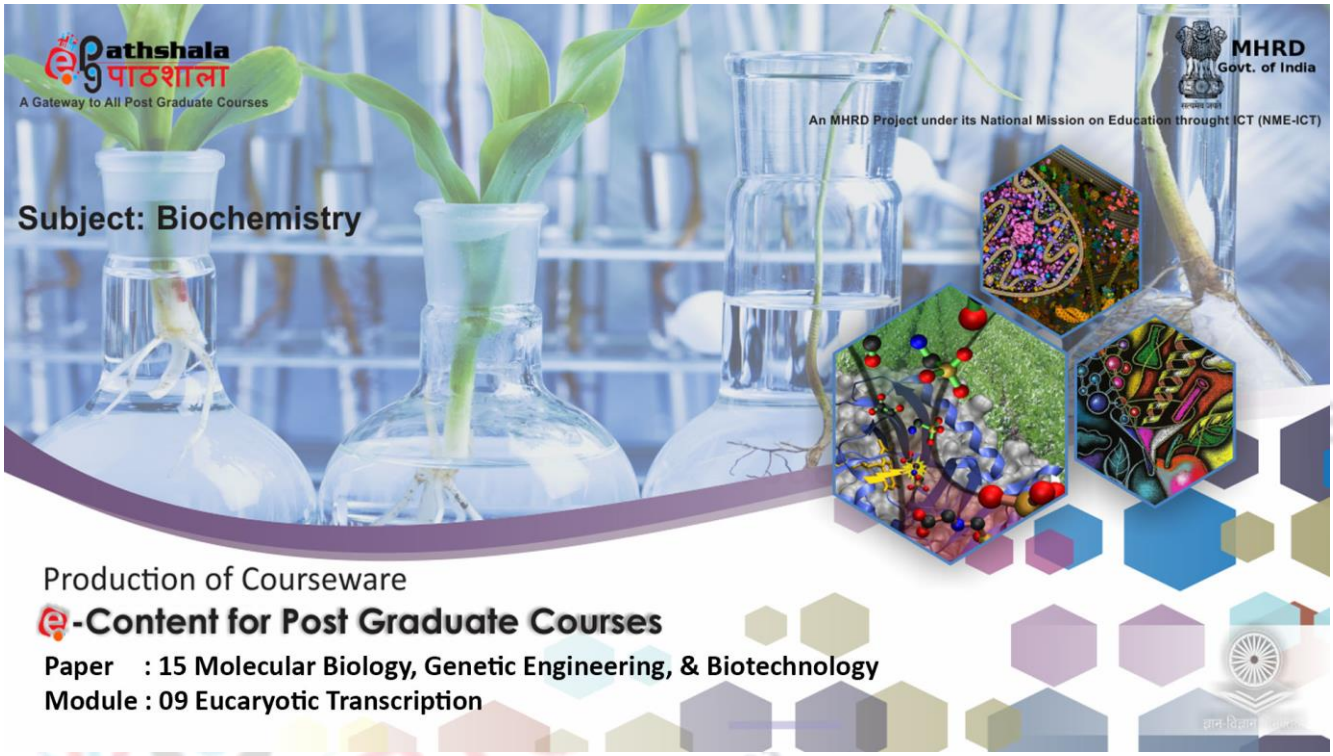


athshala पाठशाला

MHRD Govt. of India

Development Team

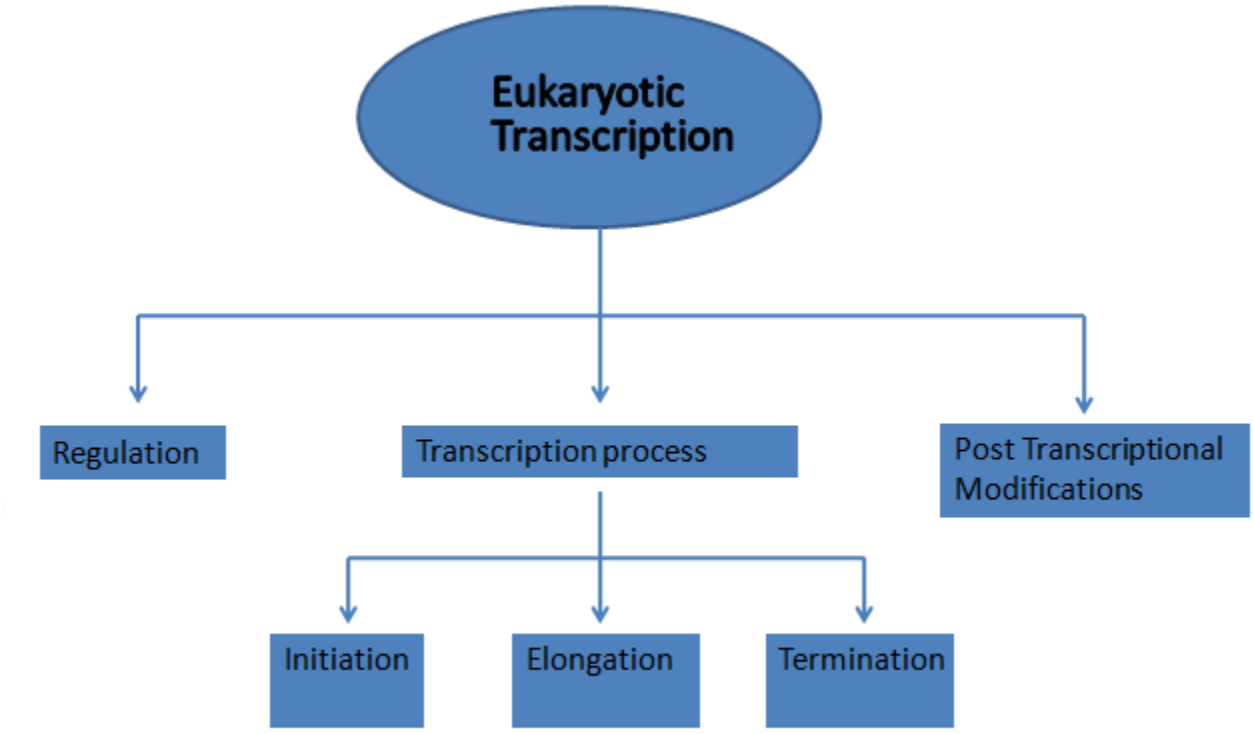
Principal Investigator	Dr. Sunil Kumar Khare, Professor, Department of Chemistry, IIT-Delhi
Content Writer:	Dr. Ramesh Kothari, Professor - Microbiology, UGC-CAS Department of Biosciences, Saurashtra University, Rajkot
Content Reviewer:	Dr. Lata, Principal Scientist, Division of microbiology, IARI, New Delhi
Paper Coordinator:	Dr. Tanmay Dutta, Department of Chemistry, Indian Institute of Technology, Delhi

Description of Module	
Subject Name	
Paper Name	
Module Name/Title	Eukaryotic Transcription

1. Objectives

1. Understand and have an overview of eucaryotic transcriptional regulation.
2. Explain about types of RNA polymerases and their role.
3. Understand the mechanism of transcription by RNA polymerase II.
4. Understanding the mechanism and types of post transcriptional mechanisms.

2. Concept Map



3.1 Introduction

Gene expression involves production of an RNA molecule making use of DNA as template. The process of transfer of information from DNA to RNA is known as transcription. RNA differs from DNA in containing base uracil instead of thymine and presence of OH group instead of H group at the 2' end of the pentose sugar. RNA is however single stranded structure and can fold back on themselves generating structural diversity. RNA is a macromolecular structures with varied functions including storage and transmission of information and catalysis. Transcription is a process of transfer of information from a double stranded DNA molecule into a single stranded RNA molecule by making use of enzyme system for the conversion. In general there are three major kinds of RNA : Messenger RNA (mRNA) which act as a template or source of information for protein synthesis process, transfer RNAs (tRNA) which act as carriers of amino acid during protein synthesis. Transfer RNAs read the information in the mRNA and transfer the appropriate aminoacid during protein synthesis. Ribosomal RNA (rRNA) are involved in formation of ribosomes which are the protein synthesizing machinery. Other additional specialized RNAs have regulatory and catalytic functions. In DNA replication process the entire chromosome or genome of the organism is replicated and two copies of DNA are synthesized. However transcription process is selective expression of particular genes or group of genes depending on the requirement of a particular cell. This expression is highly regulated process involving activity of several proteins, enzymes and regulatory sequences. The specific regulatory sequences mark the fragment of DNA to be transcribed as well as which strand to be used a template DNA strand. In the following section we will discuss about Eucaryotic transcription and its steps in details.

3.2 Eucaryotic RNA polymerase

DNA dependent RNA polymerase requires DNA as a template, four ribonucleosides 5' triphosphate ATP, GTP, CTP and UTP as building blocks of RNA and Mg2+ as a cofactor. The E.coli RNA polymerase holoenzyme consists of RNA polymerase core enzyme ($\alpha 2\beta\beta'$) + sigma factor which carry out the 4 steps of prokaryotic transcription process -template binding, initiation, elongation and termination. Eukaryotes on the contrary have three different RNA polymerases specialized for transcription of different types of RNA molecules. Eukaryotic genomes are larger in the range of approximately 10^9 base pairs. Large genome means large number of genes and hence requires more specificity for amplification. Additionally eukaryotic cells have diversity of functions, organelles and specialized cell types etc which also require specificity of gene expressions. Hence eukaryotes have different RNA polymerases specialized for transcription of different types of RNA molecules.

3.2.1 Types of RNA polymerases

Three different types of RNA polymerases are present in eukaryotes:

RNA Polymerase I : RNA polymerase I functions for synthesis of pre ribosomal RNA containing 18S, 5.8S and 28S rRNA.

RNA polymerase II is involved in synthesis of messenger RNA (mRNA). It is capable of recognizing several promoters that differ in their sequence. Many Polymerase II recognized promoters have few common sequences like the TATA box (consensus sequence TATAAA) located are – 30 position and Inr sequence (initiator) at RNA start site at +1 base pairs.

RNA polymerase III is responsible for synthesizing tRNA's, 5SrRNA and other small RNAs.
Some of the promoter sequences recognized by RNA polymerase III are located within the genes while others are located upstream of the RNA start site.

In general RNA polymerases are a large multisubunit complex consisting of 10-17 different subunits. These RNA polymerases cannot bind specifically to their respective promoters on their own but require assistance from several transcriptional factors. In the following section we will discuss in details about the structure of RNA polymerase II which is the main enzyme involved in synthesis of mRNA molecule and hence gene expression.

3.2.2 RNA polymerase II

RNA polymerase II has been studied extensively and is important with reference to gene expression. RNA polymerase II is a multimeric protein made up of 12 different subunits (RBP1-12). RBP1 is the largest subunit and is homologous to β' subunit of bacterial RNA polymerase. The RBP2 subunit is homologous to bacterial RNA polymerase β subunit, while RBP3 and RBP 11 are homologous to bacterial RNA polymerase α subunits.

The largest subunit RBP1 of PolII has a long C terminal tail made up of multiple repeats of 7 amino acid sequence –YSPTSPS-.For example it is repeated 27 times in yeast enzyme and about 52 times in humans. The CTD can become highly phosphorylated at its Ser and Thr residues. During different stages of transcription the CTD of RBP1 is cycled between phosphorylated and dephosphorylated forms. CTD is also involved in recruitment of proteins required for 5' Capping and 3' Polyadenylation reaction of mRNA. The RNA polymerase II requires interaction with several transcriptional factors in order to initiate eukaryotic transcription process. The transcription process by RNA polymerase II can be in general divided into several phases , - assembly, initiation, elongation and termination which are explained in the later section.

3.3.1 Eukaryotic promoter elements

Eukaryotic promoters extend from the transcriptional start site to approximately 200 bps upstream (Figure 1A). It contains several short sequences of approximate 10 bps in length. **The core promoter** contain a TATA box (sequence **TATAAA**), which is bound by TATA binding protein that functions in the formation of the RNA polymerase transcriptional complex. The TATA box is present upstream of the transcriptional start site (-25 bps upstream) (often within 50 bases) and an **INR sequence** at the RNA polymerase start site. Several **proximal promoter elements** approximately 70-200 bps towards the 5' end of transcription start site are also present. These includes CAAT box and GC box which are binding sites for CAAT- binding protein (CBP) and transcriptional factor SP1 respectively. These different sequences can be mixed and matched to give a functional promoter.

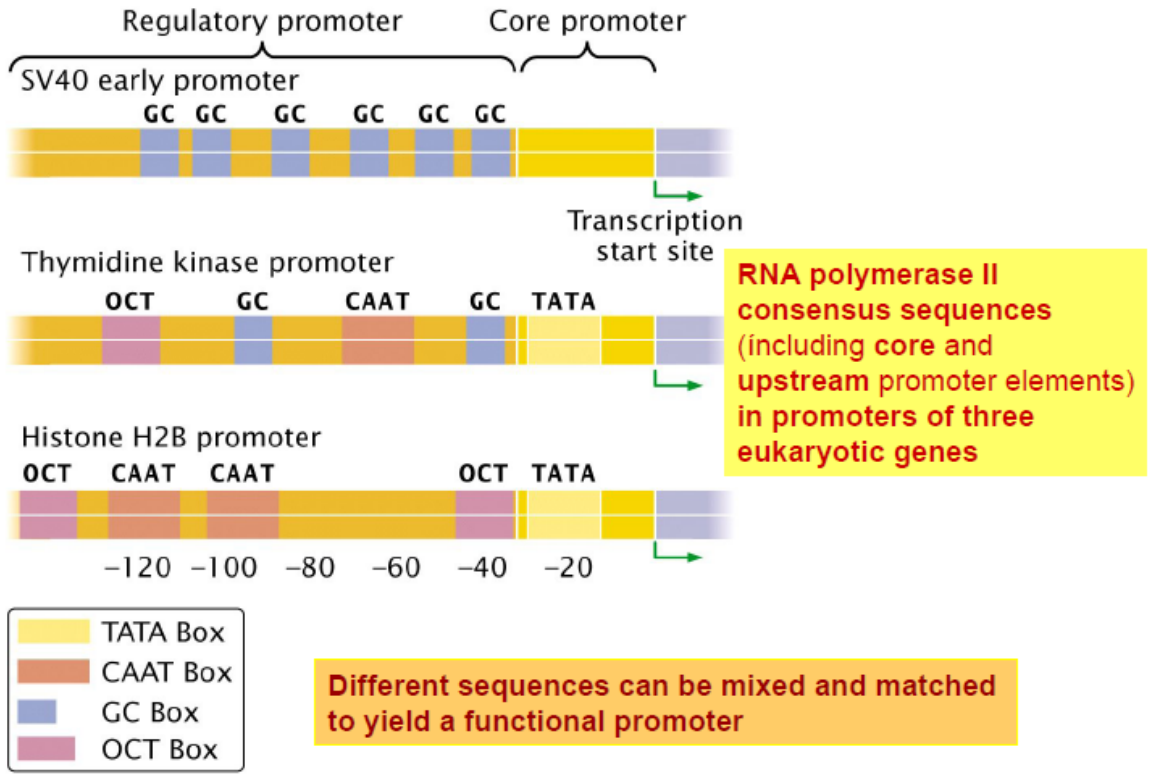


Figure 1A: Eucaryotic promoter element recognized by RNA Pol II

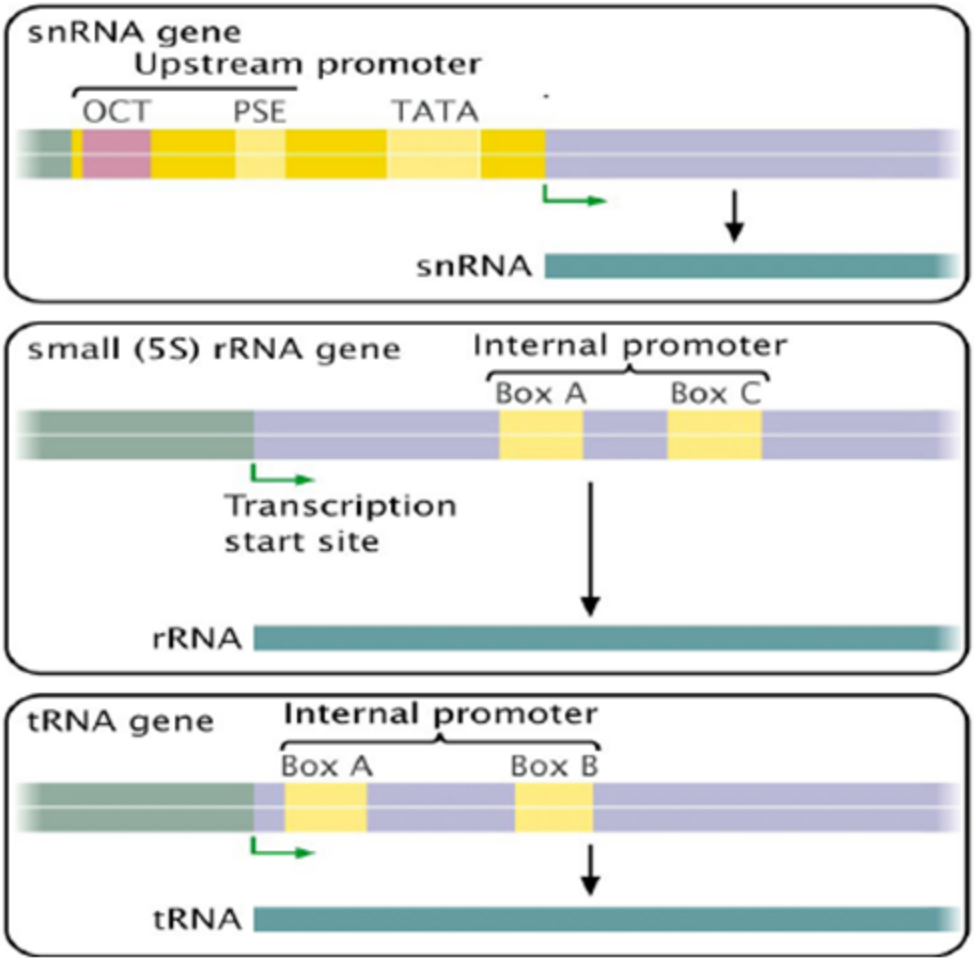


Figure 1B: Eucaryotic promoter element recognized by RNA Pol III

Promoter elements recognize by RNA polymerase I are not well conserved in sequence from one species to another. However they have a AT rich initiator conserved sequence surrounding the transcriptional start site. RNA polymerase III sequence promoters are of two types (Fig 1B). The classical polymerase III genes are present completely with in the gene eg tRNA, 5srRNA. The nonclassical Polymerase III genes like snRNA resemble the polymerase II gene promoters consisting of upstream promoter elements and TATA box upstream of the transcription start site.

3.3.2 Enhancers and silencers

Enhancers are sequences approximately 500 bps in length and contain binding sites for several different transcriptional factors. They are about 700- 1000 bps away from the transcription start site (Fig 2) can be present either downstream, upstream or within the gene sequence to be transcribed (Fig 3). DNA may be coiled, bend or rearranged such that the transcriptional factors bound at the promoter and enhancer elements interact to produce large protein complex. Enhancer increase gene promoter activity in all tissues or brings about regulated gene expression in a tissue specific or developmental stage specific manner. Similar elements that repress gene activity are called silencers.

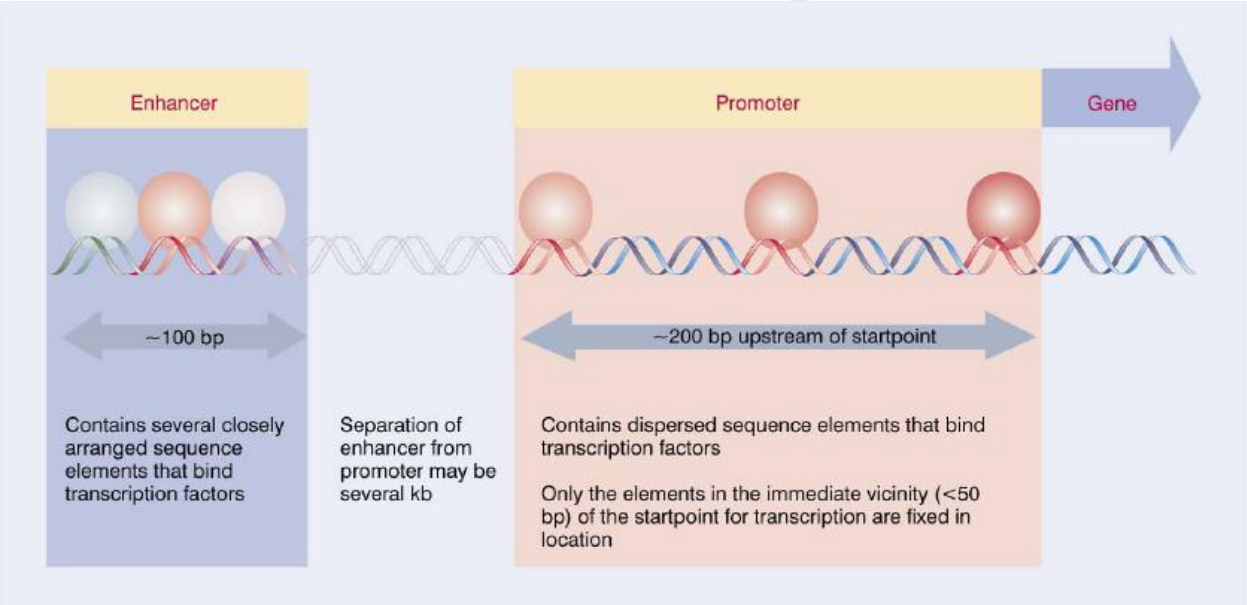


Figure 2- A Eukaryotic transcription unit

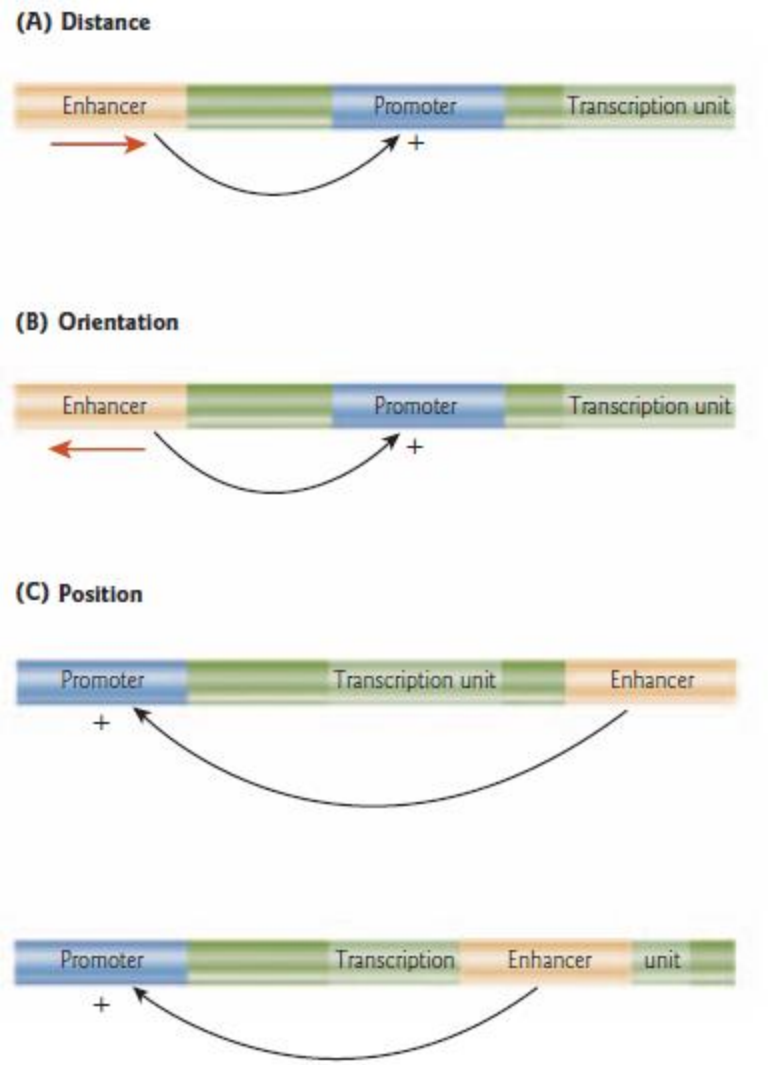


Fig3: Enhancer elements can activate a promoter at a distance (A) in either orientation (B) and upstream downstream or within a transcriptional unit (C).

3.4 Steps of Eucaryotic Transcription

The transcription process by RNA polymerase II can be in general divided into several phases , - assembly, initiation, elongation and termination which are explained in the following section.

3.4.1 Assembly of RNA Polymerase and Transcription Factors at a Promoter

The assembly of RNA polymerase and transcriptional factors at promoter begins by formation of a closed complex which begins with the binding of TATA- binding protein (TBP) to the TATA box and 8-10 TBP associated factors. TBP is universal transcriptional factor required by all three classes of RNA polymerases.

The TBP is further bound by transcriptional factor TFIIB which also binds with DNA on either side of TBP. It acts as in intermediate in the recruitment of RNA polymerase II and influences selection of transcription start site.

TFIIA at this stage stabilizes TFIIB-TBP complex on the DNA.

The TFIIB –TBP complex is then associated with another complex comprising of TFIIF and Pol II. TFIIF here recruits RNA polymerase II and targets it to its promoter via interaction with TFIIB and reducing nonspecific binding by RNA Polymerase II.

Finally TFIIE- and TFIIH bind leading to formation of an open complex. TFIIE here acts in recruitment of TFIIH and modulates its activities. TFIIH has a helicase activity which is responsible for transition from closed to open promoter complex

3.4.2 RNA Strand Initiation and Promoter Clearance

Additional function of TFIIH is required during the initiation phase. It has kinase acitivity in one of its subunits that brings about phosphorylation of Pol II at many places in its CTD. Additional protein kinase, CDK9 which is a part of transcription elongation complex pTEFb (positive transcription elongation factor b) also brings about phosphorylation of the CTD. Phosphorylation events at the CTD of RNA polymerase II largest subunit leads to a conformational change in the overall structure of the complex, initiating transcription. Phosphorylation of the CTD is important not only during the subsequent elongation phase but it also affects the interactions between the transcription complex and other enzymes involved in post transcriptional processing of the transcript (described in later section). While the initial 60-70 nucleotides of RNA are being synthesized the TFIIE followed by TFIIH is released, and Pol II continues to enter into the elongation step of transcription.

3.4.3 Elongation, Termination, and Release

TFIIF is associated with Pol II throughout elongation phase. During this step, the RNA Pol II C terminal domain is maintained in the phosphorylated stage by coordinated action of several proteins called as elongation factors. Some of these elongation factors and their function are described in table 1. The elongation factors prevent pausing of transcription process and are also involved in interaction with protein complexes that mediate post transcriptional processing of mRNA. After synthesis of the RNA molecule the process is terminated and Pol II is dephosphorylated and recycled to initiate another transcription cycle (Fig. 4).

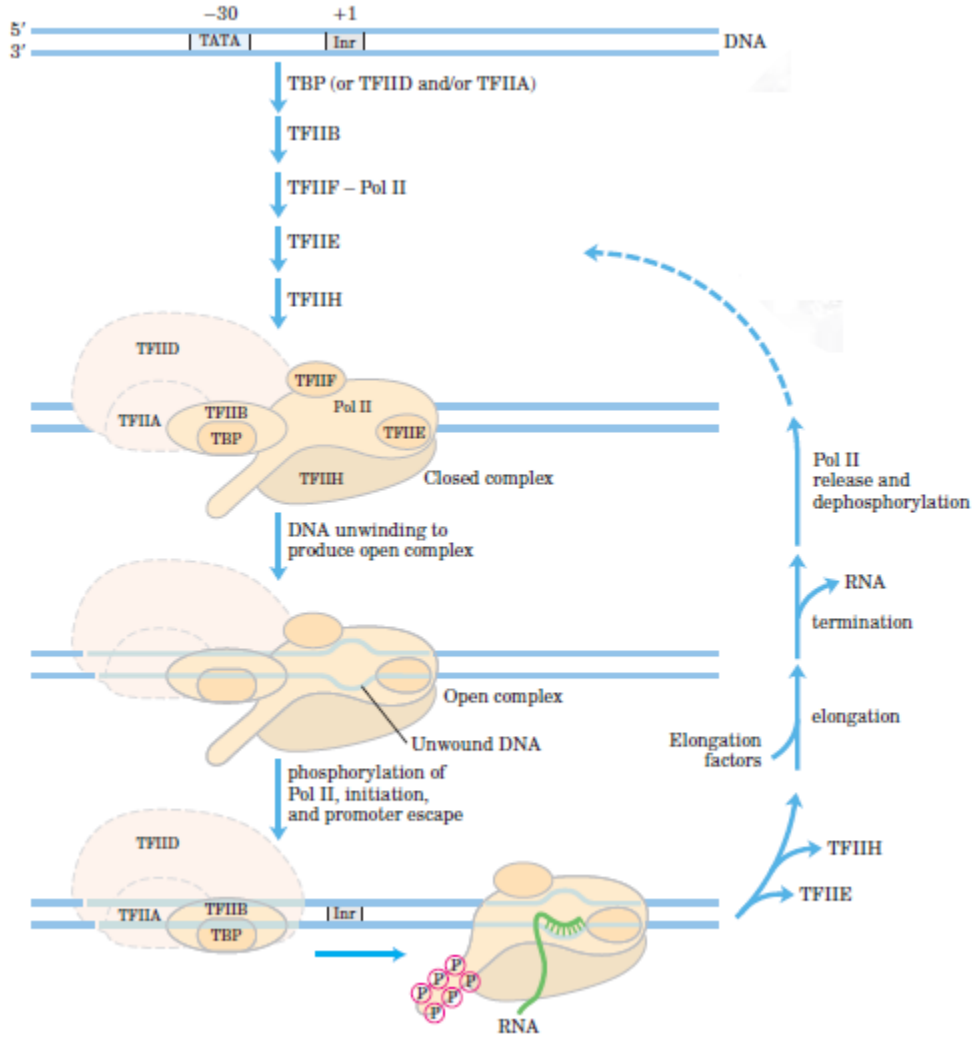


Figure 4- Transcription at RNA Pol II promoters

Elongation factor Protein	Function and activity
P-TEFb	-Positive transcription elongation factor b -Cyclin-dependent kinase -Phosphorylates CTD of large subunit, Pol II
Eukaryotic TFIIS	-may overcome pausing by the polymerase -induce cleavage of the new transcript, followed by release of the 3' terminal RNA fragment.
ELL	-increase elongation rate of RNA Pol II

Table 1-Proteins implicated in Transcription elongation:

3.5 Transcription termination:

Transcription termination process in the 3 different types of RNA polymerases differ in the details of their mechanism which are described in this section.

3.5.1 Transcription Termination by RNA polymerase I

Transcription termination by RNA polymerase I requires binding sites for Reb1p which causes pausing of transcription by RNA polymerase I and the process terminates (Figure 5).

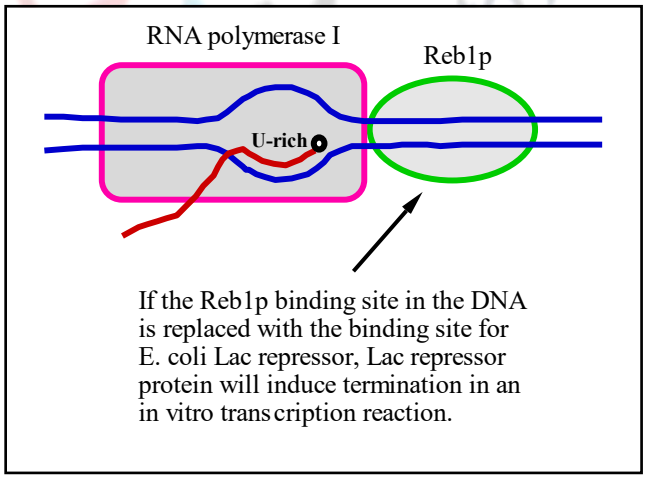


Figure 5 - Transcription termination by RNA polymerase I

3.5.2 Transcription termination by RNA polymerase III

Transcription termination by RNA Pol III does not require any protein factors. A stretch of thymine residues followed by a self complementary sequence transcription leads to formation of a unstable complex that disrupts the RNA -DNA hybrid and transcription terminates.

3.5.3 Transcription termination by RNA polymerase II

There is no discrete or clear terminator of transcription by RNA Pol II. The 3' end of mRNA is made by cleavage and polyadenylation explained later. However depending on the RNA 3' end processing signals and termination factors present at the end of the gene there are atleast 2 best known pathways for transcription termination which include the the poly(A)- dependent pathway and the Sen1-dependent pathway.

3.5.3.1 Poly A dependent termination: Transcription termination by DNA Pol II is coupled with mRNA maturation at the 3' end of mRNA molecule including cleavage and polyadenylation.

The RNA Pol II Rbp1 subunit : The extended carboxyl- terminal domain (CTD) of Rbp1 subunit of RNA Pol II has important features for recruitment of Pol II termination factors. The Pol II CTD consists of tandem heptad repeats (26 in yeast, 52 in human), which are made up primarily of the amino acid consensus sequence Tyr1 -Ser2 -Pro3 -Thr4 -Ser5 -Pro6 -Ser7 .These sequences are modified during post translational changes by phosphorylation. Changes in the phosphorylation patterns of this CTD alter the affinity of the CTD binding protein thus regulating Pol II function.

The Pol II CTD and RNA is bound by several processing factors that could act as shearing force to separate the DNA –RNA hybrid. The binding of the processing factors also causes pausing of the Pol II. Recruitment of cleavage and polyadenylation factors coincides with pausing of Pol II.

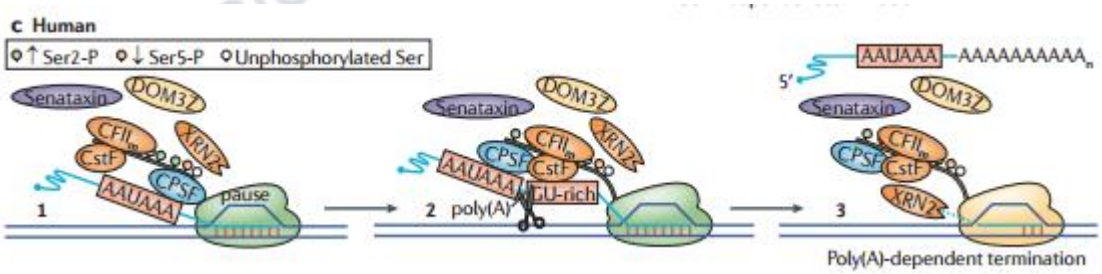


Figure 6: Poly(A)-dependent termination

During poly(A) dependent termination in humans (Fig 6), cleavage and polyadenylation specificity factors (CPSF) bind to the human Pol II and recognize the AAUAAA signal sequence that emerges in the nascent transcript (step 1). CPSF binding to this site induces pausing of Pol II. As the GU rich binding site is synthesized it is bound by cleavage stimulatory factor (CstF) that dislodges CPSF (step 2). Following cleavage at the poly(A) site, 5′–3′ exoribonuclease 2 (XRN2) degrades the downstream RNA product, leading to displacement of Pol II (Step 3).

3.5.3.2 Sen1-dependent termination.

It is an alternative pathway for most non- coding RNAs. The 3' ends of yeast snRNAs and snoRNAs are generated by endoribonucleolytic trimming by nuclear exosome TRAMP complex and do not possess a poly A tail in their mature form.

A distinct set of core factors is required for recognition and transduction of the transcription termination signal, including the RNA-binding proteins Nrd1, nuclear polyadenylated RNA-binding protein3 (Nab3) and the putative RNA and DNA helicase Sen1. In this pathway Sen1 is proposed to terminate Pol II by unwinding the RNA–DNA hybrid in the active site (Fig6).

Figure 6 - Sen1-dependent transcription termination.

3.6 Post Transcriptional Modifications:

Three types of RNA molecules which include mRNA, tRNA and rRNA are synthesized during transcription process. These molecules undergo post transciptional changes before attaining its final functional form. These post transcriptional modifications are described in the following section.

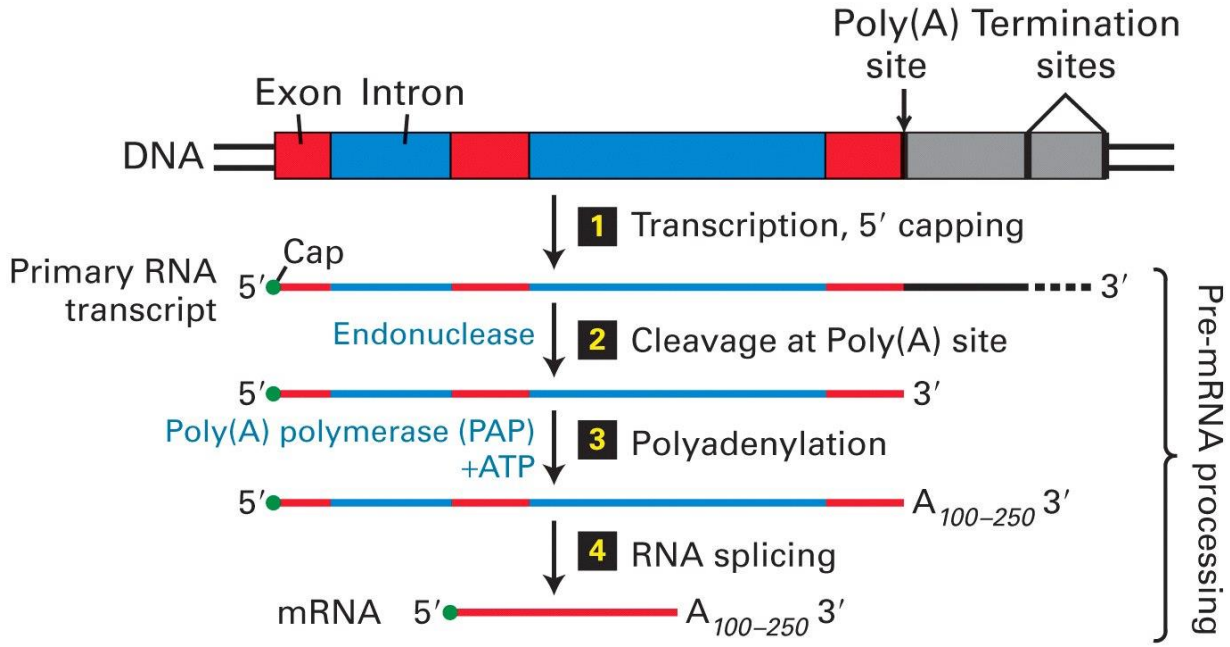


Fig 7. Mechanism of Polyadenylation

3.6.1 Polyadenylation:

Polyadenylation is addition of 80- 250 A residues at the 3' end of the mRNA molecule. It serves as a binding site for specific proteins that help protect mRNA from enzymatic degradation.

Mechanism of polyadenylation:

The mRNA is synthesized beyond the polyadenylation site, and is later cleaved by an endonuclease at the polyadenylation site which is marked by sequence (5')AAUAAA(3'), present 10 to 30 nucleotides upstream of cleavage site. Endonuclease activity generates a free 3'OH group at the end

of mRNA molecule to which a string of A's is added in a reaction catalyzed by polyadenylate polymerase generating a tail of 80-250 residues long (Fig 7).

3.6.2 5' CAPPING

At the 5' end of eukaryotic mRNA molecules is present a 7-methylguanosine residue linked to the 5' end of the mRNA by an 5'-5'-triphosphate linkage representing a cap (Fig8). It protects the mRNA from ribonuclease degradation from the 5' end of mRNA and also functions to bind the CAP binding complex that transports the mRNA across the nuclear membrane to the cytosol. The CAP binding complex further contributes in binding of mRNA to ribosomes to initiate translation.

Mechanism of 5' Capping is as detailed in Fig9. The 5'cap is formed by condensation of a molecule of GTP with the 5' end of the mRNA molecule. The Guanine is then methylated at the N-7 position by using S-adenosylmethionine as a methyl group donor. The capping reactions begins as soon as the first 20-30 nucleotides of the mRNA transcript are synthesized by the capping enzymes, Once the cap is synthesized the mRNA molecule is released and further bound by cap-binding complex.

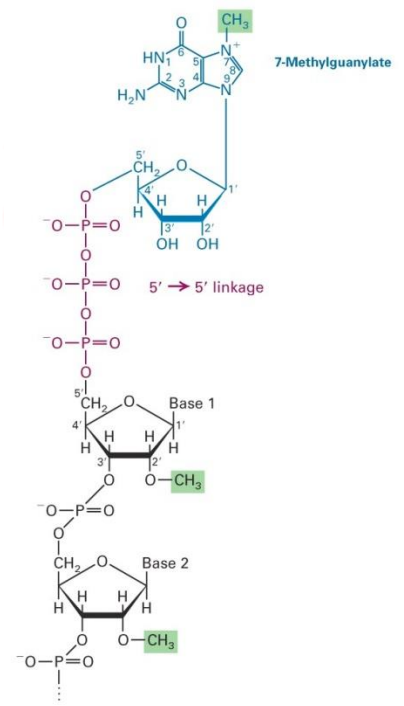


Fig 8.Structure of a CAP

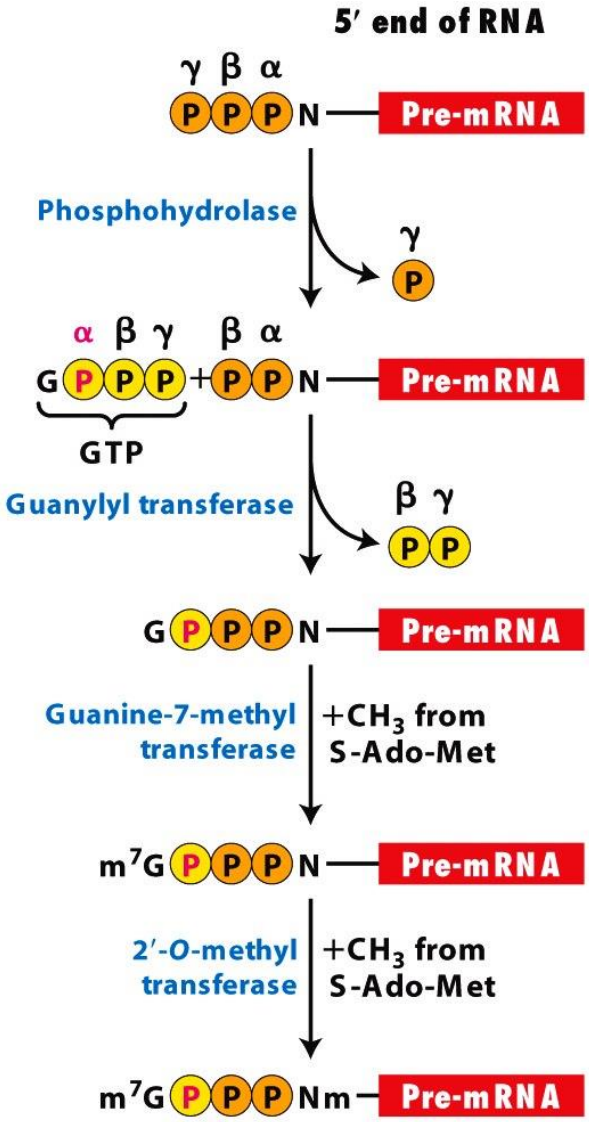


Fig 9. Mechanism of 5'Capping

3.6.3 Post transcriptional modification in Ribosomal RNA

The ribosomal RNA molecules are synthesized as longer precursors known as pre-ribosomal RNAs which in case of eucaryotes is 45 S transcript made by eukaryotic RNA PolI. The 45S precursor is methylated at several nucleotides mostly on 2-OH groups of ribose sugars. The precurssor rRNA is the cleaved by a series of enzymatic reactions in the nucleolus to the mature 18S, 28S and 5.8S rRNAs of eukaryotic ribosomes. (Fig 10). The 5S rRNA of eukaryotes is made as a separate transcript by RNA

Pol III. These rRNA molecules then associate with protein to form functional 80S eukaryotic ribosome (Fig11.). The mature Eukaryotic ribosome has sedimentation coefficient of 80S. It consist of 2 subunits the larger 60S and the smaller 40S. The larger 60S subunit is made up of 28S, 5.8S and 5S rRNA and 49 proteins, while the 40S ribosomal subunit is made up of 18SrRNA and 33 proteins.

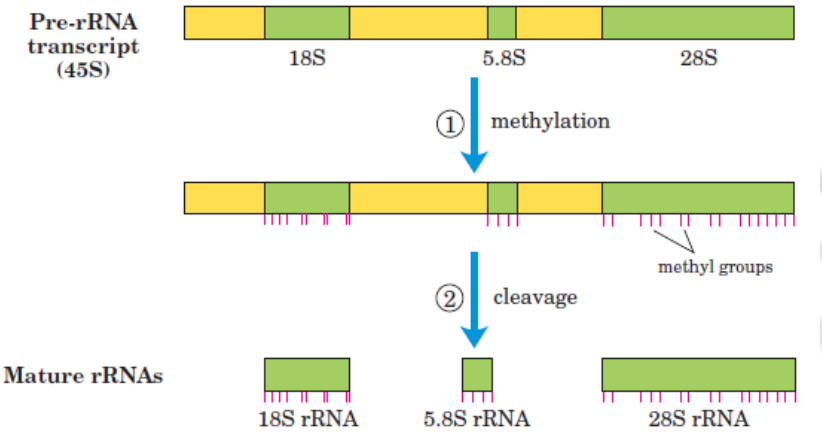


Fig 10. Pre-rRNA transcript processing

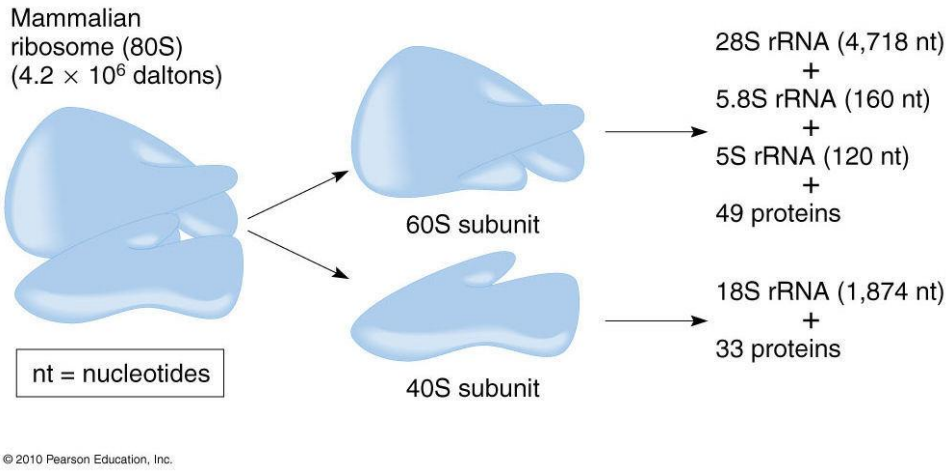


Fig 11.Eucayotic ribosome.

3.6.4 Transfer RNA (tRNA)processing

Most Eucaryotic cells have 40-50 different tRNA molecules and several copies of many of the tRNAgenes. Transfer RNA molecules are as well synthesized as longer precursors which are enzymatically processed to the final mature tRNA molecule. Several nucleotides are removed from the 5’ and 3’ ends of the tRNA precursor by endonuclease RNase P and exonuclease RNase D respectively (Fig12). Few eukaryotic tRNA precursors also contain intron sequences which are not present in the final structure. Further post transcriptional processing may include addition of 5’CAA3’ to the 3’ end of the RNase D cleaved tRNA molecule in a reaction catalyzed by tRNA nucleotidyltransferase. The enzyme binds to three ribonucleoside triphosphate precursors in separate active sites and forms phosphodiester bonds leading to synthesis of 5’ CCA 3’ sequence which is attached to the 3’ end of the mature tRNA molecule.

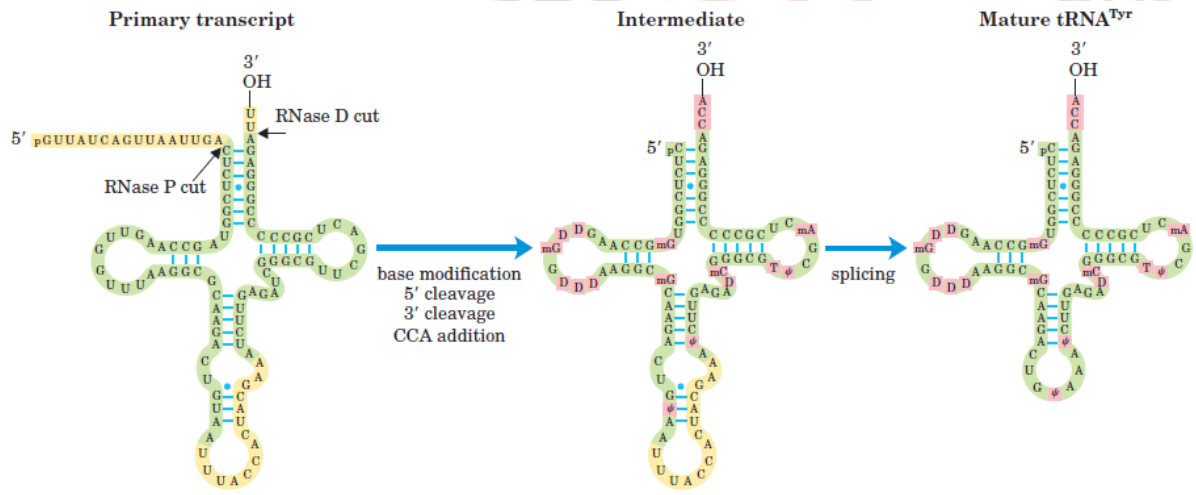


Figure 12 Transfer RNA processing

3.6.5 Splicing:

The eukaryotic precursor mRNA molecule contains both introns and exon sequences. Of these intron sequences are removed and the exons are spliced together during the post transcriptional modification. Thus the introns are not present in the mature rRNA molecules. There are in general 4 different types of introns based on the splicing mechanism which are as follows.

3.6.5.1 Group I introns

Group I introns are found in mRNAs, rRNAs and tRNAs of some nuclear, mitochondrial and chloroplast genes. The splicing reaction involves two transesterification steps. Mechanisms in both groups involve two transesterification (Fig 13) steps. A 2'or 3' hydroxyl group of the ribose sugar of the RNA acts as a nucleophile and attacks on a phosphorous of the phosphodiester bond at the exon intron junction making a new phosphodiester bond. In group I splicing reactions the 3' hydroxyl group of a guanosine nucleotide or nucleoside cofactor makes a nucleophilic attack on the phosphate of the phosphodiester bond at the exon intron junction forming a new phosphodiester bond with the 5' end of the intron. The 3' hydroxyl group of the displaced exon now similarly attacks the 3' end of the intron resulting in removal of intron and splicing of the exons together (Fig14).

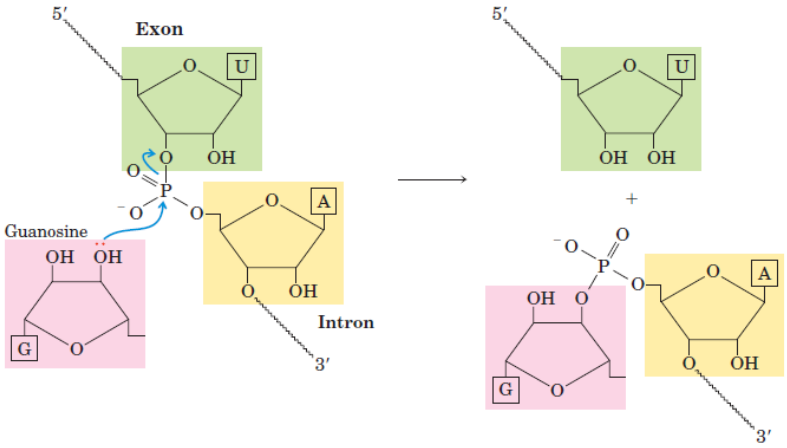


Fig 13. Transesterification reaction

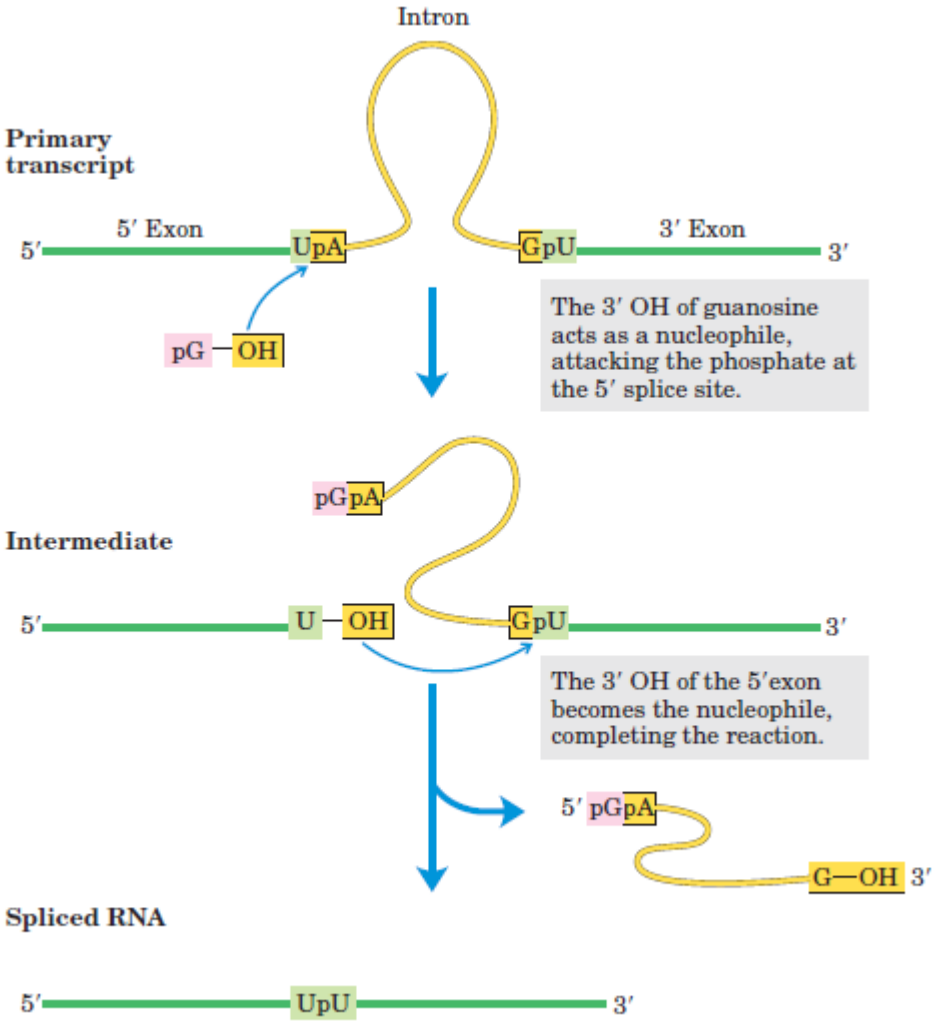


Figure 14-Group I splicing mechanism : Guanosine acts as a nucleophile in the first step. The excised intron sequence is eventually removed.

3.6.5.2 Group II introns

Groups II introns are generally found in DNA of algal fungal and plant organelles like mitochondrial or chloroplast. The groups II introns splicing mechanism is similar to the group I reactions however the 2' hydroxyl group of the A residue which resides with in the intron sequence acts as a nucleophile during the first step (Fig 15). In both Group I and II splicing reactions no external ATP energy is invested for

splicing. Formation of new phosphodiester bond utilized the energy released during cleavage of previous phosphodiester bond and hence the energy balance is maintained.

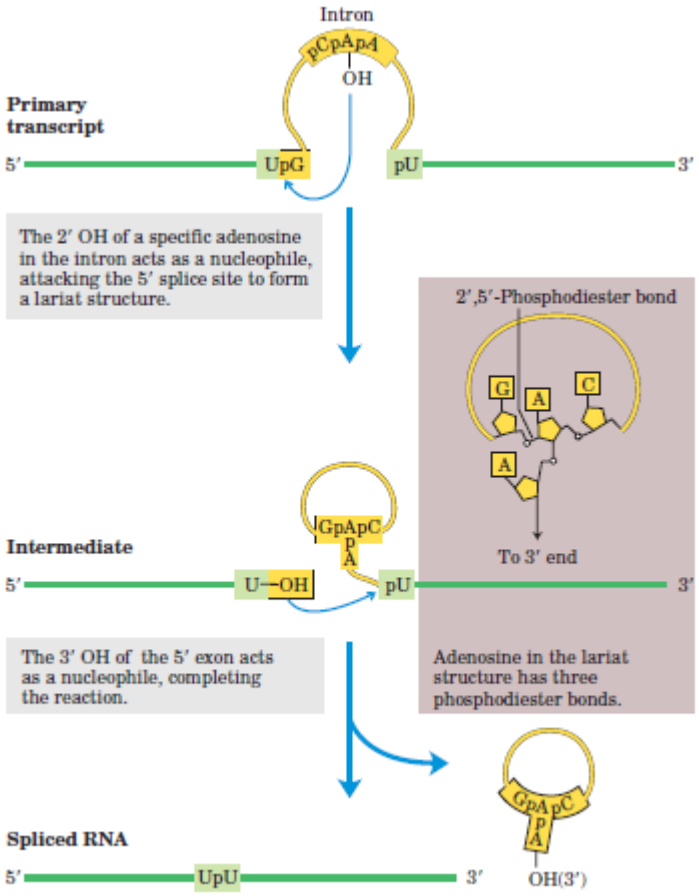


Figure 15 : Group II splicing mechanism

3.6.5.3 Spliceosomal introns:

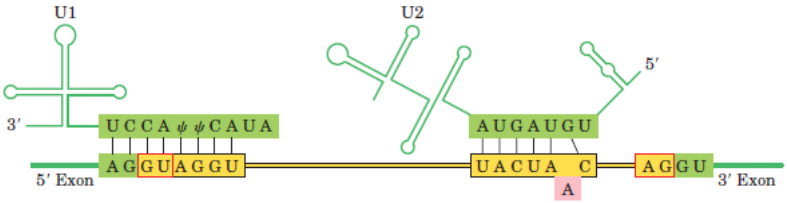
This class of introns is found in the eucaryotic nuclear premRNA. They are named so because their removal involves a large protein complex called spliceosome. It is made up of RNA and protein complexes together known as small nuclear ribonucleoproteins (snRNPs). Each snRNP consists of a small nuclear RNAs about 100- 200 nucleotides in length complexed with protein molecules. 5 different snRNAs U1, U2, U4, U5 and U6 are involved in the splicing reaction. Spliceosomal introns contains a dinucleotide sequence GU at the 5' splice site and AG at the 3' splice site. The U1snRNA

has sequence complementarities to the 5' splice site of the intron while U2snRNA has sequence complementarity toward the internal sequence of the intron containing an A residue (Fig 16 A). Binding of the U1 snRNPs actually help define the 5' splice site. The U1 snRNP binds at the 5' splice site following which the remaining snRNPs U2, U4, U5 and U6 are added leading to formation of a functional spliceosome complex consisting of 5 snRNA and 50 proteins. Formation of active spliceosome complex requires ATP energy for its assembly. Binding of the U2snRNP brings about activation of internal A residue whose 2' OH group will make a nucleophilic attack at the 5' splice site leading to formation of lariat like intermediate structure and formation of a free hydroxyl group at the end of the 5 ' splice site. This hydroxyl group now attacks the 3' splice site completing the reaction (Fig 16B).

3.6.5.4.

The fourth class of introns, found in certain tRNA molecules. It requires activity of enzyme endonuclease and energy from ATP hydrolysis. The endonuclease cleaves the phosphodiester bonds at either ends of the intron, and the 2 exons are spliced together by formation of new phosphodiester bond.

A



B

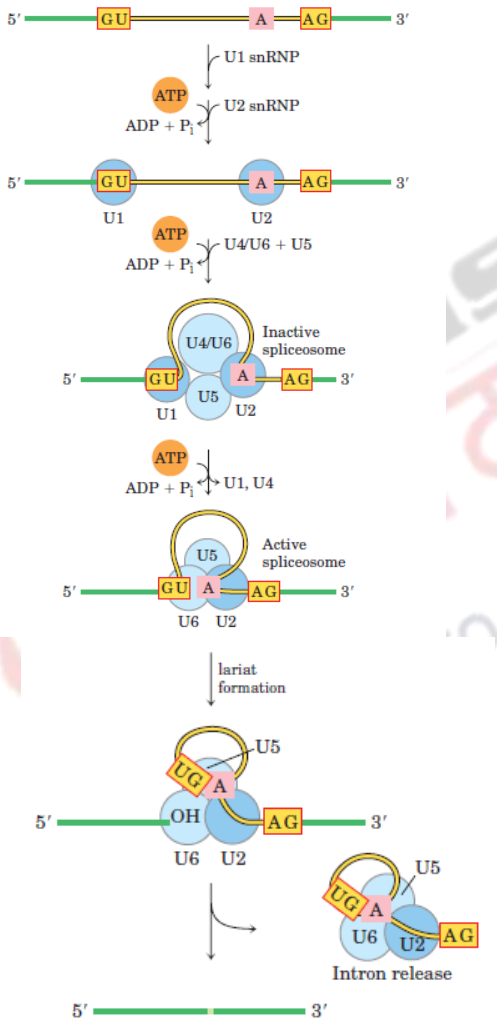


Fig 16. Spliceosomal introns A. snRNP pairing interaction during splicesome formation B.Mechanism of spliseosome formation

4. Summary

In this lecture we learnt about:

- Eukaryotic promoter elements and regulatory sequences
- Types of RNA polymerases
- Steps in eukaryotic transcription process
- Different types of post transcriptional modifications in eukaryotes