1. Details of Module and its structure

Module Detail	
Subject Name	Biology
Course Name	Biology 03 (Class XII, Semester - 1)
Module Name/Title	Transcription and Genetic code – Part 3
Module Id	lebo_10603
Pre-requisites	Knowledge about DNA structure and function
Objectives	After going through this lesson, the learners will be able to understand the following: • Transcription • Transcription Unit and Gene • Transcription in prokaryotes • Transcription in eukaryotes • Differences between Transcription in Prokaryotes (Bacteria) and Eukaryotes: • Genetic Code • Salient features of genetic code: • Mutations and Genetic Code
Keywords	Transcription, Transcription in Prokaryotes, Transcription in Eukaryotes, genetic code, mutation

2. Development Team

Role	Name	Affiliation
National MOOC Coordinator (NMC)	Prof. Amarendra P. Behera	CIET, NCERT, New Delhi
Program Coordinator	Dr. Mohd. Mamur Ali	CIET, NCERT, New Delhi
Course Coordinator (CC) / PI	Dr. Chong V Shimray	DESM, NCERT, New Delhi
Course Co-Coordinator / Co-PI	Dr. Yash Paul Sharma	CIET, NCERT, New Delhi
Subject Matter Expert (SME)	Mr. Reetesh Kumar	RPVV, Civil lines, New Delhi
Review Team	Dr. Madhumita Banerjee	Ramjas College, University of Delhi.

Table of Contents :

1. Transcription

The process by which the synthesis of RNA occurs by using a DNA template is called as transcription. In this process the genetic information stored in the DNA is copied into a RNA strand on the basis of the complementarity which exists between the nitrogenous bases. The only difference regarding the pairing is that Adenine in the DNA template strand pairs with Uracil in the mRNA instead of Thymine during transcription.

Some other differences which make the process of transcription different from the process of replication are that during transcription.

i. Only a segment of DNA is transcribed instead of the complete strand of DNA which gets copied during replication.

ii. Only one strand of DNA serves as a template for transcription of RNA at any given time. The strand of DNA which gets transcribed is called as the template strand or the non-coding strand. The other strand of DNA which does not get transcribed is called the non template or the coding strand because its sequence will be the same as that of the new RNA molecule except for the Uracil which will come in place of Thymine present on the coding strand.

It is also interesting to know that in most organisms, the strand of DNA that serves as the template for one gene may be the non-template strand (coding strand) for other genes within the same chromosome. Both the strands of DNA are not transcribed into RNA due to the following reasons-

(i) The first reason is that if both the DNA strands act a template then two RNA molecules with different sequence of bases would be formed. If both of them code for proteins then the sequence of amino acids in the proteins would be different. This will mean that one segment of

the DNA would be code for two different proteins which would complicate the machinery of the transfer of genetic information.

(ii) The other reason is that the two RNA molecules formed from the transcription of both the DNA strands would be complementary to each other. They will form double stranded RNA whereas a single stranded RNA is needed for getting translated into proteins. This would hamper the translation of RNA into proteins and the whole process of transcription would become of no use.

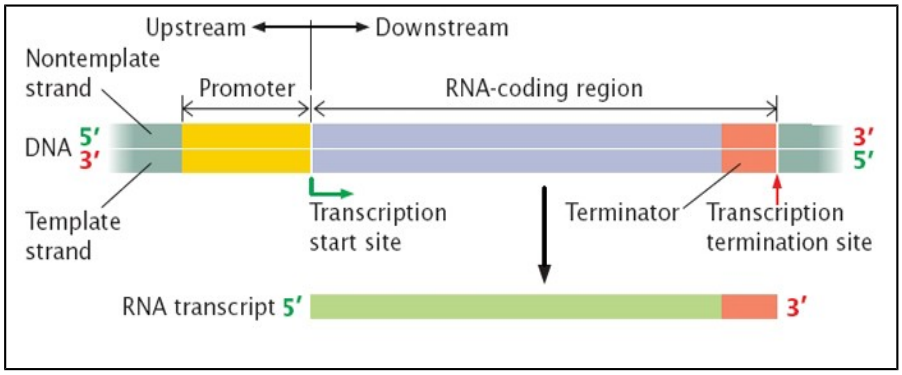
2. Transcription Unit and Gene

The specific segment of DNA which gets transcribed into RNA is called as the transcription unit. A transcription unit consists of three regions

(i) A Promoter

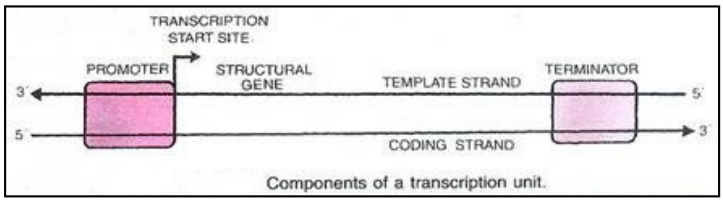
(ii) The Structural gene

(iii) A Terminator



Eukaryotes also require an enhancer apart from the promoter. The enzyme DNA-dependent RNA polymerase can catalyze the polymerization of RNA in only 5'-3' direction. So, the strand that has the polarity 3′-5′ acts as a template, and is also referred to as template strand, master strand, antisense, or (-) strand. The other strand which has the polarity (5′-3′) and the sequence same as RNA (except thymine at the place of uracil), is referred to as the coding strand, sense or plus (+) strand.

All the reference point while defining a transcription unit is made with the coding strand. It is a convention that location of promoter is considered as the 5' end of the coding strand and that of the terminator as the 3' end of the coding strand. The promoter and the terminator regions flank the structural gene. Hence, the promoter is located upstream of structural gene while the terminator region is present at the 3′ end of the coding strand, downstream of structural gene (which corresponds to the 5′ end of the template strand). By switching the position of promoter with the terminator, the definition of coding and template strands could be reversed.



Components of a transcription unit.

The RNA polymerase binds the promoter which has different parts where the various transcription factors can get attached.

Usually, the promoter has an AT rich region called TATA box which is a DNA sequence that is a type of promoter sequence which specifies to other molecules the point where the transcription begins.

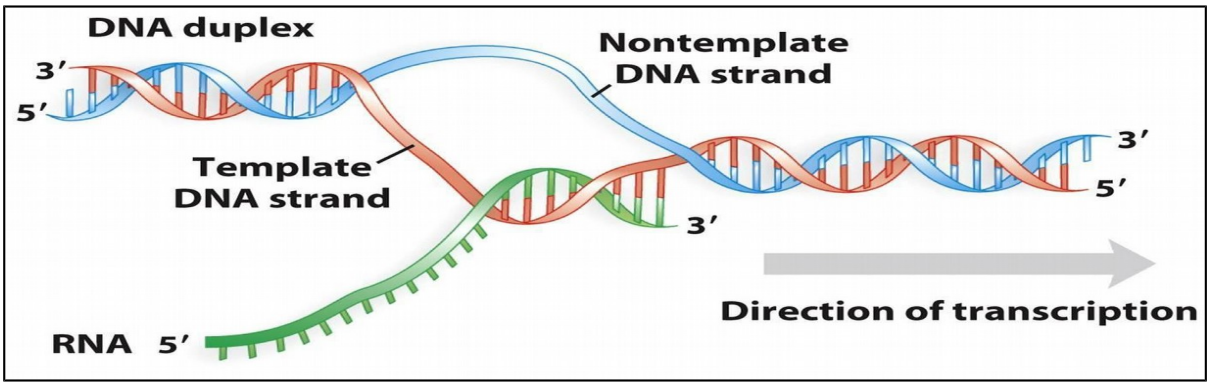
Gene: A gene is defined as the functional unit of inheritance.

Cistron: A segment of DNA which codes for a polypeptide. Depending on this the structural gene is referred to as either monocistronic or polycistronic.

Monocistronic: Found in Eukaryotes as the structural genes have interrupted coding sequences (means genes are called split). The coding sequences called Exons appear in the processed mRNA whereas the non-coding sequences called Introns are removed by process of splicing. Introns do not appear in processed RNA.

Polycistronic: Structural gene found in Prokaryotes is called as polycistronic.

Types of RNA: There are three major types of RNAs: messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). All the three RNAs are needed to synthesize a protein in a cell. The mRNA acts as a template, tRNA brings amino acids to mRNA and reads the genetic code whereas the rRNAs play structural and catalytic role during the process of translation.



3. Transcription in prokaryotes

In prokaryotes, no proper nucleus is there due to which their DNA lies in the cytoplasm and transcription occurs in cytoplasm of the cell. The process of transcription requires an enzyme called as DNA dependent RNA polymerase. The process of transcription in three steps namely initiation, elongation and termination.

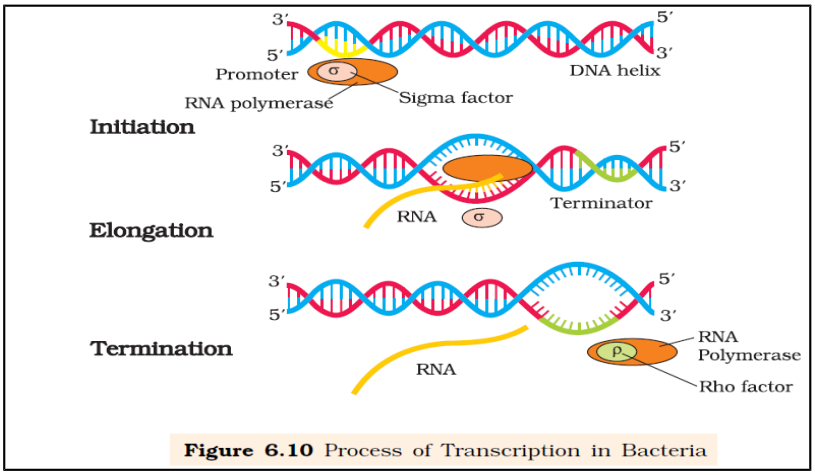
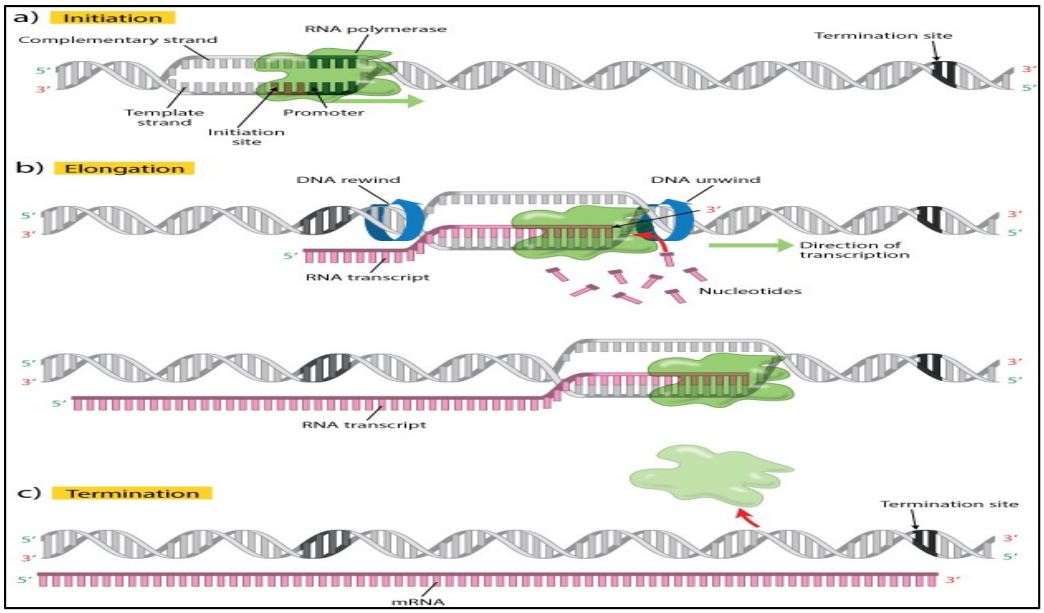


Figure 6.10 Process of Transcription in Bacteria

Initiation: RNA polymerase binds to promoter and initiates transcription. RNA polymerase enzyme transiently associates

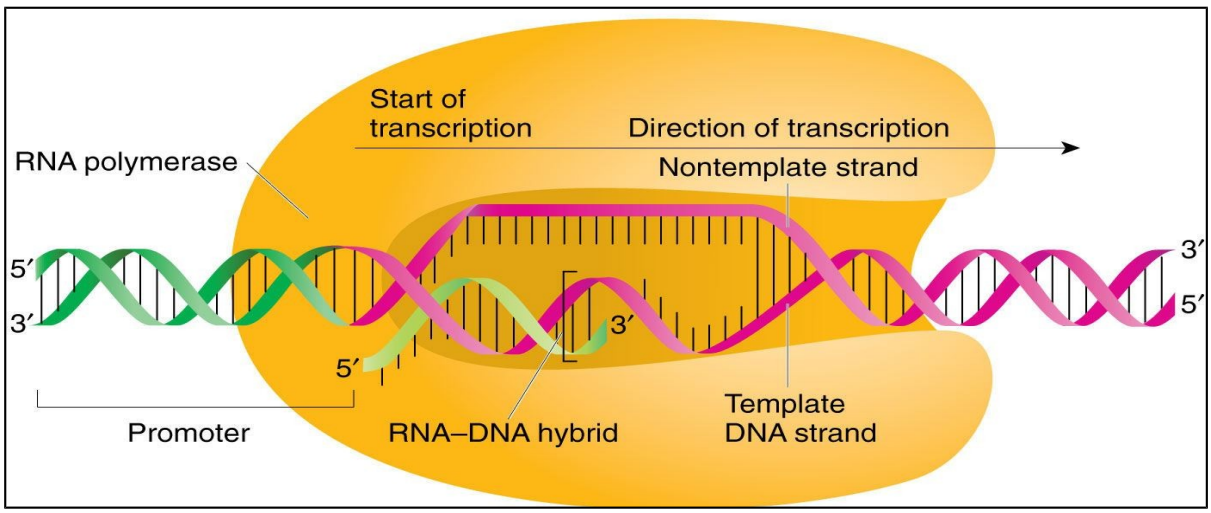


with the initiation-factor called the sigma factor (σ) to initiate transcription.

Elongation: RNA polymerase uses nucleoside triphosphates as substrate and polymerizes RNA from DNA in a template dependent fashion following the rule of complementarity. Only a short stretch of RNA remains bound to the enzyme. The RNA polymerase is only capable of catalyzing the process of elongation.

Termination: Once the RNA polymerase reaches the terminator region it associates with a termination-factor called as the rho-factor (ρ) to terminate the process of transcription. The nascent RNA as well as the RNA polymerase falls off which results in termination of transcription.

4. Transcription in eukaryotes



Transcription occurs inside the nucleus in the eukaryotes and then the products of transcription move out into cytoplasm through the nuclear pores for undergoing translation. Though the ba-

sic features of transcription of RNA are shared between prokaryotes and eukaryotes, the transcription in eukaryotes is more complex than that in prokaryotes.

5. Differences between Transcription in Prokaryotes (Bacteria) and Eukaryotes:

(A) *Difference in RNA polymerase:*

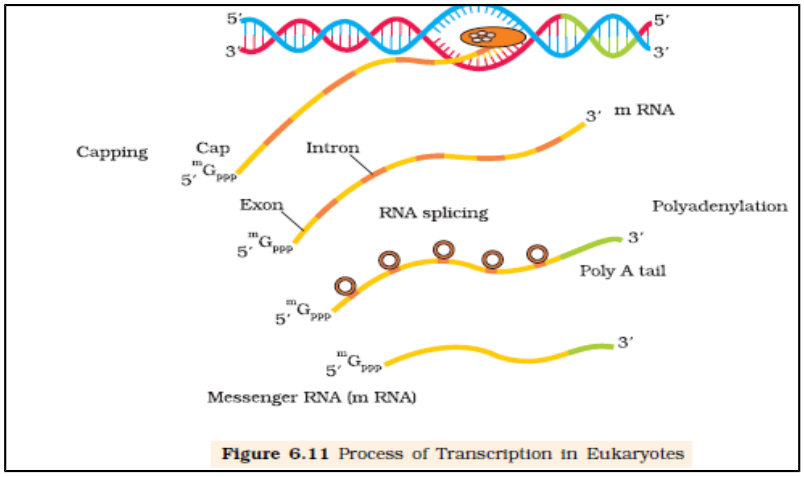


Figure 6.11 Process of Transcription in Eukaryotes

A single RNA polymerase helps in the transcription of all types of RNA in bacteria but eukaryotes have three different RNA polymerases. RNA polymerase-I helps in transcription of rRNAs (28S, 18S, and 5.8S). RNA polymerase-III catalyzes the transcription of tRNA, 5srRNA, and the snRNAs (small nuclear RNAs). The RNA polymerase-II carries out the transcription of the precursor of mRNA referred to as the heterogeneous nuclear RNA (hnRNA) hnRNA is processed(spliced) in nucleus, the introns are removed and the exons joined and sometimes reshuffled (alternate splicing). The mRNA transcriptmoves out of the nucleus into the cytoplasm.Eukaryotic RNA polymerases need the help of a set of proteins called the basic transcription factors as they cannot initiate transcription by themselves. A number of functions are provided by the basic transcription factors like binding to the promoter regions of gene, attracting the appropriate RNA polymerase to the initiation site and unwind the DNA double helix to allow access to the incoming ribonucleotides required for the polymerization of RNA.

6. Genetic Code

The process of replication and transcription are easy to conceptualise on the basis of comple mentarity. The challenge before the scientists was to explain how the sequence of nucleotide pairs dictates the sequence of amino acids in the protein because neither does any complementarity exist between nucleotides and amino acids, nor could any be drawn theoretically.

It was proposed that a genetic code could direct the sequence of amino acids during synthesis of proteins because evidences were there which showed that change in sequ ence of nucleotides of nucleic acid were responsible for change in amino acids in proteins. The involvement of scientists from several disciplines – physicists, organic chemists, biochemists and geneticists was required to understand and decipher the genetic code. The set of rules by

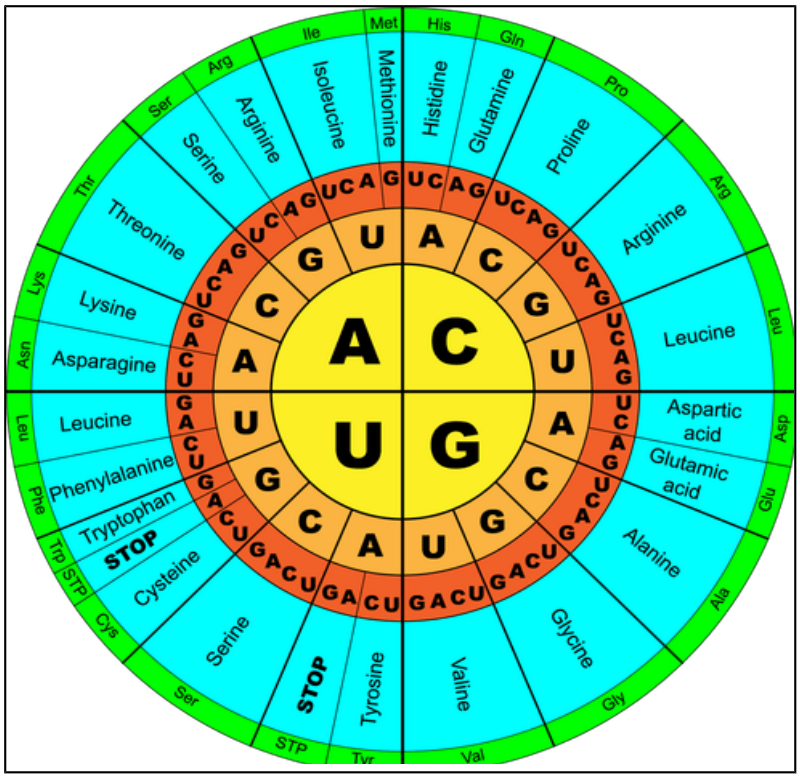
which information encoded by sequence of nucleotides present in DNA or mRNA is translated into proteins by living cells is called as genetic code.

It is observed that only one of four different bases, A, U, G, or C, can be found at each position if an mRNA molecule is read from a particular end,. Thus, if the code were one letter long, only four amino acids would be possible. This cannot be the genetic code, because cellular proteins are composed of 20 types of amino acids. If the code were two letters long, then 4^2 = 16 amino acids would be possible; for example, AU, CU, or CC. Even this would not be sufficient in order to code for the 20 amino acids.

In 1953, George Gamow made a bold proposition that in order to code for all the 20 amino acids, the code should be made up of three nucleotides (i.e., combination of three bases). He suggested that a permutation combination of 4^3 (i.e., 4 x 4 x 4), for example, UUU, ACG, or UCC would generate 64 codons which are more than enough to describe the 20 amino acids. Hence, it was concluded that the code must consist of at least three nucleotide pairs. These triplet three nitrogenous bases is called a codon.

However, if all codons are “triplets,” then we have an excess of possible codons over the 20 codons needed to code for the 20 common amino acids.

The deciphering of genetic code which involved determining the amino acid coded by each triplet was possible because of the efforts of the scientists like Har Gobind Khorana, Marshall Nirenberg, and Severo Ochoa etc. Their work also proved that codon consists of a triplet of bases.



Har Gobind Khorana devised a method to synthesise RNA molecules with defined combinations of bases (homopolymers e.g. UUU UUUUUUUUU which has one type of codon UUU and copolymers e.g. $(UC)^2$ UCUCUC UCUCUC UCU two types of codons UCU CUC). If the nucleotides of RNA are mixed with Severo Ochoa enzyme (polynucleotide phosphorylase), the nucleotides are incorporated at random to form single-stranded RNA without the

involvement of DNA strand. The ability to synthesize mRNA in a cell free systemmade it possible to create specific mRNA sequences and then see which amino acids they would specify. Hence, the Severo Ochoa enzyme (polynucleotide phosphorylase) was helpful in enzymatic synthesis of RNA with defined sequences in a template independent manner. In 1961, Marshall Nirenberg and Heinrich J. Matthaei used a cell-free system to translate a RNA sequence consisting of poly-uracil- UUU and discovered that the polypeptide synthesized consisted of only the amino acid phenylalanine. So, they deduced that the codon UUU specified the amino acid phenylalanine. The experiments in Severo Ochoa's laboratory demonstrated that the poly-adenine- AAA containing that the RNA sequence coded for the amino acid lysine and poly-cytosine-CCC containing that the RNA sequence coded for the amino acid proline. The rest of the code was identified by the experimental works of Har Gobind Khorana.

7. Salient features of genetic code

i. The codon consists of a triplet of bases. 61 codons code for amino acids whereas the remaining 3 codons (called stop codons) do not code for any amino acids.
ii. The code is said to be unambiguous and specific because one codon codes for only one amino acid.
iii. The code is said to be degenerate because some amino acids are coded by more than one codon. For example leucine (6 codons), valine(4 codons) etc.
iv. The codon is read in mRNAin a contiguous fashion as it is non-overlapping and has no punctuations.
v. The code is nearly universal as codons code for the same amino acids in all organisms, from bacteria to human. For example, UUU would code for Phenylalanine (phe) in all organisms. Some mitochondrial codons and some protozoan are exceptions to this rule of universality.
vi. AUG performs dual functions because it codes for Methionine (met) and also act as the initiator codon during the process of translation.

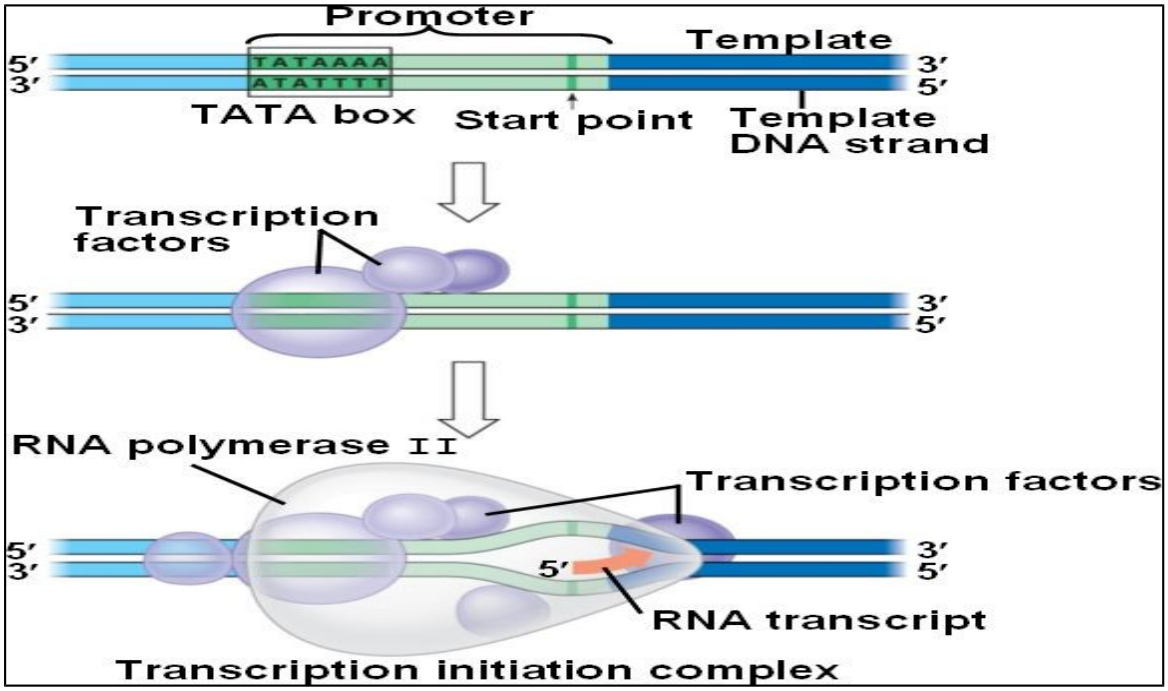
8. Mutations and Genetic Code

Mutation: A change in the DNA sequence that makes up a gene.

Point Mutation: A point mutation is caused due to a change in a single base pair. For example, a change of single base pair in the gene which codes for the beta globin chain of haemoglobin results in the substitution of amino acid residue glutamic acid by valine and results in a disease called as **sickle cell anaemia**.

Frame shift Mutation: A mutation caused by insertion or deletion of nucleotides in the sequence of nucleotides present in a DNA is called as the frame shift mutation.

Genetic basis of Proof that Codon is a triplet: It has been observed that during frame shift mutation, the reading frame gets changed from the point of insertion or deletion of one or two bases in the sequence of nucleotides of DNA. However, insertion or deletion of three or its multiple bases insert or delete one or multiple codon which results in either insertion or deletion of one or multiple amino acids due to which the reading frame remains unaltered from that point onwards. Hence, the frames shift mutations which involve insertion or deletion of three or its multiple bases forms the proof that codon is a triplet which is non-overlapping and read in a contiguous manner.

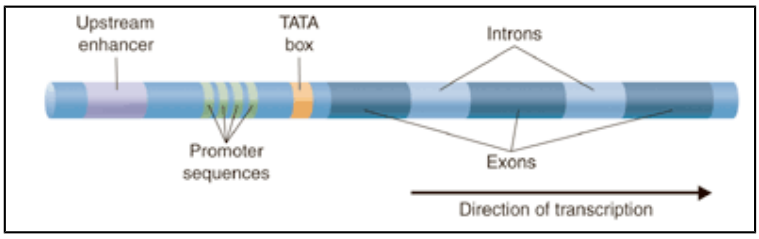


Difference in Processing of RNA:

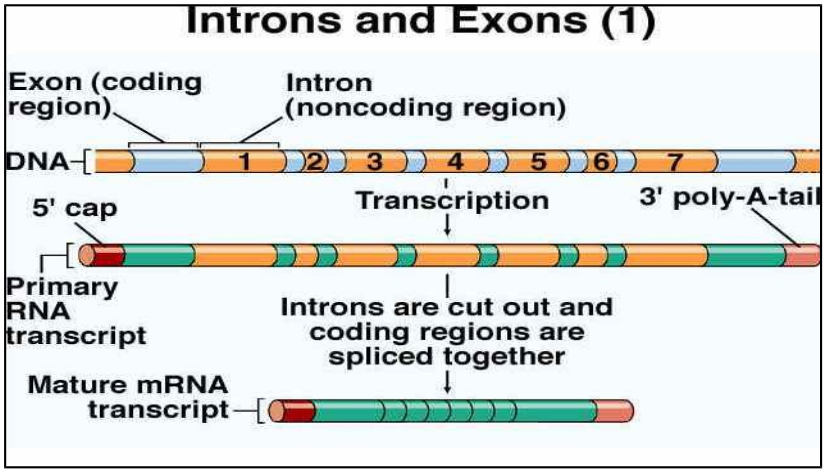
In bacteria, many times the translation can sometimes begin much before the mRNA is fully transcribed. This is because the mRNA does not require any processing to become active. Also, there is no separation of cytosol and nucleus in bacteria, so the process of transcription and translation take place in the cytoplasm of the cell and can occur simultaneously in the bacteria.

In eukaryotes the primary transcript of RNA has Introns and Exons. Most eukaryotic genes are split into segments. The stretches of DNA, which get transcribed into RNA but do not get translated into protein, are called introns. Those stretches of DNA which code for amino acids in the protein are called exons. The split-gene arrangement and the presence of introns indicate an ancient feature of the genome whereas the process of splicing reminds about the dominance of RNA-world. The introns are removed by the process of splicing and the exons are joined to

get processed mRNA from the unprocessed hnRNA.



During its processing, the hnRNA undergoes two additional process called as capping and tailing. An unusual nucleotide, methyl guanosine triphosphate is added to the 5′-end of hnRNA during the process of capping. Adenylate residues (approximately 200-300) are added at 3′-end in a template independent manner during the process of tailing.



After hnRNA gets fully processed it is called as mRNA which gets transported out of the nucleus for translation.