

Transcription

Dr. Rashmi. R

Asst. Professor

UG Dept. of Biochemistry

JSSCACS, Ooty road, Mysuru

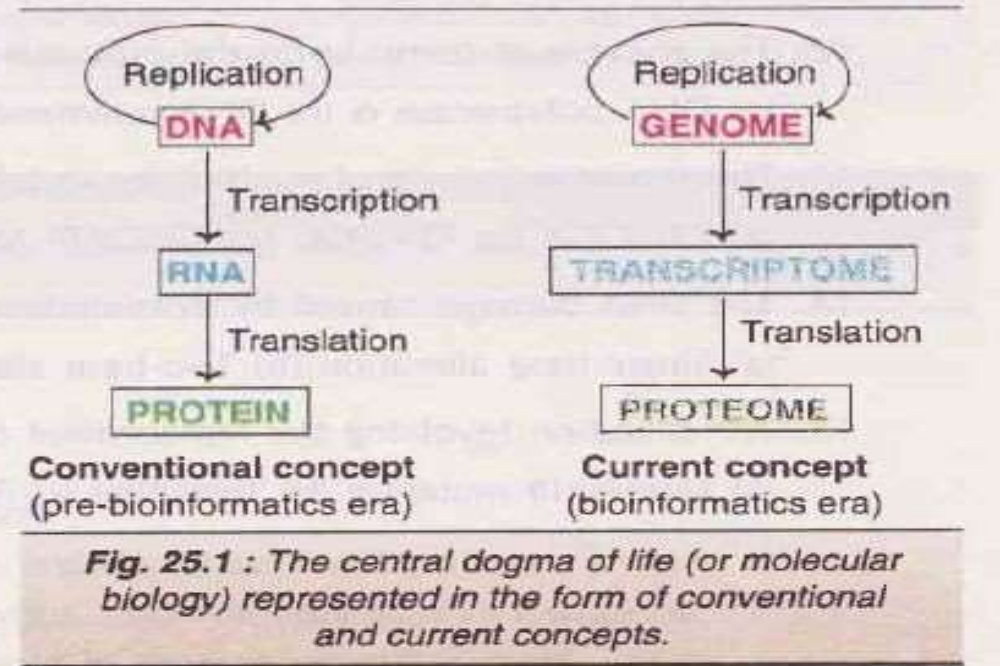
Transcription

The conventional concept of central dogma of life which in essence is "**DNA makes RNA makes protein**" is an oversimplification of molecular biology. With the advances in cell biology and rapid developments in bioinformatics, the terms genome, transcriptome and proteome are in current use to represent the central dogma of molecular biology (**Fig.25.1**). Some information on the new concepts and terminology is given hereunder.

GENOME

The **total DNA** (genetic information) contained **in an organism** or **a cell** is regarded as the genome. Thus, the genome is the storehouse of biological information. It includes the chromosomes in the nucleus and the DNA in mitochondria, and chloroplasts.

Genomics : The study of the structure and function of genome is genomics. The term **functional genomics** is used to represent the gene expression and relationship of genes with gene products. **Structural genomics** refers to the



structural motifs and complete protein structures. **Comparative genomics** involves the study of comparative gene function and phylogeny.

TRANSCRIPTOME

The **RNA copies** of the active protein coding **genes** represent transcriptome. Thus, transcriptome is the initial product of gene

Transcription: RNA biosynthesis

All RNA molecules except the RNA genomes of certain viruses are derived from information permanently stored in DNA. **During transcription, an enzyme system converts the genetic information in a segment of double-stranded DNA into an RNA strand with a base sequence complementary to one of the DNA strands.**

Three major kinds of RNA are produced

1. Messenger RNAs (mRNAs) **encode the amino acid sequence of one or more polypeptides** specified by a gene or set of genes.
2. Transfer RNAs (tRNAs) **read the information encoded in the mRNA and transfer the appropriate amino acid to a growing polypeptide chain** during protein synthesis.
3. Ribosomal RNAs (rRNAs) are **constituents of ribosomes, involved in cellular machines that synthesize proteins.**

Many additional specialized RNAs have regulatory or catalytic functions or are precursors to the three main classes of RNA.

Transcription resembles replication in its fundamental chemical mechanism, its polarity (direction of synthesis), and its use of a template.

Like replication, transcription has **initiation, elongation, and termination phases**—though in the literature on transcription, initiation is further divided into discrete **phases of DNA binding and initiation of RNA synthesis**.

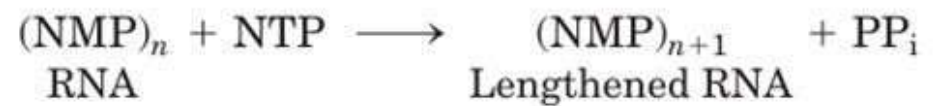
Transcription differs from replication in that it **does not require a primer** and, generally, involves only limited segments of a DNA molecule.

Additionally, within transcribed segments **only one DNA strand serves as a template**.

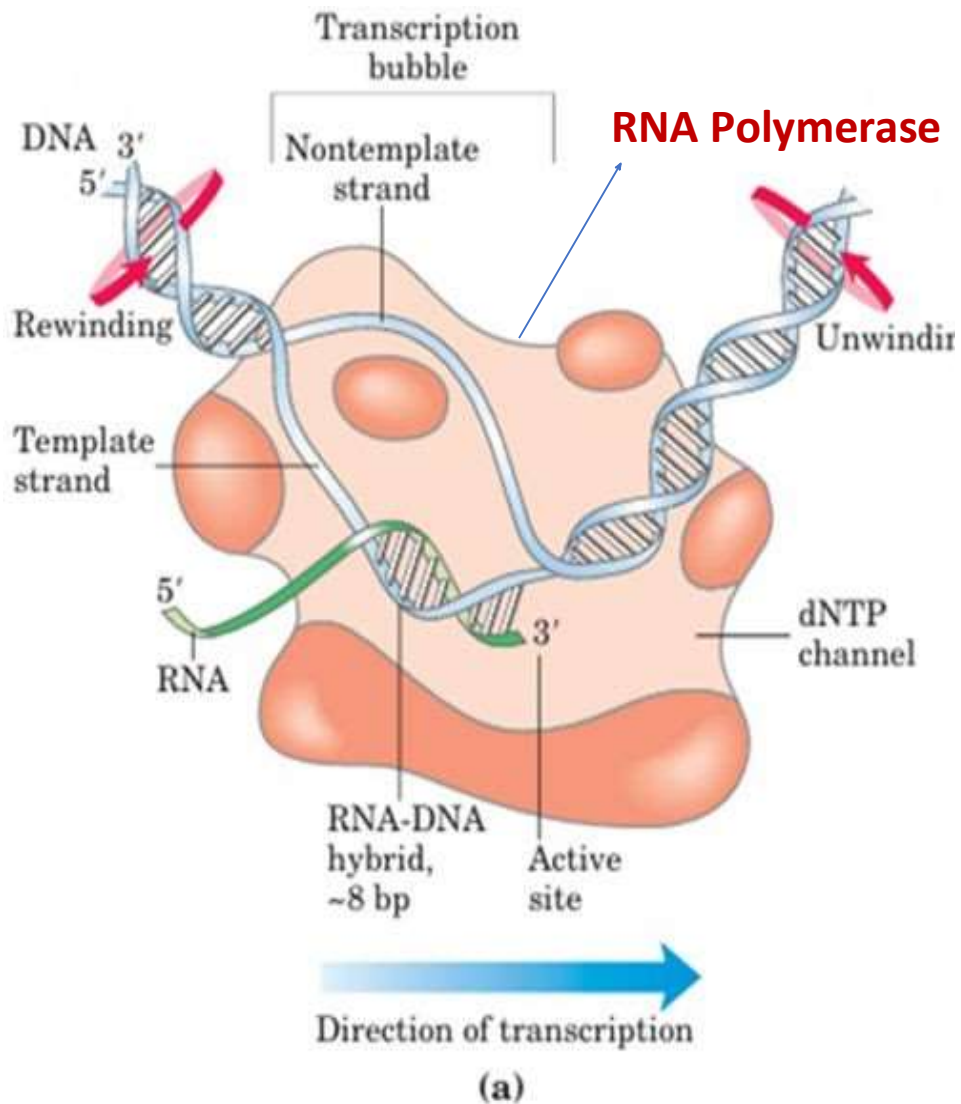
DNA-dependent RNA polymerase requires, in addition to a **DNA template**, all four **ribonucleoside 5-triphosphates (ATP, GTP, UTP, and CTP)** as precursors of the nucleotide units of **RNA**, as well as **Mg₂**. The protein also binds one **Zn₂**.

RNA polymerase elongates an RNA strand by adding ribonucleotide units to the 3'-hydroxyl end, building RNA in the 5'-3' direction.

The **3'-hydroxyl group** acts as a nucleophile, attacking the phosphate of the incoming ribonucleoside triphosphate (Fig. 26–1b) and releasing pyrophosphate.



- ✓ RNA polymerase requires DNA for activity and is most active when bound to a double-stranded DNA.
- ✓ Only one of the two DNA strands serves as a template.
- ✓ The template DNA strand is copied in the 3'-5' direction (antiparallel to the new RNA strand), just as in DNA replication.
- ✓ Each nucleotide in the newly formed RNA is selected by Watson-Crick base-pairing interactions; **U residues are inserted in the RNA to pair with A residues in the DNA template, G residues are inserted to pair with C residues, and so on.**



MECHANISM FIGURE 26-1 Transcription by RNA polymerase in *E. coli*. For synthesis of an RNA strand complementary to one of two DNA strands in a double helix, the DNA is transiently unwound. (a) About 17 bp are unwound at any given time. RNA polymerase and the bound transcription bubble move from left to right along the DNA as shown; facilitating RNA synthesis. The DNA is unwound ahead and rewound behind as RNA is transcribed. Red arrows show the direction in which the DNA must rotate to permit this process. As the DNA is rewound, the RNA-DNA hybrid is displaced and the RNA strand extruded. The RNA polymerase is in close contact with the DNA ahead of the transcription bubble, as well as with the separated DNA strands and the RNA within and immediately behind the bubble. A channel in the protein funnels new nucleoside triphosphates (NTPs) to the polymerase active site. The polymerase footprint encompasses about 35 bp of DNA during elongation.

(b) Catalytic mechanism of RNA synthesis by RNA polymerase. Note that this is essentially the same mechanism used by DNA poly-

Unlike DNA polymerase, **RNA polymerase does not require a primer to initiate synthesis.**

Initiation occurs when RNA polymerase binds at specific DNA sequences called promoters.

During the elongation phase of transcription, the growing end of the new RNA strand base-pairs temporarily with the DNA template to form a short hybrid RNA-DNA double helix, estimated to be 8 bp long .

The RNA in this hybrid duplex “peels off” shortly after its formation, and the DNA duplex re-forms.

To enable RNA polymerase to synthesize an RNA strand complementary to one of the DNA strands, the DNA duplex must unwind over a short distance, forming a transcription “bubble.”

During transcription, the *E. coli* RNA polymerase generally keeps about 17 bp unwound.

The 8 bp RNA-DNA hybrid occurs in this unwound region.

Elongation of a transcript by *E. coli* RNA polymerase proceeds at a rate of 50 to 90 nucleotides/s.

Because DNA is a helix, movement of a transcription bubble requires considerable strand rotation of the nucleic acid molecules.

DNA strand rotation is restricted in most DNAs by DNA-binding proteins and other structural barriers.

As a result, a moving RNA polymerase generates waves of positive supercoils ahead of the transcription bubble and negative supercoils behind (Fig. 26–1c).

In the cell, the topological problems caused by transcription are relieved through the action of topoisomerases.

The two complementary DNA strands have different roles in transcription. The strand that serves as template for RNA synthesis is called the **template strand**. The DNA strand complementary to the template, the **nontemplate strand, or coding strand**, is identical in base sequence to the RNA transcribed from the gene with U in the RNA in place of T in the DNA (Fig. 26-2). The coding strand for a particular gene may be located in either strand of a given chromosome (as shown in Fig. 26-3 for a virus). The regulatory sequences that control transcription are by convention designated by the sequence

(5') CGCTATAGCGTTT(3')	DNA nontemplate (coding) strand
(3') GCGATATCGCAA(5')	DNA template strand
(5') CGCUAUAGCGUUU(3')	RNA transcript

FIGURE 26-2 Template and nontemplate (coding) DNA strands. The two complementary strands of DNA are defined by their function in transcription. The RNA transcript is synthesized on the template strand and is identical in sequence (with U in place of T) to the nontemplate strand, or coding strand.

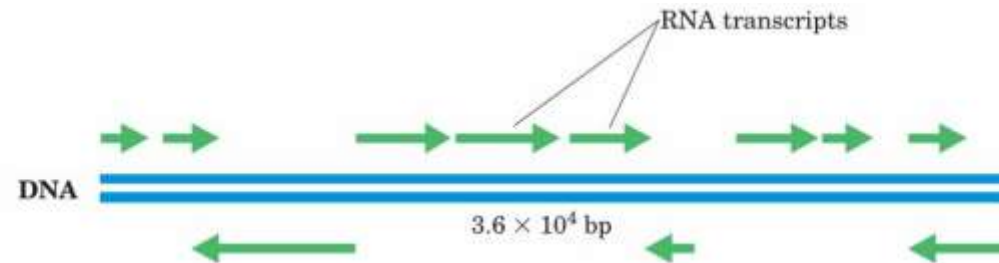


FIGURE 26-3 Organization of coding information in the adenovirus genome. The genetic information of the adenovirus genome (a conveniently simple example) is encoded by a double-stranded DNA molecule of 36,000 bp, both strands of which encode proteins. The information for most proteins is encoded by the top strand—by convention, the strand transcribed from left to right—but some is encoded by the bottom strand, which is transcribed in the opposite

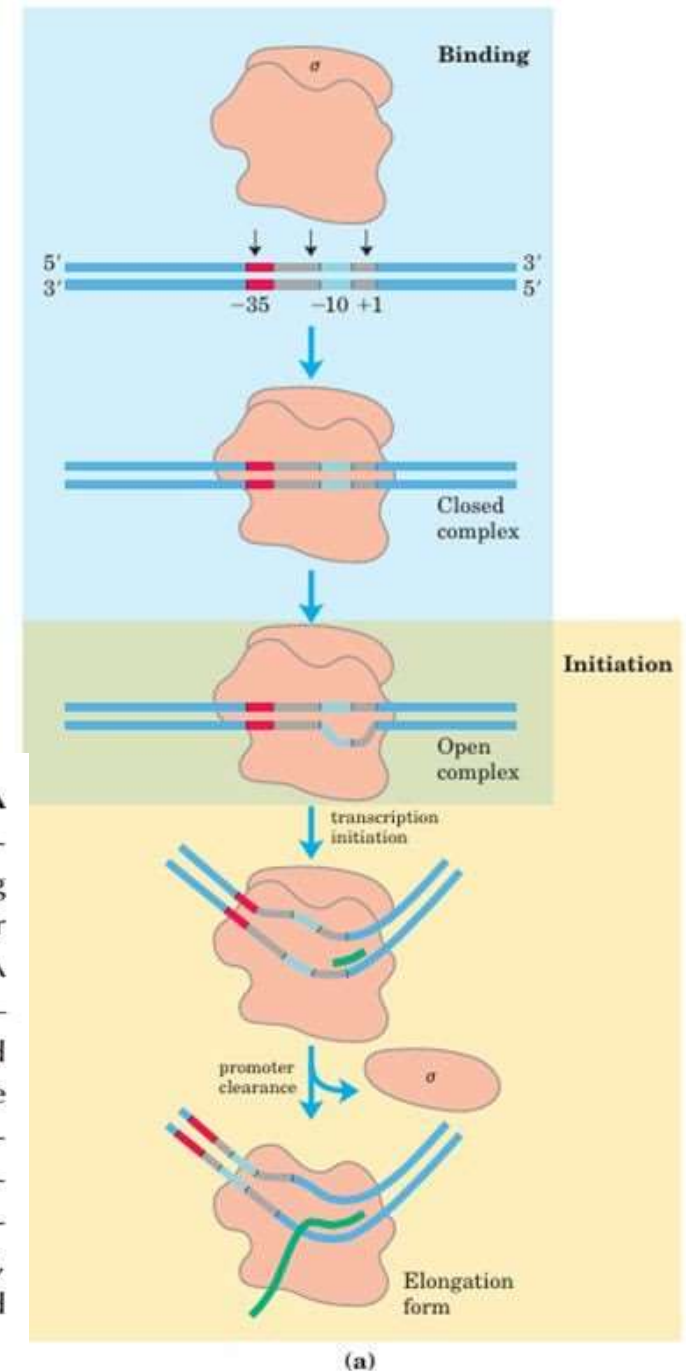
direction. Synthesis of mRNAs in adenovirus is actually much more complex than shown here. Many of the mRNAs shown for the upper strand are initially synthesized as a single, long transcript (25,000 nucleotides), which is then extensively processed to produce the separate mRNAs. Adenovirus causes upper respiratory tract infections in some vertebrates.

1. **First**, the polymerase binds to the promoter, forming, in succession, a closed complex (in which the bound DNA is intact) and an open complex (in which the bound DNA is intact and partially unwound near the -10 sequence).
2. **Second**, transcription is initiated within the complex, leading to a conformational change that converts the complex to the elongation form, followed by movement of the transcription complex away from the promoter (promoter clearance).

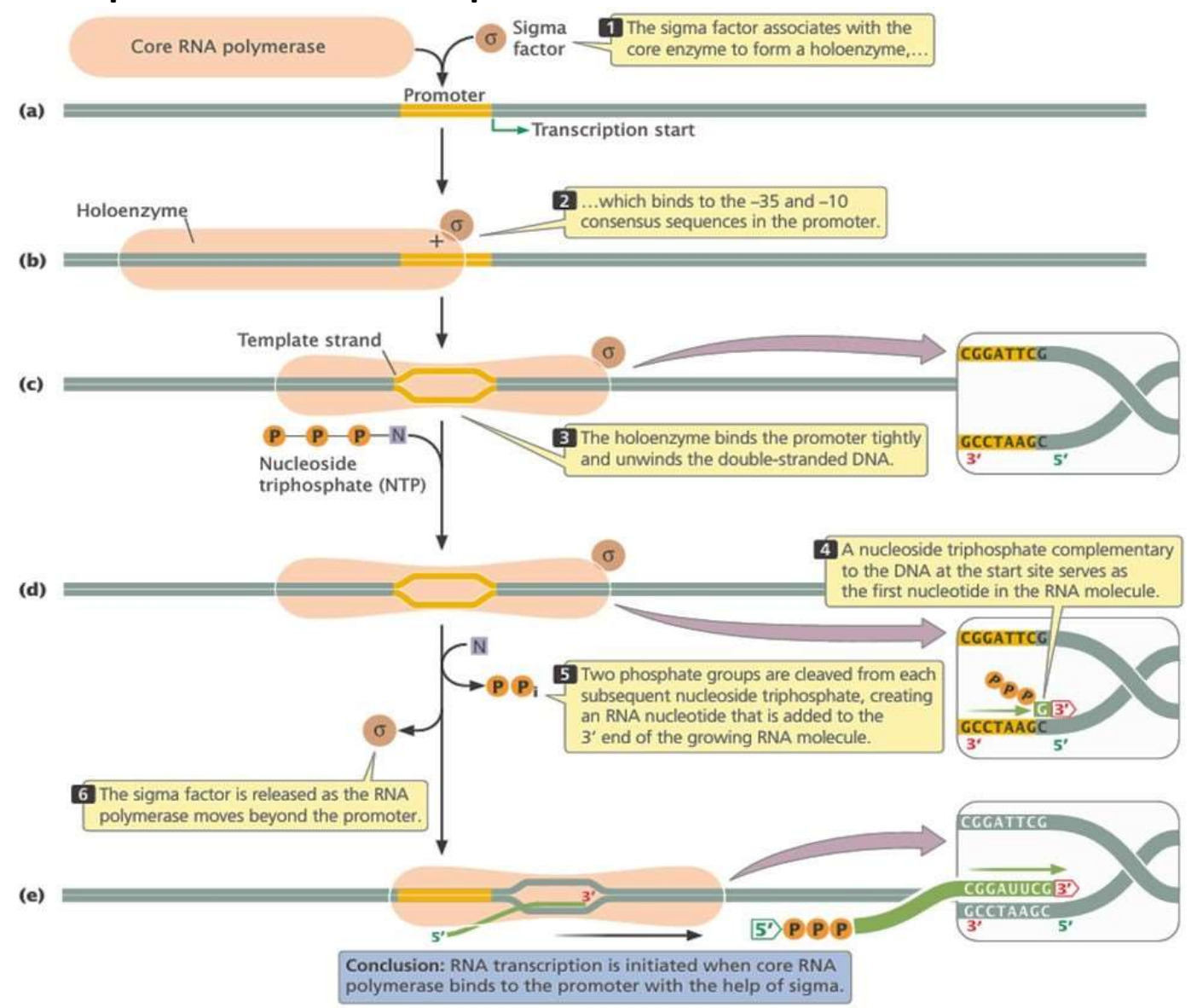
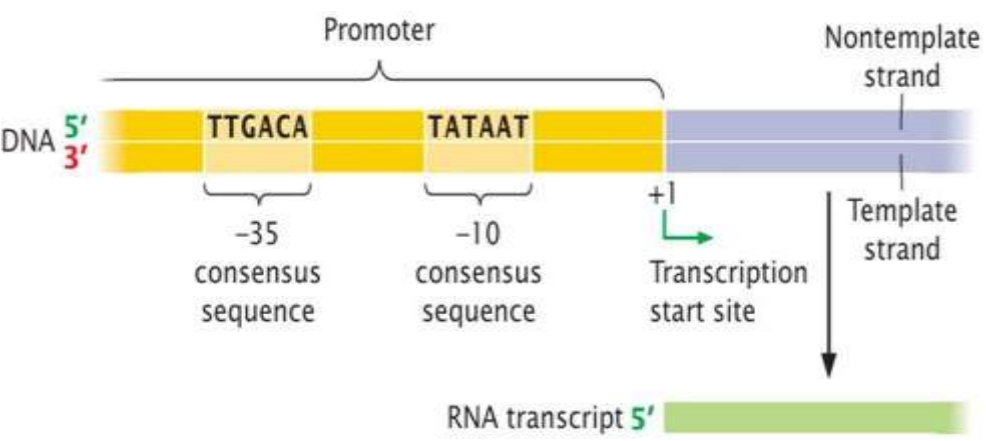
Any of these steps can be affected by the specific makeup of the promoter sequences.

The σ subunit dissociates as the polymerase enters the elongation phase of transcription (Fig. 26–6a).

FIGURE 26–6 Transcription initiation and elongation by *E. coli* RNA polymerase. (a) Initiation of transcription requires several steps generally divided into two phases, binding and initiation. In the binding phase, the initial interaction of the RNA polymerase with the promoter leads to formation of a closed complex, in which the promoter DNA is stably bound but not unwound. A 12 to 15 bp region of DNA—from within the -10 region to position $+2$ or $+3$ —is then unwound to form an open complex. Additional intermediates (not shown) have been detected in the pathways leading to the closed and open complexes, along with several changes in protein conformation. The initiation phase encompasses transcription initiation and promoter clearance. Once the first 8 or 9 nucleotides of a new RNA are synthesized, the σ subunit is released and the polymerase leaves the promoter and becomes committed to elongation of the RNA.



- 1. **Initiation**, in which the transcription apparatus assembles on the promoter and begins the synthesis of RNA;
- 2. **Elongation**, in which DNA is threaded through RNA polymerase, the polymerase unwinding the DNA 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.



Elongation

- At the end of initiation, RNA polymerase undergoes a change in conformation (shape) and thereafter is no longer able to bind to the consensus sequences in the promoter.
- This change allows the polymerase to escape from the promoter and begin transcribing downstream.
- The sigma subunit is usually released after initiation, although some populations of RNA polymerase may retain sigma throughout elongation.
- As it moves downstream along the template, RNA polymerase progressively unwinds 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 rewinds 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 1000 to 2000 nucleotides per second in bacterial cells.

Termination

- RNA polymerase adds nucleotides to the 3' end of the growing RNA molecule until it transcribes a terminator.
- Most terminators are found upstream of the site at which termination actually takes place.
- Transcription therefore does not suddenly stop when polymerase reaches a terminator, rather, transcription stops after the terminator has been transcribed.

At the terminator, several overlapping events are needed to bring an end to transcription:

RNA polymerase must stop synthesizing RNA, the RNA molecule must be released from RNA polymerase, the newly made RNA molecule must dissociate fully from the DNA, and RNA polymerase must detach from the DNA template.

Bacterial cells possess two major types of terminators.

Rho-dependent terminators are able to cause the termination of transcription only in the presence of an **protein called the rho factor**.

Rho-independent terminators (also known as intrinsic terminators) are able to cause the end of transcription in the absence of rho.

Rho-independent terminators

Rho-independent terminators, which make up about 50 percent of all terminators in prokaryotes, have two common features.

1. 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 13.13).

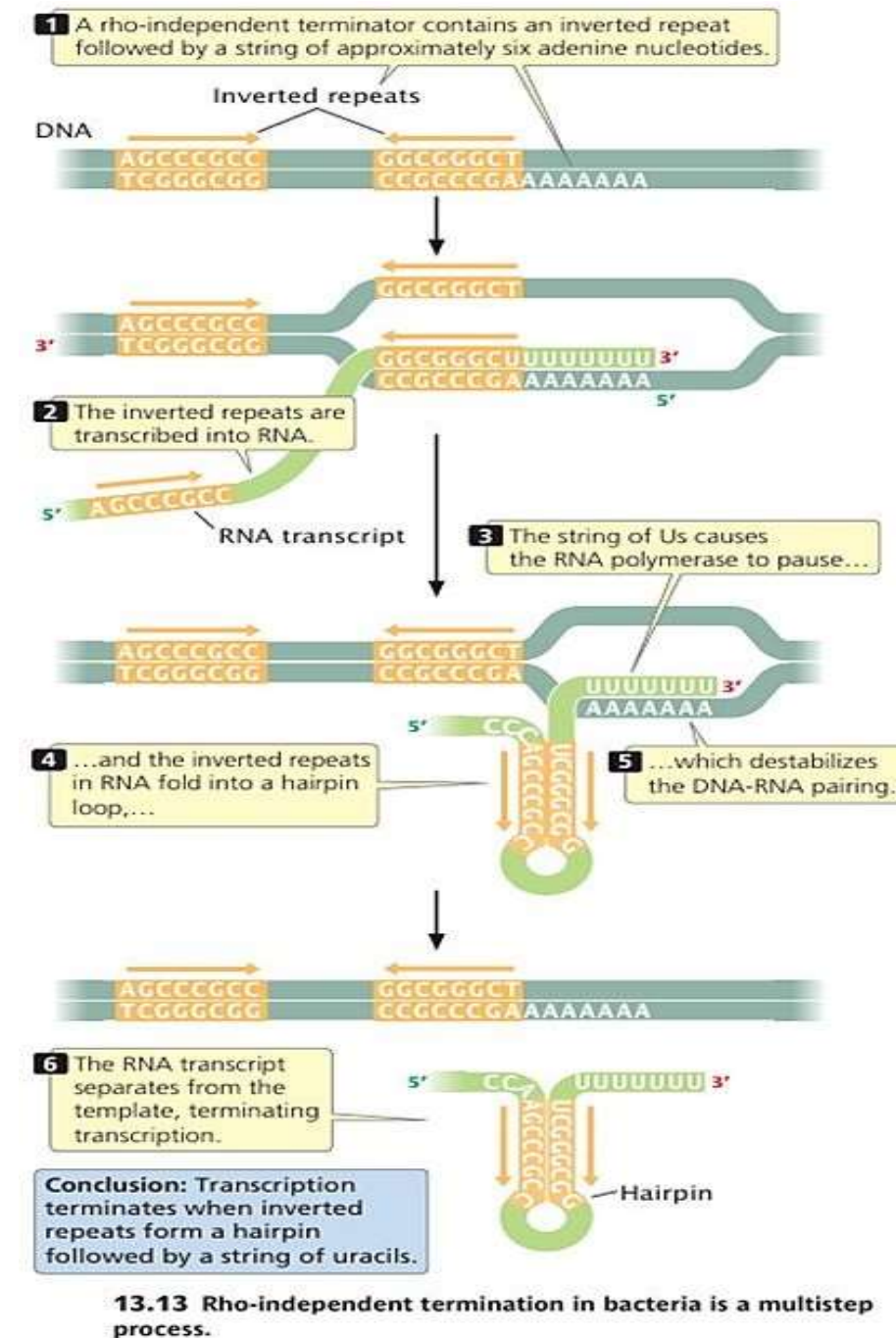
2. Second, in rho independent terminators, a string of seven to nine adenine nucleotides follows the second inverted repeat in the template DNA.

Their transcription produces a string of uracil nucleotides after the hairpin in the transcribed RNA.

The string of uracils in the RNA molecule causes the RNA polymerase to pause, allowing time for the hairpin structure to form.

Evidence suggests that the formation of the hairpin destabilizes the DNA–RNA pairing, causing the RNA molecule to separate from its DNA template. Separation may be facilitated by the adenine–uracil base pairings, which are relatively weak compared with other types of base pairings.

When the RNA transcript has separated from the template, RNA synthesis can no longer continue (see Figure 13.13).



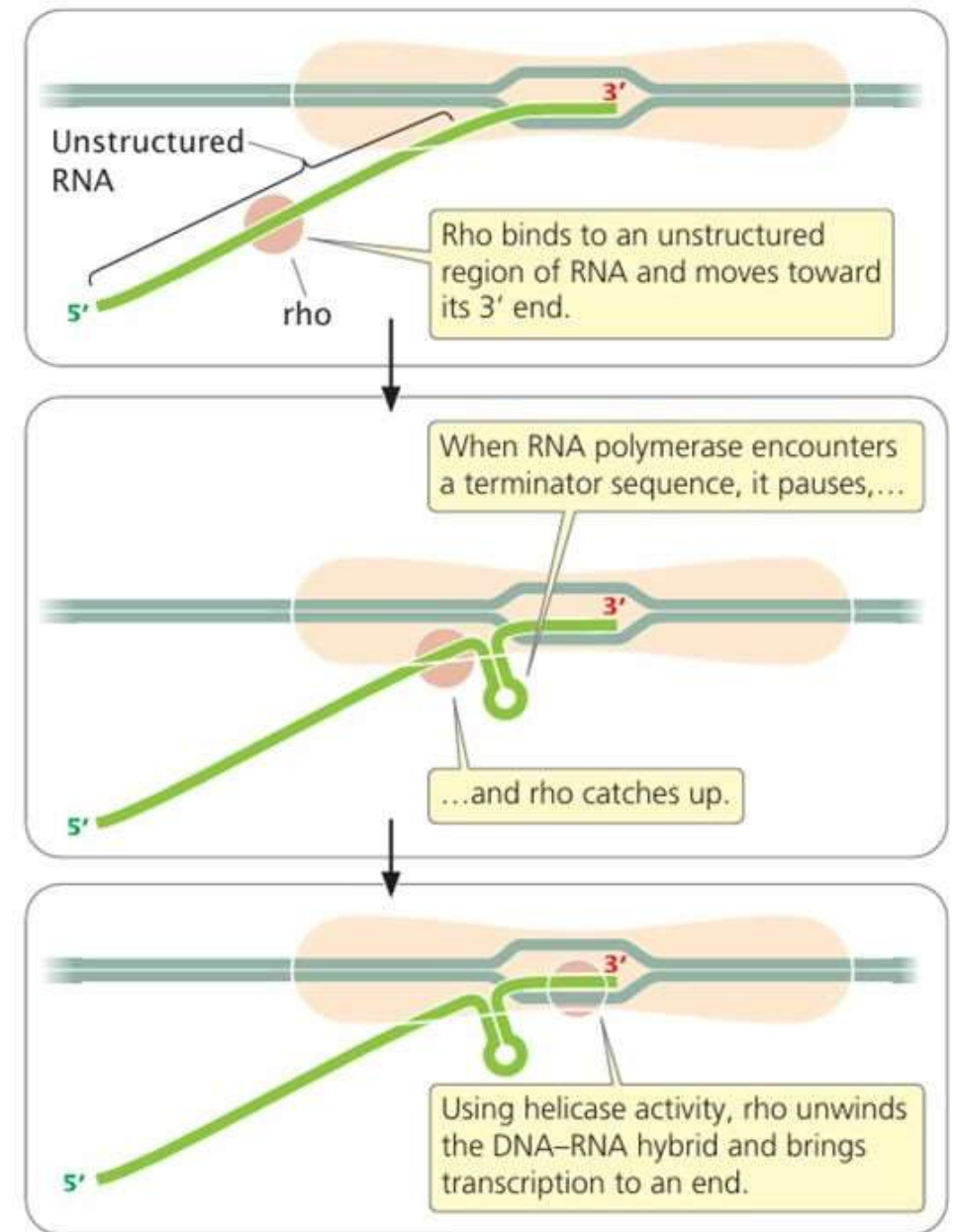
Rho-dependent terminators **Rho-dependent terminators have two features:**

- (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 a binding site for the rho protein, which binds the RNA and moves toward its 3' end, following the RNA polymerase (Figure 13.14).

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 transcription to an end.



13.14 The termination of transcription in some bacterial genes requires the presence of the rho protein.

Processing of eukaryotic mRNA:

Table 14.2 Posttranscriptional modifications to eukaryotic pre-mRNA

Modification	Function
Addition of 5' cap	Facilitates binding of ribosome to 5' end of mRNA, increases mRNA stability, enhances RNA splicing
3' cleavage and addition of poly(A) tail	Increases stability of mRNA, facilitates binding of ribosome to mRNA
RNA splicing	Removes noncoding introns from pre-mRNA, facilitates export of mRNA to cytoplasm, allows for multiple proteins to be produced through alternative splicing
RNA editing	Alters nucleotide sequence of mRNA

1. The addition at its 5' cap:

One type of modification of eukaryotic pre-mRNA is the **addition at its 5' end of a structure called a 5' cap**.

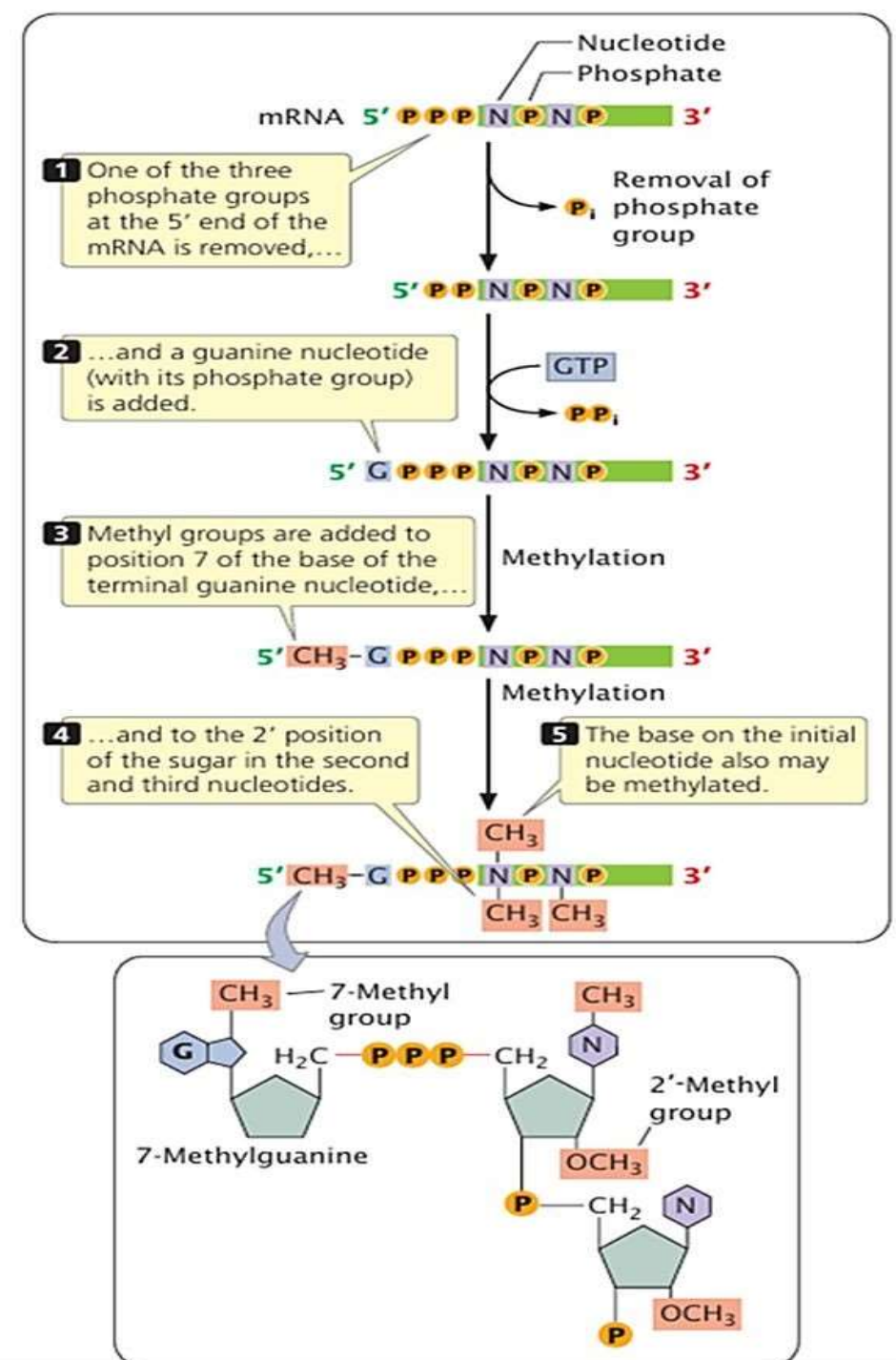
The cap consists of an extra nucleotide at the 5' end of the mRNA and methyl groups (CH_3) on the base in the newly added nucleotide and on the 2'-OH group of the sugar of one or more nucleotides at the 5' end (Figure 14.6).

The addition of the cap takes place rapidly after the initiation of transcription and, the 5' cap functions in the initiation of translation.

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.

14.6 Most eukaryotic mRNAs have a 5' cap. The cap consists of a nucleotide with 7-methylguanine attached to the pre-mRNA by a unique 5'-5' bond (shown in detail in the bottom box).



2. The Addition of the Poly(A) Tail:

A second type of modification to eukaryotic mRNA is the **addition of 50 to 250 or more adenine nucleotides at the 3' end, forming a poly(A) tail.**

These nucleotides are not encoded in the DNA but are added after transcription in a process termed **polyadenylation.**

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 and determines the point at which cleavage will take place.

A sequence rich in uracil nucleotides (or in guanine and uracil nucleotides) is typically downstream of the cleavage site.

A large number of proteins take part in finding the cleavage site and removing the 3' end.

After cleavage has been completed, adenine nucleotides are added to the new 3' end, creating the poly(A) tail.

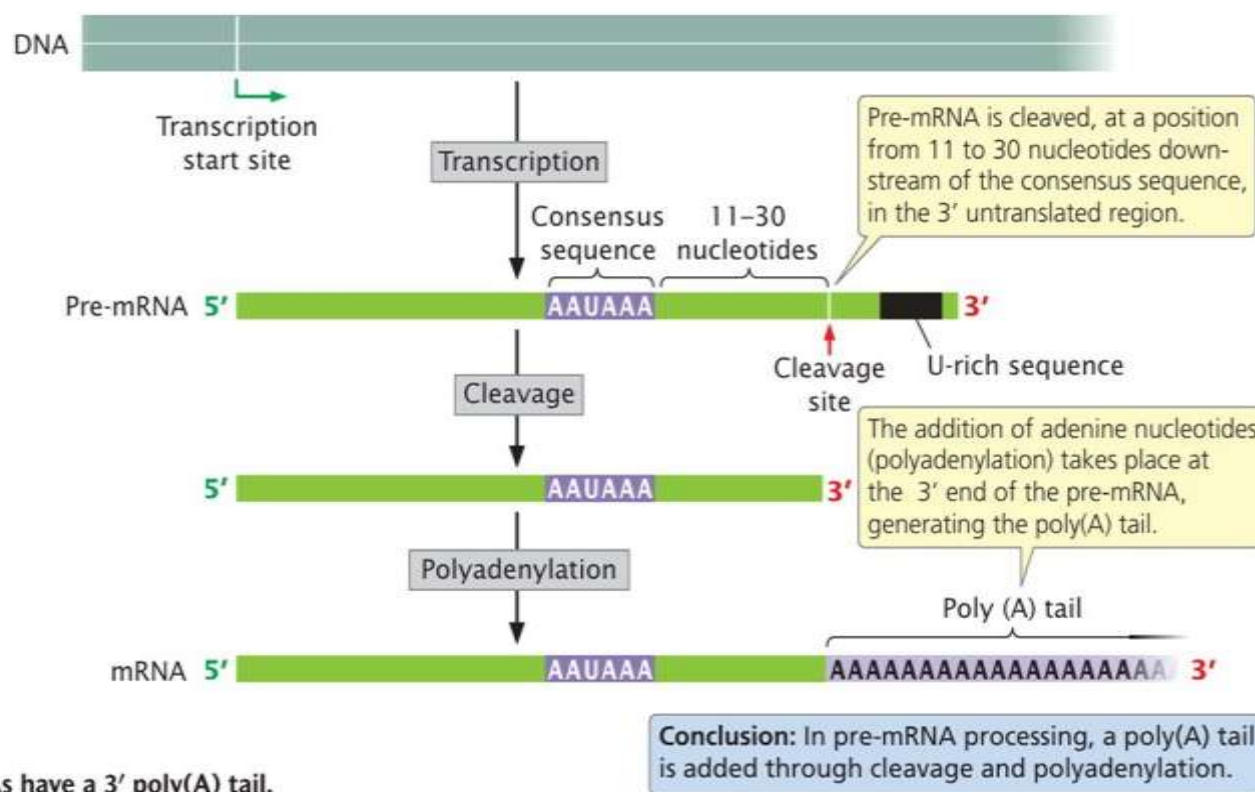
The poly(A) tail confers stability on many mRNAs, increasing the time during which the mRNA remains intact and available for translation before it is degraded by cellular enzymes.

The stability conferred by the poly(A) tail depends on the proteins that attach to the tail and on its length.

The poly(A) tail also facilitates attachment of the ribosome to the mRNA.

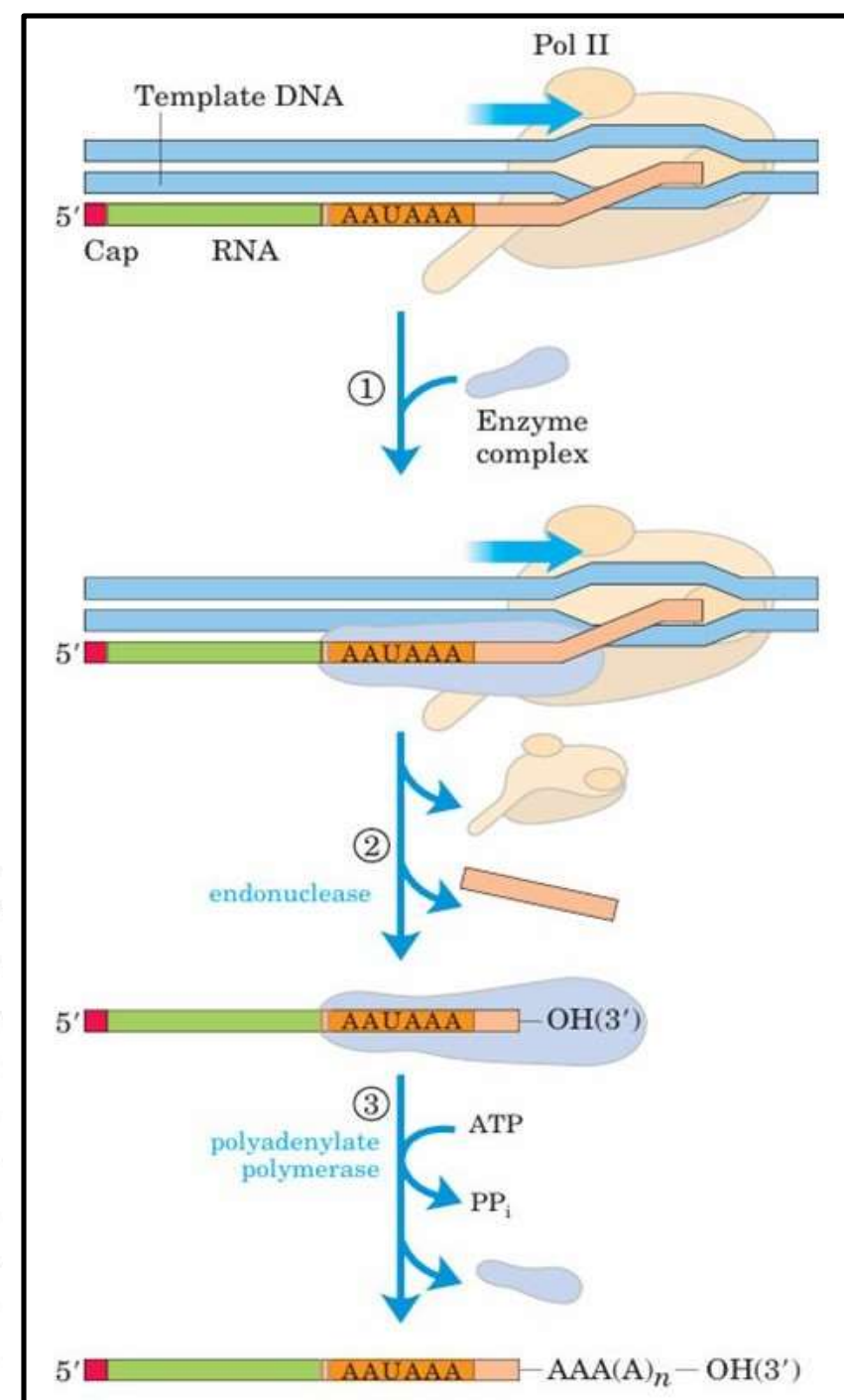
Poly(U) tails are added to the 3' ends of some mRNAs, microRNAs, and small nuclear RNAs.

Although the function of poly(U) tails is still under investigation, evidence suggests that poly(U) tails on some mRNAs may facilitate their degradation.



14.7 Most eukaryotic mRNAs have a 3' poly(A) tail.

FIGURE 26-17 Addition of the poly(A) tail to the primary RNA transcript of eukaryotes. Pol II synthesizes RNA beyond the segment of the transcript containing the cleavage signal sequences, including the highly conserved upstream sequence (5')AAUAAA. ① The cleavage signal sequence is bound by an enzyme complex that includes an endonuclease, a polyadenylate polymerase, and several other multisubunit proteins involved in sequence recognition, stimulation of cleavage, and regulation of the length of the poly(A) tail. ② The RNA is cleaved by the endonuclease at a point 10 to 30 nucleotides 3' to (downstream of) the sequence AAUAAA. ③ The polyadenylate polymerase synthesizes a poly(A) tail 80 to 250 nucleotides long, beginning at the cleavage site.



3. The other major type of modification of eukaryotic pre mRNA is the removal of introns by RNA splicing.

This modification takes place in the nucleus, before the RNA moves to the cytoplasm.

Splicing requires the presence of three sequences in the intron.

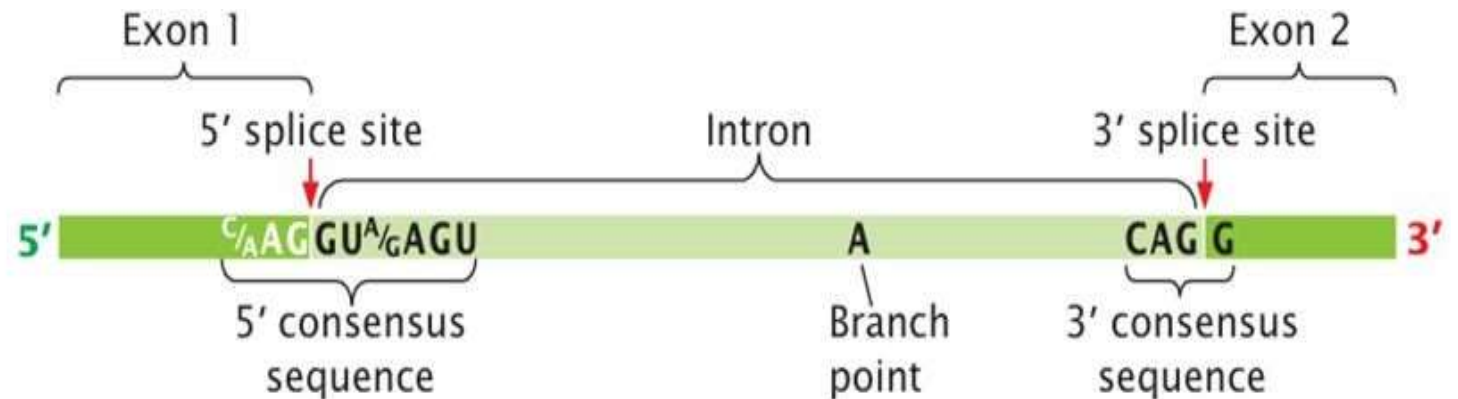
One end of the intron is referred to as the 5' splice site, and the other end is the 3' splice site (Figure 14.8); these splice sites possess short consensus sequences.

Most introns in pre mRNAs begin with GU and end with AG, indicating that these sequences play a crucial role in splicing. Indeed, changing a single nucleotide at either of these sites prevents splicing.

The third sequence important for splicing is at the branch point, which is an adenine nucleotide that lies from 18 to 40 nucleotides upstream of the 3' splice site (see Figure 14.8).

The deletion or mutation of the adenine nucleotide at the branch point prevents splicing.

14.8 Splicing of pre-mRNA requires consensus sequences. Critical consensus sequences are present at the 5' splice site and the 3' splice site. A weak consensus sequence (not shown) exists at the branch point.



Pre mRNA is spliced in two distinct steps.

An intron lies between an upstream exon (exon 1) and a downstream exon (exon 2).

In the first step of splicing, the pre-mRNA is cut at the 5' splice site.

This cut frees exon 1 from the intron, and the 5' end of the intron attaches to the branch point; that is, the intron folds back on itself, forming a structure called a lariat.

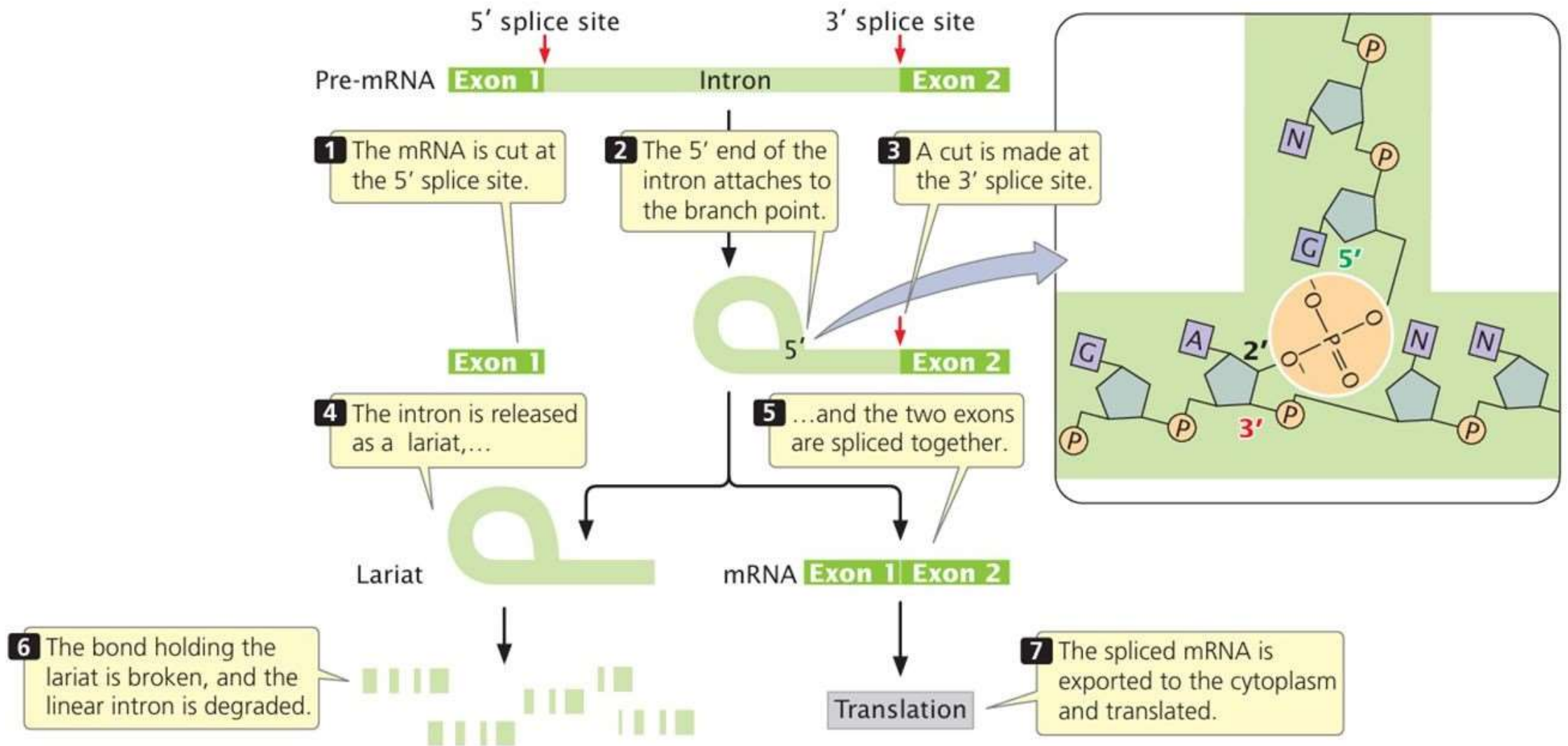
In this reaction, the guanine nucleotide in the consensus sequence at the 5' splice site bonds with the adenine nucleotide at the branch point through a transesterification reaction.

The result is that the 5' phosphate group of the guanine nucleotide is now attached to the 2'-OH group of the adenine nucleotide at the branch point.

In the second step of RNA splicing, a cut is made at the 3' splice site and, simultaneously, the 3' end of exon 1 becomes covalently attached (spliced) to the 5' end of exon 2. The intron is released as a lariat.

The intron becomes linear when the bond breaks at the branch point and is then rapidly degraded by nuclear enzymes.

The mature mRNA consisting of the exons spliced together is exported to the cytoplasm, where it is translated.



14.9 The splicing of nuclear introns requires a two-step process.

1. **INTRONS**- Non-coding sequences that are removed from mRNA before they are translated into protein.
2. **Exons**- Coding sequences- contain information for making proteins.