three different RNA polymerases, each of which transcribes a different class of RNA and recognizes a different type of promoter. (ii) Another difference is in the nature of promoter recognition and initiation. Many proteins take part in the binding of

eukaryotic RNA polymerases to DNA templates, and the different types of promoters require different proteins.

7.4.1 Transcription and Nucleosome Structure

Transcription requires that sequences on DNA are accessible to RNA polymerase and other proteins. However, in eukaryotic cells, DNA is complexed with histone proteins in highly compressed chromatin.

How can the proteins necessary for transcription gain access to eukaryotic DNA when it is complexed with histones?

The answer to this question is that, before transcription, the chromatin structure is modified so that the DNA is in a more open configuration and is more accessible to the transcription machinery. Several types of proteins have roles in chromatin modification. (i) Acetyltransferases add acetyl groups to amino acids at the ends of the histone proteins, which destabilizes the nucleosome structure and makes the DNA more accessible. (ii) Other types of histone modification also can affect chromatin packing. (iii) In addition, proteins called chromatin-remodeling proteins may bind to the chromatin and displace nucleosomes from promoters and other regions important for transcription.

7.4.2 Transcription Initiation

7.4.2.1 DNA Sequences

The initiation of transcription is a complex processes in eukaryotic cells because of the variety of initiation sequences and because numerous proteins bind to these sequences. Two broad classes of DNA sequences are important for the initiation of transcription: (i) promoters and (ii) enhancers. A **promoter** is always found adjacent to (or sometimes within) the gene that it regulates and has a fixed location with regard to the transcription start point. An **enhancer**, in contrast, need not be adjacent to the gene; enhancers can affect the transcription of genes that are thousands of nucleotides away, and their positions relative to start sites can vary.

7.4.2.2 RNA Polymerases

A significant difference between bacterial and eukaryotic transcription is the existence of three different eukaryotic RNA polymerases, which recognize different types of promoters. In bacterial cells, the holoenzyme (RNA polymerase plus sigma)

recognizes and binds directly to sequences in the promoter. In eukaryotic cells, promoter recognition is carried out by accessory proteins that bind to the promoter and then recruit a specific RNA polymerase (I, II, or III) to the promoter.

7.4.2.3 Transcription Factors

- (1) One class of accessory proteins comprises **general transcription factors**, which, along with RNA polymerase, form the **basal transcription apparatus** that assembles near the start site and is sufficient to initiate minimal levels of transcription.
- (2) Another class of accessory proteins consists of **transcriptional activator proteins**, which bind to specific DNA sequences and bring about higher levels of transcription by stimulating the assembly of the basal transcription apparatus at the start site.

Concepts: Two classes of DNA sequences in eukaryotic cells affect transcription: enhancers and promoters. A promoter is near the gene and has a fixed position relative to the start site of transcription. An enhancer can be distant from the gene and variable in location.

7.4.3 RNA Polymerase II Promoters

We will focus most of our attention on promoters recognized by RNA polymerase II, which transcribes the genes that encode proteins. A promoter for a gene transcribed by RNA polymerase II typically consists of two primary parts: (i) the core promoter and (ii) the regulatory promoter.

7.4.3.1 Core Promoter

The core promoter is located immediately upstream of the gene (**Figure 7.13**) and typically includes one or more **consensus sequences**.

- (1) The most common of these consensus sequences is the **TATA box**, which has the consensus sequence TATAAA and is located from –25 to –30 bp upstream of the start site. Mutations in the sequence of the TATA box affect the rate of transcription, and changing its position alters the location of the transcription start site.
- (2) Another common consensus sequence in the core promoter is the **TFIIB recognition element (BRE)**, which has the consensus sequence G/C G/C G/C C G C C and is located from –32 to –38 bp upstream of the start site. (TFIIB is the abbreviation for a transcription factor that binds to this element).

- (3) Instead of a TATA box, some core promoters have an **initiator element (Inr)** that directly overlaps the start site and has the consensus Y Y A N T/A Y Y.
- (4) Another consensus sequence called the **downstream core promoter element (DPE)** is found approximately -30 bp downstream of the start site in many promoters that also have Inr; the consensus sequence of DPE is RGA/TCGTG.

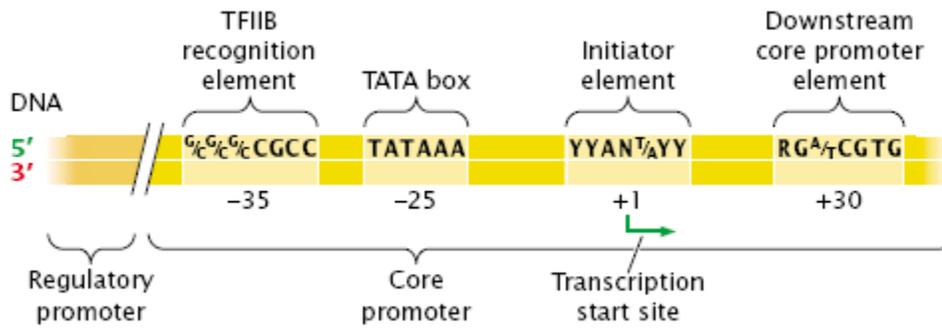


Figure 7.13: The promoters of genes transcribed by RNA polymerase II consist of a core promoter and a regulatory promoter that contain consensus sequences. Not all the consensus sequences shown are found in all promoters.

All of these consensus sequences in the core promoter are recognized by transcription factors that bind to them and serve as a platform for the assembly of the basal transcription apparatus.

[Y means that the nucleotide in the consensus sequence is usually a pyrimidine, but either pyrimidine is equally likely, while (R) means that the nucleotide in the sequence is most likely one of the purines, but both are equally frequent.]

7.4.3.2 Assembly of the Basal Transcription Apparatus

The basic transcriptional machinery that binds to DNA at the start site is called the basal transcription apparatus and is required to initiate minimal levels of transcription. It consists of RNA polymerase, a series of general transcription factors, and a complex of proteins known as the mediator (**Figure 7.14**). The general transcription factors include TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIF, in which TFII stands for transcription factor for RNA polymerase II and the letter designates the individual factor.

TFIID binds to the TATA box and positions the active site of RNA polymerase II so that it begins transcription at the correct place. TFIID consists of at least nine polypeptides. One of them is the **TATA-binding protein (TBP)**, which recognizes and binds to the TATA box on the DNA template. The TATA-binding protein binds to

the minor groove and straddles the DNA as a molecular saddle, bending the DNA and partly unwinding it. Other proteins, called **TBP-associated factors (TAFs)**, combine with TBP to form the complete TFIID transcription factor.

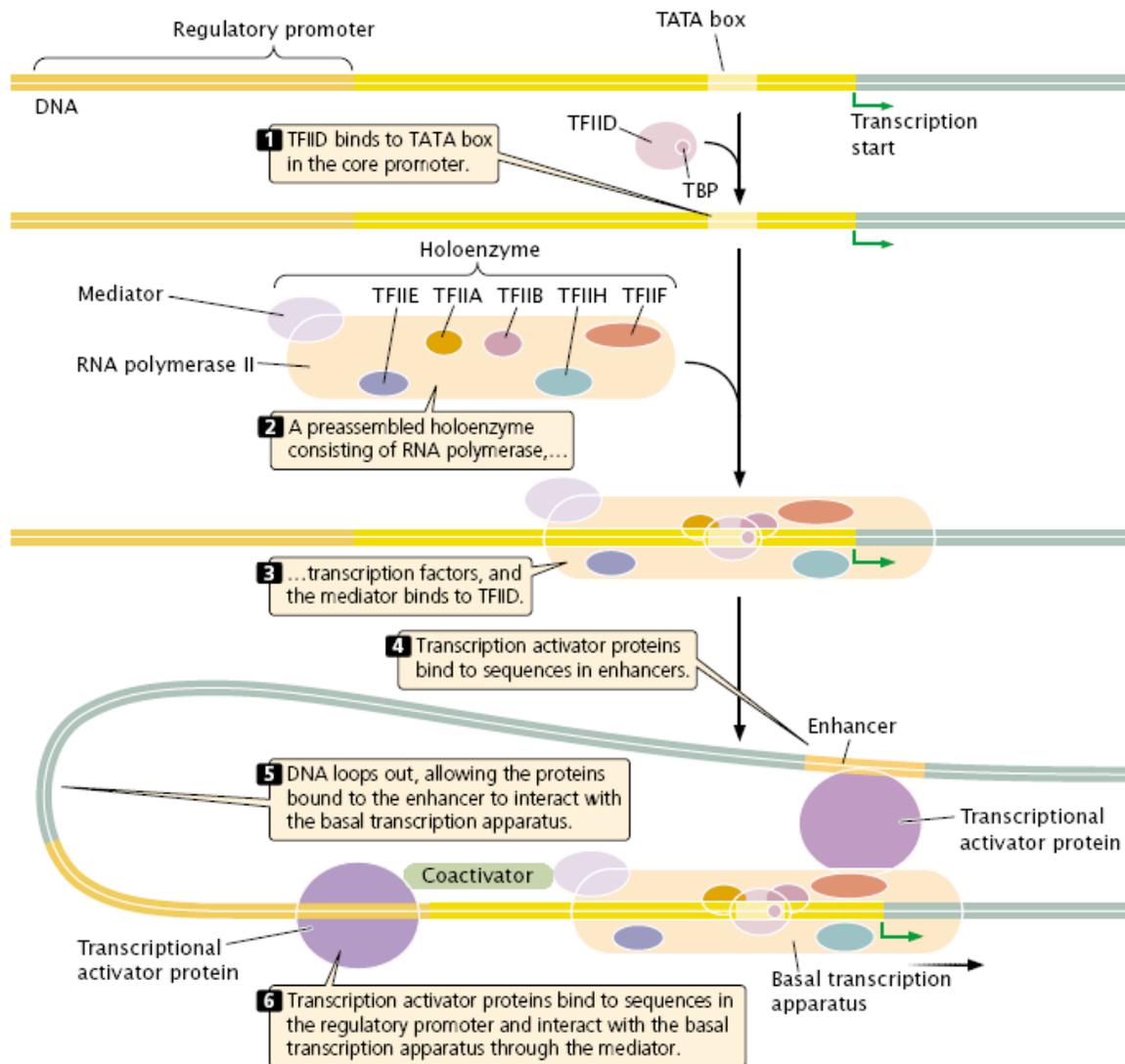


Figure 7.14: Transcription is initiated at RNA polymerase II promoters when the TFIID transcription factor binds to the TATA box, followed by the binding of a preassembled holoenzyme containing general transcription factors, RNA polymerase II, and the mediator.

The large holoenzyme consisting of RNA polymerase, additional transcription factors, and the mediator are thought to preassemble and bind as a unit to TFIID. The other transcription factors provide additional functions: (i) **TFIIA** helps to stabilize the interaction between TBP and DNA, (ii) **TFIIB** plays a role in the selection of the start site, and (iii) **TFIIH** has helicase activity and unwinds the DNA during transcription.

The **mediator** plays a role in communication between the basal transcription apparatus and transcriptional activator proteins (see next subsection).

7.4.3.3 Regulatory Promoter

The **regulatory promoter** is located immediately upstream of the core promoter. A variety of different consensus sequences may be found in the regulatory promoters, and they can be mixed and matched in different combinations (**Figure 7.15**). Transcriptional activator proteins bind to these sequences and, either directly or indirectly (through the mediation of coactivator proteins), make contact with the mediator in the basal transcription apparatus and affect the rate at which transcription is initiated. Some regulatory promoters also contain repressing sequences, which are bound by proteins that lower the rate of transcription through inhibitory interactions with the mediator.

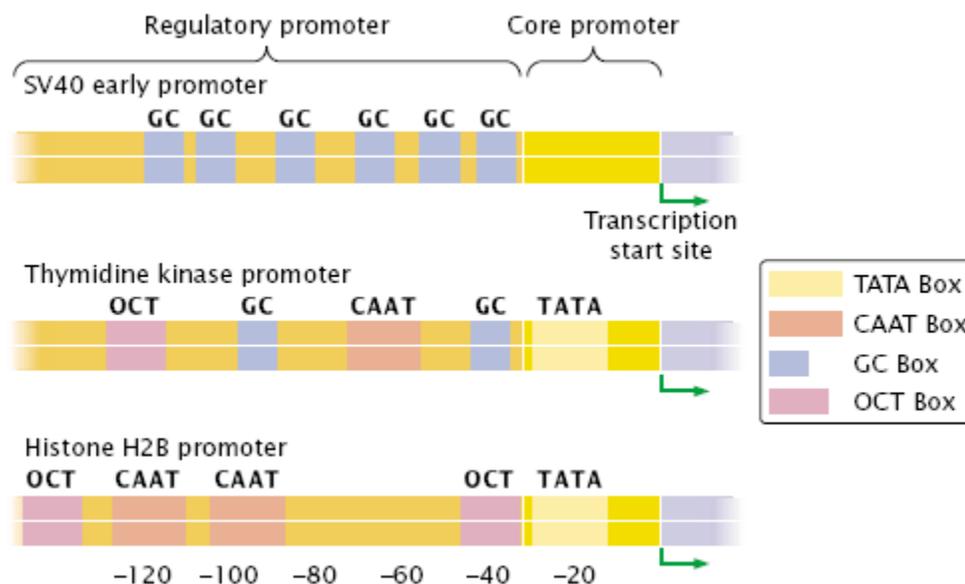


Figure 7.15: The consensus sequences in promoters of three eukaryotic genes illustrate the principle that different sequences can be mixed and matched to yield a functional promoter.

7.4.3.4 Enhancers

DNA sequences that increase the rate of transcription at distant genes are called **enhancers**. Furthermore, the precise position of an enhancer relative to a gene's transcriptional start site is not critical; most enhancers can stimulate any promoter in their vicinities, and an enhancer may be upstream or downstream from the affected

gene or, in some cases, within an intron of the gene itself. Enhancers also contain sequences that are recognized by transcriptional activator proteins.

How does the binding of a transcriptional activator protein to an enhancer affect the initiation of transcription at a gene thousands of nucleotides away?

The answer is that the DNA between the enhancer and the promoter loops out, allowing the enhancer and the promoter to lie close to each other. Transcriptional activator proteins bound to the enhancer interact with proteins bound to the promoter and stimulate the transcription of the adjacent gene (see **Figure 7.14**). The looping of DNA between the enhancer and the promoter explains how the position of an enhancer can vary with regard to the start site—enhancers that are farther from the start site simply cause a longer length of DNA to loop out.

Sequences having many of the properties possessed by enhancers sometimes take part in *repressing* transcription instead of enhancing it; such sequences are called **silencers**. Although enhancers and silencers are characteristic of eukaryotic DNA, some enhancer-like sequences have been found in bacterial cells.

Concepts: General transcription factors assemble into the basal transcription apparatus, which binds to DNA near the start site and is necessary for transcription to take place at minimal levels. Additional proteins called transcriptional activators bind to other consensus sequences in promoters and enhancers, and affect the rate of transcription.

7.5 Characteristics of Eukaryotic Promoters and Transcription Factors

Some general principles of eukaryotic promoters and transcription factors:

1. Several types of DNA sequences take part in the initiation of transcription in eukaryotic cells. These sequences generally serve as the **binding sites for proteins** that interact with RNA polymerase and influence the initiation of transcription.
2. Some sequences that affect transcription, called **promoters**, are adjacent to or within the RNA coding region and are relatively fixed with regard to the start site of transcription. Promoters consist of a core promoter located adjacent to the gene and a regulatory promoter located farther upstream.
3. Other sequences, called **enhancers**, are distant from the gene and function independently of position and direction. Enhancers stimulate transcription.

4. **General transcription factors** bind to the core promoter near the start site and, with RNA polymerase, assemble into a basal transcription apparatus. The TATA-binding protein (TBP) is a critical transcription factor that positions the active site of RNA polymerase over the start site.
5. **Transcriptional activator proteins** bind to sequences in the regulatory promoter and enhancers and affect transcription by interacting with the basal transcription apparatus.
6. Proteins binding to enhancers interact with the basal transcription apparatus by causing the DNA between the promoter and the enhancer to loop out, bringing the enhancer into close proximity to the promoter.

7.6 Termination

The termination of transcription in eukaryotic genes is less well understood than in bacterial genes. The three eukaryotic RNA polymerases use different mechanisms for termination.

- (1) RNA polymerase I requires a termination factor, like the rho factor utilized in termination of some bacterial genes. Unlike rho, which binds to the newly transcribed RNA molecule, the termination factor for RNA polymerase I binds to a DNA sequence downstream of the termination site.
- (2) RNA polymerase III ends transcription after transcribing a terminator sequence that produces a string of Us in the RNA molecule, like that produced by the rho-independent terminators of bacteria. Unlike rho-independent terminators in bacterial cells, RNA polymerase III does not require that a hairpin structure precede the string of Us.
- (2) In many of the genes transcribed by RNA polymerase II, transcription can end at multiple sites located within a span of hundreds or thousands of base pairs.

Concepts: The different eukaryotic RNA polymerases utilize different mechanisms of termination.

7.7 Pre-mRNA Processing

In bacterial cells, transcription and translation take place simultaneously; while the 3' end of an mRNA is undergoing transcription, ribosomes attach to the Shine-Dalgarno sequence near the 5' end and begin translation. Because transcription and translation are coupled, there is little opportunity for the bacterial mRNA to be modified before protein synthesis. In contrast, transcription and translation are separated in both time and space in eukaryotic cells. Transcription takes place in the nucleus, whereas most translation takes place in the cytoplasm; this separation provides an opportunity for eukaryotic RNA to be modified before it is translated. Indeed, eukaryotic mRNA is extensively altered after transcription. Changes are made to the 5' end, the 3' end, and the protein-coding section of the RNA molecule. The initial transcript of protein-encoding genes of eukaryotic cells is called pre-mRNA, whereas the mature, processed transcript is mRNA. We will reserve the term mRNA for RNA molecules that have been completely processed and are ready to undergo translation.

7.7.1 The Addition of the 5' Cap

Almost all eukaryotic pre-mRNAs are modified at their 5' ends by the addition of a structure called a **5' cap**. This capping consists of the addition of an extra nucleotide at the 5' end of the mRNA and methylation by the addition of a methyl group (CH_3) to the base in the newly added nucleotide and to the 2'-OH group of the sugar of one or more nucleotides at the 5' end. Capping takes place rapidly after the initiation of transcription and the 5' cap functions in the initiation of translation.

Early in the elongation process, the 5' ends of eukaryotic pre-mRNAs are modified by the addition of 7-methylguanosine (7-MG) caps by three enzymatic steps (**Figure 7.16**). These 7-MG caps are added when the growing RNA chains are only about 30 nucleotides long.

We know that three phosphates are present at the 5' end of all RNA molecules, because phosphates are not cleaved from the first ribonucleoside triphosphate in the transcription reaction. The 5' end of pre-mRNA can be represented as 5'-pppNpNpN, in which the letter N represents a ribonucleotide and p represents a phosphate.

- (1) Shortly after the initiation of transcription, one of these phosphates is removed and a guanine nucleotide is added (**Figure 7.16**).

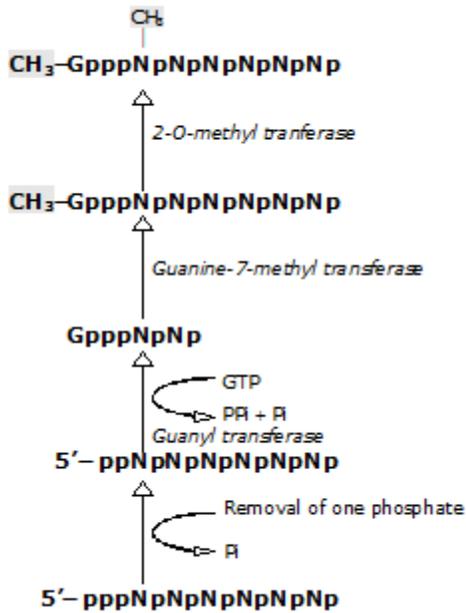


Figure 7.16: Enzymatic modification of 5' ends of eukaryotic pre-mRNAs to form 7-methylguanosine (7-MG) caps.

- (2) This guanine nucleotide is attached to the pre-mRNA by a unique **5'–5' triphosphate linkage**, which is quite different from the usual 5'–3' phosphodiester bond that joins all the other nucleotides in RNA.
- (3) One or more methyl groups are then added to the 5' end; the first of these methyl groups is added to position 7 of the base of the terminal guanine nucleotide, making the base 7-methylguanine. Next, a methyl group may be added to the 2' position of the sugar in the second and third nucleotides, as shown in **Figure 7.17**. Rarely, additional methyl groups may be attached to the bases of the second and third nucleotides of the pre-mRNA.

Cap-binding proteins recognize the cap and attach to it; a ribosome then binds to these proteins and moves downstream along the mRNA until the start codon is reached and translation begins. The presence of a 5' cap also increases the stability of mRNA and influences the removal of introns.

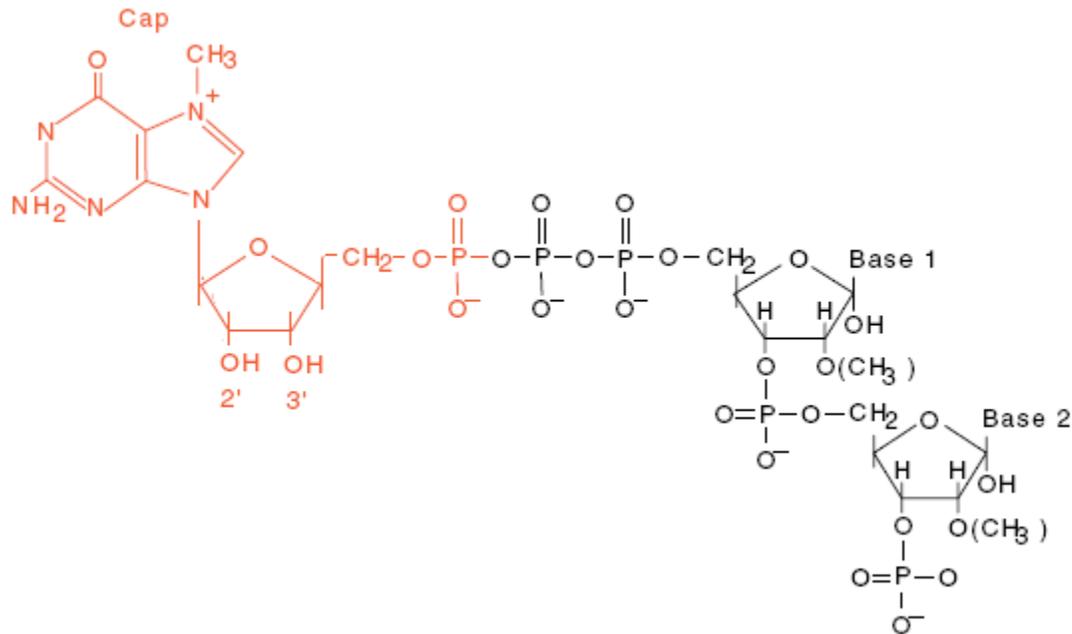


Figure 7.17: The structure of the cap found on eukaryotic messenger RNAs. The first base is 7-methylguanylate connected by a 5'-5' triphosphate linkage to the next base. The 2' positions on bases 1 and 2 may or may not be methylated.

7.7.2 The Addition of the 3'-Poly(A) Tail

Most mature eukaryotic mRNAs have from 50 to 250 adenine nucleotides at the 3' end (a **poly(A) tail**). These nucleotides are not encoded in the DNA but are added after transcription (**Figure 7.18**) in a process termed polyadenylation. Many eukaryotic genes transcribed by RNA polymerase II are transcribed well beyond the end of the coding sequence; the extra material at the 3' end is then cleaved by endonuclease and the poly(A) tail is added by the action of the enzyme poly(A) polymerase. For some pre-mRNA molecules, more than 1,000 nucleotides may be cleaved from the 3' end.

Processing of the 3' end of pre-mRNA requires sequences both upstream and downstream of the cleavage site. The consensus sequence **AAUAAA** is usually from 11 to 30 nucleotides upstream of the cleavage site (**Figure 7.18**) and determines the point at which cleavage will take place. A sequence rich in Us (or Gs and Us) is typically downstream of the cleavage site.

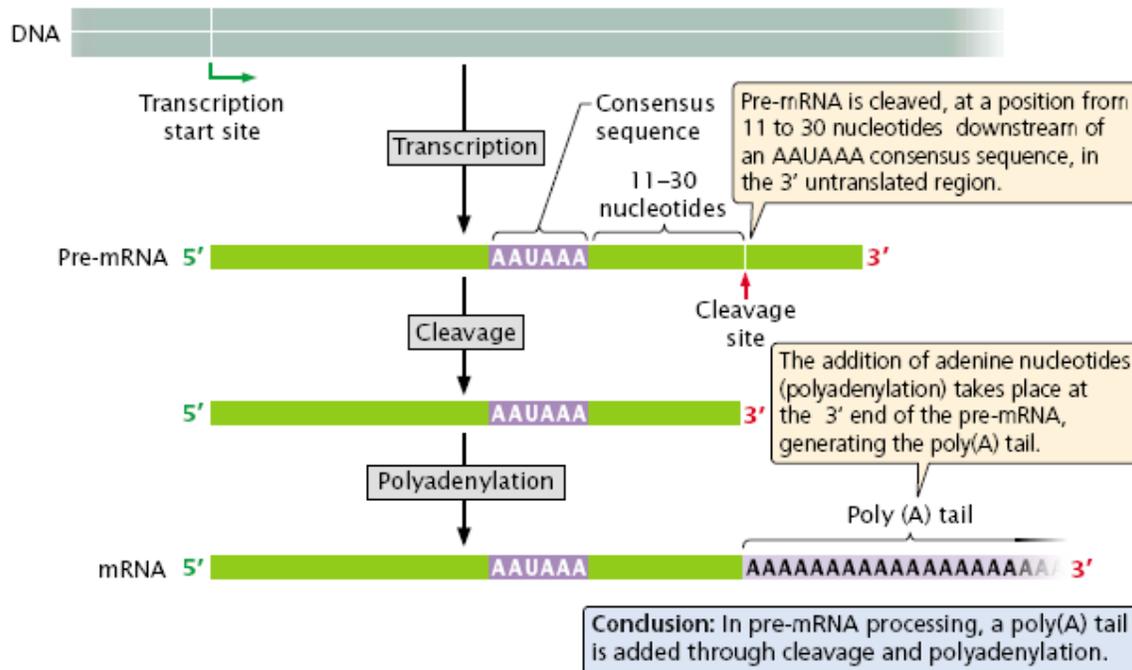


Figure 7.18: Most eukaryotic mRNAs have a 3' poly(A) tail.

Concepts: Eukaryotic pre-mRNAs are processed at their 5' and 3' ends. A cap, consisting of a modified nucleotide and several methyl groups, is added to the 5' end. The cap facilitates the binding of a ribosome, increases the stability of the mRNA, and may affect the removal of introns. Processing at the 3' end includes cleavage downstream of an AAUAAA consensus sequence and the addition of a poly(A) tail.

7.7.3 RNA Splicing

The other major type of modification that takes place in eukaryotic pre-mRNA is the removal of introns by **RNA splicing**. This occurs in the nucleus following transcription but before the RNA moves to the cytoplasm. Most nuclear genes that encode proteins in multicellular eukaryotes contain introns (**Figure 7.19**). Fewer, but still many of the genes of multicellular eukaryotes such as the yeasts contain noncoding introns. Rare genes of a few viruses of prokaryotes and an archaebacteria also contain introns. In the case of these "split gene", with coding sequences interrupted by noncoding sequences, the primary transcript contains the entire sequence of the gene and noncoding sequences are spliced out during RNA processing.

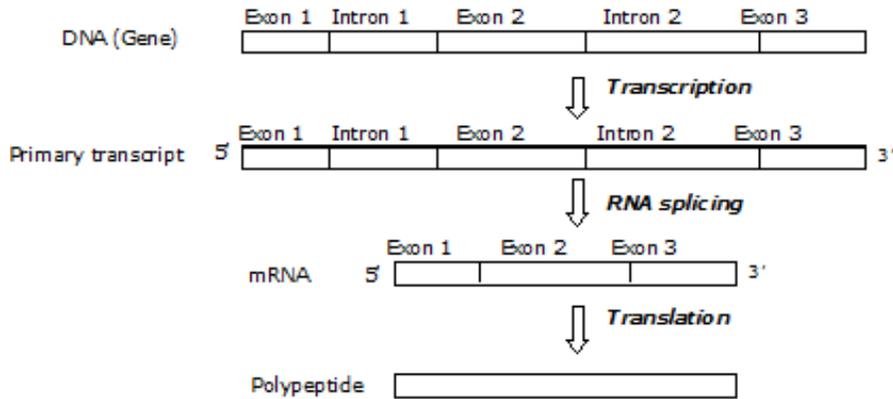
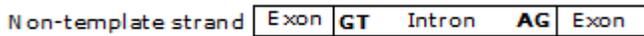


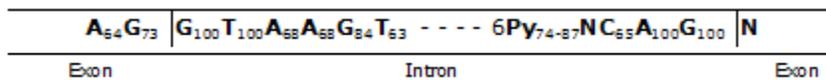
Figure 7.19: The excision of intron sequences from primary transcripts by RNA splicing.

For genes that encode proteins, the splicing mechanism must be precise; it must join exon sequences with accuracy to the single nucleotide to assure that codons in exons distal to introns are read correctly. Accuracy to this degree would seem to require precise splicing signals, presumably nucleotide sequences within introns and at the exon-intron junctions. However, in the primary transcripts of nuclear genes, the only completely conserved sequences of different introns are the dinucleotide sequences at the ends of introns, namely:



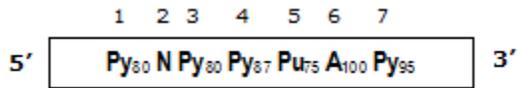
The sequences shown here are for the DNA nontemplate strand (equivalent to the RNA transcript, but with T rather than U).

In addition, there are short consensus sequences at the exon-intron junctions. For nuclear genes, the consensus junctions are:



[The exon-intron junctions are different for tRNA genes and structural genes in mitochondria and chloroplast, which utilize different RNA splicing mechanism.]

There is only one short conserved sequence, the **TACTAAC Box**, located about 30 nucleotides upstream from the 3' splicing site of introns in nuclear genes, and it is rather poorly conserved. The TACTAAC Box does exhibit a strong preference for either a purine or a pyrimidine at each site as follows:



The adenine residue at position six in the TACTAAC Box is completely conserved and is known to play a key role in the splicing reaction. With the exception of the terminal dinucleotides and the TACTAAC Box, the intron sequences of nuclear genes are highly divergent, apparently random sequences.

7.8 RNA Splicing Mechanisms

There are three distinct types of intron excision from RNA transcripts, presented here in the order of increasing complexity, not in the order of importance.

1. The introns of tRNA precursors are excised by precise endonucleolytic cleavage and ligation reaction catalyzed by special **splicing endonuclease and ligase activities**.
2. The introns of some rRNA precursors are removed **autocatalytically** in a unique reaction mediated by the RNA molecule itself (ribozyme).
3. The introns of nuclear pre-mRNA (hnRNA) transcripts are spliced out in two-step reactions carried out by complex ribonucleoprotein particles called **spliceosomes**.

7.8.1 tRNA Precursor Splicing: Unique Nuclease and Ligase Activities

The tRNA precursor splicing reaction has been worked out in detail in the yeast *Saccharomyces cerevisiae*. The excision of introns from yeast tRNA precursors occurs in two stages.

In **Stage I**, a nuclear membrane bound **splicing endonuclease** makes two cuts precisely at the ends of the intron.

In **Stage II**, a **splicing ligase** joins the two halves of the tRNA to produce the mature form of the tRNA molecule. The specificity of these reactions resides in conserved three-dimensional features of the tRNA precursors, not in the nucleotide sequence. Cleavage of the tRNA precursor by splicing endonuclease yields 5'-OH termini and 2'-3' cyclic phosphate groups at the 3' termini (**Figure 7.20**).

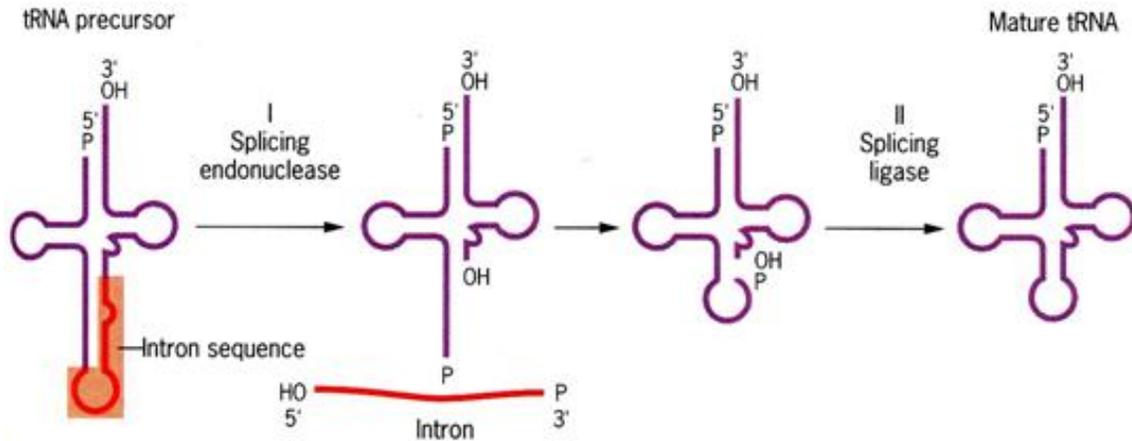


Figure 7.20: Splicing by nuclease and ligase.

The Stage II ligation process involves four to five separate reactions.

- (i) The first reaction is the addition of a phosphate group to 5'-OH terminus; this reaction requires **kinase** activity and a **phosphate** donor (ATP).
- (ii) Then, the 5'-phosphate group is activated by the transfer of an AMP group to the terminus from an **AMP-ligase** intermediate.
- (iii) The 2'-3'-cyclic phosphate is opened by a **cyclic phosphodiesterase** that produces a 2'-phosphate and a free 3'-hydroxyl.
- (iv) The final ligation reaction occurs via a nucleophilic attack of the free 3'-OH on the interior 5'-phosphate with the release of AMP. All four of these reactions are catalyzed by the **splicing ligase**.
- (v) Finally, the 2'-phosphate group (remaining from the 2'-3'-cyclic phosphate) is removed by a **phosphatase** activity to yield the mature tRNA molecule.

7.8.2 rRNA Precursor Splicing: Autocatalytic Splicing

A general theme in biology is that metabolism occurs via sequences of enzyme-catalyzed reactions. These all-important enzymes are generally proteins, sometimes single polypeptides and sometimes complex heteromultimers. Occasionally, enzymes require nonprotein cofactors to perform their functions. When covalent bonds are being altered, it is usually assumed that the reaction is being catalyzed by an enzyme. Thus, the 1982 discovery by Thomas Cech and his coworkers that the intron in the rRNA precursor of *Tetrahymena thermophila* was excised without the involvement of any protein catalytic activity was quite surprising. However, it is now clearly established that the splicing activity that excises the intron from this rRNA precursor is intrinsic to the RNA molecule itself. Indeed, Cech and Sidney Altman

shared the 1989 Nobel Prize in Chemistry for their discovery of catalytic RNAs. Moreover, such *self-splicing* or *autocatalytic activity* has been shown to occur in rRNA precursors of several lower eukaryotes and in a large number of rRNA, tRNA, and mRNA precursors in mitochondria and chloroplasts of many different species. In the case of many of these introns, the self-splicing mechanism is the same as or very similar to that utilized by the *Tetrahymena* rRNA precursors (see **Figure 7.21**). For others, the self-splicing mechanism is similar to the splicing mechanism observed with nuclear mRNA precursors, but without the involvement of the spliceosome.

The autocatalytic excision of the intron in the *Tetrahymena* rRNA precursor and certain other introns requires no external energy source and no protein catalytic activity. Instead, the splicing mechanism involves a series of phosphoester bond transfers, with no bonds lost or gained in the process. The reaction requires a guanine nucleoside or nucleotide with a free 3'-OH group (GTP, GDP, GMP, or guanosine all work) as a cofactor plus a monovalent cation and a divalent cation. The requirement for the G-3'-OH is absolute; no other base can be substituted in the nucleoside or nucleotide cofactor. The intron is excised by means of two phosphoester bond transfers, and the excised intron can subsequently circularize by means of another phosphoester bond transfer. These reactions are diagrammed in **Figure 7.21**.

The autocatalytic circularization of the excised intron suggests that the self-splicing of these rRNA precursors resides primarily, if not entirely, within the intron structure itself. Presumably, the autocatalytic activity is dependent on the secondary structure of the intron or at least the secondary structure of the RNA precursor molecule. The secondary structures of these self-splicing RNAs must bring the reactive groups into close juxtaposition to allow the phosphoester bond transfers to occur. Since the self-splicing phosphoester bond transfers are potentially reversible reactions, rapid degradation of the excised introns or export of the spliced rRNAs to the cytoplasm may drive splicing in the forward direction.

Note that the autocatalytic splicing reactions are intramolecular in nature and thus are not dependent on concentration. Moreover, the RNA precursors are capable of forming an active center in which the guanosine-3'-OH cofactor binds. The autocatalytic splicing of these rRNA precursors demonstrates that catalytic sites are not restricted to proteins; however, there is no *trans* catalytic activity as for

enzymes, only *cis* catalytic activity. Some scientists believe that autocatalytic RNA splicing may be a relic of an early RNA-based world.

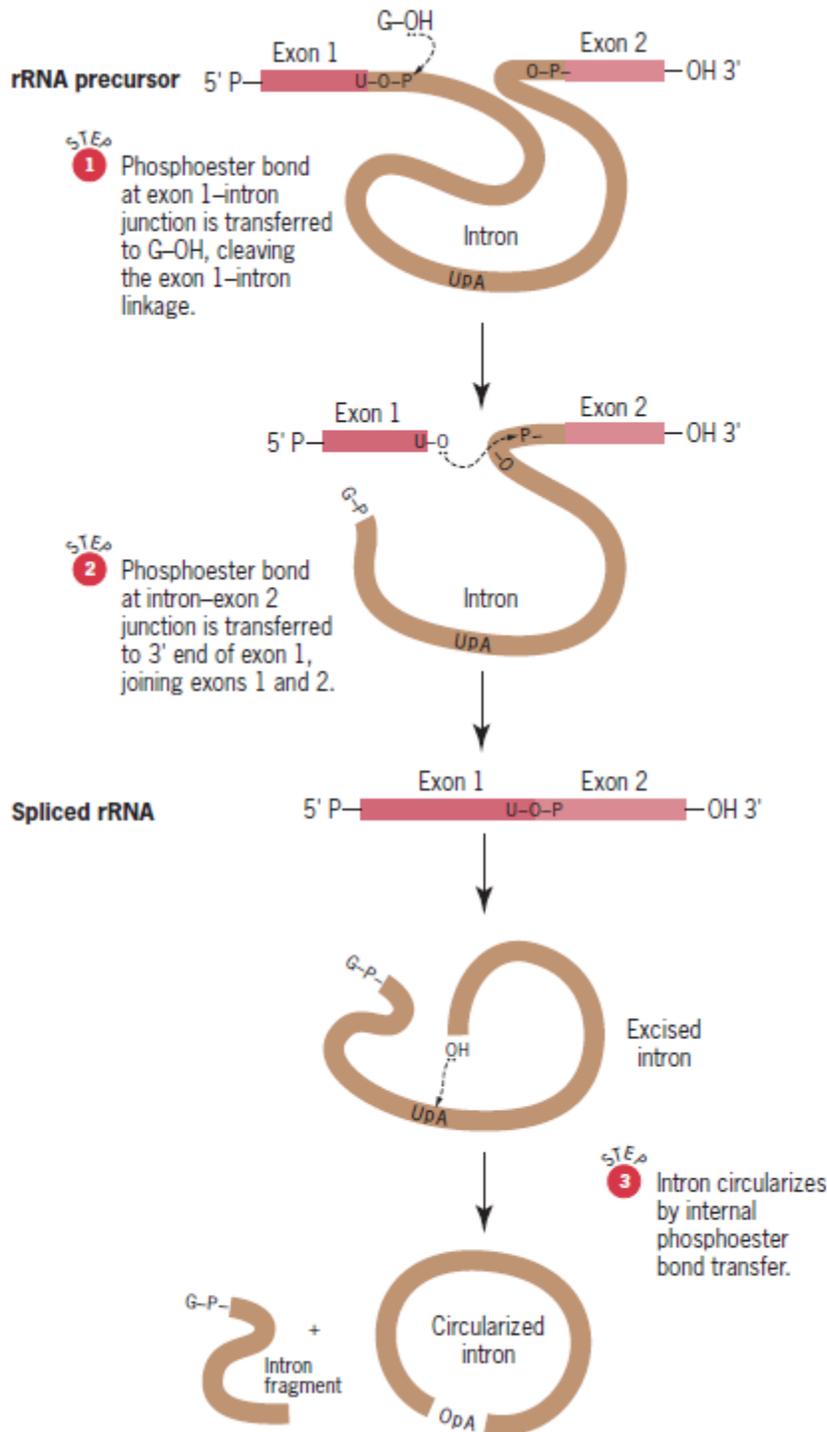


Figure 7.21: Diagram of the mechanism of self-splicing.

7.8.3 Pre-mRNA Splicing: snRNA, snRNP and the Spliceosome

The introns in nuclear pre-mRNAs are excised in two steps by complex RNA/protein structures called **spliceosomes**. These structures are in any ways like small ribosomes. They contain a set of small RNA molecules called snRNA (small nuclear RNAs) and a set of proteins that are still not completely defined. The two steps in nuclear pre-mRNA splicing are known; however, some of the details of the splicing process are still uncertain.

7.8.3.1 snRNA, snRNP and Spliceosome

(1) *Small Nuclear RNA (snRNA)*: Five snRNAs, called U1, U2, U4, U5 and U6, are involved in nuclear pre-mRNA splicing as components of the spliceosomes. [snRNA U3 is located in the nucleolus and probably is involved in the formation of ribosomes.] In mammals, these snRNAs range in size from 100 nucleotides (U6) to 215 nucleotides (U3). Some of the snRNAs in the yeast *Saccharomyces cerevisiae* are much larger.

(2) *Small Nuclear Ribonucleoprotein (snRNP)*: snRNAs do not exist as free RNA molecules. Instead, they are present in small nuclear RNA-protein complexes called snRNPs (small nuclear ribonucleoproteins).

(3) *Spliceosomes*: Spliceosomes are assembled from four different snRNPs and protein splicing factors during the splicing process.

[Characterization of snRNPs has been facilitated by the discovery that some patients with an often fatal autoimmune disease called **systemic lupus erythematosus** produce antibodies that react with many of their own cellular components including snRNP proteins. These antibodies are called autoantibodies because they react with the patient's own proteins; normally, the human immune system will produce only antibodies that react with foreign proteins. In lupus patients, the autoantibodies cause inflammations of tissue and organs that often result in heart, kidney, or liver malfunction and even death. The autoantibodies from patients with systemic lupus erythematosus can be used to precipitate snRNPs; thus greatly facilitate the purification of snRNPs for structural and functional studies.]

7.8.3.2 Splicing by Spliceosomes

Each of the snRNAs, U1, U2 and U5 is present by itself in a specific snRNP particle. snRNA U4 and U6 are present together in a fourth snRNP; U4 and U6 snRNAs contain two regions of intermolecular complementarity that probably are base-paired in the U4/U6 snRNP. Each of the four types of snRNP particle contains a subset of seven well-characterized snRNP proteins plus one or more protein unique to the particular type of snRNP particle. All four snRNP complexes are present in the isolated spliceosomes (**Figure 7.22**). The exact protein composition of intact spliceosomes still is not established.

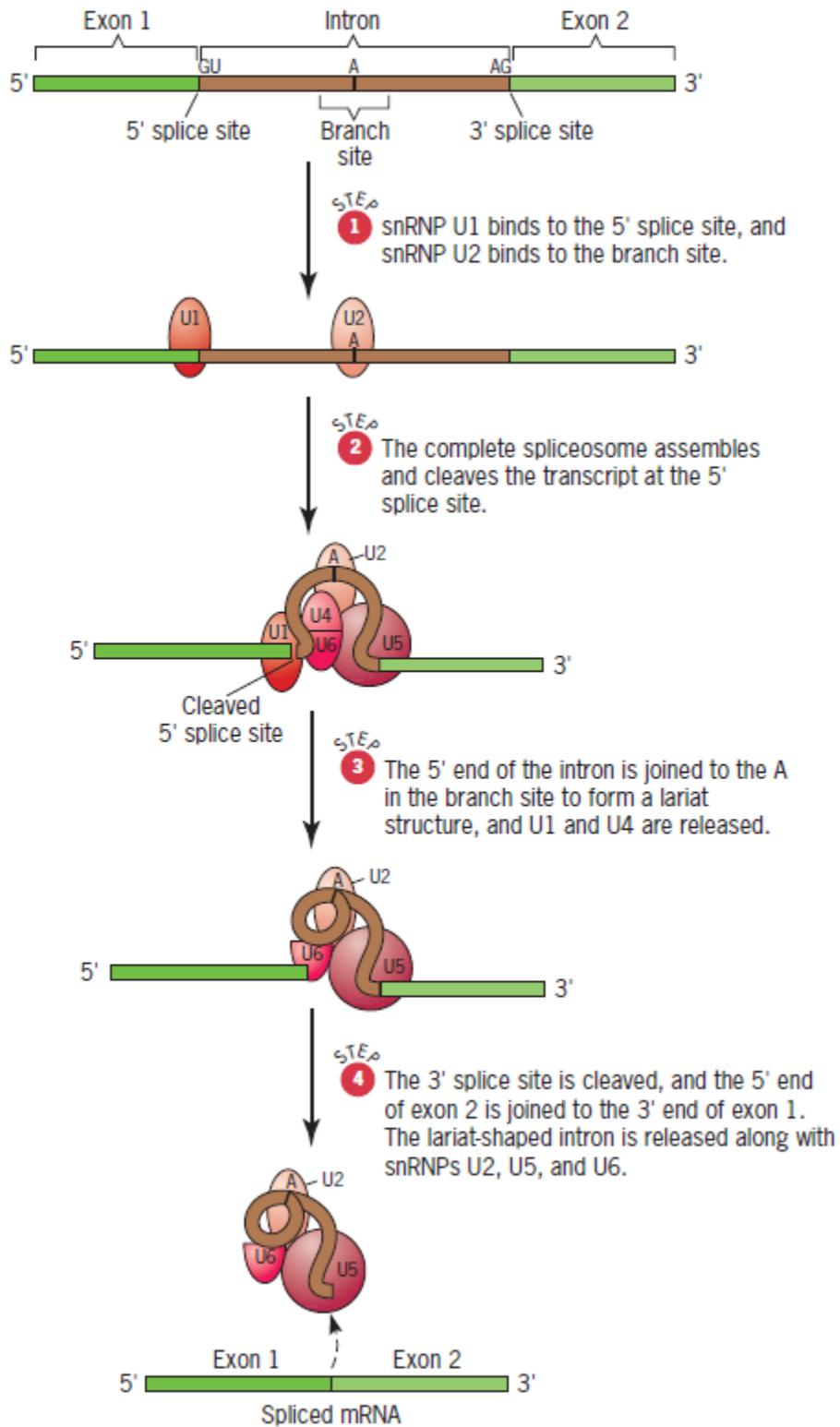
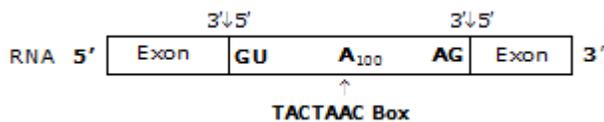


Figure 7.22: The postulated roles of the snRNA-containing snRNPs in nuclear pre-mRNA splicing.

The **first step** in nuclear pre-mRNA splicing involves cleavage at the 5'-intron splice site (\downarrow GU-intron) and the formation of an intramolecular phosphodiester linkage between the 5'-carbon of the **G** at the cleavage site and the 2'-carbon of a conserved A residue near the 3' end of the intron. This step occurs on complete spliceosomes and requires the hydrolysis of ATP. Evidence indicates that the U1 snRNP must bind at the 5'-splice site prior to the initial cleavage reaction. Recognition of the cleavage site at the 5'-end of the intron probably involves base-pairing between the consensus sequence at this site and a complementary sequence near the 5'-terminus of snRNA U1.



The second snRNP to be added to the splicing complex appears to be the U2 snRNP; it binds at the consensus sequence that contains the conserved **A** residue that forms the branch point in the lariat structure of spliced intron. Thereafter, the U5 snRNP binds at the 3' splice site, and the U4/U6 snRNP is added to the complex to yield the complete spliceosome. When the 5' intron splice site is cleaved in step 1, the U4 snRNA is released from the spliceosome.

In **step 2** of the splicing reaction, the 3' splice site of the intron is cleaved, and the two exons are joined by a normal 5' to 3' phosphodiester linkage. The spliced mRNA is now ready for export to the cytoplasm and translation on ribosomes.

Concepts: Intron splicing of nuclear genes is a two-step process: (1) the 5' end of the intron is cleaved and attached to the branch point to form a lariat and (2) the 3' end of the intron is cleaved and the two ends of the exon are spliced together. These reactions take place within the spliceosome.

References

1. **Genetics: A Conceptual Approach**, First Edition. 2007. Benjamin A Pierce. WH Freeman & Company, New York.
2. **Principles of Genetics**, Sixth Edition. 2012. Snustad P and Simmons MJ. John Wiley and Sons Ltd., New York.

7.9 Review Questions

1. Draw an RNA nucleotide and a DNA nucleotide, highlighting the differences. How is the structure of RNA similar to that of DNA? How is it different?
2. What are the major classes of cellular RNA? Where would you expect to find each class of RNA within eukaryotic cells?
3. What parts of DNA make up a transcription unit? Draw and label a typical transcription unit in a bacterial cell.
4. What is the substrate for RNA synthesis? How is this substrate modified and joined together to produce an RNA molecule?
5. Describe the structure of bacterial RNA polymerase.
6. Give the names of the three RNA polymerases found in eukaryotic cells and the types of RNA that they transcribe.
7. What are the four basic stages of transcription? Describe what happens at each stage.
8. Draw and label a typical bacterial promoter. Include any common consensus sequences.
9. What are the two basic types of terminators found in bacterial cells? Describe the structure of each.
10. How is the process of transcription in eukaryotic cells different from that in bacterial cells?
11. How are promoters and enhancers similar? How are they different?
12. How can an enhancer affect the transcription of a gene that is thousands of nucleotides away?
13. Compare the roles of general transcription factors and transcriptional activator proteins.
14. What are some of the common consensus sequences found in RNA polymerase II promoters?
15. What protein associated with a transcription factor is common to all eukaryotic promoters? What is its function in transcription?
16. Compare and contrast transcription and replication. How are these processes similar and how are they different?
17. (a) What is the 5' cap? (b) How is the 5' cap added to eukaryotic pre-mRNA? (c) What is the function of the 5' cap?
18. How is the poly(A) tail added to pre-mRNA? What is the purpose of the poly(A) tail?
19. What makes up the spliceosome? What is the function of the spliceosome?
20. Explain the process of pre-mRNA splicing in nuclear genes.
21. Summarize the different types of processing that can take place in pre-mRNA.
22. What are some of the modifications in tRNA processing?
23. Describe the basic structure of ribosomes in bacterial and eukaryotic cells
24. Explain how rRNA is processed.