# Module 2

# TRANSCRIPTION

#### **SYLLABUS:**

#### **TRANSCRIPTION:**

Structure and function of RNA polymerases (prokaryotes & eukaryotes), mechanism of transcription in prokaryotes and eukaryotes, transcription factors, Fidelity of RNA synthesis, post-transcriptional processing (RNA editing, siRNA, splicing, poly A tail and 5' capping), Ribozymes, transcription inhibitors.

# TRANSCRIPTION:

- It is the synthesis of a RNA which carries the genetic information present in the DNA.
- The principal enzyme involved: **DNA dependent RNA polymerase**
- RNA polymerase is **similar to DNA polymerase** in most of its requirements like template, Mg ions and 4 NTP's.
- Similar to DNA polymerase, RNA polymerase also cannot use NDP or NMP as substrate.
- It **differs from DNA polymerase** in the fact that the RNA polymerase doesnot require a primer to initiate RNA synthesis
- It does not have exonuclease activity. Therefore it cannot carry out proofreading function
- As a result, the fidelity of RNA polymerase is less than that of of DNA polymerase.

- It is a multimeric enzyme having 5 different types of subunits :  $\alpha$  (alpha),  $\beta$  (beta),  $\beta'$ ,  $\omega$  (omega) and  $\sigma$  (sigma)
- The complete enzyme (holo enzyme) has 2  $\alpha$  subunits and one each of other 4 subunits
- MW of holoenzyme: 450 kD
- Round in shape, diameter: 100 A
- It can cover only about 30bp of DNA at the gene but during initiation of transcription about 60 bases of template DNA are covered. This is possible only if the enzyme is elongated in its shape
- σ factor: is an ancillary factor which is not required for polymerization of nucleotides but it is necessary for the initiation of transcription.
- Enzyme without **o** factor is known as **core enzyme**
- Complete enzyme (core enzyme +  $\sigma$  factor) is known as **Holo enzyme**
- For specific binding of RNA polymerase at the promoter site holoenzyme is needed because **σ** factor bears certain degree of promoter



### Eukaryotic RNA polymerase:

- Eukaryotes have more than one type of RNA polymerase
- 3 classes of RNA Polymerases based on their sensitivity to **α** aminitin- an antibiotic which inhibits mRNA synthesis
  - RNA polymerase I
  - RNA polymerase II
  - RNA polymerase III

### **RNA polymerase I:**

- Not sensitive to  $\alpha$  aminitin
- Present in nucleus
- Synthesizes rRNA

# **RNA polymerase II:**

- Highly sensitive to  $\alpha$  aminitin
- Present in nucleoplasm
- Synthesizes mRNA

# **RNA polymerase III:**

- Partially sensitive to  $\alpha$  aminitin
- Present in nucleolus
- Synthesizes tRNA

In general:

- **Eukaryotic RNA polymerase** is a large molecule of >500 kD in size
- Has 2 large subunits of  $\approx$  200 kD and  $\approx$  140 kD respectively
- 200 kD subunit: similar to the  $\beta$ ' subunit of E.coli RNA polymerase and have similar function i.e template binding
- Besides these 2 subunits, it also has upto 10 different small subunits
- There is a subunit of RNA polymerase II which has similarity with one of the subunits present in RNA polymerase I and also in RNA polymerase III-- which is similar to α subunit of E.coli RNA polymerase and help in enzyme assembly
- A sigma like factor is also present
- The tertiary structure of RNA polymerase of both prokaryotes and eukaryotes is such that it forms a groove and about 20-25 nucleotides of DNA fit within this groove making a tight fit.

Events:

- 1. Template recognition
- 2. Initiation
- 3. Elongation
- 4. Termination

# 1. Template recognition

- The promoter directs RNA polymerase to recognize the correct region of gene and to bind at this site
- **-35 region of the gene** serves this function and is recognized by the enzyme
- Size of RNA polymerase is such that 60 nucleotides in gene are involved in binding of enzyme to template
- Sigma factor plays an important role in this specific binding and is necessary for the formation of **promoter-enzyme complex**

# 2. Initiation (in prokaryotes)

- Once RNA polymerase is bound to the promoter, the first transcription complex is formed which is known as the **Binary complex** (DNA + enzyme)
- At this stage, gene is still ds and complex is called as the **closed promoter complex**
- Specific binding triggers the melting of -10 region of the gene resulting in the opening of small region of DNA formation of ssDNA. Now the complex gets converted to **open promoter complex**
- Melting also results in tight binding of the enzyme
- Without sigma factor melting does not occur and transcription does not proceed
- Now the enzyme RNA polymerase proceeds with RNA synthesis.
- RNA polymerase does not require a primer. It initiates denovo synthesis of new RNA chain
- To start with 1st 2 nucleotides are incorporated and **1st phosphodiester bond** is formed. This complex becomes the **ternary complex**.
- 1st nucleotide to be added are complementary to TSS of the gene
- For proper initiation, RNA polymerase should cover both -10 region and TSS.
- More nucleotides keep on getting incorporated and upto 9 nucleotides are added to the growing RNA chain before there is any movement of enzyme along with the DNA template
- After successful initiation, sigma factor is released

# **Initiation (in eukaryotes)**

- Is more complex
- Involves no of specific transcription factors
- For initiation, it requires a no of trans acting factors along with RNA polymerase
- All transcription factors involved with RNA poymerase II are called as **TF II**
- 1st the factor **TF IID** binds to TATA box covering 25 nucleotides
- Now **TF IIA** associates itself to the complex further extending the protected region towards upstream
- On the other hand, **TF IIB** associates itself protecting the region at -10 to +10. This complex prepares the stage for the binding of RNA polymerase II
- Finally **TF IIE** joins, extending the protection upto **+30** region
- Once the entire complex is assembled, the incorporation of 1st nucleotide takes place.
- 1st 2 nucleotides are incorporated and **1st phosphodiester bond** is formed.
- 1st nucleotide to be added are complementary to TSS of the gene
- More nucleotides keep on getting incorporated and upto 9 nucleotides are added to the growing RNA chain before there is any movement of enzyme along with the DNA template





# 3. Elongation

- Relatively simple process
- RNA polymerase moves along the template DNA
- Melting of DNA takes place ahead of the enzyme and simultaneous renaturation of DNA occurs in the region left by the enzyme
- At any given time, about 17bp of DNA remains open
- During elongation, the RNA polymerase keeps moving along the gene and area of gene which is being transcribed is covered by it.
- Every time a new base is added, the back of the enzyme (facing 5' side of the gene) moves one base along the gene. The front remains stationary
- Thus there is change in the configuration of the enzyme. This continues to happen until the addition of 9 nucleotides has taken place.
- By then the enzyme is compressed to its limit and front portion of the enzyme moves several bases ahead.
- The rate of elongation is approximately 40 nucleotides per second.



### 4. Termination

- Once the enzyme RNA polymerase hits the terminator, it falls off the template and the transcription stops.
- Termination takes place in one of the **two manners in prokaryotes**:
  - a. Rho-dependent termination
  - b. Rho-independent termination

# **Rho-dependent termination:**

- The RNA contains a binding site for a protein called **Rho** factor.
- Rho factor 46 kD protein binds to this sequence and starts "climbing" up the transcript towards RNA polymerase.
- When it catches up with the polymerase at the transcription bubble, Rho pulls the RNA transcript and the template DNA strand apart, releasing the RNA molecule and ending transcription



#### Rho-independent termination:

- Depends on specific sequences in the DNA template strand.
- As the RNA polymerase approaches the end of the gene being transcribed, it hits a region rich in C and G nucleotides.
- The RNA transcribed from this region folds back on itself, and the complementary C and G nucleotides bind together.
- This results in the formation of 7-20 bp long intra molecular **hairpin structure**. This region is followed by a small stretch of U residues.
- Such structure is thermodynamically highly unstable and cause the displacement of newly synthesized RNA (Transcript) from DNA template
- Once RNA is detached, RNA polymerase falls off and the transcription terminates



#### **Termination in eukaryotes:**

- The termination of transcription is different for the three different eukaryotic RNA polymerases
- **RNA Polymerase I and RNA Polymerase III** terminate transcription in response to specific termination sequences in either the DNA being transcribed (RNA Polymerase I) or in the newly-synthesized RNA (RNA Polymerase III).
- **RNA Polymerase II** terminates transcription at random locations past the end of the gene being transcribed. The newly-synthesized RNA is cleaved at a sequence-specified location and released before transcription terminates.

#### Transcription and translation are coupled in prokaryotes

- In prokaryotes, the transcription and translation of newly synthesized mRNA are coupled
- Translation of mRNA starts even before the entire mRNA chain has been synthesized and there is hardly any free mRNA present in the cell
- The half life of bacterial mRNA is very short
- The rate of bacterial translation is about 12-15 aa per second, which is almost same as the rate of transcription.

#### In eukaryotes transcription and translation are not coupled

- They are independent to each other
- Transcription takes place in the nucleus while translation in cytoplasm
- mRNA has to be transported to cytoplasm from nucleus
- The primary transcript is unstable in its original form and undergoes a number of post transcriptional modifications before mRNA attains the mature form and is translated.

#### Post transcriptional modifications

- 1. Splicing
- 2. Polyadenylation
- 3. Cap formation
- 4. RNA editing

# Splicing:

- Majority of eukaryotic genes have introns Which are present within the coding region of the gene but are not the part of ORF
- Introns are not required for biological function of mRNA and are excised by the process of Splicing.
- Splicing is one the **first and the most important modifications** which takes place during the maturation of mRNA
- Prokayotes donot have introns
- In eukaryotes the precursor form of mRNA is many times bigger than the mature mRNA.

# Splicing signals:

- To facilitate the cell machinery to recognise the precise place for cutting and joining of exons
- All introns start with GT at their 5' end and finish with AG at their 3'end. This is oftern referred to as **the** GT-AG rule of intron structure. (GU-AG rule in mRNA)
- These conserved regions are further extended by a few more nucleotides.
- An exon ends with trinucleotide **C/AAG** followed by intron beginning with **GT** which is further extended to **A/GAGT**.
- Similarly AG at 3'end of intron may be **preceded** by a sequence consisting of upto **11 T/C, N (any of 4nt) followed by C/T**
- 5' end of exon may often **begin with G**.
- The **5'end of intron** is known as **donor site** while the **3' end** is referred as the **Acceptor site**.
- About 18-40 nucleotides upstream of 3' splice site or acceptor site there is a region of conserved sequence knowm as **Branch site**.

Splice donor site		Branch site				Splice acceptor sit		
A/C A G	<mark>G U</mark> Pu A G U	CUPu	I <mark>A</mark> Py	F	<sup>p</sup> y rich	N C <mark>A G</mark>	G	
			←	20	- 50 bas	ses ———————————————————————————————————	•	
	<	 Intron				:	•	

. The consensus sequence for splicing. Pu = A or G; Py = C or U.

# **Splicing factors**

- Nuclear splicing of mRNA is mediated by RNA itself without the involvement of other enzyme or proteins
- These RNA's with catalytic activity are known as **Ribozymes**
- Nuclear splicing requires the bringing of donor site to the branch site and to the 3' end of 1st exon to 5'end of 2nd exon for their interaction.
- This is facilitated by the participation of number of small RNA molecules in the nucleus. These are known as **small nuclear RNA (Sn RNA)**
- Usually these are not present as free RNA molecules but are present in a protein associated form and are known as **small nuclear ribonucleoprotein particles (Sn RNP)**.
- These SnRNP's act as splicing factors
- Different types of Splicing factors:
  - U1 SnRNP
  - U2 SnRNP
  - U4 SnRNP
  - U5 SnRNP
  - U6 SnRNP

#### Mechanism of splicing/ Lariat formation:

- splicing requires the bringing of donor site to the branch site and to the 3' end of 1st exon to 5'end of 2nd exon for their interaction. This results in **Lariat formation**.
- SnRNP's play important role in the entire processes.



- U1/U2 snRNPs recognize 5' and 3' splice sites (GU-AG) and adenosine branch point (A) by base-pairing. Results in **Complex E.**
- U1 and U2 snRNPs interact to bring together 5' and 3' splice sites. Results in **pre-spliceosomal Complex A**
- **Complex B (pre-catalytic spliceosome)** is created by entry of preformed U4/U6.U5 tri-snRNP
- **Catalytic Complex B\*** is formed after dissociation of U1 and U4 snRNPs and other conformational and compositional changes.
- In **complex B\***, the U2, U5, and U6 snRNPs are positioned to execute the first step of splicing, which releases the 5' exon and creates an intron-3' exon **lariat structure**.
- In catalytic Complex C\*, the U2, U5 and U6 snRNPs are positioned to carry out the second step of splicing, which excises the intron and joins the two exons.
- After formation of the **Post-spliceosomal complex P**, the spliced mRNA is released



2. Cap formation:

- It takes place in the nucleus
- Cap is formed at 5' end of mRNA
- Cap is a very specific structure with unusual features
- The **most important feature** is the **addition of a 7- methylguanosine residue at 5' end of hnRNA** (heterogeneous nuclear RNA)
- This addition doesn't take place by the ususal 3'-5' phosphodiester bond rather **7mG is attached** to the terminal nucleotide of primary transcript by a **5'-5' condensation of phosphates** present at the 5'end of hnRNA and 5'end of 7mG.
- In this condensation, not only alpha phosphate, but **all 3 phosphates are preserved**
- Another striking feature of cap is that either one or two terminal nucleotides of the transcript are also methylated (2nd and 3rd nt after 7mG)
- In multicellular eukaryotes and some viruses, cap formation not only occurs in the base but also at the 2'OH of the ribose sugar.
- Cap structure plays an important role in recognition of initiation codon by the eukaryotic ribosome and is essential for translation
- It also provides stability to mRNA



Fig: 5' Cap structure of eukaryotic mRNA

# 3. Polyadenylation:

- Addition of a long stetch of A residues (commonly known as **poly A tail)** at **3'end of mRNA**
- It takes place in nucleus and hnRNA is said to be polyadenylated.
- It is carried out by **enzyme poly(A) Polymerase**, which is a template dependent enzyme
- It is directed by the poly(A) signal (AAATAA) present 10-20 nucleotides upstream of the end of transcript
- Poly(A) tail plays **important role in the stability of mRNA**
- The **length of poly(A) tail is shortened as the mRNA ages** and when it is short enough mRNA is enzymatically degraded
- The average length is about 100-200 bases.

Once these modifications have taken place, mRNA is transported to cytoplasm through nuclear pores in an energy dependent active transport manner and is translated.

• Poly(A) tail is important for nuclear transport, translation and stability of mRNA.

# SiRNA:

- short interfering RNA or silencing RNA, is a class of double-stranded RNA molecules, 20–25 nucleotides in length
- siRNA is involved in the RNA interference pathway, where it interferes with the expression of a specific gene.
- siRNAs are generated from cytoplasmic long double-stranded RNAs (dsRNAs) that are cleaved in the cytoplasm by Dicer into functional siRNAs.



**Figure 1.** General structure of siRNA. Two RNA strands form a duplex 21 bp long with 3' dinucleotide overhangs on each strand. The antisense strand is a perfect reverse complement of the intended target mRNA.

# Antisense strand also known as guide strand

# Sense strand also knowns as passenger strand

#### The mechanism of gene silencing

- RNAi-mediated gene silencing is executed by siRNAs.
- The process of silencing begins with the cleavage of long dsRNAs into 21–25 -nt fragments of siRNAs in cytoplasm.
- The process is catalyzed by Dicer enzyme.
- These siRNAs are inserted into multiprotein silencing complex, which is known as RNA-induced silencing complex (RISC).
- Subsequent unwinding of siRNA duplex, in turn, leads to active confirmation of RISC complex (RISC\*).
- Next, target mRNA (mRNA to be degraded) is recognized by antisense or guide RNA
- Target mRNA is cleaved in the centre of the region that is recognized by complimentary guide siRNA, which is 10–12 -nt away from the 5' terminus of siRNA.
- The RNAi process is completed by the last step of siRNA molecule amplification



#### **RNA EDITING:**

- It is the post transcriptional process through which some cells can make discrete changes to specific nucleotide sequences within a RNA molecule such that it differs from sequence of the DNA template.
- It is seen in mRNA, tRNA and rRNA of eukaryotes
- RNA editing occurs in the nucleus, as well as mitochondria and plastids (chloroplasts)
- Editing occurs in the RNA after it has been transcribed from DNA but before it is translated into protein

#### 2 mechanism of RNA editing:

- 1. Substitution editing
- 2. Insertion/ deletion editing

# Substitution editing:

- Involves chemical alteration of individual nucleotides
- Catalyzed by enzymes that recognise a specific target sequence of nucleotides.
- Example:
  - **Cytidine deaminases**: convert C in the RNA to U
  - Adenosine deaminase: convert A to I (Inosine), which the ribosome translates as G
    - Thus a CAG codon (for Gln) can be converted to CGG codon (for Arg).

# Insertion/ deletion editing:

- Involves insertion or deletion of nucleotides in the RNA
- These alterations are **mediated by Guide RNA molecules** that
  - Base pair with RNA to be edited and
  - Serve as template for the addition or deletion of nucleotides in the target RNA

- Guide RNAs (gRNAs) direct editing
  - gRNAs are small and complementary to portions of the edited mRNA
  - Base-pairing of gRNA with unedited RNA gives mismatched regions, which are recognized by the editing machinery
  - Machinery includes an Endonuclease, a Terminal UridylyITransferase (TUTase), and a RNA ligase
- Editing is directional, from 3' to 5'



#### **Transcription factors**

- Transcription factors are proteins that help turn specific genes "on" or "off" by binding to nearby DNA.
- Transcription factors that are activators boost a gene's transcription. Repressors decrease transcription.
- Groups of transcription factor binding sites called enhancers and silencers can turn a gene on/off in specific parts of the body.
- Gene expression is when a gene in DNA is "turned on," that is, used to make the protein it specifies.
- 2 types of TF: General or Basal TF and specific TF
- General, or basal, transcription factors simply assist in the binding of RNA polymerase to the promoter.
- Other types of transcription factors include activators and repressors. These transcription factors affect transcription in different ways; activators assist in the binding of RNA polymerase and repressors stop transcription

- A typical transcription factor binds to DNA at a certain target sequence. Once it's bound, the transcription factor makes it either harder or easier for RNA polymerase to bind to the promoter of the gene.
- In eukaryotes RNA polymerase can attach to the promoter only with the help of proteins called **basal (general) transcription factors**
- Activators: Some transcription factors activate transcription. For instance, they may help the general transcription factors and/or RNA polymerase bind to the promoter
- **Repressors**: These transcription factors repress transcription. This repression can work in a variety of ways. As one example, a repressor may get in the way of the basal transcription factors or RNA polymerase, making it so they can't bind to the promoter or begin transcription.
- **Binding sites:** The binding sites for transcription factors are often close to a gene's promoter. However, they can also be found in other parts of the DNA, sometimes very far away from the promoter, and still affect transcription of the gene.
- The flexibility of DNA is what allows transcription factors at distant binding sites to do their job.

- In eukaryotes, an important class of transcription factors called general transcription factors (GTFs) are necessary for transcription to occur.
- Many of these GTFs do not actually bind DNA, but rather are part of the large transcription preinitiation complex that interacts with RNA polymerase directly.
- The most common GTFs are TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH.
- The preinitiation complex binds to promoter regions of DNA upstream to the gene that they regulate

#### Mechanism:

- Transcription factors bind to either enhancer or promoter regions of DNA adjacent to the genes that they regulate.
- Depending on the transcription factor, the transcription of the adjacent gene is either up- or down-regulated
- Transcription factors use a variety of mechanisms for the regulation of gene expression
- These mechanisms include:
  - stabilize or block the binding of RNA polymerase to DNA
  - catalyze the acetylation or deacetylation of histone proteins.
    - histone acetyltransferase (HAT) activity acetylates histone proteins, which weakens the association of DNA with histones, which make the DNA more accessible to transcription, thereby up-regulating transcription
    - histone deacetylase (HDAC) activity deacetylates histone proteins, which strengthens the association of DNA with histones, which make the DNA less accessible to transcription, thereby down-regulating transcription
  - recruit coactivator or corepressor proteins to the transcription factor DNA complex

#### Ribozymes

- A ribozyme is a ribonucleic acid (RNA) enzyme that catalyzes a chemical reaction. The ribozyme catalyses specific reactions in a similar way to that of protein enzymes.
- Also called catalytic RNA, ribozymes are found in the ribosome where they join amino acids together to form protein chains.
- Ribozymes also play a role in other vital reactions such as RNA splicing, transfer RNA biosynthesis, and viral replication.
- The first ribozyme was discovered in the early 1980s and led to researchers demonstrating that RNA functions both as a genetic material and as a biological catalyst.
- It is thought that RNAs used to catalyse functions such as cleavage, replication and RNA molecule ligation before proteins evolved and took over these catalytic functions,
- **Examples of ribozymes** include the hammerhead ribozyme, the VS ribozyme, Leadzyme and the hairpin ribozyme, Group I self-splicing intron, Group II self-splicing intron Spliceosome is likely derived from Group II self-splicing ribozymes.

- <u>RNA can also act as a hereditary molecule, which encouraged Walter Gilbert to propose that in the distant past,</u> the cell used RNA as both the genetic material and the structural and catalytic molecule rather than dividing these functions between DNA and protein as they are today; this hypothesis is known as the "**RNA world hypothesis**" of the origin of life.
- The discovery of catalytic activity of RNA solved the "chicken and egg" paradox of the origin of life, solving the problem of origin of peptide and nucleic acid central dogma. According to this scenario, in earliest life all enzymatic activity and genetic information encoding was done by one molecule, the RNA

#### **Applications:**

- Ribozymes have been proposed and developed for the treatment of disease through gene therapy
- One area of ribozyme gene therapy has been the inhibition of RNA-based viruses.
- A type of synthetic ribozyme directed against HIV RNA called gene shears has been developed and has entered clinical testing for HIV infection
- Similarly, ribozymes have been designed to target the hepatitis C virus RNA, SARS coronavirus (SARS-CoV), Adenovirus and influenza A and B virus RNA
- The ribozyme is able to cleave the conserved regions of the virus's genome which has been shown to reduce the virus in mammalian cell culture. These projects have remained in the preclinical stage

# **Fidelity of RNA synthesis**

- Faithful transcription of DNA into RNA is essential to the flow of biological information.
- Relatively little is known about what controls the fidelity of transcription in vivo.
- The fidelity of RNA synthesis depends on both accurate template-mediated nucleotide selection and proper maintenance of register between template and RNA.
- Loss of register, or **transcriptional slippage**, can alter the coding capacity of mRNAs and is used as a regulatory mechanism.
- RNA polymerase (RNAP) misincorporates only one wrong nucleotide per ~100 000 bases
- Maintaining high fidelity during transcription is essential for the accurate transfer of genetic information from DNA to RNA.

- The discovery of a new **catalytic domain, the Trigger Loop,** revealed that RNA polymerase can actively choose the correct substrates.
- Also, the **intrinsic proofreading activity** was found to proceed via a ribozyme-like mechanism, whereby the erroneous nucleoside triphosphate (NTP) helps its own excision
- **Factor-assisted proofreading** takes place via exchange of RNAP active centres.
- Misincorporation is a major source of transcription pausing that may cause conflicts with fellow RNA polymerases and the replication machinery
- Therefore fidelity is not only required for production of errorless RNAs, but also for prevention of frequent misincorporation-induced pausing

# **Transcription inhibitors**

#### 1. Mitomycin

- Intercalates with DNA strands
- Blocks transcription
- Used as anti cancer drug

# 2. Adriamycin

• Inhibits the initiation phase by

https://www.slideshare.net/adurganaveen/transcription-57863890

https://slideplayer.com/slide/13029107/