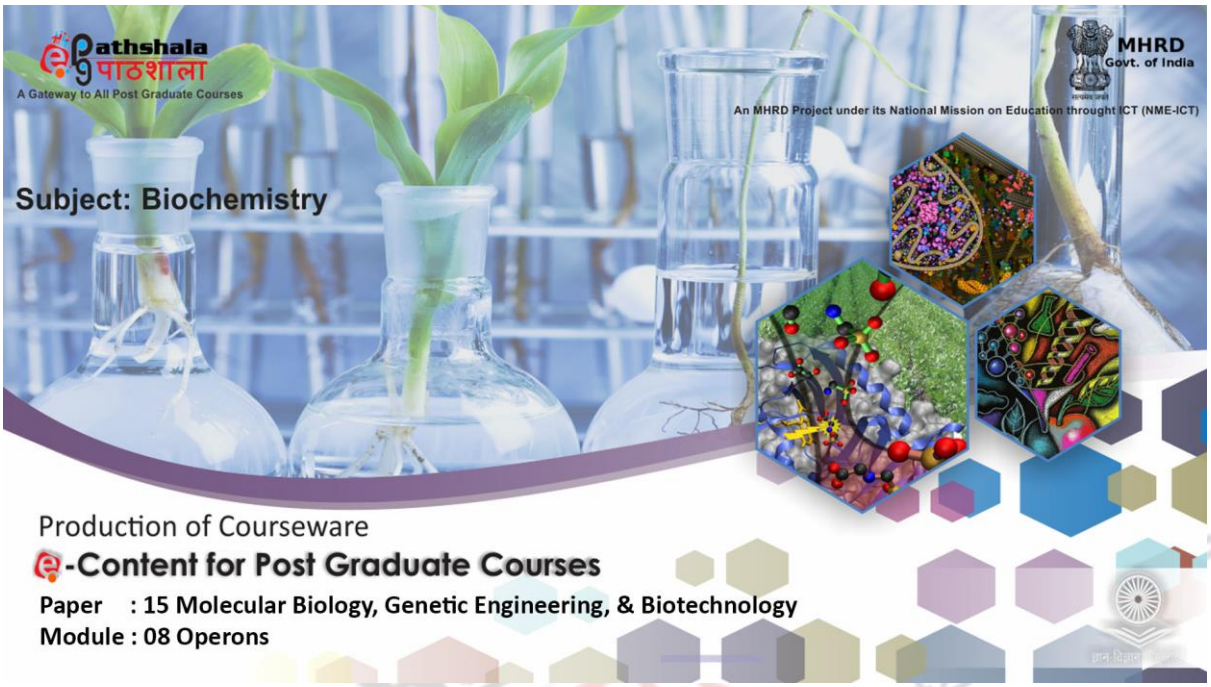


Development Team

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	XV-Molecular Biology, Genetic Engineering & Biotechnology
Module Name/Title	04: Operons

OPERONS

OBJECTIVES

- To understand how gene is expressed and regulated in prokaryotic cell
- To understand the regulation of Lactose utilization in *E.coli*
- To understand regulation of Tryptophan biosynthesis
- To understand regulation of Arabinose metabolisms *E.coli*

INTRODUCTION

OPERON

- In addition to encoding the information for the specific polypeptide sequences of proteins, the genome of the cell codes the information that regulates its own expression

- The genome is divided into sequences of DNA, known as **genes** that have specific functions

- Some genes code for the synthesis of RNA and proteins, determining, respectively, the sequences of the submit ribonucleotide bases and amino acids in these macromolecules

- Genes that **code for proteins** are known as **structural genes** or **cistrons**

- Other genes have regulatory functions and act to control the expression of the structure genes

- Together the structure and regulatory genes constitute the genotype and determine the phenotype, that is, the actual appearance and activities of the organism

- By controlling which of the genes of the organism are to be converted into functional enzymes, the cell regulates its metabolic activities

- Some regions of DNA are specifically involved in regulating transcription, and these regulatory genes can control the synthesis of specific enzymes

- In some cases, gene expression is not subject to specific genetic regulatory control, and the enzymes coded for by such regions of the DNA are **constitutive**, that is, they are continuously synthesized

- Some enzymes are synthesized only when the cell requires them. Such enzymes are **inducible**, or **repressible**

- Often, several enzymes that have related functions are controlled by the same regulatory genes

- Control of the expression of the structural genes for coordinated metabolic activity in bacterial cells is explained in part by the operon model, which demonstrates how the transcription of mRNA directing the synthesis of these enzymes is regulates

- An **operon** is a "DNA sequence that codes for one or more polypeptides, usually of related function, and a DNA sequence that regulates the expression of these genes"

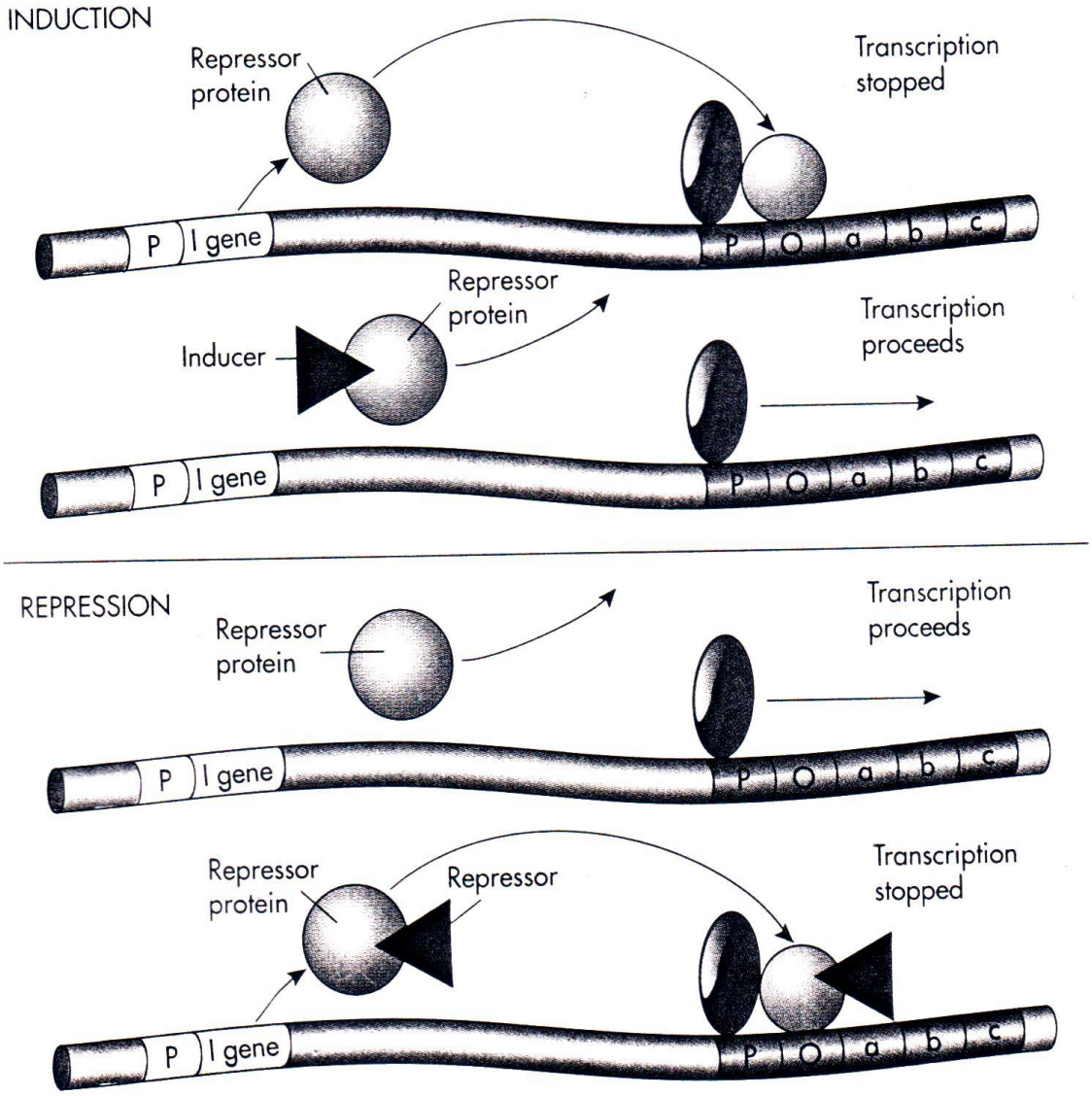


Fig. 6-40 Operons—Induction and Repression of Transcription. Some genes are inducible and others repressible. Induction occurs when a repressor protein, encoded by the I gene, that normally binds to the operator region of the DNA reacts with an inducer substance so that it no longer binds to the operator region. This results in the ability of RNA polymerase to move past the operator region so that structural genes under the control of that operator are transcribed. Repression occurs when a repressor protein, encoded by the I gene, that normally does not bind to the operator region of the DNA reacts with a repressor substance so that it then binds to the operator region. This results in the inability of RNA polymerase to move past the operator region so that structural genes under the control of that operator no longer are transcribed.

- Induction and repression are based on regulatory genes producing a regulator protein that controls transcription by binding to a specific site of DNA

- Regulation of transcription may be under **negative control**, where mRNA for a particular sets of genes are synthesized unless it is turned off by the regulatory protein

- Alternately, transcriptional regulation of a different set of genes may be under **positive control**, where mRNA is synthesized only in the presence of a regulatory protein that binds to the DNA

- Archaeal genes with coordinated functions are clustered as they are in bacterial cells

- There are differences from bacterial operons, however, in that clusters of genes in the archaeal chromosome contain internal promoters and terminators

- The promoters in bacterial cells are upstream of the operator region and not within the operons

- Also the transcripts produced from an operon in bacterial cells are of uniform length and always codes for the full complement of proteins of the operons

- The transcripts of archaeal operons are of varying lengths and may code for only some of the proteins in the operons

REGULATING THE METABOLISM OF LACTOSE: THE LAC OPERON

- Base on data arising from genetic studies, Jacob and monad proposed the operon model in 1961 to explain the phenomenon of induction.
- One of the best studied regulatory system concerns the enzymes produced by ***E.coli*** **strain K-12** for the metabolism of lactose

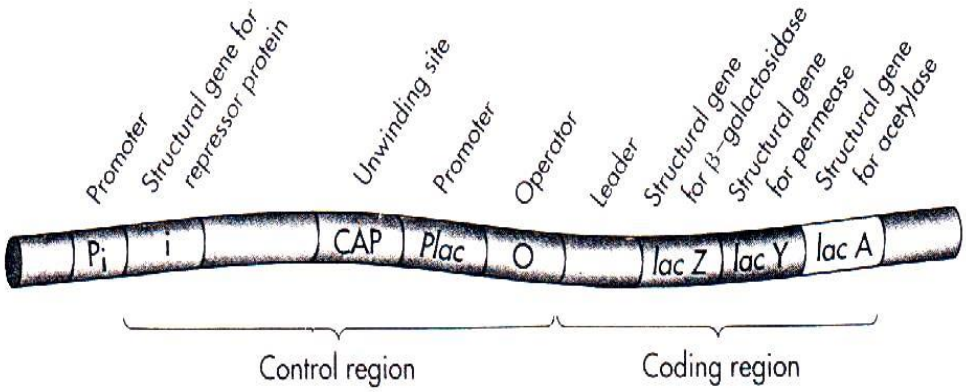


Fig. 6-41 The *lac* Operon. The *lac* operon controls the utilization of lactose. Three structural genes under the control of the *lac* promoter *(Plac)* code for the synthesis of the enzymes needed for lactose utilization. These enzymes are made only when lactose is present.

- Three enzymes are specifically synthesized by *E.coli* for the metabolism of lactose **(Fig. 6.41)**

- These enzymes are……

    1. β-galactosidase
    2. Galactosides permease
    3. Thiogalactoside transacetylase

- The β-galactosidase cleaves the disaccharide lactose into the monosaccharides galactose and glucose

- Permease is required for the active transport of lactose across the bacterial cytoplasmic membrane

- Transacetylase acetylates galactosides, allowing them to escape from the cell so they do not accumulate to toxic levels

- The structural genes that code for the production of these three enzymes occur in a contagious segment of the DNA, which code for **polycistronic DNA**

- Although there is some basal level of gene expression in the absence of an inducer, these structural genes are appreciably transcribed only in the presence of an inducer

- The operon for lactose metabolism, the *lac* operon, includes….

1. A promoter region **(P)** where RNA polymerase binds
2. A regulatory gene **(I)** that code for the synthesis of a repressor protein
3. An operator region **(O)** that occurs between the promoter and the three structural genes involved in lactose metabolism

- The regulatory gene codes for a repressor protein ,which in the absence of lactose binds to the operator region of DNA
- The binding of the repressor protein at the operator region blocks the transcription of the structural genes under the control of that operator region
- In some operons the operator region nucleotide sequence and the promoter region nucleotide sequence overlap each other
- Other operons may have more than one, or multiple, promoters
- In the case of the *lac* operon, the **three structural *lac* genes** that codes for the three enzymes involved in lactose metabolism are not transcribed in the absence of lactose
- The operator region of the lac operon is adjacent to or overlaps the promoter region
- Binding of the repressor protein at the operator region interferes with binding of RNA polymerase at the promoter region

- The inducer of the lac operon, **allolactose** (a derivative of lactose), binds to the repressor protein and alters the conformation of the repressor protein

- It acts as an **allosteric effector**, and therefore it is unable to interact with and bind at the operator region

- Thus in the presence of an inducer that binds with the repressor protein, transcription of lac operon is derepressed, and the synthesis of three structural proteins needed for lactose metabolism proceeds

*(Note: repressed = **suppress**, and derepress = **express**)*

- As the lactose is metabolized and its concentration diminished, the concentration of the derivative allolactose, produced from lactose by low levels of β-galactosidase, also declines, making it unavailable for binding with the repressor protein

- Thus active repressor protein molecules are again available for binding at the operator region and the transcription of the lac operon is repressed, ceasing further production of the enzymes involved in lactose metabolism that are controlled by this regulatory region of the DNA

- Under normal conditions the lac operon is repressed and transcription of the lactose-utilization gens occurs

- The **lac operon is typical of negatively-controlled inducible operons** that **regulate catabolic pathway**, wherein the presence of an appropriate inducer, the system is derepreesed

- Recent studied on the structure of the lactose operon repressor and its complexes with DNA and inducer have greatly increased understanding how the lac operon functions

- It is now known that the lac operon has three lac repressor recognition sites in a **500 base pair region** of the DNA

- In the absence of lactose the functional repressor protein, which is a tetramer, binds tightly to regions of the DNA and blocks transcription

- Each monomeric subunit has….
  - An amino terminal region that specifically binds to the DNA
  - A hinge region
  - A sugar binding domain and
  - A carboxyl terminal helix

- The DNA binding region of the lac repressor contains a helix motif that is similar to other regulatory proteins

- The tetrameric lac repressor proteins has **360 amino acids** and a molecular weight of 154,520 daltons

- Protease digestion of the tetramer produced a single carboxyl terminal core that binds to the inducer and four amino terminal fragments each of which can bind to operator DNA

- Control of the lac operon is complex
- These are three operator regions in the lac operon designated as **O_1 , O_2** and **O_3**

- The principal operator is **O_1**; **O_2** is located **401 bp** downstream from **O_1**; **O_3** is located **93 bp** upstream from **O_1**

- All three operators are required for maximal gens expression

- Binding of the repressor to the operator distorts the conformation of the DNA so that it bends away from the repressor

- This alters the grooves of the DNA double helix and is responsible for the change in efficiency of transcription

- Additionally activation of the lac operon involves a cyclic AMP-dependent activator protein that in presence of high concentrations of cyclic AMP increases transcription by binding to a recognition site on the DNA adjacent to the location of RNA polymerase attachment

- The structure of the tetrameric repressor suggests that it acts with the catabolite gene activator protein and forms repression loops in which one tetrameric repressor interacts simultaneously with two sites on the DNA

CATABOLITE REPRESSION

- When *E.coli* grows in a medium that contains glucose and lactose, it does not utilize both sugars simultaneously (at the same time)

- Instead, it preferentially utilizes glucose first until that sugar is depleted and only then switches to utilization of lactose as the carbon source

- This results in a biphasic (two phase) pattern of growth known as **diauxie** or **diauxic growth (figure 6.43)**

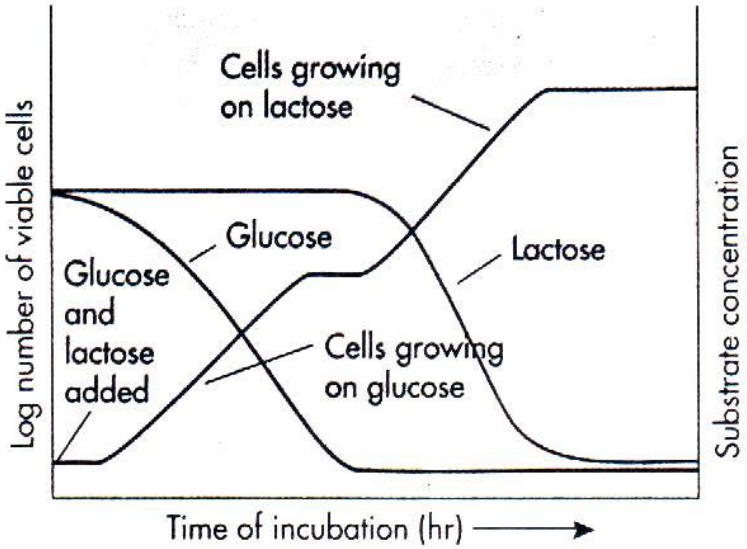


Fig. 6-43 Diauxie. *Escherichia coli* preferentially utilizes glucose and uses lactose only when the glucose supply is exhausted. This results in biphasic growth and the phenomenon of diauxie.

- In first phase of growth on glucose, the genes that code for the enzymes that metabolize lactose are "***shut off***"

- After glucose depletion, there is a lag phase in growth during which the genes that code for the enzymes that metabolize lactose are turned on and are transcribed and translated into proteins

- Then the cells can resume optimal growth using lactose the mechanism that allows bacteria to discriminate between utilization of two different carbon sources is largely due to catabolite repression

- **Catabolite repression** is a generalized type of repression that simultaneously shut off several operons. **(figure 6.44)**

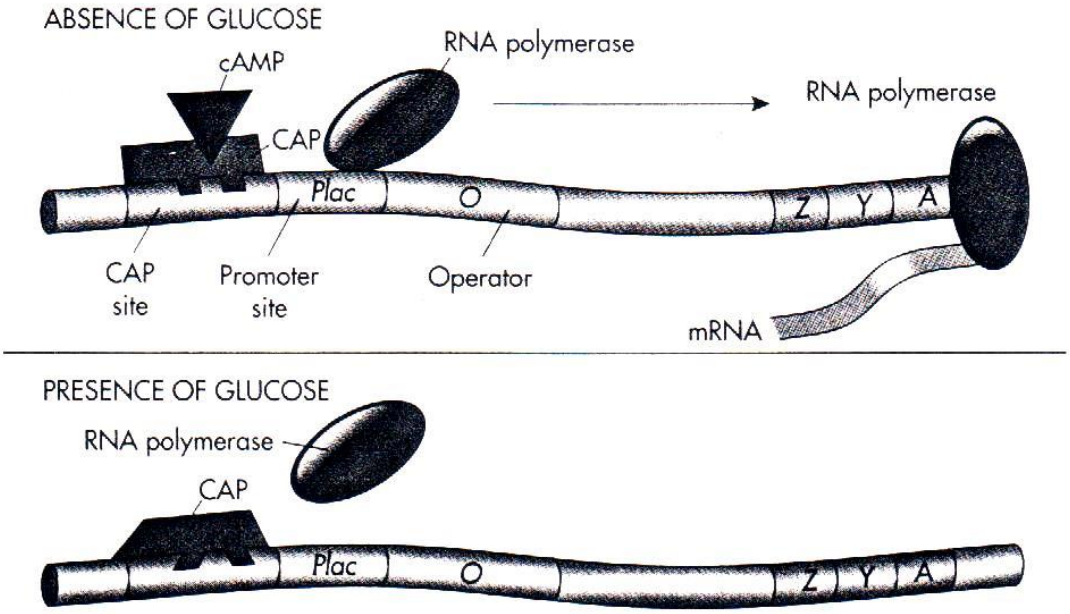


Fig. 6-44 Catabolite Repression. Catabolite repression explains why several catabolic pathways are shut off in the presence of glucose. Catabolite repression is based on the need for cyclic AMP (cAMP) to form an activated complex with catabolite activator protein (CAP) at the promoter site that enhances the binding of RNA polymerase. When glucose is metabolized, there is inadequate cAMP to facilitate RNA polymerase binding. Therefore transcription at several promoters ceases. When there is inadequate glucose, there is enough cAMP to bind with CAP, and so that transcription occurs at those promoters.

- In the presence of an adequate concentration of glucose, for example, some catabolic pathways are repressed by catabolite repression, including those involved in the metabolism of lactose, galactose, and arabinose

- When glucose is available for catabolism in the glycolytic pathway, other monosaccharide and disaccharides need not be used by the cell to generate ATP, and by blocking the metabolism of these other carbohydrates, the cell conserves its metabolic resources

- Catabolite repression is an example of regulation by positive control. It acts via the promoter region of DNA, and by doing so it complements the control exerted by the operator region

- The efficient binding of RNA polymerase to promoter regions subject to catabolite repression requires the presence of a catabolite activator protein, also called **cAMP** receptor protein. In the absence of the **CAP**, the RNA polymerase had a greatly decreased affinity to bind to the promoter region

- The **CAP**, in turn, cannot bind to the promoter region unless it is bound to **cAMP**

- There is an inverse relationship between the concentration of **cAMP** and **ATP**, in which levels of **cAMP** respond to the state of cellular metabolism

- Molecules of **cAMP** are formed from **ATP** by the enzyme **adenyl cyclase (figure 6.45)**

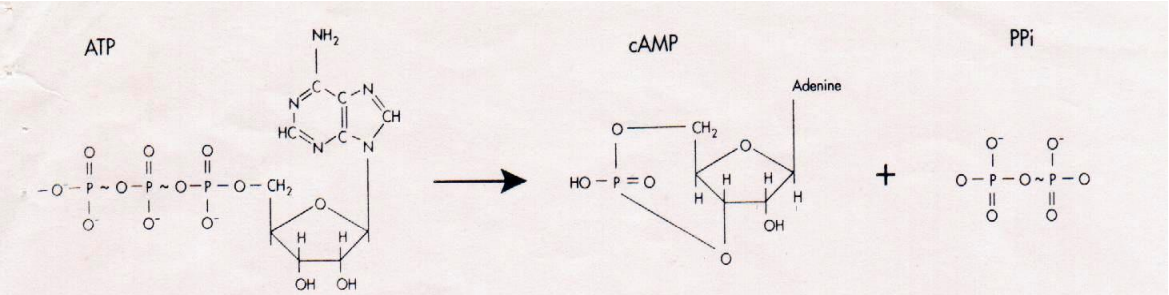


Fig. 6-45 Catabolite Repression—ATP Conversion to Cyclic AMP. Conversion of ATP to cyclic AMP by adenyl cyclase is important in the regulation of gene expression and control by catabolite repression.

- Intracellular level **cAMP** are low when rapidly metabolizable substrate such as glucose are used
- Under these conditions, the CAP is unable to bind at the promoter region
- Consequently, RNA polymerase are unable to bind to catabolite repressible promoters and transcription at a number of regulated structural genes ceases in a coordinated manner
- In the absence of glucose, there is an adequate supply of **cAMP** to permit the binding of RNA polymerase to the promoter region
- Thus, when glucose levels are low, **cAMP** stimulates the initiation of many inducible enzymes
  - **Adenyl cyclase activity**, which affects the concentration of intracellular **cAMP**, is partly regulated by the phosphoenolpyruvate: phosphotransferase system (PEP:PTS)

- Enzyme III^{glc} of the PEP:PTS function to shuttle a phosphate group from phospharylated HPr to enzyme II^{glc}

- Therefore, enzyme III^{glc} can exists in two different forms, either phosphorylated ($EIII^{glc}$ ~P) or non-phosphorylated ($EIII^{glc}$)

- When glucose is present outside the cell, enzymes III^{glc} continually transfers a phosphate group to enzyme II^{glc} and then to glucose as the sugar is transported through the cytoplasmic membrane

- Therefore, when glucose is present, $EIII^{glc}$ from predominates. When glucose is absent, phosphate groups are not transferred and the $EIII^{glc}$ ~P from predominates

- $EIII^{glc}$ but not ($EIII^{glc}$ ~P) is a allosteric inhibitor of adenyl cyclase

- This means that when glucose is present, and $EIII^{glc}$ levels are high, adenyl cyclase activity is inhibited and cAMP levels become reduced

- Conversely, when glucose is absent and lactose is present, $EIII^{glc}$ ~P levels are high, adenyl cyclase activity is not inhibited, and cAMP levels in the cell to transcribe its catabolite repressible operons, such as the *lac* operon

The Operon Model : Regulation of tryptophan biosynthesis:

The trp operon

Regulatory genes can be shut off under specific conditions. Such **repressible operons** control specific biosynthetic pathways. For example, the **trp operon**, which contains the genes that code for the enzymes required for the biosynthesis of the amino acid tryptophan is repressible (fig.6.46)

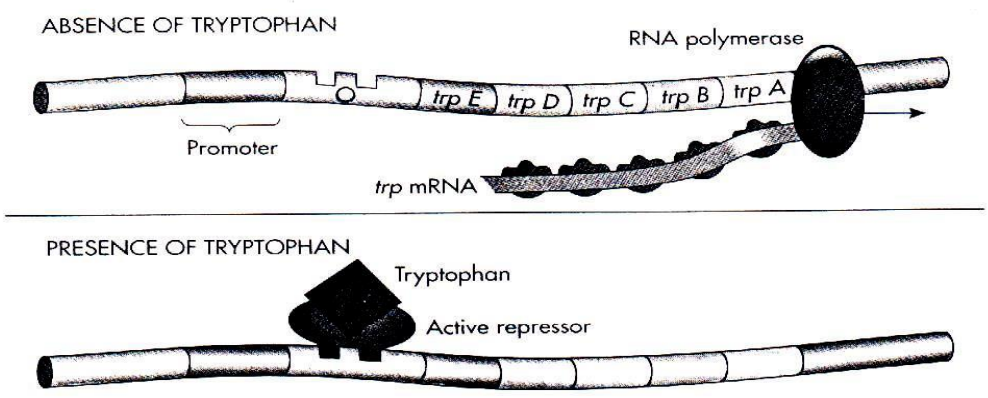


Fig. 6-46 The *trp* Operon. The operation of the *trp* operon permits the cell to stop synthesizing the enzymes involved in the biosynthesis of tryptophan when there is a sufficient concentration of this amino acid. Tryptophan interacts with a repressor protein, altering its conformation so that it can bind to the operator gene controlling the synthesis of several enzymes needed for the biosynthesis of tryptophan. Thus when there is enough tryptophan the enzymes needed for tryptophan biosynthesis are not made, thereby regulating this pathway.

There are five structural genes in the trp operon that are responsible for the synthesis of five enzymes. As with other operons there is also an operator region, a promoter region, and a gene that codes for a regulator protein in the trp operon. The trp repressor preotein is normally inactive and unable to bind

at the operator region, but trptophan can act as an allosteric effector or corepressor. In the presence of excess tryptophan, the trp repressor protein binds with tryptophan and, as a result, is also then able to bind to the trp operator region. When the trp repressor protein tryptophan complex binds at the trp operator region, the transcription of the enzymes involved in the biosynthesis of tryptophan is repessed.

In the case of negative control by the trp operon, tryptophan acts as the repressor substance that shuts off the biosynthetic pathway for its own synthesis when there is a sufficient supply of tryptophan. This is an example of end product repression-the process of shutting off transcription by the genes in that operon. When the level of tryptophan in the cell declines, there is insufficient tryptophan to act as co repressor, and the transcription of the genes for the biosynthesis of tryptophan therefore resumes. The trp operon is typical of anabolic pathways., where in the presence of a sufficient supply of the biosynthetic product of the pathway, the system is repressed.

Attenuation: regulation of transcription by translation

Attenuation is a mode of regulating gene expression in which the events that occur during translation affect the transcription of an operon region of the DNA (fig.6.47)

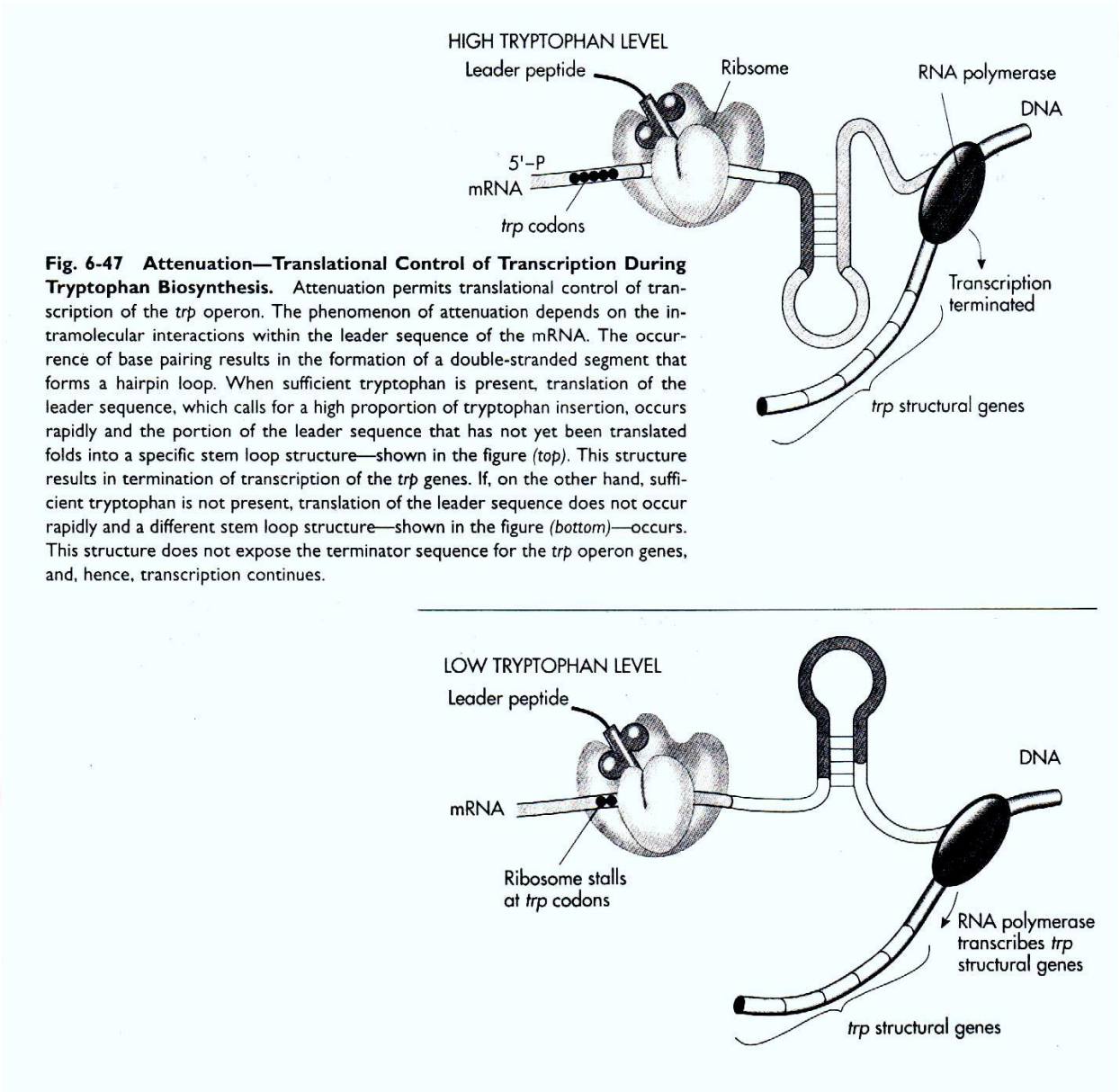


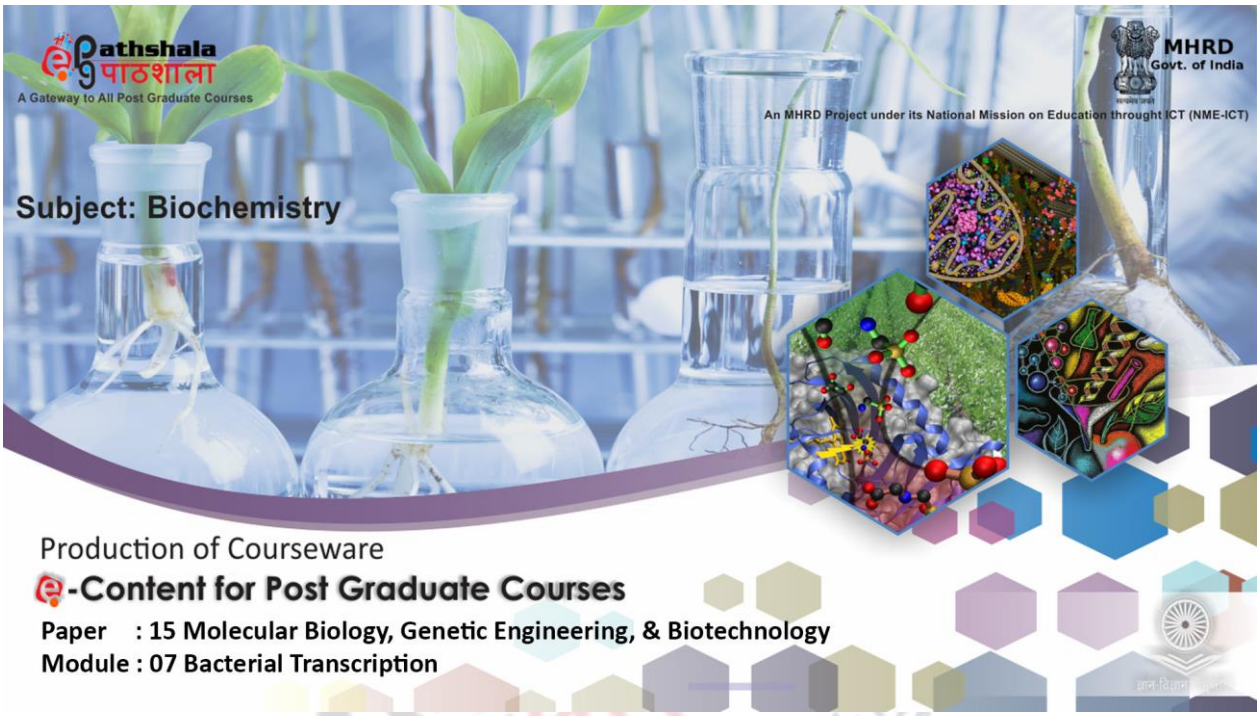
Fig. 6-47 Attenuation—Translational Control of Transcription During Tryptophan Biosynthesis. Attenuation permits translational control of transcription of the *trp* operon. The phenomenon of attenuation depends on the intramolecular interactions within the leader sequence of the mRNA. The occurrence of base pairing results in the formation of a double-stranded segment that forms a hairpin loop. When sufficient tryptophan is present, translation of the leader sequence, which calls for a high proportion of tryptophan insertion, occurs rapidly and the portion of the leader sequence that has not yet been translated folds into a specific stem loop structure—shown in the figure *(top)*. This structure results in termination of transcription of the *trp* genes. If, on the other hand, sufficient tryptophan is not present, translation of the leader sequence does not occur rapidly and a different stem loop structure—shown in the figure *(bottom)*—occurs. This structure does not expose the terminator sequence for the *trp* operon genes, and, hence, transcription continues.

As indicated earlier, translation in bacteria and archea normally begins before transcription of the mRNA is completed, with the leader potion of the mRNA attaching to the ribosome and translation beginning before the complete mRNA is transcribed. The leader sequence in bacterial cells occurs between the

operator and the first structural genes. In the case of the trp operon there is an attenuator site between the operator region and the first structural gene of the operon where transcription can be interrupted.

Whether or not the termination occurs is determined by the secondary structure of the mRNA at the attenuator region. There are two possible structures of mRNA. In one form, mRNA folds to establish a double stranded region, known as the terminator hairpin, that causes termination of transcription. The particular structure that forms depends on the availability of tryptophan because tryptophan is one the amino acids coded for in the leader sequence. When tryptophan is available for incorporation into proteins, the peptide sequence coded for by this leader sequence can be successfully translated. When tryptophan reaches very low concentrations, however, translation is delayed at the leader codons that code for the insertion of tryptophan into the polypeptide. This is sufficient tryptophan and charged tryptophyl-tRNA are not available. When the translational process is slowed, the mRNA is in the form that permits transcription to proceed through the entire sequence of the trp operon. However, if there is sufficient tryptophan to permit rapid translation to proceed through the attenuator site,the mRNA forms the terminator hairpin structure and further transcription of the trp operon ceases, so that none of the structural genes for tryptophan metabolism are transcribed.

In addition to the tryptophan operon, the histidine and phenylalanine operons in E.coli also contain attenuator regions. In histidine, there is a sequence of 7 contigouus codons in the leader sequence. In phenylalanine, there amino acids are very low does translation stall, allowing transcription to proceed through the attenuator site. The attenuator complements the regulation of genes expression by the operator genes. Thus there is a redundancy in the control mechanisms for the biosynthesis of amino acids such as tryptophan . the leader sequence and the associated attenuator site provide a mechanism for



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	XV-Molecular Biology, Genetic Engineering & Biotechnology
Module Name/Title	03: Bacterial transcription

BACTERIAL TRANSCRIPTION

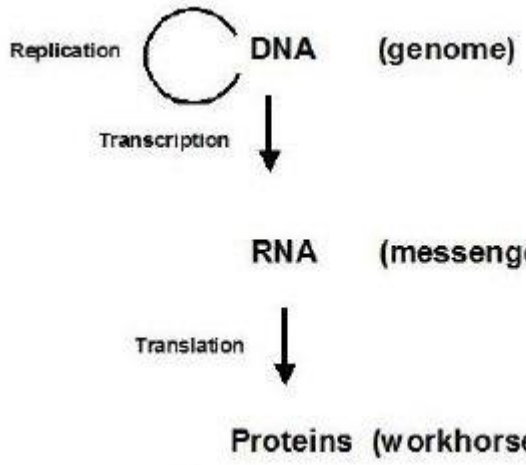
OBJECTIVES

- To understand central dogma of life
- To understand about synthesis of RNA molecule
  - ✓ Initiation
  - ✓ Elongation
  - ✓ Termination
- Post transcriptional modification of RNA molecules
- Inhibitors of transcription

INTRODUCTION

RIBONUCLEIC ACIDS

- In normal cell the flow of information occurs as …



- The information carried in DNA is expressed as specific sequence of amino acids in polypeptide
- RNA is an intermediate in this process and functions as a template in protein synthesis.

A.THE THREE KINDS OF RNA

- Messenger RNA (mRNA) molecules are transcripts of DNA segments (genes).
- Ribosomal RNA (rRNA) is part of the ribosome structure, which is the site of protein synthesis.
- Transfer RNA (tRNA) molecules are adaptor molecules being an activated amino acid and processing a recognition site for specific base triplets on mRNA.

B.THE SIZE OF RNA MOLECULES

- The length of mRNA is related to gene size and so varies greatly.
- rRNA in *E.coli* has three components varying in size from 120 to 3700 bases
- tRNA in *E.coli* average about 80 bases in length.

C. RNA MOLECULES

- RNA molecules are usually single-stranded but may have region of double helical structure produced by hairpin loops which are caused by Watson-crick base pairing.

I. SYNTHESIS OF RNA: TRANSCRIPTION

- The synthesis of RNA occurs during ***transcription***. This process involves unwinding the double helical DNA molecule for a short sequence of nucleotide bases, alignment of complementary ribonuceotides by base pairing opposite the nucleotide of the DNA strand being transcribed, and linkage of these nucleotide with phosphodiester bonds by a DNA-dependent RNA polymerase

- The process begins at a site where RNA binds and proceeds downstream toward the 3`-OH end until termination occurs. Binding of RNA polymerase to DNA may also be involved in localized unwinding and proper alignment of complementary RNA nucleotides

- RNA polymerases are able to link nucleotides only to the 3'-OH free end of the polymer. Thus the synthesis of RNA, like that of DNA, occurs in a 5'-P→3'-OH direction

- The molecule of RNA that is synthesized in transcription is anti-parallel to the strand of DNA that serves as a template

II. RNA POLYMERASE

- RNA polymerase synthesizes all cellular RNA on DNA templates.

- Requirements for synthesis of cellular RNA include:
  - ✓ A template of double stranded DNA (or occasionally of single-stranded DNA)
  - ✓ All four ribonucleoside triphosphate ,i.e.,ATP,GTP,UTP, and CTP.
  - ✓ Magnesium or manganese ions.

- The reaction catalyzed by RNA polymerase is represented by :

$$(RNA)_n \text{ bases} + \text{ribonucleoside triphosphate} \Longleftrightarrow (RNA)_{n+1 \text{ bases}} + PP_i$$

  - ✓ The hydrolysis of PP_i by phosphate drives the reaction to the right as in the DNA polymerase reaction.
  - ✓ Also as in DNA polymerization, there is a nucleophilic attack by the 3'-OH of the terminal nucleotide at the growing end of the chain on the incoming nucleoside triphosphate.
  - ✓ Synthesis is in the 5' to 3' direction.
  - ✓ Enzymes that synthesize RNA from ribonuceotides are DNA-dependent ***RNA Polymerases***. These enzymes can form phosphodiester bonds between two ribonuceotides only as long as they are aligned opposite the complementary DNA template nucleotides. Unlike DNA replication, RNA synthesis does not require a primer(Table-1)

- In *E.coli*, one enzyme polymerizes all three types of RNA, but in mammalian cells several different RNA polymerase perform these function.

- The RNA polymerase of *E.coli* is a complex enzyme consisting of several different kinds of subunits. The subunit composition of the complete enzyme (holoenzyme) is $\alpha_2\beta\beta'\sigma$. A form of the enzyme known as the core enzyme lack the sigma subunit.

Component	**Bacteria**	**Archaea**	**Eukaryotes**
Types of RNA polymerase	One type	Several types	Three types
RNA polymerase composition	4 subunits	8-12 subunits	12-14 subunits
Inhibited by Anisomycin	-	+	+
Inhibited by Rifampicin or Streptolydigin	+	-	-
Inhibited by heparin	+	-	-

- Bacteria have one basic type of RNA polymerase that synthesizes all three classes of RNA molecules
- In ***E.coli*** there is only one form, although other bacteria may possess several variants of the basic type of RNA polymerase
- In ***E.coli***, RNA polymerase is actually a complex of four protein subunits that form the core enzyme. These subunits are labelled **α**, ***β*** and ***β`***
- There are two copies of the ***α*** subunit in the core enzyme and one copy each of the ***β*** and ***β`***.
- In addition to the core proteins there is a sigma (***σ***) factor, which is involved in the initiation of RNA synthesis, and an omega (***ω***) factor, whose function in transcription is not clear at this time
- Bacterial RNA polymerases are inhibited by rifampicin and streptolydigin, which bind to the ***β*** subunit and prevent the RNA polymerase from initiating RNA synthesis

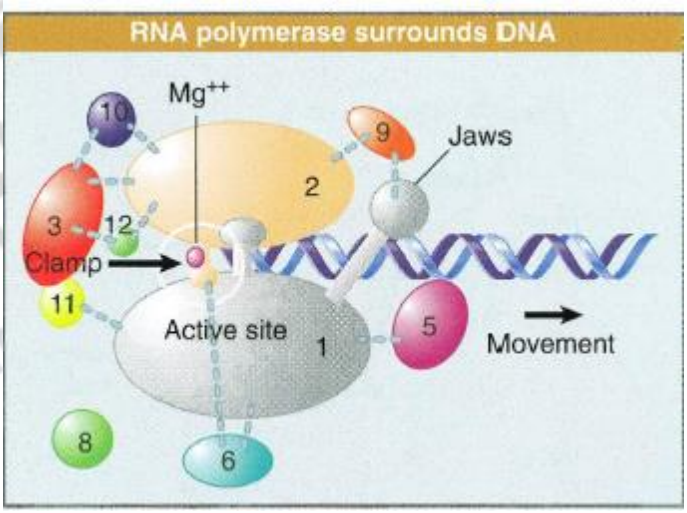


Figure1-Binding of RNA polymerase to DNA.

- Archaea appear to have their own unique RNA polymerases. These contain 8 to 12 polypeptides but differ in the size and number of copies of each subunit in the core enzyme in different archaea

- All archaeal RNA polymerases examined so far seem to be insensitive to the antibiotic rifampicin and streptolydigin. Archaeal RNA polymerases show a greater similarity to eukaryotic RNA polymerases than bacterial RNA polymerases. Each archaeal species has only one type of RNA polymerase, but different archaea have RNA polymerases that are similar to all three types of RNA polymerases found in eukaryotic cell based on comparisons of nucleotide sequence of their genes.

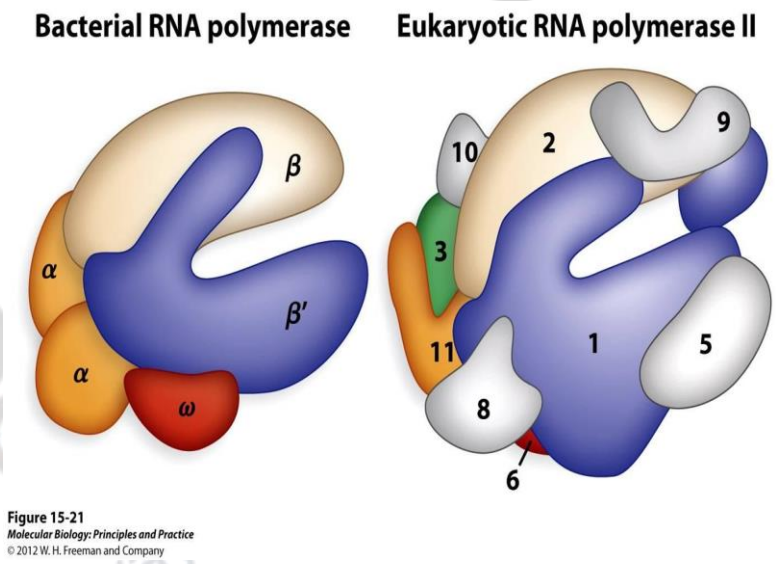


Figure -2. Comparison of Bacterial and Eukaryotic RNA polymerase.

- There is a characteristic organization of the genes coding for RNA polymerases in archaeal cells that is distinct from bacterial and eukaryotic cells .

- There are clusters of genes containing one small component gene, *rpoH* , followed by the genes of the large components *rpoB1* and *rpoB2* that codes for the *B* ` and *B* `` subunits followed by *rpoA1* and *rpoA2* that codes for the *A*` and *A*`` subunits.

- The archaeal RNA polymerases (B subunits) are insensitive to rifampicin and streptolydigin.

- Eukaryotic cells have three distinct RNA polymerase enzymes that are responsible for the synthesis of the three different classes of RNA
- These enzymes are quite complex and are composed of 9 to 12 subunits or more. RNA polymerase- 1 synthesizes rRNA, RNA polymerase- 2 synthesizes mRNA and polymerase- 3 synthesizes tRNA and 5s rRNA
- RNA polymerase- 1 is insensitive to α-amanitin, whereas RNA polymerase- 2 has a low sensitivity, and RNA polymerase- 3 has a high sensitivity to chemicals produced by some fungi
- All eukaryotic RNA polymerases are insensitivity to rifampicin and streptolydigin (antibiotics that inhibit bacterial RNA polymerase).

III. INITIATION AND TERMINATION OF TRANSCRIPTION

INITIATION OF TRANSCRIPTION

- ✓ The transfer of information from DNA to RNA requires that transcription begin at precise locations
- ✓ There are multiple initiation sites for transcription along the DNA molecule in bacterial, archaeal, and eukaryotic cells. Different initiation sites are needed to begin the synthesis of different classes of RNA and the synthesis of RNA for different polypeptide sequences
- ✓ There are also specific sites for the termination of transcription. By examining the DNA sequence for specific transcription start and stop signals it is possible to locate a region called an open reading frame (nucleotide sequence coding for a polypeptide). The open reading frame is equivalent to a gene.

Promoters

- ✓ What in the DNA molecule signals where to start reading a specific gene? DNA contains specific sequence of nucleotides, known as ***promoter regions,*** which RNA polymerase recognize and also that serve as signals for the initiation of transcription
- ✓ The promoter region of DNA is the site here RNA polymerase initially binds for transcription. The presence of the promoter region specifies (1) the site of transcription initiation and (2) which of the two DNA strands is to serve as the sense strand for transcription in that region.

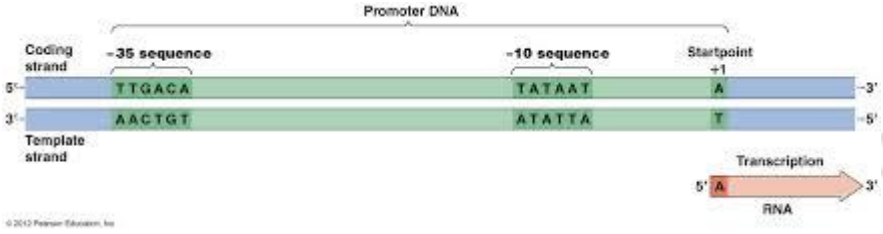


Figure -3 structure of prokaryotic promoter region.

(Source: http://www.mun.ca/biology/desmid/brian/BIOL2060/BIOL2060-21/CB21.html)

- ✓ The promoter region in the DNA of bacteria that have been examined consist of about 40 nucleotides ,that occurs just before the message that defines the m RNA, and also contain a seven-nucleotide sequence, known as ***pribnow sequence***, that appears to be a key part of the recognition signal (fig-3) .
- ✓ The ***pribnow sequence*** occurs about 5 to 8 nucleotide bases upstream (in the 5`- P direction) from the actual start of transcription.
- ✓ The designation "up-downstream" indicates that it is transcribed prior to later downstream nucleotides. Since the ***pribnow sequence*** has seven nucleotides, it overlaps the – 10 position, that is, a location 10 nucleotides upstream from the initial nucleotides of the gene that is transcribed.
- ✓ The ***pribnow sequence*** contains a sequence of nucleotide that is the same or almost the same as TATAAT for many of the bacterial promoters that have been examined
- ✓ This type of conserved DNA sequence is called a ***consensus sequence*** (meaning region of general agreement, that is, high nucleotide sequence homology).

- ✓ The pribnow consensus sequence starts at the -10 position on the DNA (counting nucleotides backward, or upstream, from the site of transcription, which is 11 and excluding 0). A second consensus sequence, TTGACA, is located on the promoter at about position --35. (fig-4).

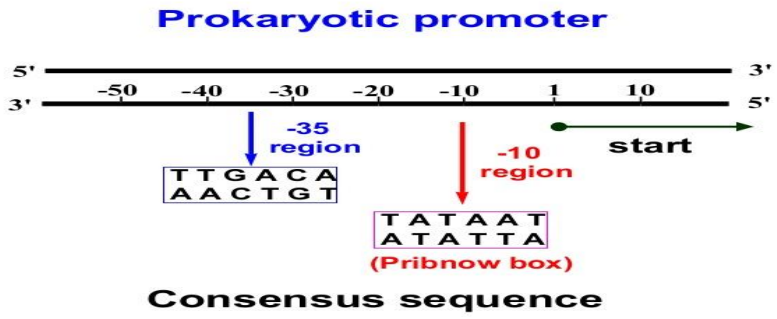


Figure-4 structure of prokaryotic promoter showing consensus sequences.

Source: http://www.slideshare.net/TapeshwarYadav1/transcription-56053031

- ✓ A highly conserved (virtually identical) consensus sequence for initiation of transcription has been found in all eukaryotic cells.
- ✓ ***Transcription factors (TFs)*** bind to DNA at specific promoter sites independently of the RNA polymerase. RNA polymerase -2 requires four transcription factors: TF -2A, TF-2B, TF -2D, and TF -2E.
- ✓ The transcription of RNA polymerase -2 promoters in eukaryotic cells requires the binding of TF -2D, also called TATA factor.
- ✓ The TATA factor is a protein transcription factor that preferentially binds to a conserved A-T rich DNA sequence called the TATA box:

$$5` - TATA(\frac{T}{A})A(\frac{T}{A})$$

- ✓ This conserved consensus sequence is centered about -25 nucleotides upstream from the start nucleotide and is analogous to the -10 consensus sequence in bacterial cells

- ✓ In eukaryotic cells, transcription factor –2B (TF -2B) plays a role early in transcription initiation by RNA polymerase -2 (Pol -2)
- ✓ The first transcription factor -2D (TF -2D) binds to the TATA box of the promoter and then TF -2B is added
- ✓ Since RNA polymerase -2 synthesizes mRNA, the TATA box is an important recognition site for initiation of transcription that leads to synthesis of the proteins of the eukaryotic cells.

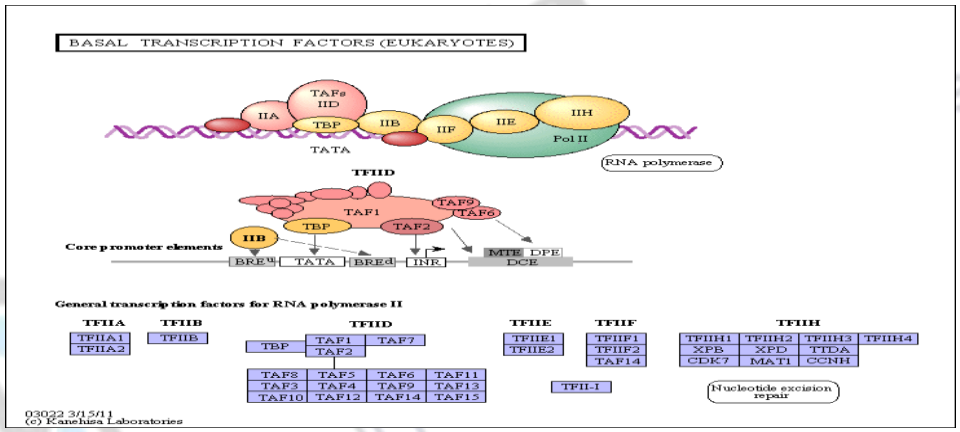


Figure- 5. Transcription complexes in Eukaryotic transcription process

Link-http://employees.csbsju.edu/hjakubowski/classes/ch331/bind/olbindtransciption.html

- ✓ There is a significant difference in the consensus sequence of bacterial versus archaeal and eukaryotic cells. Archaeal RNA polymerases recognize initiation sites (promoters) that are very similar to those of eukaryotic cells
- ✓ The major element determining transcription initiation by archaeal methanogens is:

$$5` - TTTA(\frac{T}{A})ATA$$

- ✓ This is very similar to the TATA box of eukaryotic cells. In archaea there is a second conserved element with the consensus sequence 5` -ATGC is located approximately 25 bp (base pair) downstream from the TATA box

- ✓ This region is the actual site of transcription initiation. Transcription is activated by the interaction of specific transcription factors with a general transcription complex that binds to the TATA box promoter elements
- ✓ The sequence surrounding these TATA boxes may not contain conserved regulatory regions in bacteria. Replacement of a standard TATA box of an archaeal cell with the eukaryotic polymerase -2 TATA box causes a 31% reduction in transcription efficiency
- ✓ These findings suggest that the mechanism of initiation of transcription in archaea is more like that of eukaryotic cells and less like transcription initiation in bacterial cells
- ✓ Another similarity between archaeal and eukaryotic transcription is the finding of a transcription factor analogous to the eukaryotic transcription factor -2B in the archaean ***pyrococcus woesei***
- ✓ The archaeal transcription factor includes nucleotide sequence that encodes a protein similar to TF -2B of eukaryotic cells
- ✓ In summary, archaeal transcription appears to be similar to eukaryotic cellular transcription.

Bacterial sigma factors

- ✓ The core enzyme ($\alpha_2\beta\beta`$) alone cannot recognize the promoter region.The sigma factor is essential for this function
- ✓ The sigma factor also play a essential role in unwinding the DNA helix
- ✓ The initial binding of bacterial RNA polymerase core enzyme ($\alpha_2\beta\beta`$) to the promoter region depends on the presence of sigma factor (σ factor) .
- ✓ Without the sigma subunit, the RNA polymerase fails to exhibit the necessary specificity for recognizing the initiation sites for transcription. The sigma factor thus ensures that RNA synthesis begins at the correct site.
- ✓ The complete RNA polymerase (core + sigma unit) is the ***holoenzyme***. The RNA polymerase holoenzyme first binds to the DNA promoter at the -35 consensus sequence, forming a closed complex

- ✓ The RNA polymerase holoenzyme then shifts its binding to the -10 **pribnow sequence**. As it does so. The DNA helix is unwound to form a single-stranded region and an *open complex*
- ✓ The RNA polymerase holoenzyme is now poised to begin transcription. The first nucleotide added is usually a purine (adenosine or guanosine)
- ✓ After formation of about 10 phosphodiester bonds between ribonucleotides, the sigma subunit dissociates from the RNA polymerase and the remainder of the RNA molecule is synthesized or elongated by the core RNA polymerase
- ✓ The sigma subunit is then free to associate with another RNA polymerase molecule, completing that molecule and establishing the necessary specificity for the binding to a new transcriptional site.

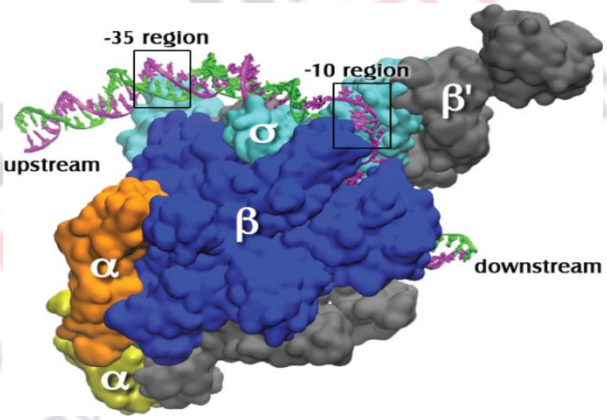


Figure-6. The open complex of RNA polymerase holoenzyme with promoter DNA.

(Source: Biomolecules 2015,5,668-678;doi:10.3390/biome5020668)

- ✓ Bacteria actually have multiple σ factors; each is responsible for the recognition of specific promoter initiation sequences
- ✓ The main σ factor in *E.coli* is σ^{70} with σ^{54}, σ^{32}, and σ^{28} normally present in lower concentration. The superscript associated with each σ factor represents the molecular weight of the protein × 10^{-3}
- ✓ Under certain changes in environmental conditions, σ^{54} or σ^{32} increase in concentration and direct the RNA polymerase to bind at other promoter consensus

sequence (TTGCA for σ^{54} and CCCCAT for σ^{32}), which are different than the **pribnow sequence** recognized by σ^{70}

- ✓ As a result of this control mechanism, regulation of the concentrations of the different σ factors in the cell leads to the specific or preferential transcription of certain genes and not others.

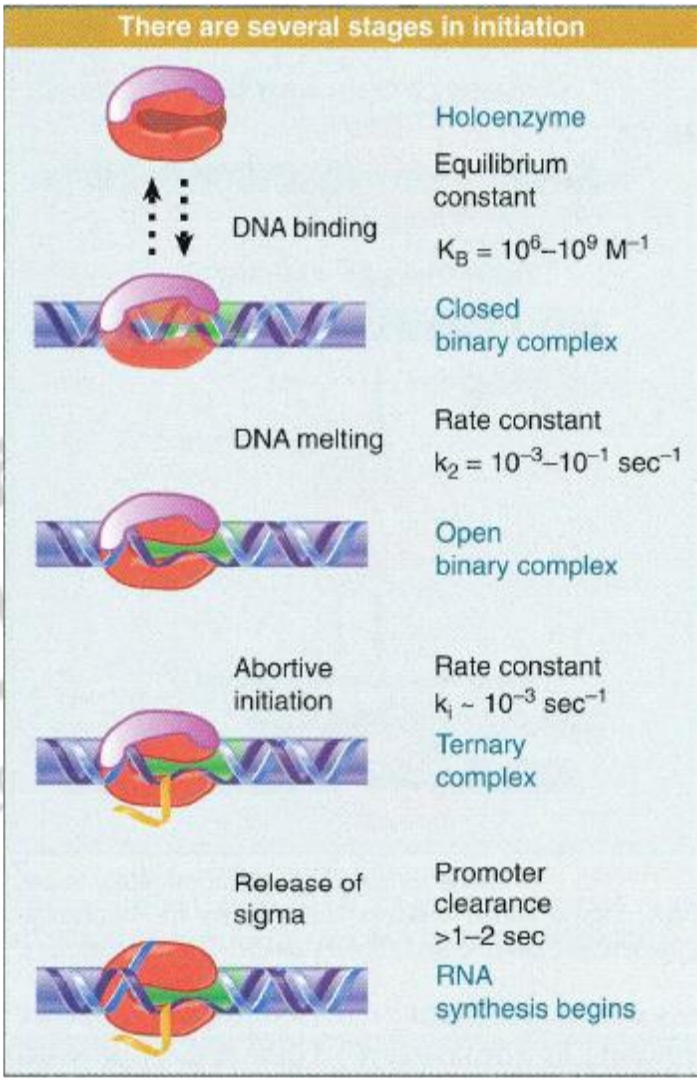


Figure-7. RNA polymerase passes through the several step prior to elongation. a closed binary complex is converted to an open form and then into ternary complex .

ELONGATION

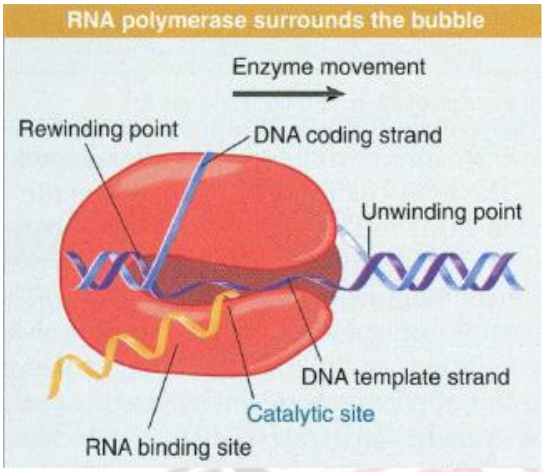


Figure-8. during transcription bubble maintained within bacterial RNA polymerase, which unwind and rewind DNA and synthesize RNA

- ✓ In elongation process enzyme moves with the DNA and extends the growing RNA chain
- ✓ As the enzyme moves it unwind the DNA helix to expose a new segment of the template in single stranded condition
- ✓ Nucleotide is added at 3' end to growing RNA chain, to form RNA –DNA hybrid in the unwound region
- ✓ Behind unwound region, the DNA template pair with its original partner to reform double helix
- ✓ The RNA emerges as a free single stranded.

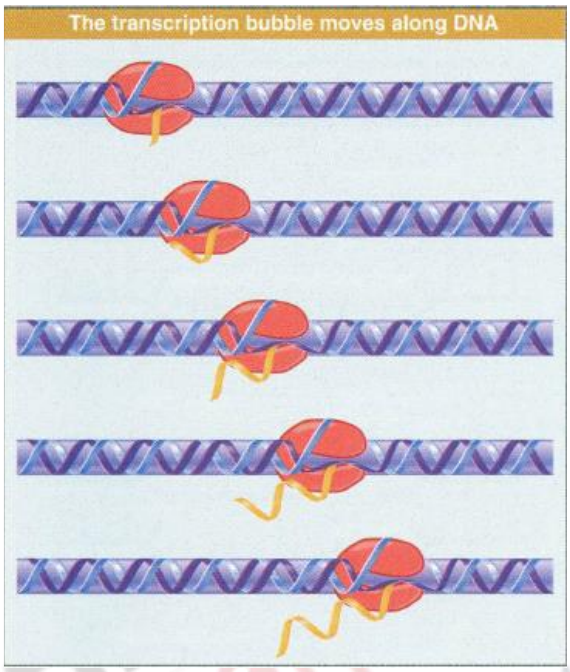


Figure- 9. Transcription take place in bubble, in which RNA is synthesized by base pairing with one strand of DNA in the transiently unwound region. As the bubble progresses the DNA duplex reform behind it, displacing RNA in the form of a single polynucleotide chain.

TERMINATION

- ✓ The DNA message also contains stop signals or point where transcription ceases, more base should not be added to chain
- ✓ The stop regions have a two fold symmetry in base composition that allows the complimentary single – stranded RNA to form base pairs and produce a hairpin loop
- ✓ Some termination sites also require the presence of protein called rho protein to ensure chain termination
- ✓ When last base added to the RNA chain the transcription bubble is collapse as the RNA- DNA hybrid is disrupted , the DNA reform in duplex state and RNA and enzyme both are both released.

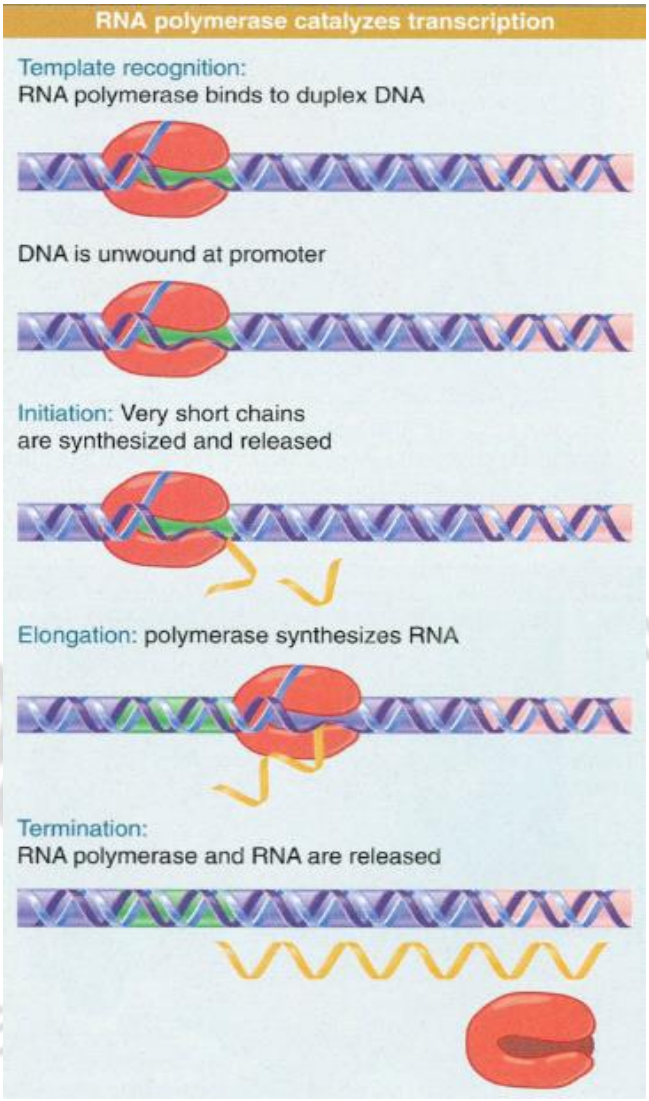


Figure-10. Transcription has a four stage: the enzyme bind to the promoter and melt DNA , remain stationary during initiation ,moves along the templates during elongation and diassociation at termination

IV. POST –TRANSCRIPTIONAL MODIFICATION OF RNA MOLECULES

- The r RNA and tRNA in prokaryotes and all RNA in eukaryotes undergoes modification after transcription to form functional molecules. The mRNA of prokaryotes, on the other hand, is rarely modified

- Three types of modification occurs in prokaryotic RNA

1. Cleavage of a primary transcript results in the formation of functional molecules of rRNA and tRNA . in E.coli, three type of rRNA and tRNA are excised from a single primary transcript,which also include a spacer region that do not become a part of any final functional molecules.

2. The addition of specific nucleotides to the terminus of RNA and also produce functional molecules, as in the case of tRNA molecules which all have a CCA sequence added to their 3' terminus.

3. Finally, in bacteria , bases can be methylated using s- adenosylmethionine as the methyl donor.

Note:

- ✓ All tRNA molecules contains unusual bases, such as pseudouridine and ribothymine, which are formed from uridine and thymine, respectively, after the tRNA chain has been synthesized.
- ✓ In prokeryotes, mRNA usually codes for a number of peptides, i.e.,it is polycistronic, in eukaryotes mRNA is monocistronic.

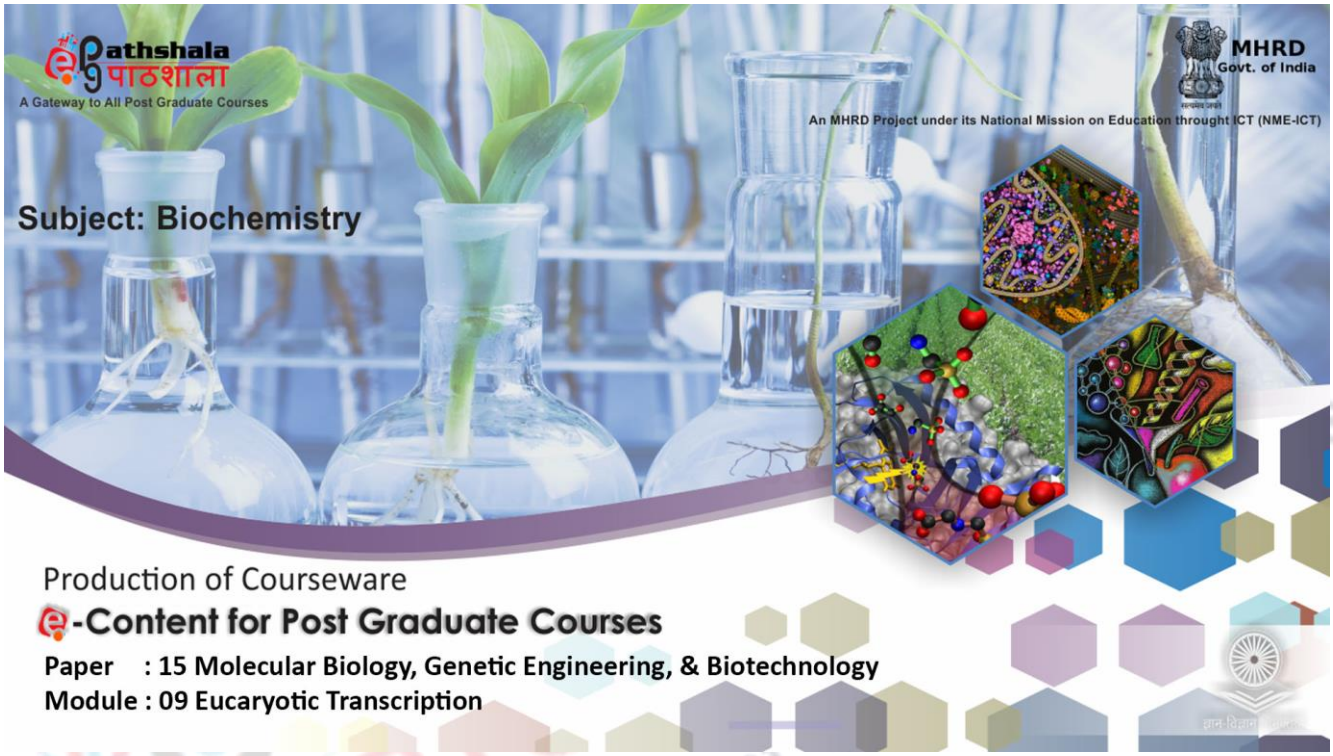
V. INHIBITORS OF TRANSCRIPTION

- ***RIFAMPIN***

  - ✓ Rifampin is a semi synthetic derivatives of rifamycine (from Streptomyces mediterranei)
  - ✓ Binds to the β subunit of the bacterial RNA polymerase, blocking the formation of first phosphodiester bond in the RNA chian
  - ✓ Dose not prevent binding of RNA polymerase to DNA or block the elongation of the chains already initiated

- ✓ Does not inhibit most eukaryotic nuclear RNA polymerase and some viral enzymes
- ✓ Rifampin is used in the therapy of tuberculosis.

- ***ACTINOMYCIN D***
  - ✓ Binds to duplex DNA and prevents the DNA from acting as a template.
  - ✓ Interaction between bases, articularaly in G-rich regions.
  - ✓ At low concentration inhibits transcription without producing a major effect on the replication of DNA or protein synthesis.
- ***STREPTOLYDIGINE.***
  - ✓ In contrast to rifampin, block RNA elongation in prokaryotes even though site of its action also is at β subunit of RNA polymerase.
- ***α AMANITIN***
  - ✓ α amanitin is derived from the mushroom Amanita phalloides
  - ✓ Inhibits eukaryotic RNA polymerase II and III (polymerase II is involved with the mRNA synthesis ; polymerase III is involved with tRNA and 5S rRNA synthesis
  - ✓ Affects polymerase II at lower concentration than polymerase III
  - ✓ Does not inhibit synthesis of nucleolar rRNA
  - ✓ Does not affect bacterial, mitochondrial, or chloroplast RNA polymerase.



athshala पाठशाला

MHRD Govt. of India

Development Team

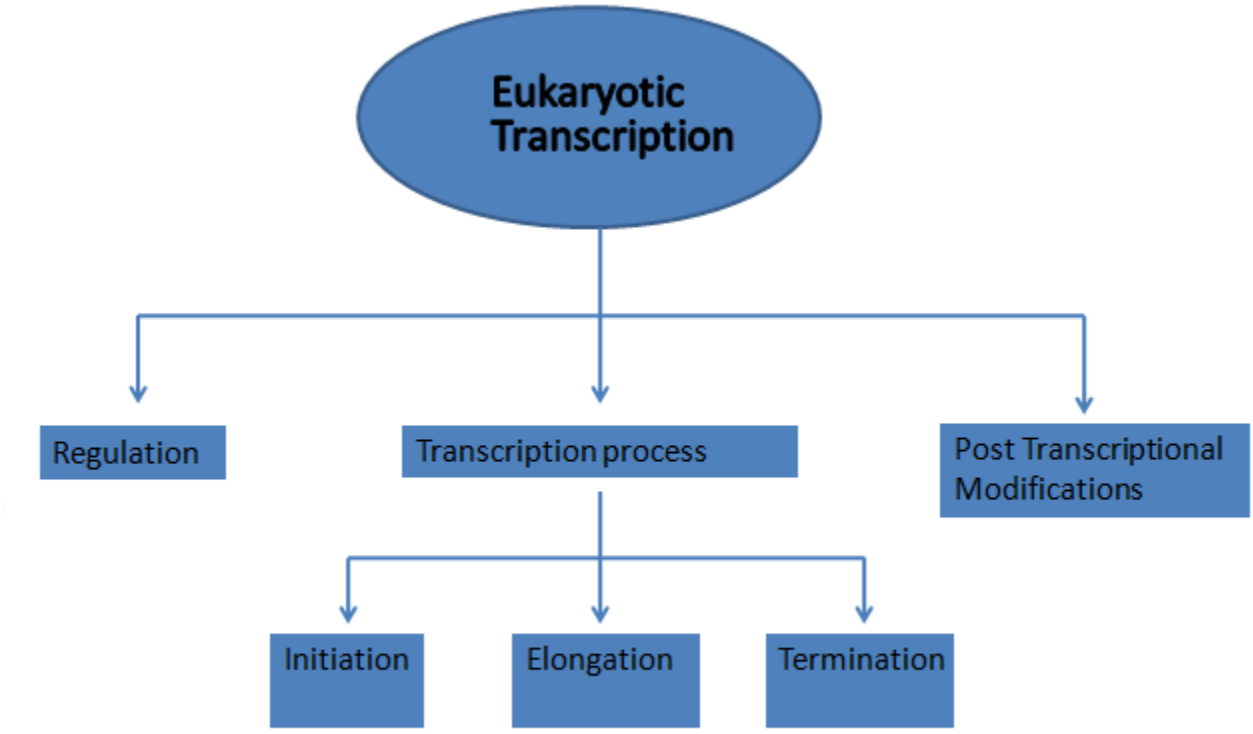
Principal Investigator	Dr. Sunil Kumar Khare, Professor, Department of Chemistry, IIT-Delhi
Content Writer:	Dr. Ramesh Kothari, Professor - Microbiology, UGC-CAS Department of Biosciences, Saurashtra University, Rajkot
Content Reviewer:	Dr. Lata, Principal Scientist, Division of microbiology, IARI, New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

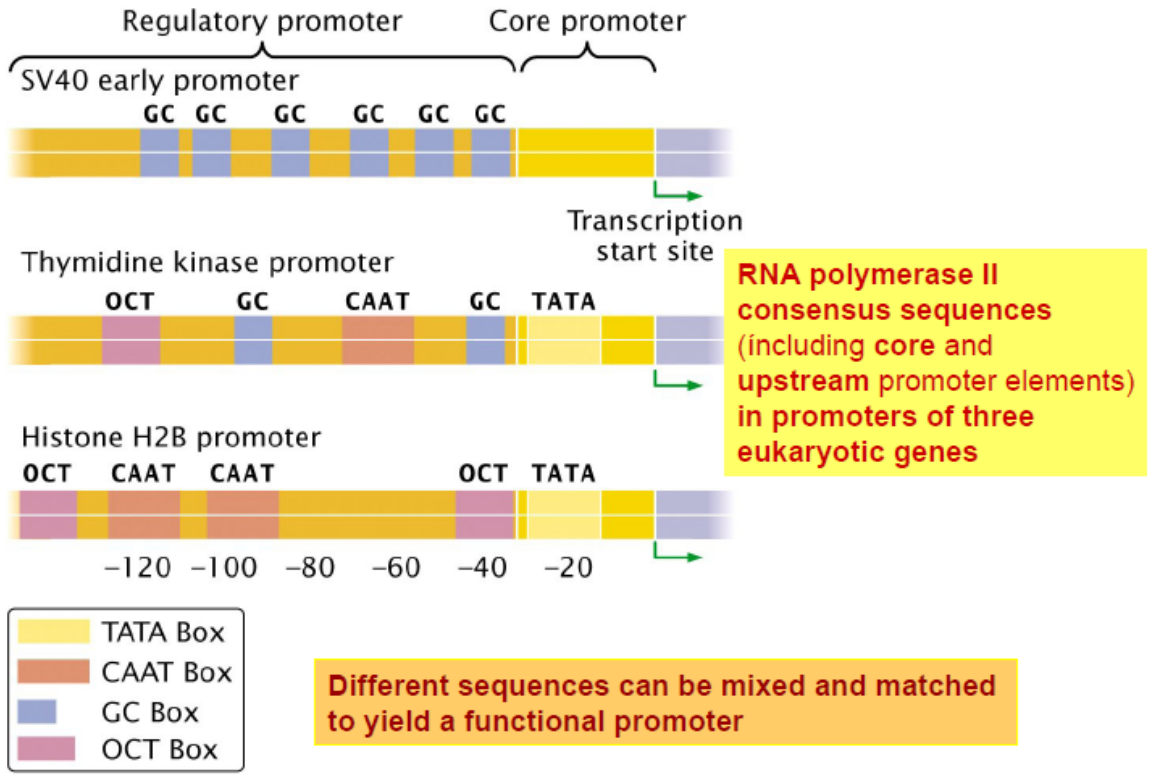


Figure 1A: Eucaryotic promoter element recognized by RNA Pol II

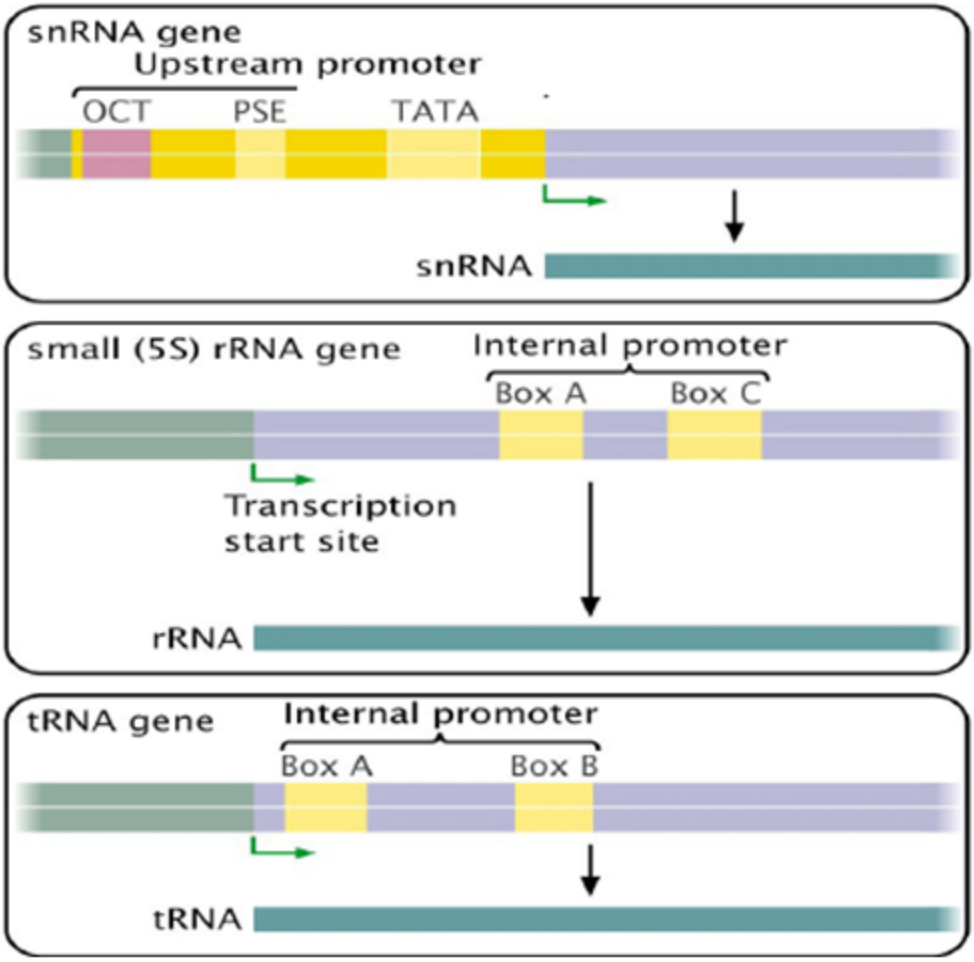


Figure 1B: Eucaryotic promoter element recognized by RNA Pol III

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.

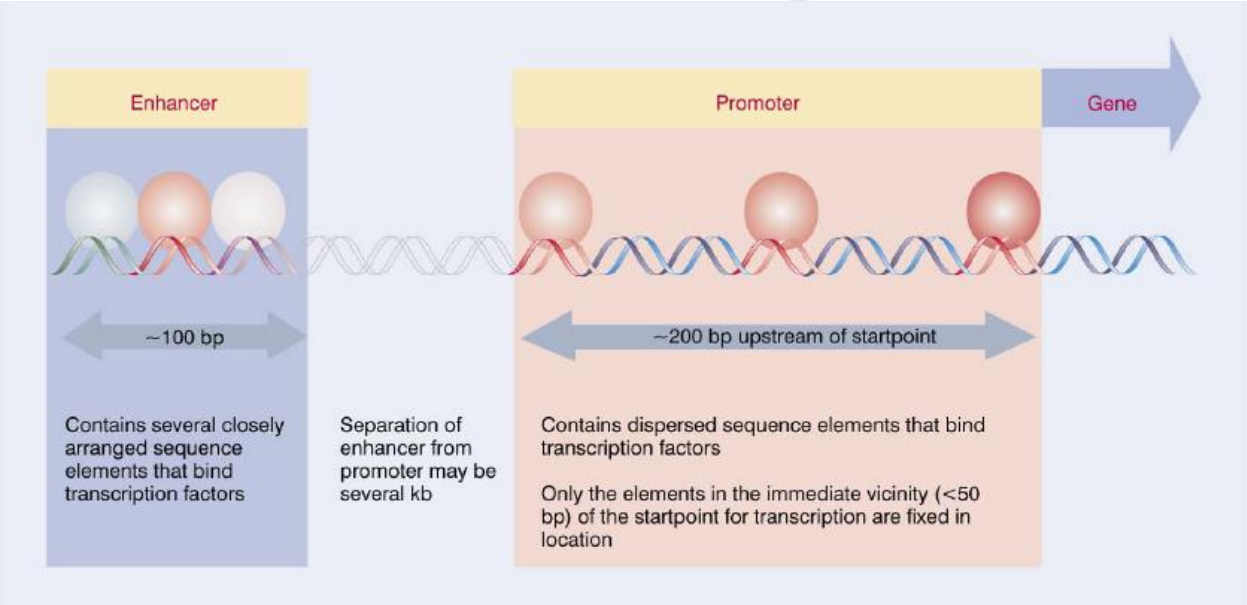


Figure 2- A Eukaryotic transcription unit

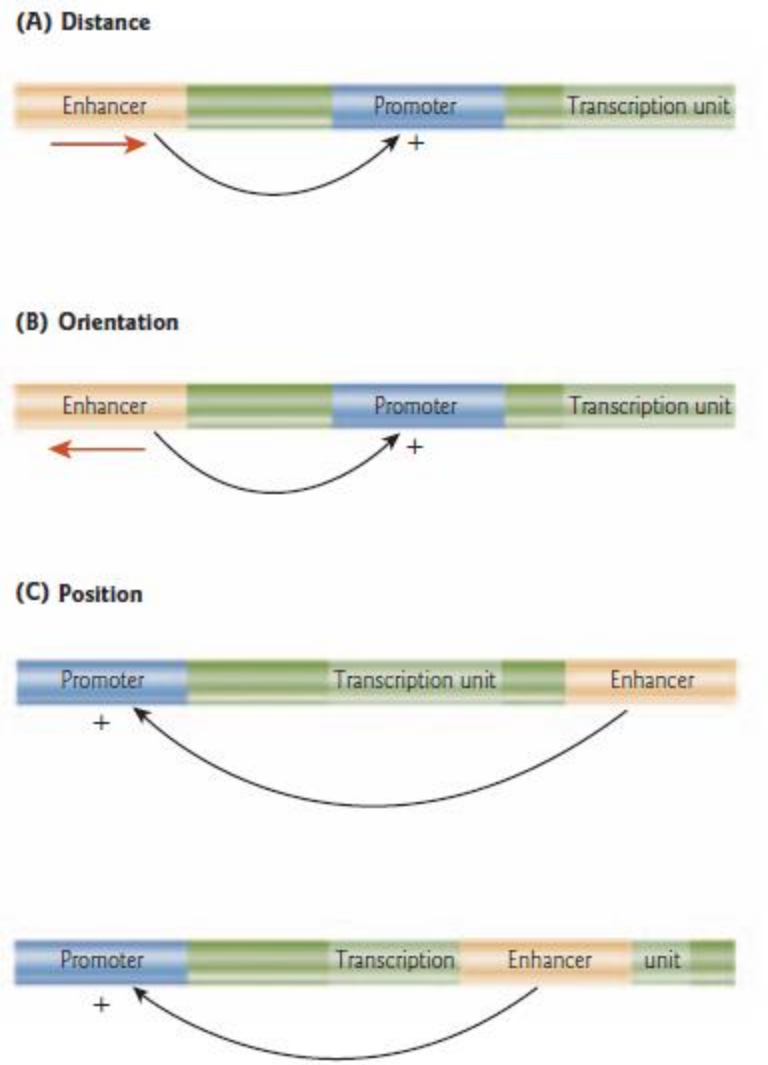


Fig3: Enhancer elements can activate a promoter at a distance (A) in either orientation (B) and upstream downstream or within a transcriptional unit (C).

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).

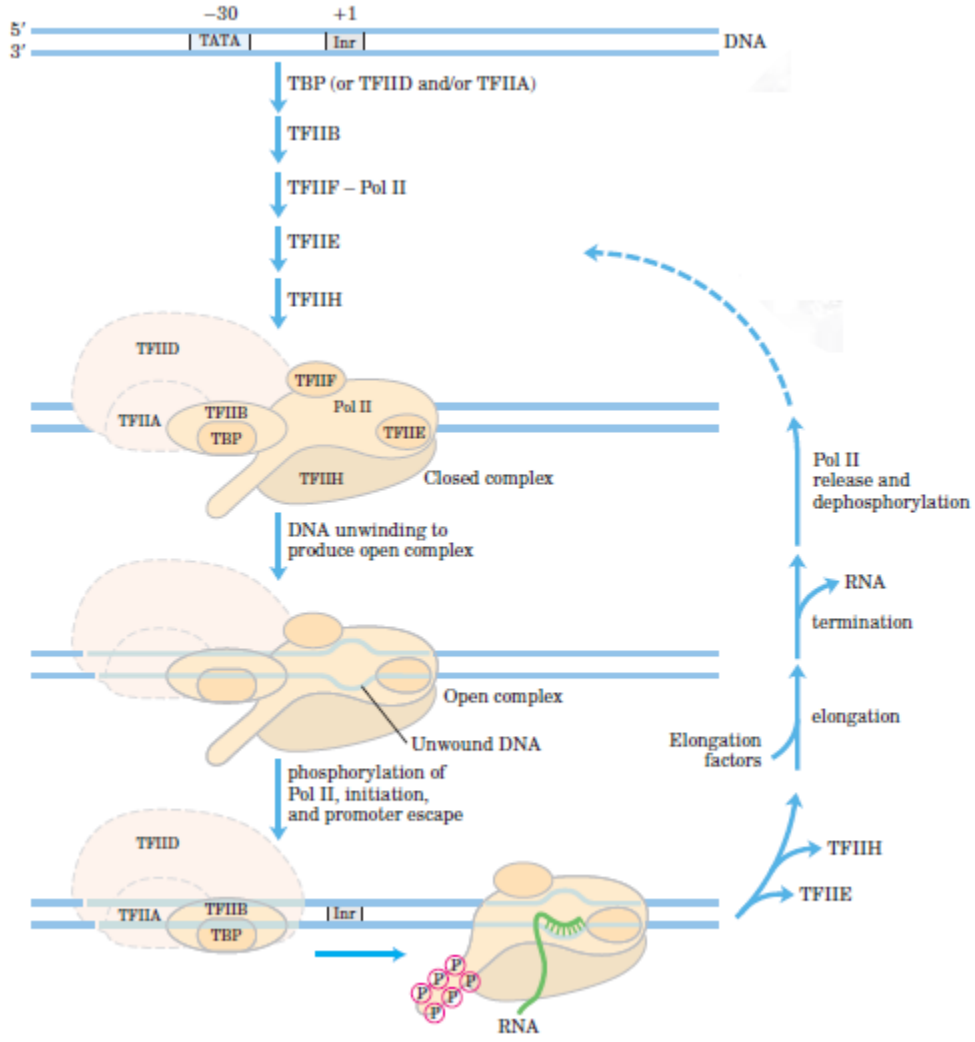


Figure 4- Transcription at RNA Pol II promoters

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).

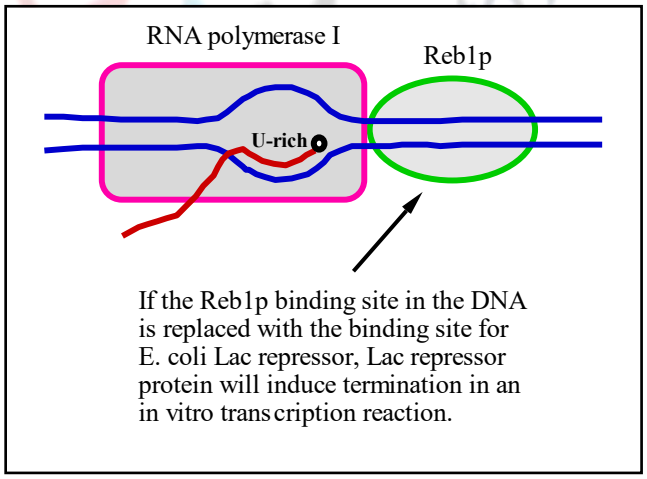


Figure 5 - Transcription termination by RNA polymerase I

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.

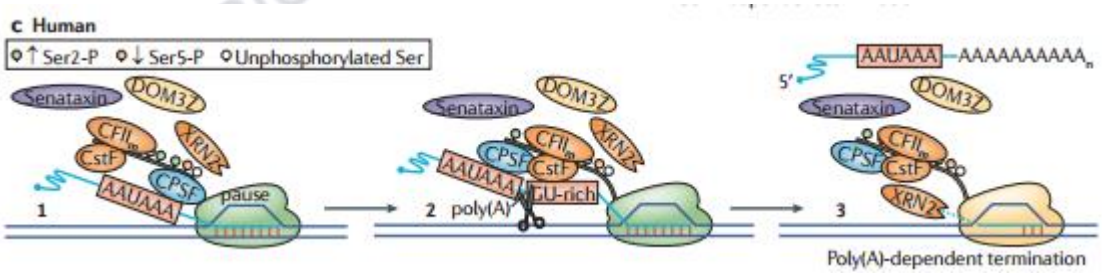


Figure 6: Poly(A)-dependent termination

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.

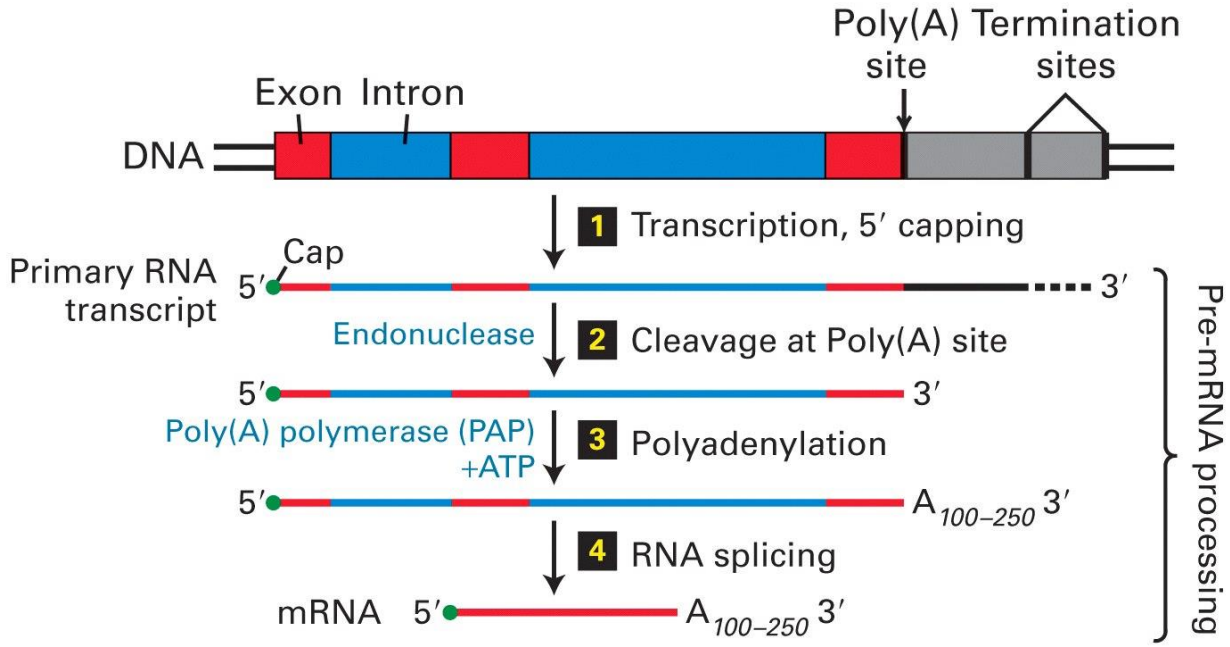


Fig 7. Mechanism of Polyadenylation

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.

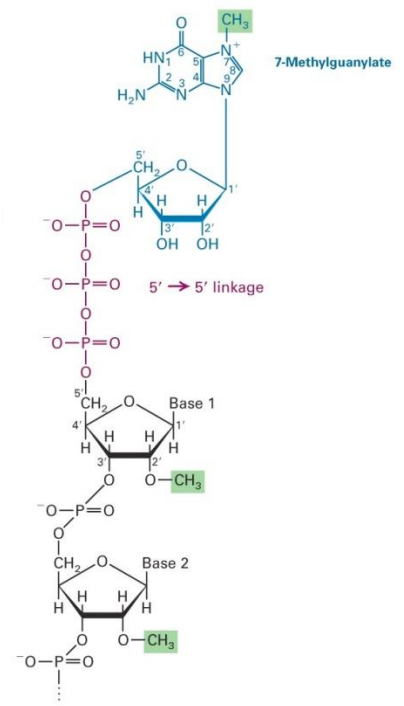


Fig 8.Structure of a CAP

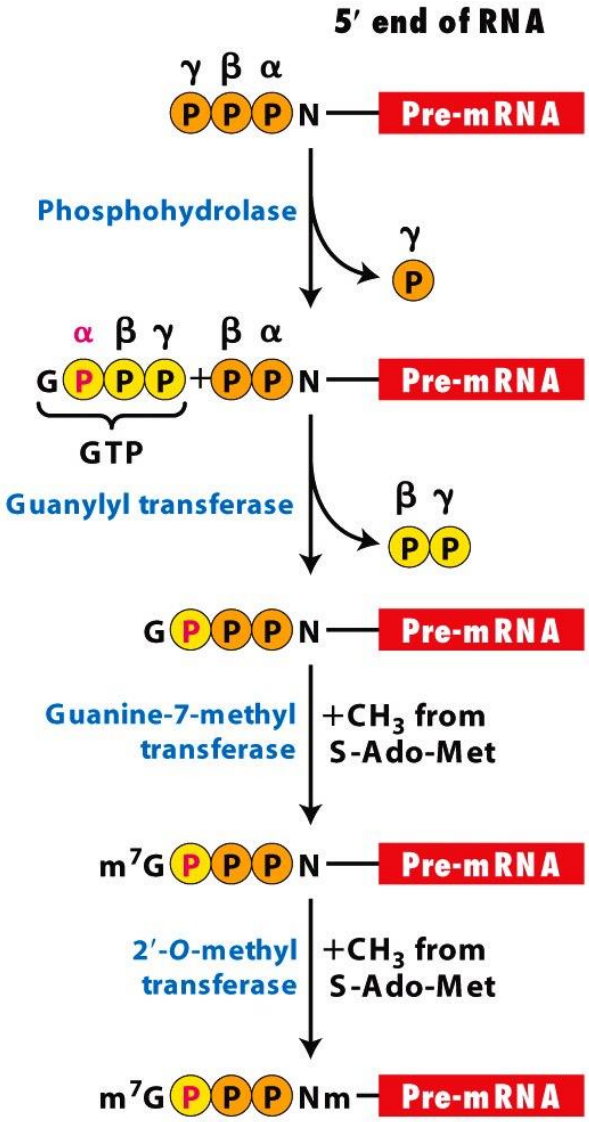


Fig 9. Mechanism of 5'Capping

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.

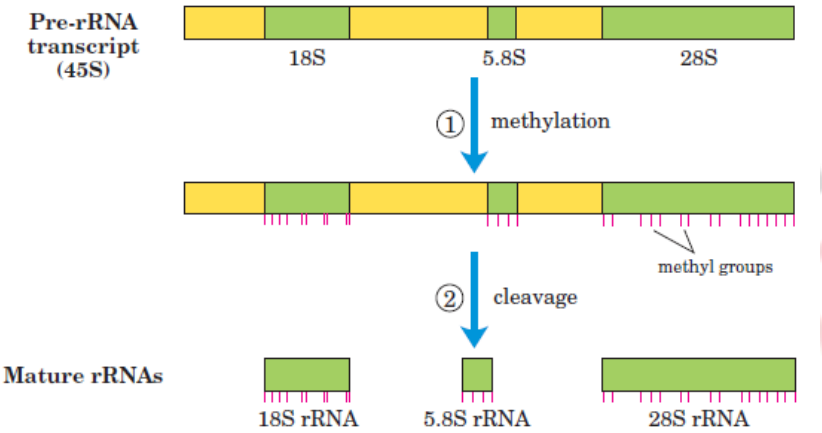


Fig 10. Pre-rRNA transcript processing

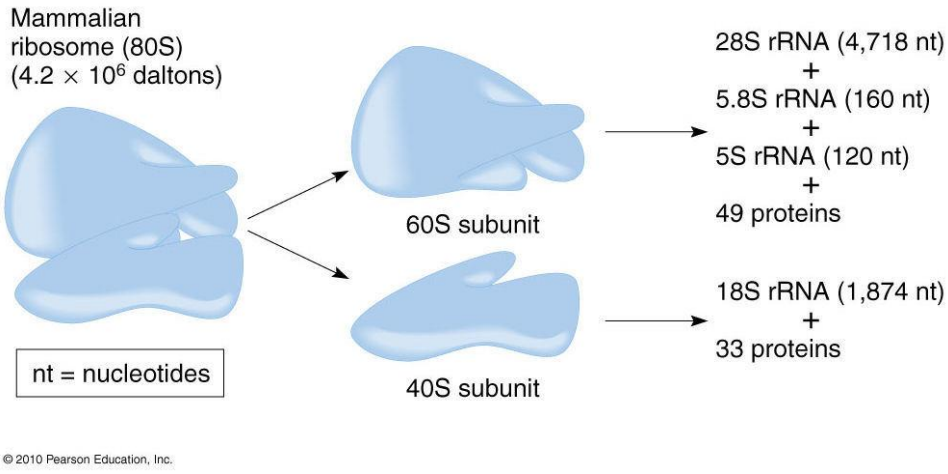


Fig 11.Eucayotic ribosome.

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.

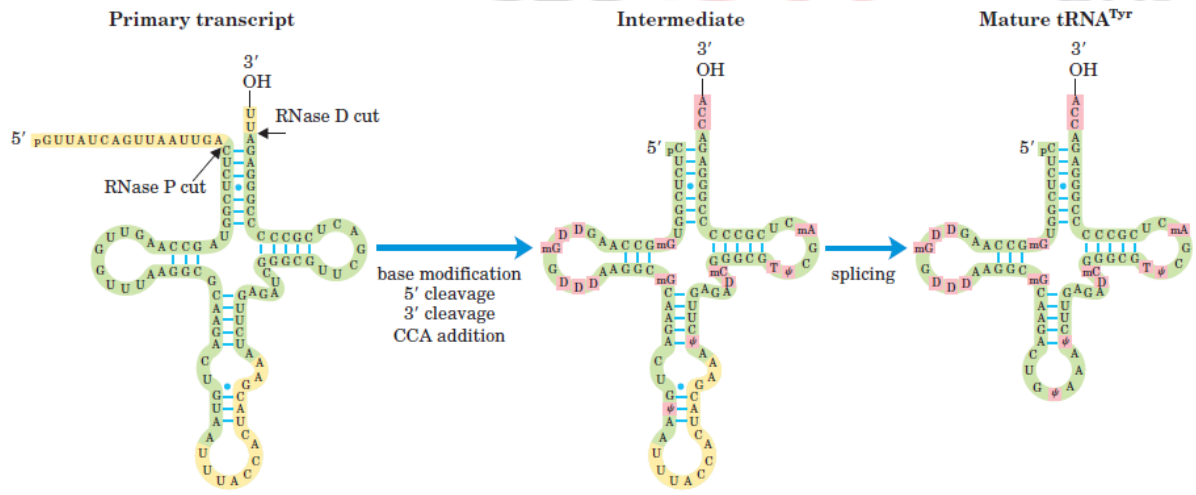


Figure 12 Transfer RNA processing

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).

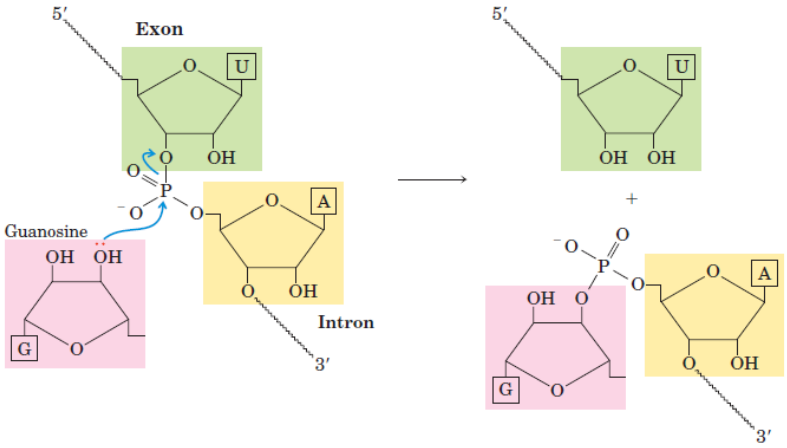


Fig 13. Transesterification reaction

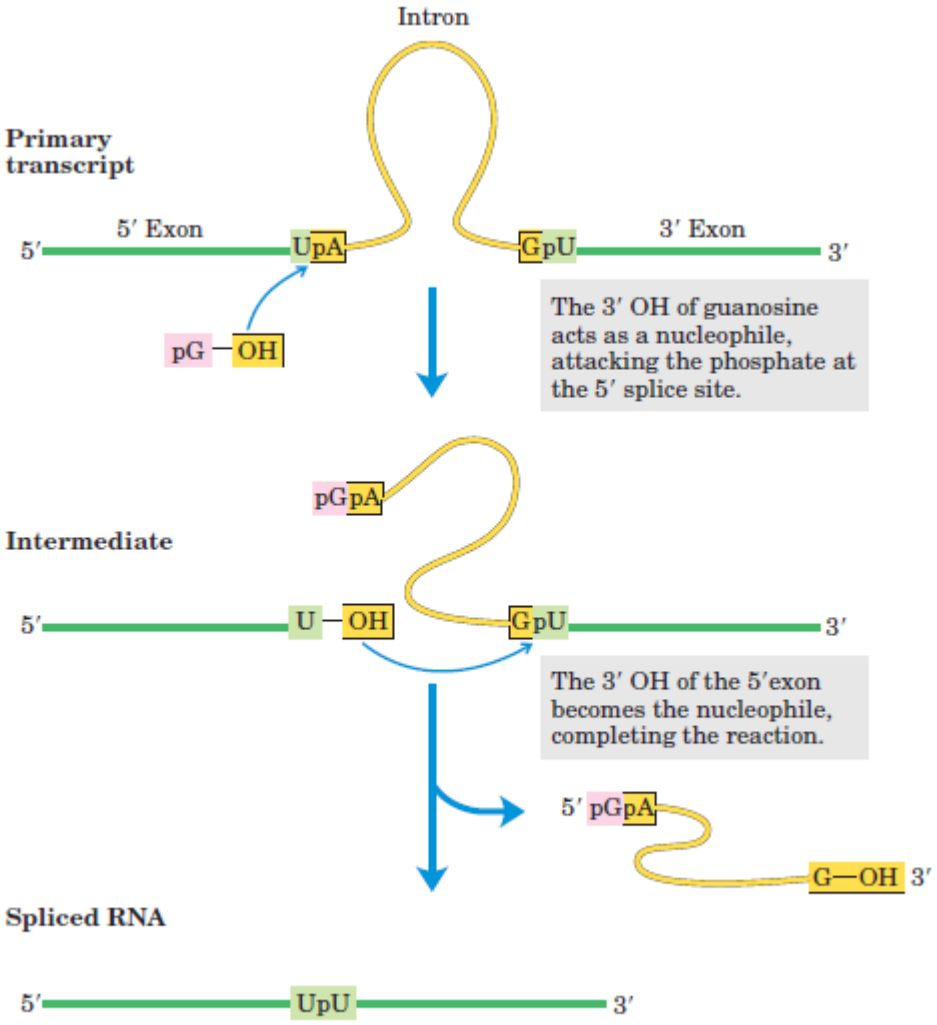


Figure 14-Group I splicing mechanism : Guanosine acts as a nucleophile in the first step. The excised intron sequence is eventually removed.

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.

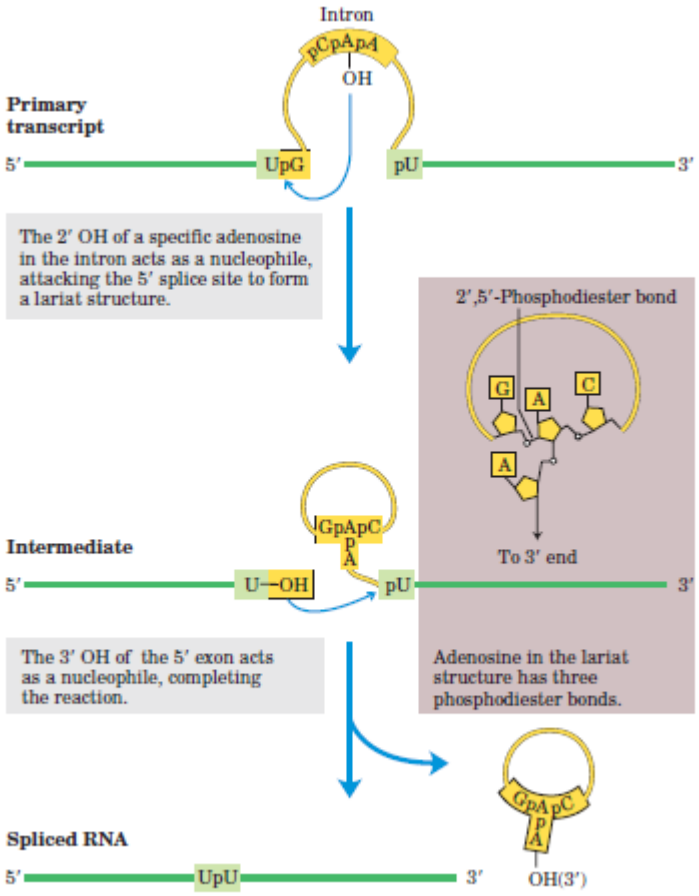


Figure 15 : Group II splicing mechanism

3.6.5.3 Spliceosomal introns:

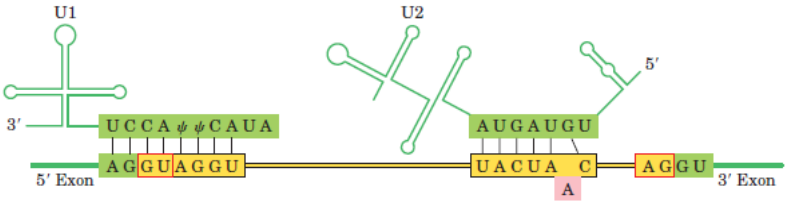
This class of introns is found in the eucaryotic nuclear premRNA. They are named so because their removal involves a large protein complex called spliceosome. It is made up of RNA and protein complexes together known as small nuclear ribonucleoproteins (snRNPs). Each snRNP consists of a small nuclear 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

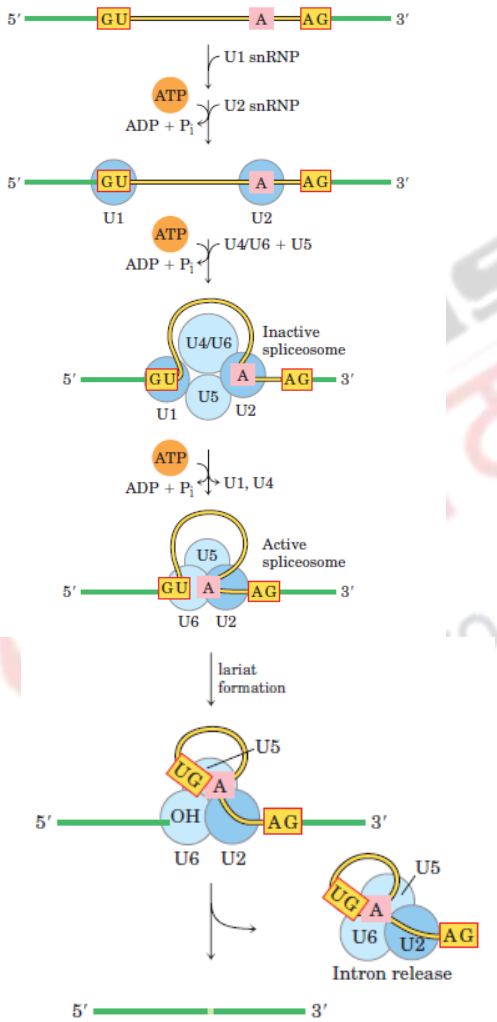
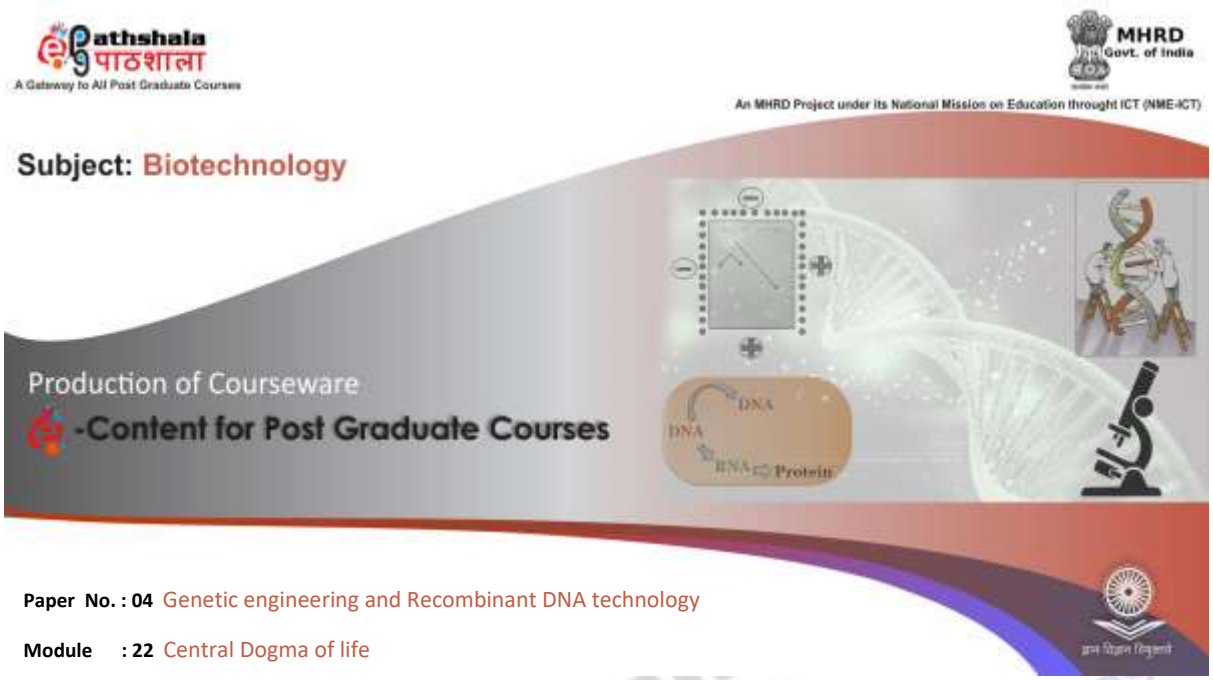


Fig 16. Spliceosomal introns A. snRNP pairing interaction during splicesome formation B.Mechanism of spliseosome formation

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



Paper No. : 04 Genetic engineering and Recombinant DNA technology

Module : 22 Central Dogma of life

Principal Investigator:	Dr Vibha Dhawan, Distinguished Fellow and Sr. Director The Energy and Resouurces Institute (TERI), New Delhi
Co-Principal Investigator:	Prof S K Jain, Professor, of Medical Biochemistry Jamia Hamdard University, New Delhi
Paper Coordinator:	Dr Mohan Chandra Joshi, Assistant Professor, Jamia Millia Islamia, New Delhi
Content Writer:	Dr. Pravir Kumar, Professor, Department of Biotechnology, Delhi Technological University
Content Reviwer:	Dr Rohini Muthuswami, Associate Professor, Jawaharlal Nehru University

Description of Module

Subject Name	Biotechnology
Paper Name	Genetic engineering and Recombinant DNA technology
Module Name/Title	Central Dogma of life
Module Id	22
Pre-requisites	
Objectives	
Keywords	

HISTORY:

The seminal work of Francis Crick in 1958, reveals the concept of central dogma. Initially, it has been hypothesized that any possible transfer between DNA, RNA and protein can be done, and formation is reversible. However, these postulations proved to be wrong, and later it was established DNA to RNA and RNA to protein formation is convincing, but RNA to DNA formation was still ambiguous. However, this theory has gained more popularity but without much experimental proof. Later, unidirectional flow of genetic information is validated. However, a reverse transcription mechanism has proven to be correct and established that indeed RNA can be reverse transcribed to form DNA (F. Crick, NATURE VOL. 227 AUGUST 8 1970).

INTRODUCTION:

In the molecular biology, central dogma is an important phenomenon which gives us information regarding the flow of genetic information in the cells where DNA is transcribed and produced messenger RNA (mRNA) which is ultimately translated and produced protein. This phenomenon also tells us how gene recognizes the sequence of mRNA molecules and thus specifies protein sequence. Since DNA is the storage of all vital information and required to perform all keys cellular function and therefore, cells have to preserve all the information intact in the DNA and copy it in the form of RNA. In short, DNA information is in the form of specific sequences of bases along the DNA strands. The DNA leads to specific traits by dictating the synthesis of proteins where, proteins are the links between genotype and phenotype. On contrary, DNA comprised of entire genetic information, which provides the shape and function of an organism. Proteins are important for the execution of genetic information in the term of expression with the help of genetic code of DNA through translation. Taken together, these three individual processes are regulating the transfer of genetic information and for its conversion from one polymer to another. These three functions are:

Replication (dsDNA to ssDNA): is a process where double-stranded DNA to duplicate an identical copy where this process perpetuates the genetic information.

Transcription (DNA to RNA): In this process a segment of DNA that comprised a gene which read and transcribed into a single-stranded sequence of RNA where RNA translocate from the nucleus to the cytoplasm for the further function. Both DNA and RNA are come under nucleic acids that use base pairs of nucleotides as a complementary sequence and with the help of enzymes, it is producing DNA to RNA and vice versa. In the process of transcription, a DNA sequence is first read by the RNA polymerase enzyme that produces an antiparallel complementary strand of RNA. For the study of gene expression, transcription is the first process, where uracil (U) of RNA is replaced by thymine (T) of DNA wherever it is required. Further, a DNA molecule is transcribed into an RNA strand is called transcript and sometime uracil has been used as regulatory RNA whereas others encode one or many proteins. Moreover, if a gene is transcribed which that encode a protein, messenger RNA (mRNA) is the resultant product of transcription, which is, then further used during the generation of protein by a process called translation.

Translation (RNA to protein): Translation is the process through which mRNA sequence is translated into amino acid sequence and joined together to form a polypeptide chain,

otherwise known as a protein. The synthesis of proteins is initiated by mRNA with the help of large complex ribosome RNA (rRNA). During the process of translation, ribosome recognizes a codon at a time from the RNA and then translates it into one amino acid with the help of transfer RNA (tRNA). The function of tRNA in the translation process is to translate the codon's sequence on the mRNA strand where tRNA is to transfer a free amino acid from the cytoplasm to a ribosome and attached to the growing polypeptide chain until they reach a stop codon on the mRNA.

CENTRAL DOGMA AND "OMICS"

The genesis of "omics" theory is derived from central dogma of biology. How to analyze the conserved and consensus sequence among the entire genome was a baffling issue. However, by the deciphering of genetic flow, it was clearly indicated that genomics is a promising to tool to analyze genome through gene flow, complexities in the genome and of course population structure. We can also use genomics approach to understand many anomalies, for instance, single and double gene disorder, mutational analysis, conservative genomics, functional and comparative genomics across the multiple species.

The peculiarity of genome is that base pair that code all genetic information of an individual. Further, after conversion of information from DNA to RNA, through transcription process, transcriptome produced, which carry all the information coding for a particular cell, for example, human kidney and heart have the similar genome but varied at transcriptome, and this is why kidney cells' express genes that are required in the kidney function whereas cardiac cells are expressing different genes to perform heart function properly. Transcriptomics are tissue-specific study based on the expression of selective or tissue-specific genes. However, transcriptomics can be regulated based on the physiological and pharmacological stimuli and thus expression of genes are also modulated.

The powerful approach of system biology is associated with other omics levels, which is "proteomics". To understand the gene based on its protein expression. Proteins are the building-block unit for the cells and very important for the execution of function of a cell. In the end, it is also necessary to understand the end product of genetic flow, which is enabling us to deduce many interpretation and functionality of a gene, i.e. metabolomics. In this step we can validate that what has happened during the genetic flow and also predict what could be the flow of information of a particular gene.

ONE GENE-ONE ENZYME HYPOTHESIS

In order to understand the function of a gene at protein level, in 1941, Beadle proposed a hypothesis which is called one gene one enzyme hypothesis. This theory proposed that each gene produces one enzyme and that enzyme which is produced at first step can trigger the downstream cascade of a metabolic pathway.

NUCLEIC ACID:

Nucleic acids are considered to be the most essential unit of life. These molecules enable an organism to transfer genetic information from parent to offspring, and the process is continued. Nucleic acids are of two type; deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA possessing all the hereditary information that passed from one generation to the next. Nucleotide consists of nitrogenous base, five carbon sugar and a phosphate group. The composition of nucleic acids is nucleotide monomer which is linked together to form a polynucleotide chain and joint together through covalent bound between the phosphate of one and another one with sugar. This linkage is called phosphodiester bond that forms a sugar-phosphate backbone in the nucleic acid. The chemical reaction behind this

association is dehydration synthesis where, nitrogenous bases are linked together, and a water molecule is released.

Deoxyribonucleic acid (DNA): DNA is an important building-block component to perform cellular function and during cell cycle division, DNA is copied and genetic information passed from one generation to the next generation. In the cell, DNA is present inside the nucleus and organized on chromosome. Structurally DNA is double stranded and present in a double helix. As discussed, DNA is comprised of a phosphate-deoxyribose sugar backbone, purine and pyrimidine bases viz. Adenine (A), Guanine (G), Cytosine (C) and Thymine (T). These bases are joined together with hydrogen bond, which is A-T and C-G.

Ribonucleic acid (RNA): An important function of RNA is to synthesize protein, which is essential to maintain the genetic flow and thus regulating cellular homeostasis. Broadly, RNA is subdivided into three types, which are mRNA, tRNA and rRNA and all three components are involved in the protein synthesis. Usually, RNA is present in the form of single strand. RNA is consisted of a phosphate ribose sugar backbone and bases Adenine (A), Guanine (G), Cytosine (C) and Uracil (U). During the process of transcription, guanine paired with cytosine and adenine pairs with uracil.

MESSENGER RNA (mRNA):

By definition, mRNA is a large family of RNA molecules, which transfer genetic information from DNA to the ribosome, where they are encoding amino acid sequence, which is essential for gene expression. In the total RNA, messenger RNA (mRNA) is a subtype, and it carries a part of DNA code for the cell processing. During the transcription process, mRNA is created and a single-stranded DNA is decoded by RNA polymerase and thus mRNA is synthesized. Structurally, nucleotide strand of mRNA is known as ribonucleic acid, which is single stranded. During the transcription process, primary transcripts are formed with the help of RNA polymerase, which is known as pre-mRNA whereas, mature mRNA is translated and a chain of amino acids are formed.

TRANSFER RNA (tRNA):

Initially, transfer RNA was known as sRNA or soluble RNA, an adaptor molecule, which is consisted of 76 to 90 nucleotides in length, and plays as the physical link between the mRNA and the amino acid sequence of proteins. The transfer RNA is small RNA that has a very specific secondary and tertiary structure such that it can bind an amino acid at one end, and mRNA at the other end. It acts as an adaptor to carry the amino acid elements of a protein to the appropriate place as coded for by the mRNA. During the translation process, tRNAs function at specific sites in the ribosome. Further, the yield of protein which is a chain of amino acid is recognized by three nucleotide sequence, which is called a codon on mRNA. Each codon is decoding one amino acid and therefore, each codon is specified by a specific tRNA. Structurally, tRNA has a typical folded architecture with three hairpin loops, which looks like a three leafed clove. Out of these three hairpin loops, one loop has specific sequence that represents anticodon and this is decoding the mRNA codon. Furthermore, each tRNA molecule has its own amino acid, which is attached to its end. Once the tRNA is recognized and attached to its specific codon in the ribosome, the t-RNA transfers a specific amino acid to the end of the growing chain of amino acid, and this process is to be continued until the stop codon is recognized by the tRNA and a full-length protein is translated.

RIBOSOMAL RNA (rRNA):

This molecule form part of the protein synthesizing organelle called as a ribosome which is translocated to the cytoplasm and enable mRNA to translate mRNA into the protein. The place where rRNA is synthesized is called the nucleolus (inside nucleus), that appears to be a dense area, and it consists of genes, which encode rRNA. However, encoded rRNA is differed in the size and in the nucleolus, large and small subunits combined with ribosomal proteins to form a separate large and small subunit of the ribosome. Initially, ribosomal proteins are synthesized in the cytoplasm but translocated to the nucleus for assembly. Once the subunits are formed it again returned to the cytoplasm for final assembly. Another hallmark feature of rRNA is to form an extensive secondary structure and involve in a crucial role in identifying conserved sequence of mRNA and tRNA. Eukaryotic cell has a well-defined nucleus and 50 to 5000 sets of rRNA and approximately 10 million ribosomes can be present in an individual cell. On contrary to that, in prokaryotes which are devoid of the nucleus usually have fewer sets of rRNA gene and ribosome per cell.

SUMMARY:

To conclude this chapter, the first step of genetic flow is transcription which occurs in the nucleus where one strand acts as a template for the synthesis of a complementary RNA strand, whereas the end product is mRNA that carries the message from the nucleus to the ribosome. On contrary, Translation, which the last step of genetic flow occurs in the cytoplasm with the help of ribosome and the end product of translation, is protein. Furthermore, the concept of gene is classically defined based on the Mendellian inheritance.