Transcription

Key points of Transcription

1- **Transcription** is the process in which a gene's DNA sequence is copied (transcribed) to make an RNA molecule.
2- **RNA polymerase** is the main transcription enzyme.
3- Transcription begins when RNA polymerase binds to a **promoter** sequence near the beginning of a gene (directly or through helper proteins).
4- RNA polymerase uses one of the DNA strands (the **template strand**) as a template to make a new, complementary RNA molecule.
5- Transcription ends in a process called **termination**. Termination depends on sequences in the RNA, which signal that the transcript is finished.

Transcription is the process by which the synthesis of RNA molecules is initiated and terminated representing one strand of DNA duplex. By 'representing' means that the RNA is ***identical*** in sequence with one strand of the DNA, it is ***complementary*** to the other strand, which provides the ***template*** for its synthesis. It takes place by the usual process of complementary base pairing, catalysed by the enzyme ***RNA polymerase***.

The transcription of eukaryotic genes is a far more complicated process than transcription in prokaryotes. Eukaryotic transcription involves separate polymerases for the synthesis of rRNA, tRNA, and mRNA. In addition, a large number of proteins called transcription factors (TFs) are involved. TFs bind to distinct sites on the DNA either within the core promoter region, close (proximal) to it, or some distance away (distal). They are required both for the assembly of a transcription complex at the promoter and the determination of which genes are to be transcribed. [Note: Each eukaryotic RNA pol has its own promoters and TFs that bind core promoter sequences.]

It actually consists of two processes: **transcription** and **translation**. Transcription takes place in the nucleus. It uses DNA as a template to make an RNA molecule. RNA then leaves the nucleus and goes to a ribosome in the cytoplasm, where translation occurs. Translation reads the genetic code in mRNA and makes a protein.

Transcription is the first part of the central dogma of molecular biology: *DNA → RNA*. It is the transfer of genetic instructions in DNA to messenger RNA (mRNA).

RNA Polymerase

RNA polymerase being the key enzyme in transcription**,** it is worthwhile to study the details of its structure and mode of its action. A single type of RNA polymerase is responsible for synthesis of m-RNA, r-RNA and t-RNA in bacteria. However, **in eukaryotes** several different enzymes are required to synthesise the different types of RNA. They are called as:

• *RNA polymerase I,* is responsible for synthesis of rRNA (ribosomal)**.**

• *RNA polymerase II,* is the main enzyme synthesizing mRNAs.

• *RNA polymerase III,* is responsible for production of tRNA

Action of RNA Polymerase

1- Like DNA polymerases, RNA polymerases catalyze the formation of ester bonds between nucleotides that base-pair with the complementary nucleotides on the DNA template.

2- RNA polymerases differ from DNA polymerases in that they can **initiate** the **synthesis** of new strands in the absence of a **primer.**

3- In addition to catalyzing the polymerization of **ribonucleotides,** RNA polymerases must be able to recognize:

   a- the appropriate gene to transcribe,

   b- the appropriate strand of the double-stranded DNA to copy,

   c- and the **start point** of transcription

   d- control the frequency of transcription.

A- Sequences of Genes

Before transcription can take place, the DNA double helix must unwind near the gene that is getting transcribed. The region of opened-up DNA is called a **transcription bubble**. In transcription, a region of DNA opens up. One strand, the template strand, serves as a template for synthesis of a complementary RNA transcript. The other strand, the coding strand, is identical in base sequence and direction to the RNA transcript, except that it has uracil (U) bases in place of thymine (T) bases (Figure 1).

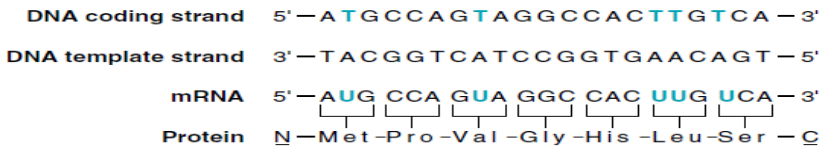

Figure 1: Relationship between the coding strand of DNA, the DNA template strand, the mRNA transcript, and the protein produced from the gene.

The site on the DNA from which the first RNA nucleotide is transcribed is called the +1 or the **initiation site**. Nucleotides that come before the initiation site are given negative numbers and said to be **upstream**. Nucleotides that come after the initiation site are marked with positive numbers and said to be **downstream**.

A gene consists of the transcribed region and the regions that regulate transcription of the gene (e.g., promoter and enhancer regions) (Figure 2). The base in the coding strand of the gene serving as the start point for transcription is numbered +1. This nucleotide corresponds to the first nucleotide incorporated into the RNA at the 5`-end of the transcript. Subsequent nucleotides within the transcribed region of the gene are numbered +2, +3, etc., toward the 3` -end of the gene. Untranscribed sequences to the left of the start point, known as the 5` -flanking region of the gene, are numbered -1, -2, -3, etc., starting with the nucleotide (-1) immediately to the left of the start point (+1) and moving from right to left. By analogy to a river, the sequences to the left of the start point are said to be upstream from the start point and those to the right are said to be downstream.

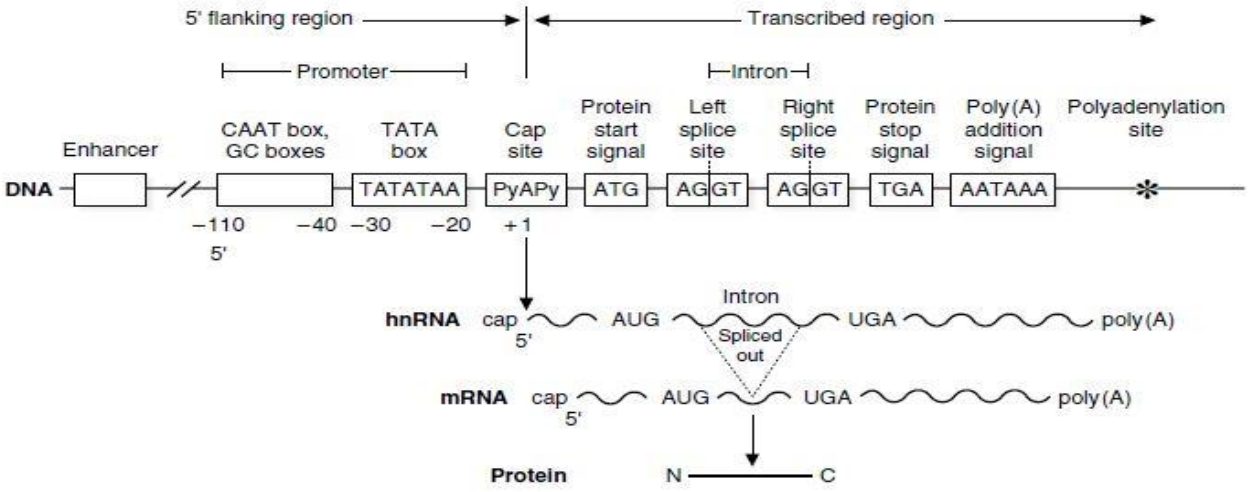

Figure 2**.** A schematic view of a eukaryotic gene, and steps required to produce a protein product. The gene consists of promoter and transcribed regions. The transcribed region contains introns, which do not contain coding sequence for proteins, and exons, which do carry the coding sequences for proteins. The first RNA form produced is heterogenous nuclear RNA (hn RNA), which contains both intronic and exonic sequences. The hnRNA is modified such that a cap is added at the 5' end (cap site), and a poly-A tail added to the 3_ end. The introns are removed (a process called splicing) to produce the mature mRNA, which leaves the nucleus to direct protein synthesis in the cytoplasm. Py is pyrimidine (C or T).

Stages of Transcription

1-Formation of transcription complex:

To begin transcribing a gene, the enzyme RNA polymerase needs to bind with specific sequences on a DNA. These sequences recognised by RNA polymerase are called as **promoter.** Basically, the promoter tells the polymerase where to "sit down" on the DNA and begin transcribing. In eukaryotes like humans, the main RNA polymerase in your cells does not attach directly to promoters like bacterial RNA polymerase. Instead, helper proteins called **basal** (**general**) **transcription factors** bind to the promoter first, helping the RNA polymerase in your cells get a foothold on the DNA (Figure 3).

a- **Promoter:** sequence of nucleotides (TATAAA) is found centred about 25 nucleotides upstream of the transcription start site. This core promoter consensus sequence is called the TATA, or Hogness, box.

b- **Regulatory elements and transcriptional activators**: Promoter proximal elements “CAT Box” = CAAT and “GC Box” GGGCGG

c- **General transcription factors (GTF):** These are the minimal requirements for recognition of the promoter, recruitment of RNA pol II to the promoter, and initiation of transcription at a basal level, includes:

- TFIID, a containing TATA-binding protein and TATA-associated factors, recognizes and binds the TATA box.
- TFIIF, brings the polymerase to the promoter.
- TFIIH, the helicase activity melts the DNA, and its kinase activity phosphorylates polymerase, allowing it to clear the promoter.
- TFIIE and TFIIB.

Complete complex (RNA polymerase + GTFs) is called a pre-initiation complex (PIC).

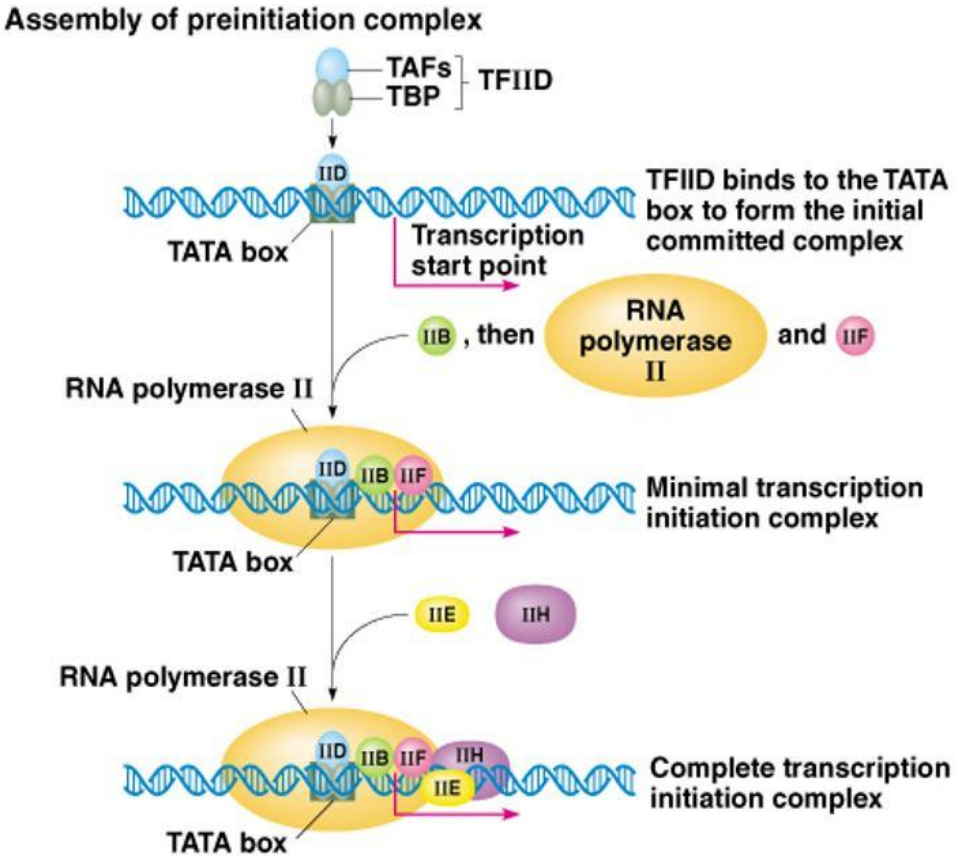

Figure 3: Transcription apparatus. The TATA-binding protein (TBP), a component of TFIID, binds to the TATA box. Transcription factors TFII A and B bind to TBP. RNA polymerase binds, then TFII E, F, and H bind. This complex can transcribe at a basal level. Some coactivator proteins are present as a component of TFIID, and these can bind to other regulatory DNA binding proteins (called specific transcription factors or transcriptional activators).

2-Initiation (Figure 4)

a- RNA polymerase binds to the transcription factor complex in the promoter region and to the DNA.

b- the helix unwinds within a region near the start-point of transcription, DNA strand separation occurs,

c- synthesis of the RNA transcript is initiated

3-Elongation ((Figure 4)

Once RNA polymerase is in position at the promoter, the next step of transcription-elongation-can begin. Basically, elongation is the stage when the RNA strand gets longer, due to the addition of new nucleotides.

During elongation, RNA polymerase "walks" along one strand of DNA, known as the **template strand**, in the 3' to 5' direction. For each nucleotide in the template, RNA polymerase adds a matching (complementary) RNA nucleotide to the 3' end of the RNA strand.

4-Termination (Figure 4)

In eukaryotes, extensive processing of the primary transcript occurs before the mature mRNA is formed and can migrate to the cytosol, where it is translated into a protein product.

RNA polymerase II synthesizes a large primary transcript from the template strand that is:

a- capped at the 5` end as it is transcribed.

b- The transcript also rapidly acquires a poly (A) tail at the 3` end.

c- Pre-mRNAs thus contain untranslated regions at both the 5` and 3’ ends (the leader and trailing sequences, respectively). These untranslated regions are retained in the mature mRNA. The coding region of the pre-mRNA, which begins with the start codon for protein synthesis and ends with the stop codon, contains both exons and introns.

d- Exons consist of the nucleotide codons that dictate the amino acid sequence of the eventual protein product. Between the exons, interspersing regions called introns contain nucleotide sequences that are removed by splicing reactions to form the mature RNA. The mature RNA thus contains a leader sequence (that includes the cap), a coding region comprising exons, and a tailing sequence that includes the poly(A) tail. This mature mRNA complexes with the poly(A) binding protein and other proteins. It travels through pores in the nuclear envelope into the cytoplasm. There it combines with ribosomes and directs the incorporation of amino acids into proteins.

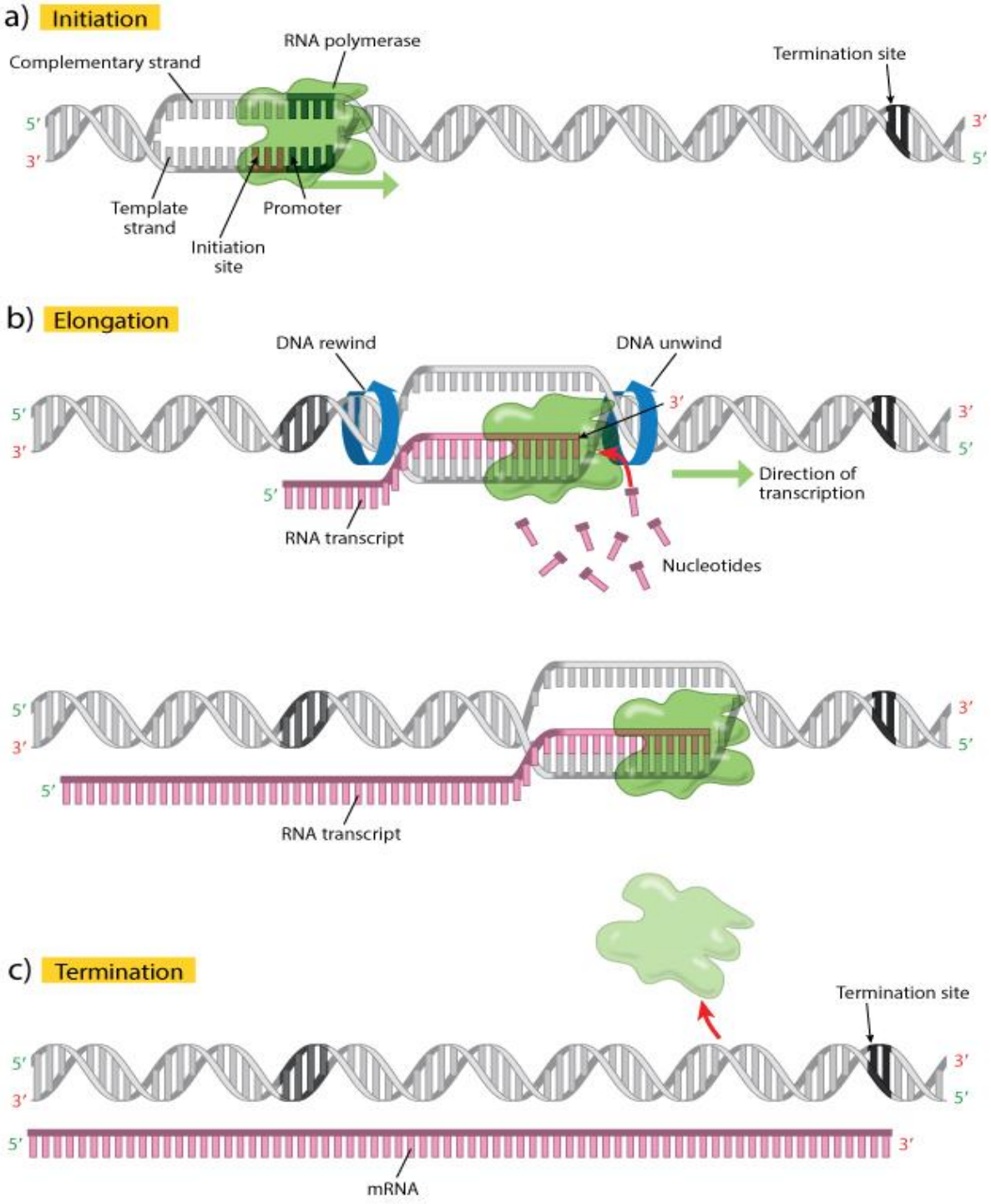

Figure 4: Transcription Steps

Transcription regulatory proteins [Activators]

Enhancers (or activators) are special DNA sequences that increase the rate of initiation of transcription by RNA pol II. Enhancers are typically on the same chromosome as the gene whose transcription they stimulate. They can:

1- Be located upstream (to the 5'-side) or downstream (to the 3'-side) of the transcription start site.
2- Be close to or thousands of base pairs away from the promoter
3- Occur on either strand of the DNA (Figure 5).

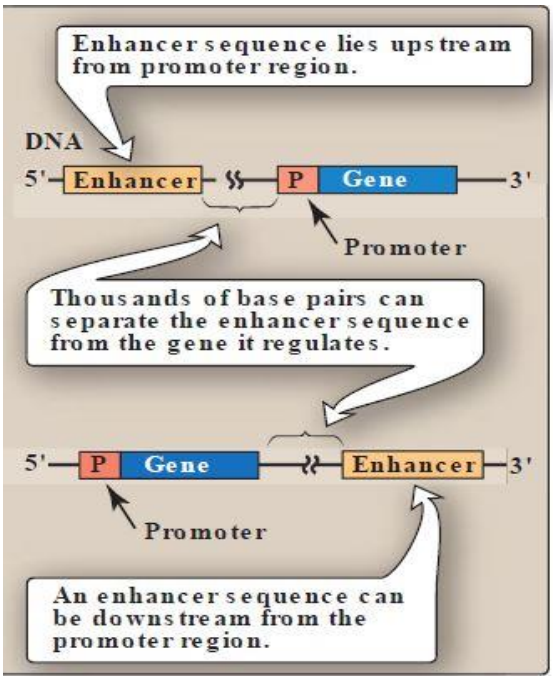

Figure 5: Some possible locations of enhancer sequences.

Inhibitors of transcription:

α-Amanitin, a potent toxin produced by the poisonous mushroom Amanita phalloides (some times called "the death cap"), forms a tight complex with RNA pol II, thereby inhibiting mRNA synthesis.

Messenger RNA (mRNA) is Modified at the 5` & 3` Ends:

1- 5' Capping:

This is the first of the processing reactions for pre-mRNA.

- The cap is a 7-methylguanosine attached to the 5'-terminal end of the mRNA through an unusual 5'→5' triphosphate linkage that is resistant to most nucleases.
- Creation of the cap requires removal of the phosphoryl group from the 5'-triphosphate of the pre-mRNA,
- followed by addition of guanosine monophosphate (GMP) (from GTP) by the nuclear enzyme guanylyltransferase.
- Methylation of this terminal guanine occurs in the cytosol and is catalyzed by guanine-7-methyltransferase.
- S-adenosylmethionine is the source of the methyl group. Additional methylation steps may occur.
- The addition of this 7-methylguanosine cap helps stabilize the mRNA and permits efficient initiation of translation (see the figure 6).

2- A Poly (A) Tail:

Poly(A) tails are added to the 3′ end of mRNA molecules in a posttranscriptional processing step. The mRNA is first cleaved about 20 nucleotides downstream from an AAUAA recognition sequence. Another enzyme, poly(A) polymerase, adds a poly(A) tail which is subsequently extended to as many as 200 A residues. ATP serves as the precursor for the sequential addition of the adenine nucleotides. They are added one at a time, with poly(A) polymerase catalyzing each addition.

The **poly(A) tail** appears to protect the 3′ end of mRNA from 3′ → 5′ exonuclease attack.

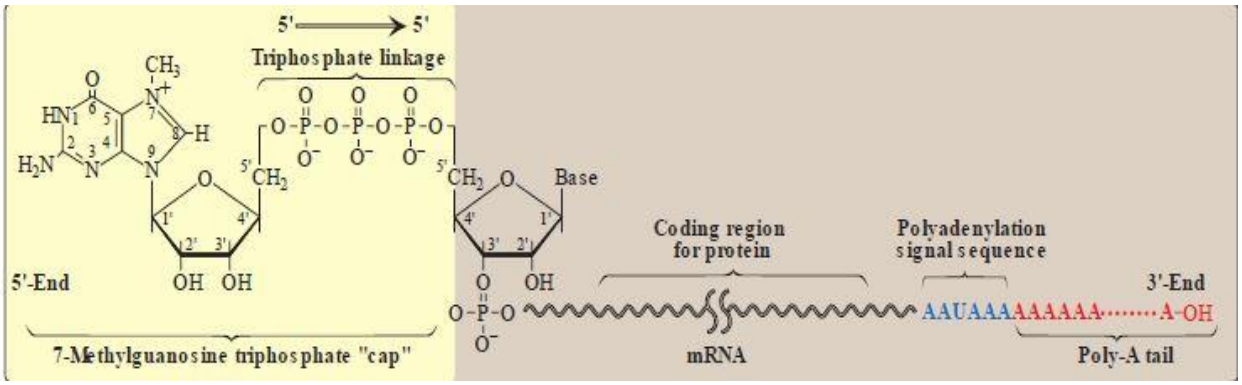

Figure 6: Posttranscriptional modification of messenger RNA (mRNA) showing the 7-methylguanosine cap and poly-A tail.

Splicing: *Reactions remove introns and connect the exons (Figure 7).*

Maturation of eukaryotic mRNA usually involves removal from the primary transcript of RNA sequences introns, or intervening sequences that do not code for protein. The remaining coding sequences, the exons, are joined together to form the mature mRNA.

a- The splice point at the 5′ end of an intron usually has the sequence GU; it is preceded by an invariant AG at the 3′ end of the exon adjacent to it. At the 3′ end of the intron, an invariant AG is frequently followed by GU at the 5′ end of the adjacent exon.

b- The molecular complex ribonucleoproteins (snRNPs) known as a **spliceosome** are involved in the cleavage and splicing process. A lariat structure is generated during the splicing reaction.

It is essential that all the intron-encoded RNA sequences are completely and precisely removed from a pre-mRNA before protein synthesis so that the exon-encoded RNA sequences are properly joined together to code for a functional polypeptide. Mutations at splice sites can lead to improper splicing and the production of an aberrant proteins. It is estimated that over 15% of all genetic diseases are a result of mutations that affect RNA splicing. For example, mutations that cause the incorrect splicing of β-globin mRNA are responsible for some cases of β-thalassemia, a disease in which the production of the β-globin protein is defective.

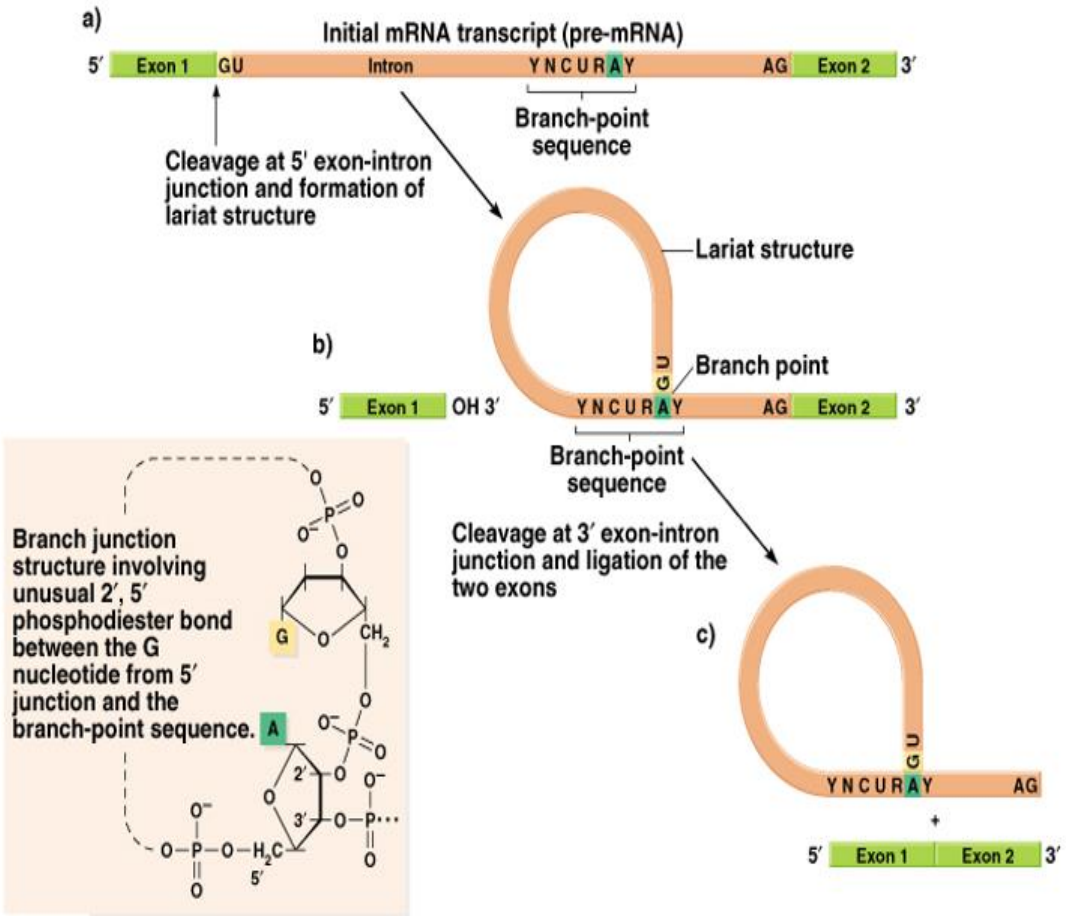

Figure 7: Splicing process.

Alternative splicing:

The pre-mRNA molecules from over 50% of human genes can be spliced in alternative ways in different tissues. This produces multiple variations of the mRNA and, therefore, of its protein product (Figure 8), and thus is a mechanism for producing a large, diverse set of proteins from a limited set of genes. The advantage of alternative splicing is that different types of mRNA transcripts can be generated, all derived from the same DNA sequence.

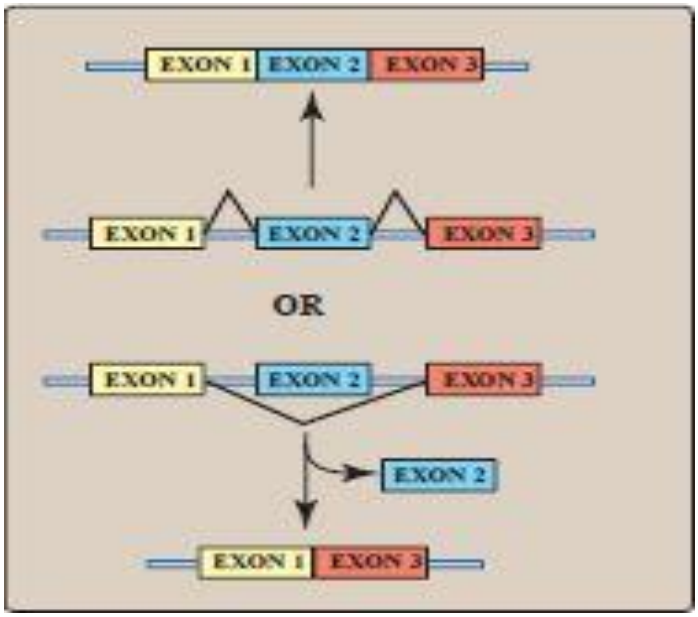

Figure 8: Alternative splicing patterns in eukaryotic messenger RNA.