RNA Transcription and Modifications

Learning Outcomes:

1. Describe the organization of a protein-encoding gene and its mRNA transcript.
2. Outline the three stages of transcription

The primary function of the genetic material, which is DNA, is to store the information necessary to create a living organism. The information is contained within units called genes. At the molecular level, **a gene** is defined **as a segment of DNA that is used to make a functional product, either a RNA molecule or a polypeptide**.

Francis Crick (1958) named the principle of genetic information flow from DNA to mRNA to polypeptide or protein by **the central dogma of molecular biology.** The first step in genetic information flow is called **transcription,** this term refers to the process of synthesizing a copy of RNA from a DNA template. The total RNA transcripts called **transcriptome** refers to all types of RNA produced by a cell, which includes the **protein coding RNA also called messenger RNA** (mRNA) and **the non-coding RNA** which includes: the RNA involving in protein synthesis **Transfer RNA(t RNA), Ribosomal RNA(r RNA**), and those regulate the gene expression like (**long non-coding RNA (lncRNA) and small non-coding RNA(ncRNA).**

The **Protein-encoding genes** (also called **structural genes**) carry the information for the amino acid sequence of a polypeptide. When a protein-encoding gene is transcribed, the first product is an RNA molecule known as **messenger RNA (mRNA**). During polypeptide synthesis a process called **translation**. The sequence of nucleotides within the mRNA determines the sequence of amino acids in a polypeptide. One or more polypeptides then assemble into a functional protein. The structures and functions of proteins ultimately determine an organism's traits **(phenotype).**

The Stages of Transcription Are three: **Initiation, Elongation, and Termination.**

the transcription process of a gene is started at specific sequence of the gene called **promoter** which provides a site for beginning transcription, and the **terminator** specifies the end of transcription.

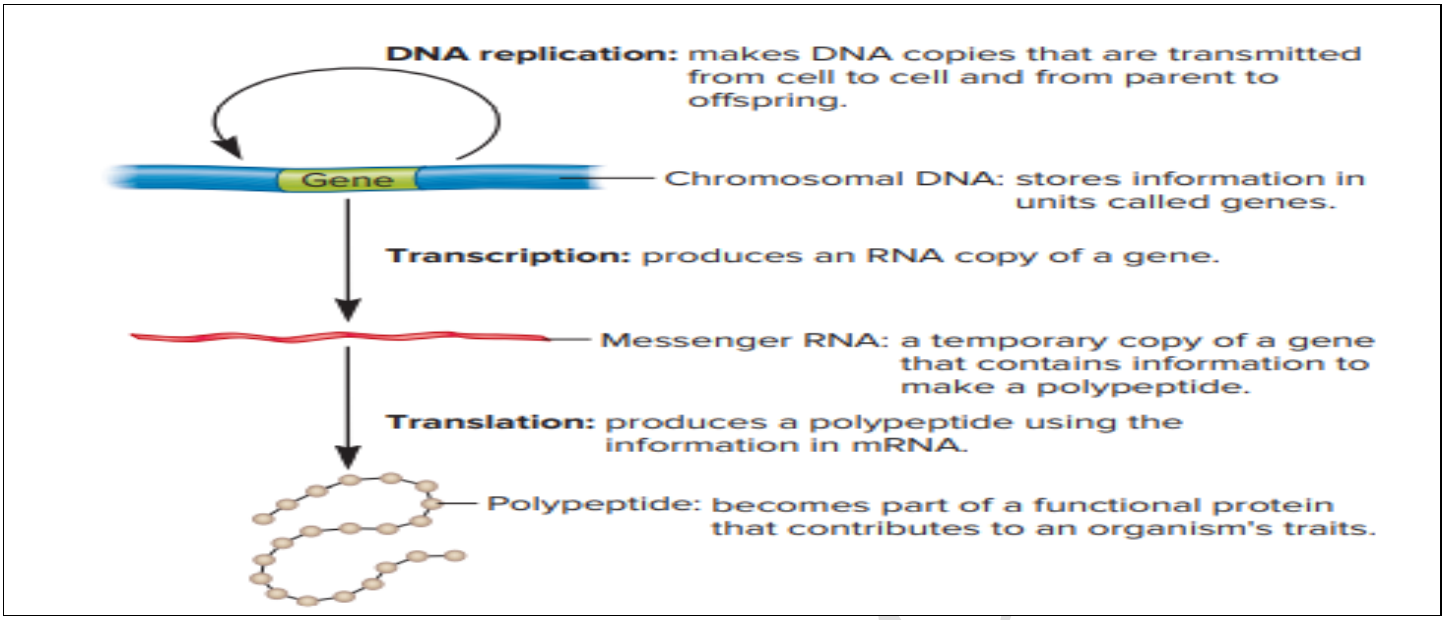
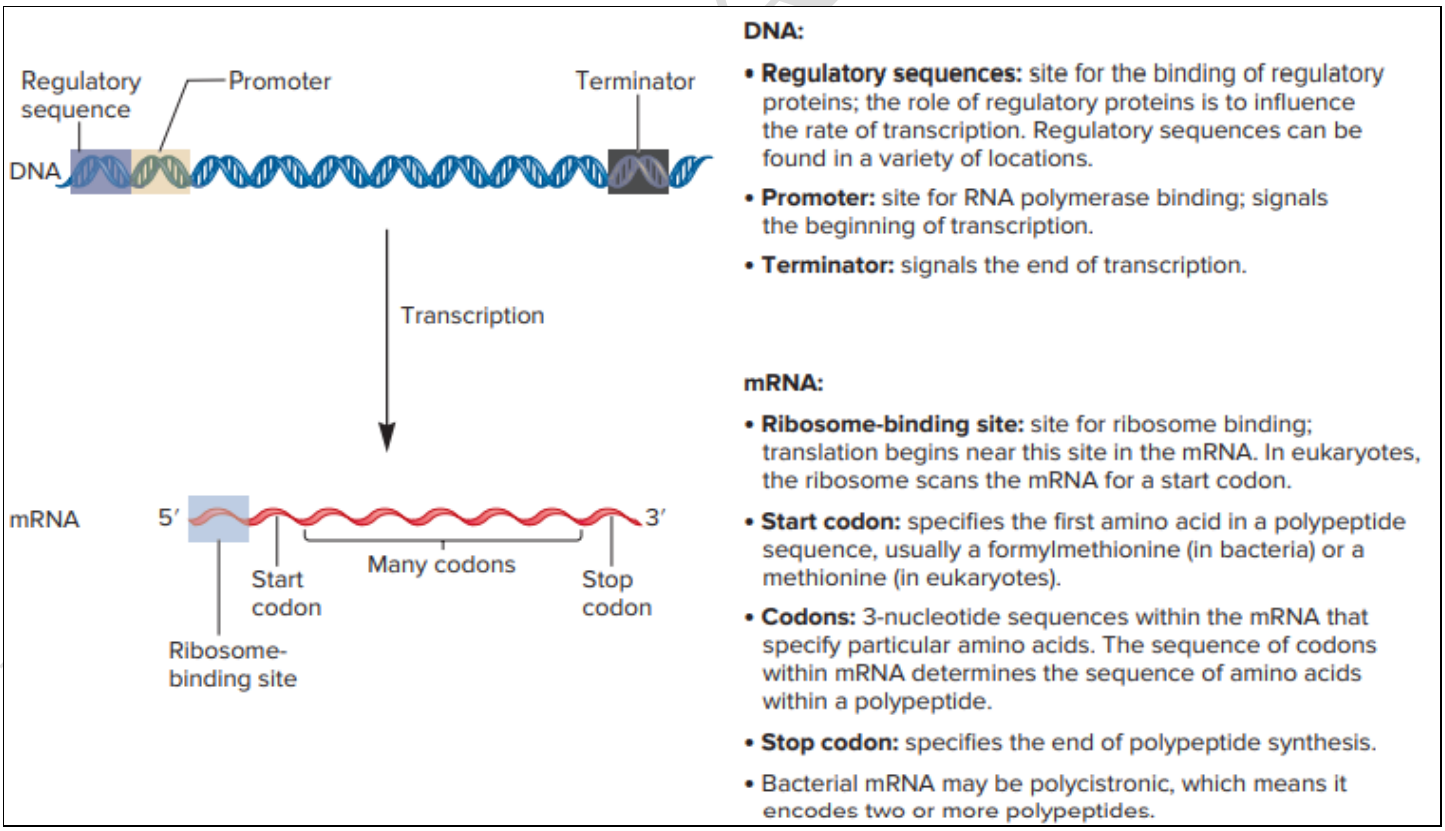


Figure (1): the flow of genetic information (the central dogma of molecular biology) From DNA to m RNA to Polypeptide chain (Brooker, RJ, 2018, Genetics: Analysis & Principles, 6th Edi. McGraw-Hill Education, US).



Figure(2): The organization of bacterial gene and its mRNA transcript (Brooker, RJ, 2018, Genetics: Analysis & Principles, 6th Edi. McGraw-Hill Education, US).

The base sequence in the RNA transcript is complementary to the template strand of DNA. The opposite strand of DNA is the non-template strand. genes, the **non-template** strand is also called the **coding strand**, or sense strand, because its sequence is the same as the transcribed mRNA that encodes a polypeptide, except that the DNA has T's in places where the mRNA contains U's. By comparison, the template strand is also called the non-coding strand, or antisense strand.

to start the **transcription factors** (specific proteins bind directly to the promoter or binds to a regulatory region to regulate the rate of transcription) and **RNA polymerase** first bind to the promoter when the DNA is in the form of a double helix. For transcription to occur, the DNA strands must be separated. This allows one of the two strands to be used as a template for the synthesis of a complementary strand of RNA. This synthesis occurs as RNA polymerase slides along the DNA (**elongation stage**), forming a small bubble-like structure known as the **open promoter complex**, or simply the **open complex**. Then, RNA polymerase reaches **a terminator,** which causes both RNA polymerase and the newly made RNA transcript to dissociate from the DNA.

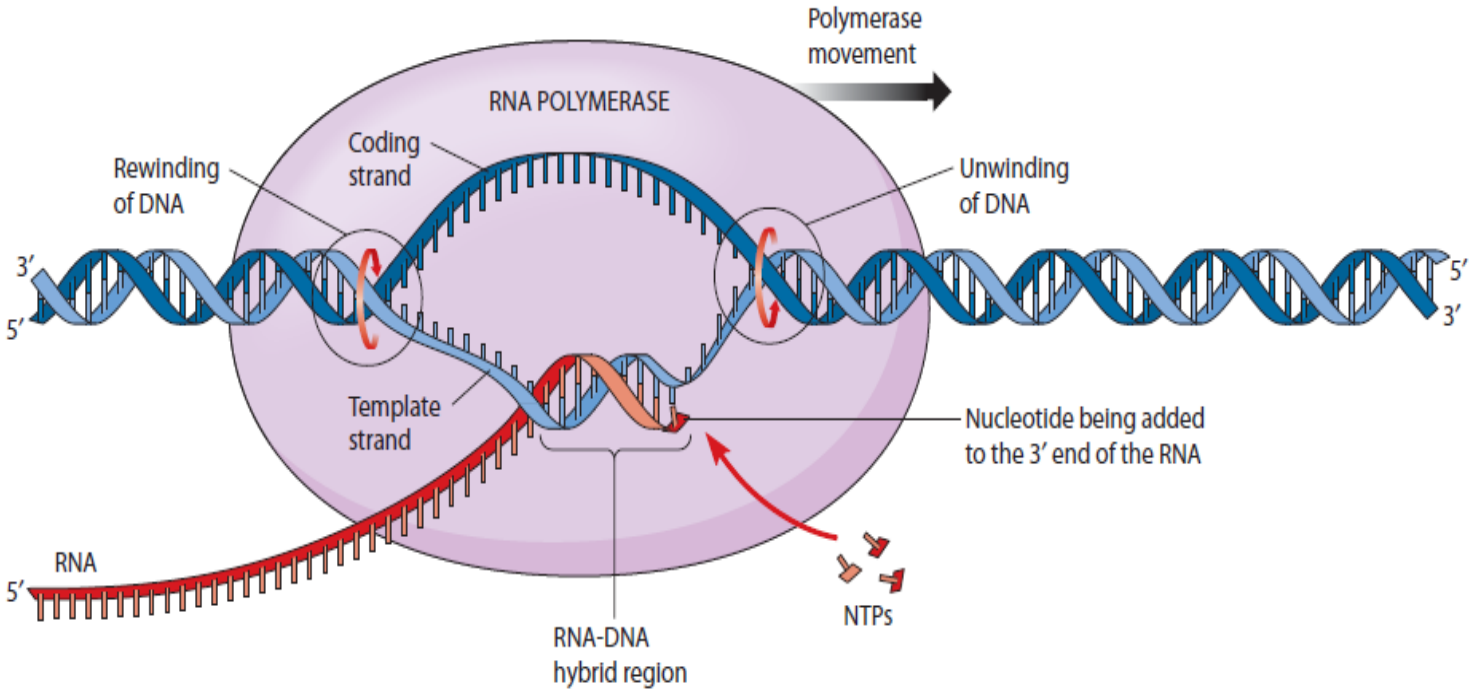


Figure (3): Transcription process. (Hardin,J. ; G. Bertoni And L. J. Kleinsmith, World Of The Cell, 8 Th Edition, 2012, Beakers)

The Initiation Stage:

The promoter gets its name from the idea that it "promotes" gene expression. this sequence of bases directs the exact location for the initiation of transcription. Most of the promoter is located just **ahead of, or upstream from**, the site where transcription of a gene actually begins. By convention, the bases in a promoter sequence are numbered in relation to the transcriptional start site. This site is the first base used as a template for transcription and is denoted +1. The bases preceding this site are numbered in a negative direction. No base is numbered zero. Therefore, most of the promoter is labeled with negative numbers that describe the number of bases preceding the beginning of transcription. Two important sequences within the promoter sequence, which are located at approximately the −35 and −10 site. The sequence in the top DNA strand at the −35 site is 5′–TTGACA–3′, and the one at the −10 site is 5′–TATAAT–3′. The TATAAT sequence is called the Pribnow box after David Pribnow, who initially discovered it in 1975.

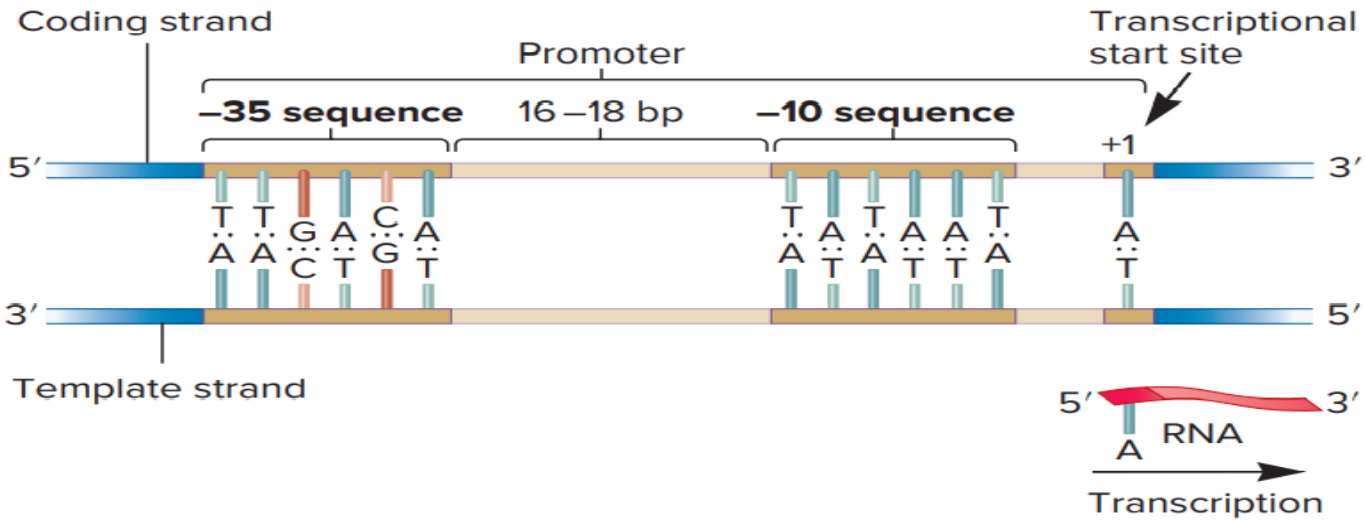


Figure (4): The conventional numbering system of promoters (Brooker, RJ, 2018, Genetics: Analysis & Principles, 6th Edi. McGraw-Hill Education, US).

The enzyme that catalyzes the synthesis of RNA is **RNA polymerase**. In *E. coli*, **the core enzyme** is composed of **five subunits, α2ββ′ω**. The association of a **sixth subunit, sigma (σ) factor, with the core enzyme creates what is referred to as RNA polymerase holoenzyme. The two α** subunits are important in the proper assembly of the holoenzyme and in the process of binding to DNA. **The β and β′** subunits are also needed for binding to the DNA, and they carry out the catalytic synthesis of RNA. **The ω (omega)** subunit is important for the proper assembly of the core enzyme. **The holoenzyme is required to initiate transcription;** the primary role of **σ factor is to recognize the promote**r. Proteins such as σ factor that influence the function of RNA polymerase are types of transcription factors. After RNA polymerase holoenzyme is

assembled into its six subunits, it binds loosely to the DNA and then slides along the DNA, such as a train rolls down the tracks. **the holoenzyme encounters a promoter, σ factor recognizes both the −35 and −10 sequences.** the process of transcription is initiated when σ factor within the holoenzyme has bound to the promoter to form a closed complex.

then σ factor is released from the core enzyme. The release of σ factor marks the transition to the elongation phase of transcription. The core enzyme may now slide down the DNA to synthesize a strand of RNA.

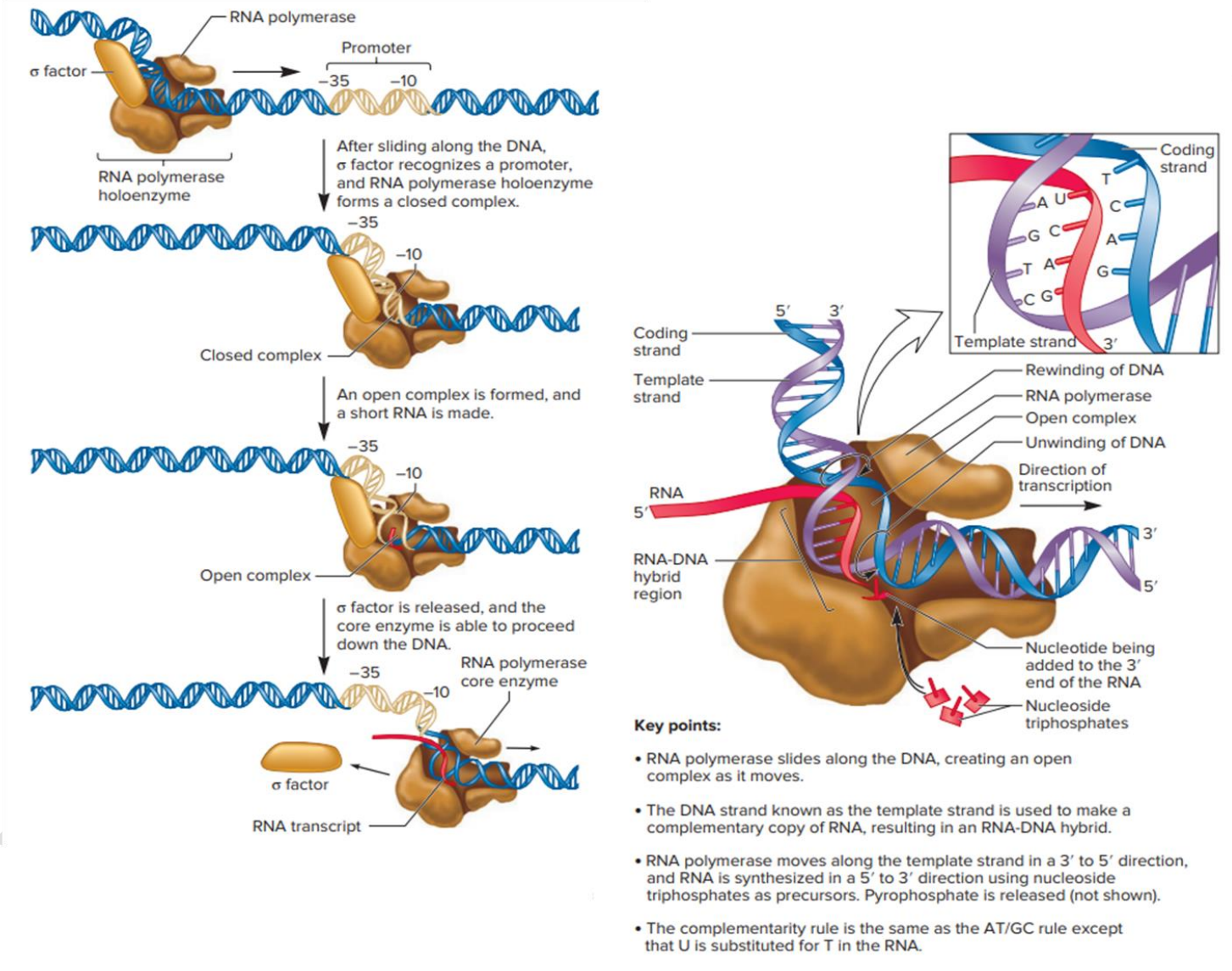


Figure (5): The initiation stage of transcription in bacteria (Brooker, RJ, 2018, Genetics: Analysis & Principles, 6th Edi. McGraw-Hill Education, US).

The Elongation Stage:

After the initiation stage of transcription is completed, the RNA transcript is made during the elongation stage. During the synthesis of the RNA transcript, RNA polymerase moves along the DNA, causing it to unwind. RNA polymerase always connects nucleotides in the 5′ to 3′ direction. During this process, RNA polymerase catalyzes the formation of a bond between the 5′ PO4 2− group on one nucleotide and the 3′−OH group on the previous nucleotide. The complementarity rule is similar to the AT/GC rule, except that uracil substitutes for thymine in the RNA. the transcription of multiple genes within a chromosome, the direction of transcription and the DNA strand used as a template vary among different genes.

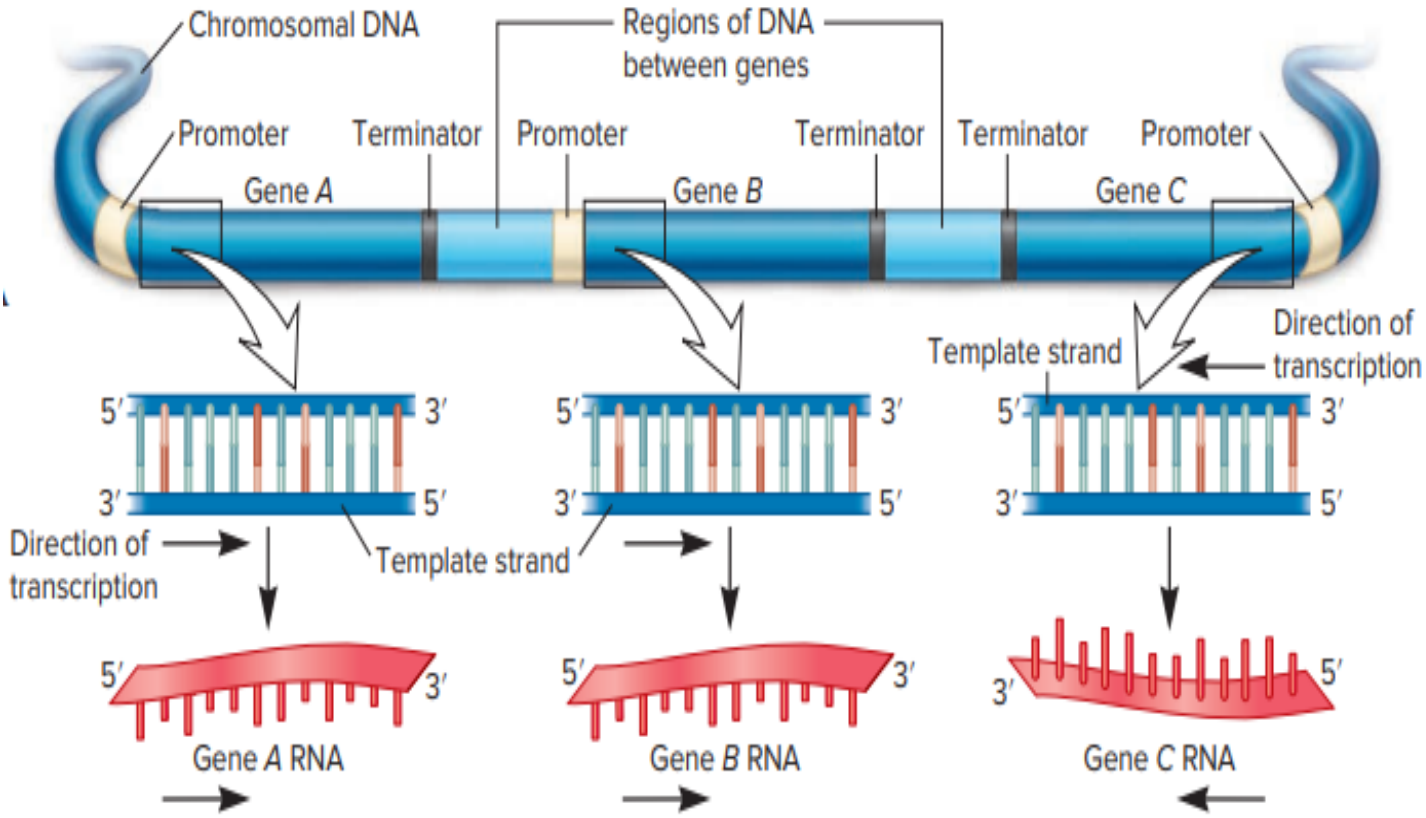


Figure (6): The transcription of three different genes found in the same chromosome. RNA polymerase synthesizes each RNA transcript in a 5′ to 3′ direction, sliding along a DNA template strand in a 3′ to 5′ direction. However, which strand is used as the template strand varies from gene to gene. For example, genes A and B use the bottom strand, but gene C uses the top strand (Brooker, RJ, 2018, Genetics: Analysis & Principles, 6th Edi. McGraw-Hill Education, US).

The Termination Stage

Termination occurs when this short RNA-DNA hybrid region is forced to separate, thereby releasing RNA polymerase as well as the newly made RNA transcript. In *E. coli*, two different mechanisms for termination have been identified. For certain genes, an RNA-binding protein known as ρ (rho) is responsible for terminating transcription, in a mechanism called **ρ-dependent termination**. For other genes, termination does not require the involvement of the ρ protein and in these cases, it is referred to as **ρ-independent termination.**

a. Rho-dependent (ρ-dependent) or type II terminators lack the poly(U) region, and many also lack the palindrome. The protein ρ is required for termination. It has two domains, one binding RNA and the other binding ATP. ATP hydrolysis provides energy for ρ to move along the transcript and destabilize the RNA-DNA hybrid at the termination region.

b.Rho-independent (ρ-independent) or type I terminators have two-fold symmetry that would allow a hairpin loop to form. The palindrome is followed by 4-8U residues in the transcript, and together these sequences cause termination, possibly because rapid hairpin formation destabilizes the RNA-DNA hybrid.

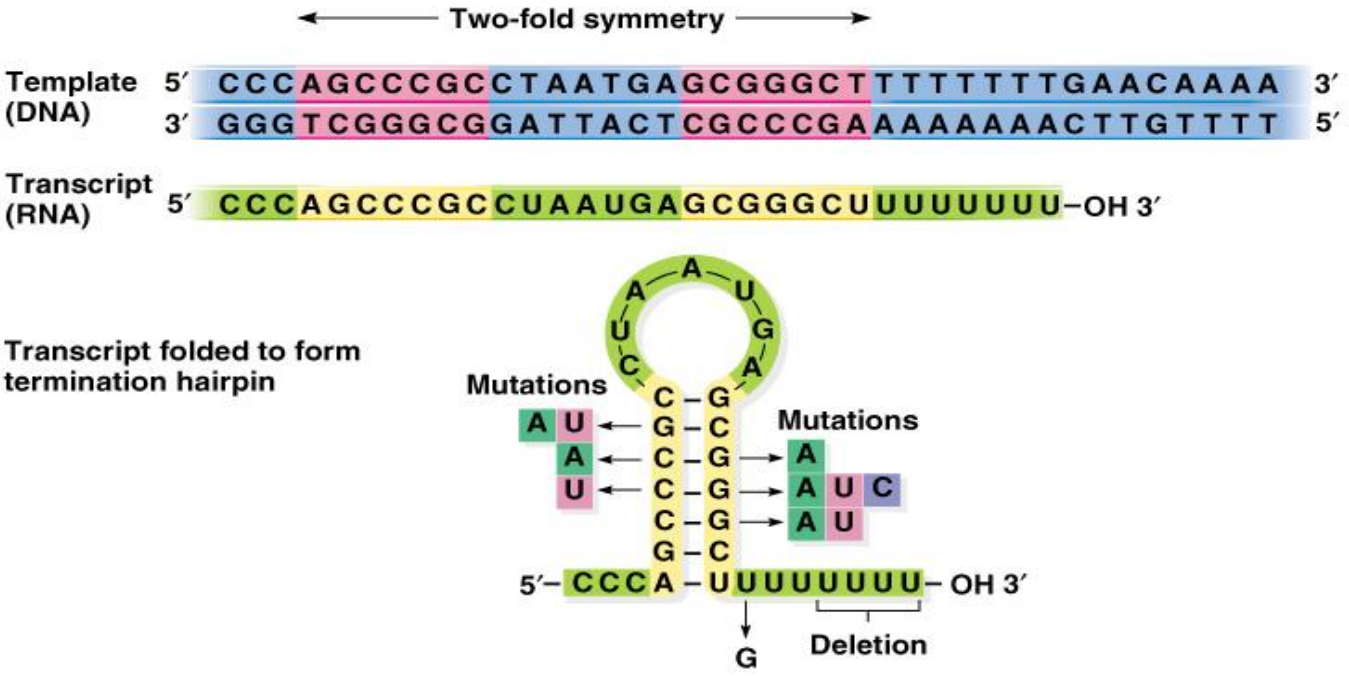


Figure (7): Sequence of a ρ-independent terminator and structure of the terminated RNA. (Russell, P., *I* Genetics, a molecular approach, 3rd edition, 2009, edited by Yue-Wen Wang. Pearsons,USA).

Transcription in Eukaryotes:

Eukaryotes have three different polymerases, each transcribing a different class of RNA. Processing of transcripts is also more complex in eukaryotes.

a. RNA polymerase I, located in the nucleolus, synthesizes three of the four rRNAs found in ribosomes: three of the RNAs (the 28S, 18S, and 5.8S rRNA molecules).

b. RNA polymerase II, located in the nucleoplasm, synthesizes messenger RNAs (mRNAs; translated to produce polypeptides) and
some small nuclear RNAs (snRNAs), some of which are involved in
R NA processing events.

c. RNA polymerase III, also located in the nucleoplasm, synthesizes the transfer RNAs (tRNAs), which bring amino acids to the ribosome; 5S rRNA, the fourth rRNA molecule found in each ribosome; and the small nuclear RNAs (snRNAs) not made by RNA polymerase II.

Transcription of Protein-Coding Genes by RNA Polymerase II.

Promoters control the expression of protein-coding genes. The promoter located upstream each gene and near the transcription start site. Examples include:
i. The TATA box: its full sequence is TATAAAA. This element aids in local DNA denaturation, and sets the start point for transcription.
ii. The initiator element (Inr), a pyramiding-rich sequence near the transcription start site.
Higher levels of transcription are induced by activator factors that bind DNA sequences called **enhancers.**

Characteristics of enhancers:

a. They are found in single or multiple copies.
b. They function in either orientation.
c. They function upstream, downstream or within the gene, although they are usually located upstream.
d. They may be several kb from the gene they control.

Eukaryotic mRNAs

a. Prokaryotes use the RNA transcript as mRNA without modification. Transcription and translation are coupled in the cytoplasm. Messages may be polycistronic.

b. Eukaryotes modify pre-RNA into mRNA by RNA processing. The processed mRNA migrates from nucleus to cytoplasm before translation. Messages are always monocistronic.

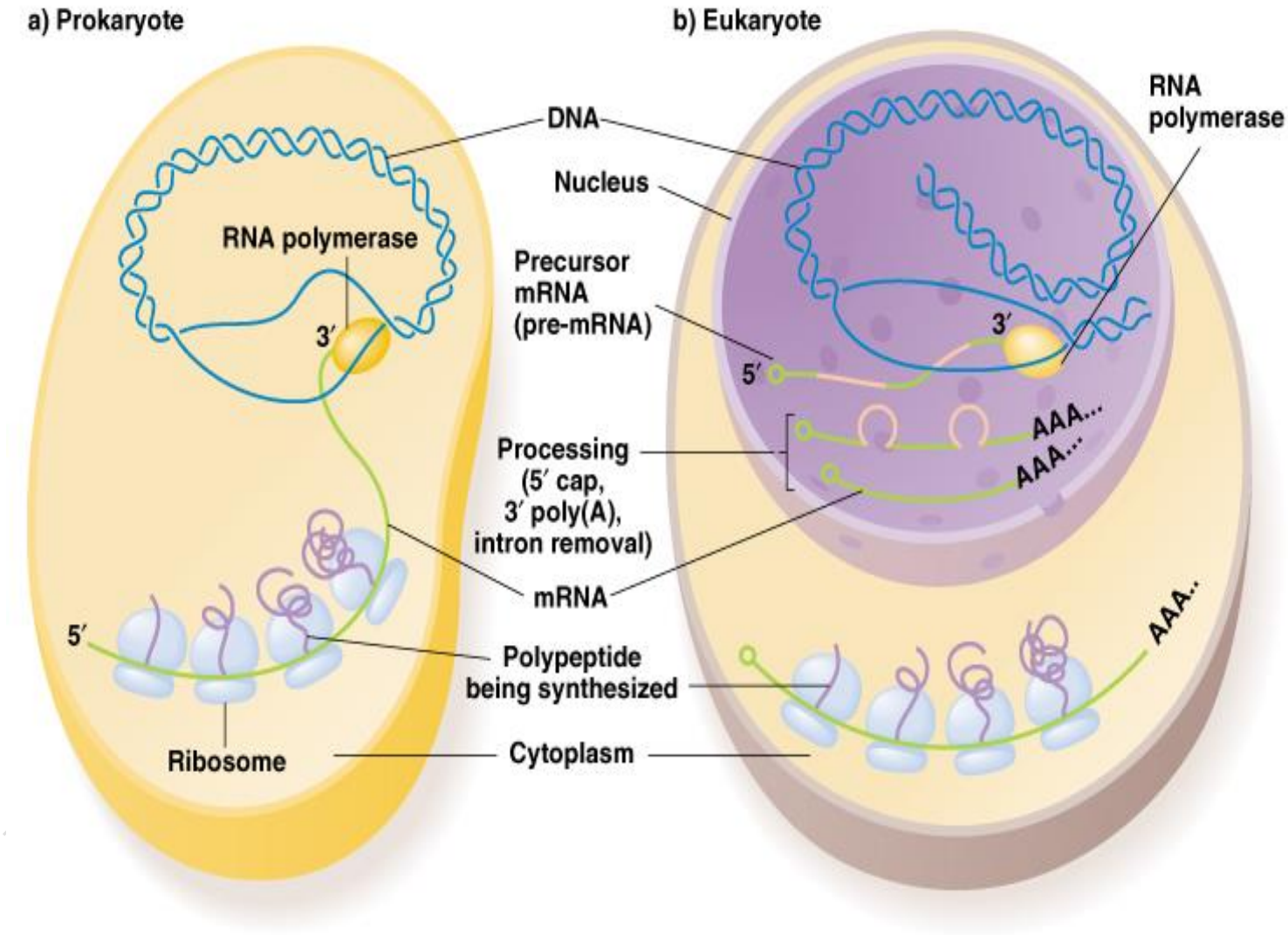


Figure (8);Processes for synthesis of functional mRNA in prokaryotes and eukaryotes. (Russell ,P. , *I* Genetics, a molecular approach , 3rd edition , 2009, edited by Yue-Wen Wang. Pearsons,USA).

Production of Mature mRNA in Eukaryotes:

Events in eukaryotic mRNA production are summarized in. They include:

a. Transcription of the gene by RNA polymerase II.
b. Addition of the 5' cap.
c. Addition of the poly(A) tail.
d. Splicing to remove introns.

In eukaryotes, the transcription of protein-encoding genes produces a long transcript known **as pre-mRNA,** which is made in the nucleus. This pre-mRNA is usually altered by **splicing** and other modifications before it exits the nucleus. pre-mRNA splicing requires the aid of a complex known as a spliceosome. The spliceosome is needed to recognize the boundaries of the intron and to properly remove it. the spliceosome is a large complex that splices pre-mRNA in eukaryotes. It is composed of five **subunits (U1, U2, U4, U5, and U6) known as snRNPs.** Each snRNP contains small nuclear RNA and a set of proteins. During splicing, the subunits of a spliceosome carry out several functions. **First, spliceosome subunits bind to an intron sequence and precisely recognize the intron-exon boundaries. In addition, the spliceosome must hold the pre-mRNA in the correct configuration to ensure the splicing together of the exons. And finally, the spliceosome catalyzes the chemical reactions that cause the intron to be removed and the exons to be covalently linked.**

Eukaryotic pre-RNAs often have introns (intervening sequences) between the exons (expressed sequences) that are removed.

- **5'- 3' Modifications by** adding 7-methyl guanosine (m^7G), to the 5' end using a 5'-to-5' linkage. The cap is used for ribosome binding to the mRNA during translation initiation.

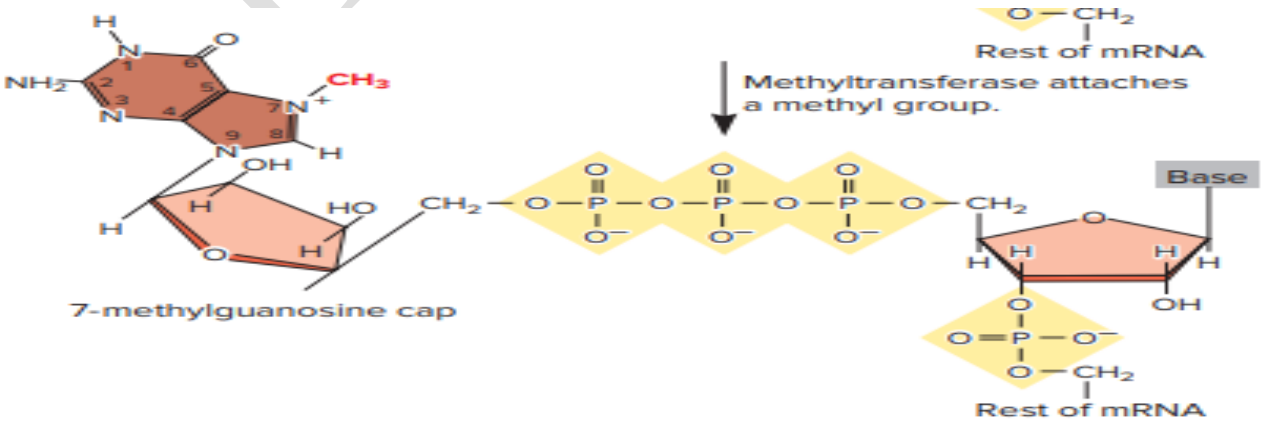


Figure (9): attachment of 7-methylguanosine cap to the 5' end of the pre m RNA.

- **The 3' end** of the pre-RNA has 50–250 adenines added enzymatically to **form a poly(A) tail**. The poly(A) tail is important in mRNA stability, and also plays a role in transcription termination.

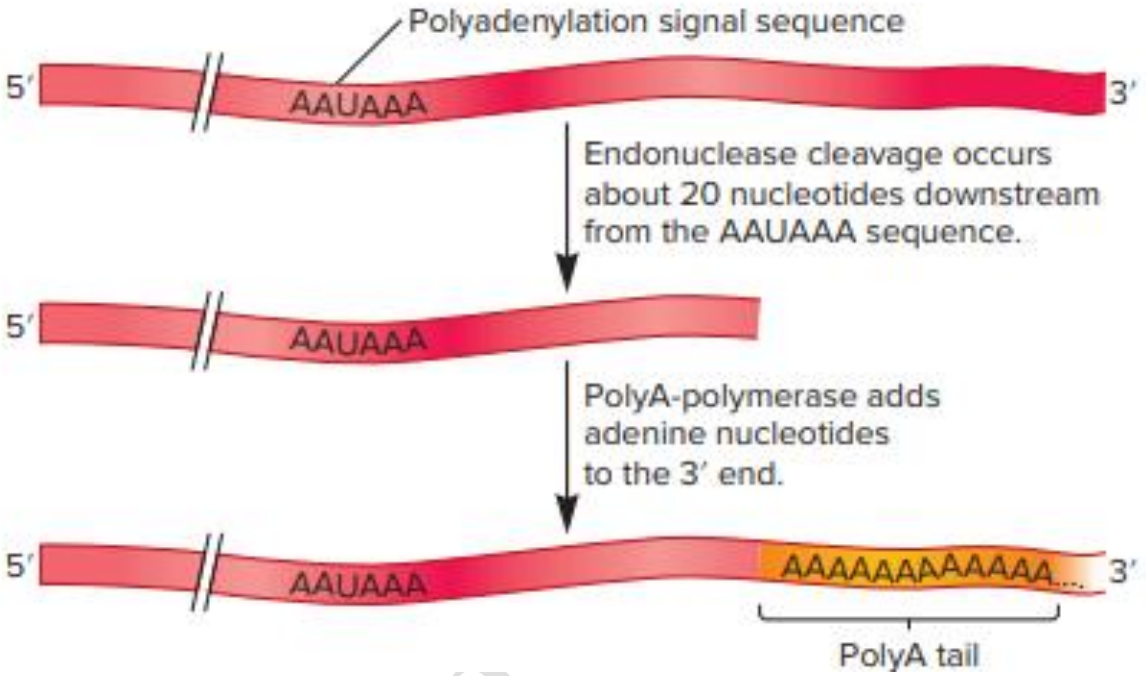


Figure (10): adding the poly A tail to the 3' end of the pre-RNA.

- **Removal of introns** is necessary for mRNA maturation by splicing.

Events in splicing together two exons (designated 1 and 2):
a. cleavage occurs at the 5' splice junction of exon 1 and the intron.
b. The G nucleotide at the free 5' end of the intron joins with a specific A nucleotide in the branch-point sequence of the intron, forming an RNA lariat structure.

c. The bond forming the lariat is a 2'-5' phosphodiester linkage between the 5' phosphate of the free guanine nt at the end of introns, and the 2' OH of the adenine nt in the branch-point sequence.

d. The introns lariat is excised, and the exons are joined to form a spliced mRNA. The introns RNA is degraded by the cell.

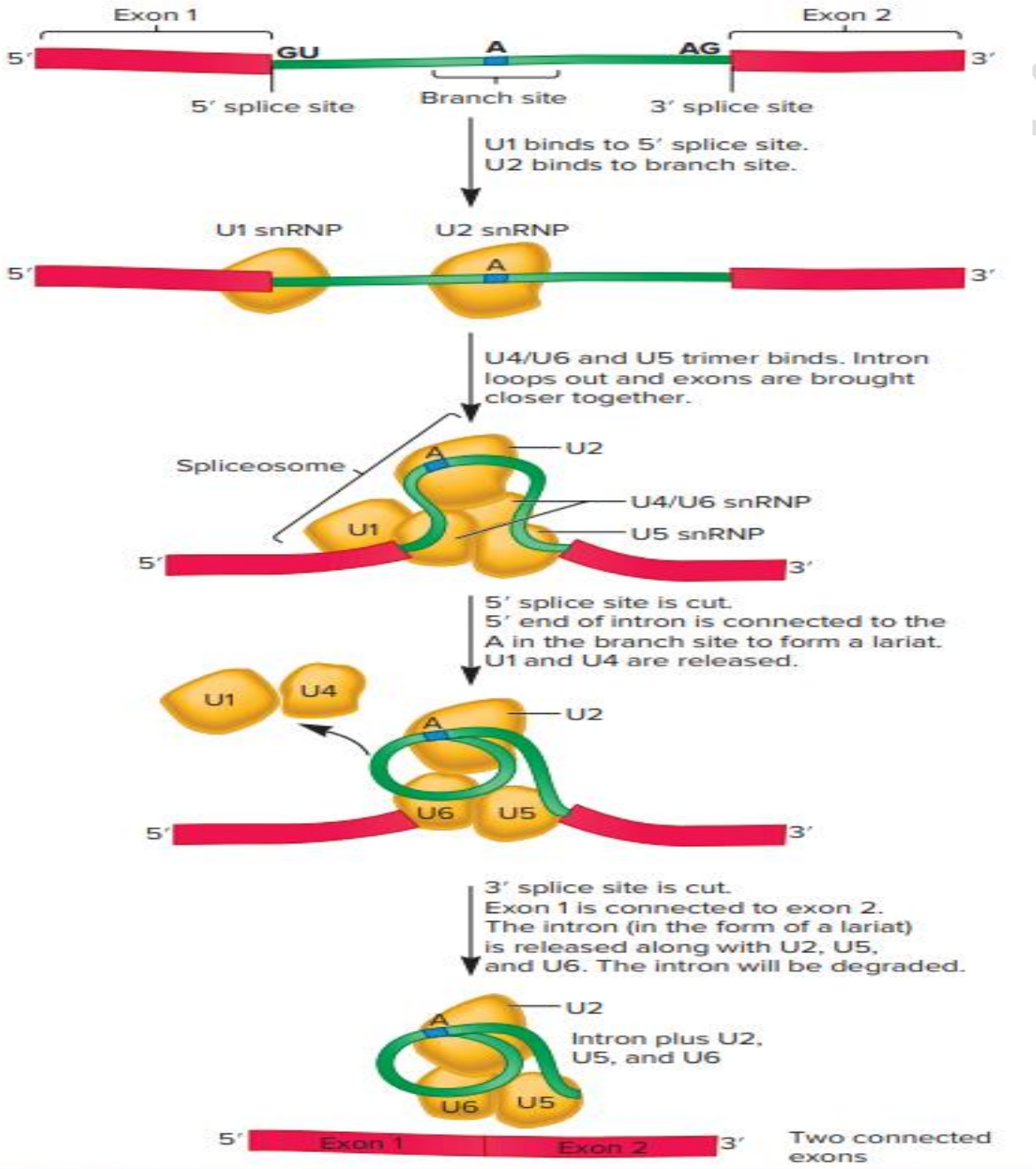


Figure (11): the splicing of intron. (Brooker, RJ, 2018, Genetics: Analysis & Principles, 6th Edi. McGraw-Hill Education, US).