UNIT 6

MOLECULAR BIOLOGY

Karry Mullis

Modern biology can be divided into two epochs, the one before the invention of the Polymerase Chain Reaction (PCR) procedure and the one after it. This method is now routinely used in all molecular biology labs and also for medical diagnosis and forensic investigation. Karry Mullis is an American biochemist who was responsible for this invention and was duly awarded Nobel Prize in chemistry in 1993.

QO67M

Learning Objectives

After studying this unit the students will be able to

- Explain the central dogma of Molecular Biology.
- Explain the replication process.
- Describe the mechanism of RNA synthesis.
- Explain the structure of activated tRNA.
- Describe the mechanism of protein synthesis.

INTRODUCTION

Molecular biology is a discipline of biology which aims to understand the basic cellular functions such as growth, division, specialization, movement, interaction etc... in terms of the cell's macro molecular components. The term molecular biology was coined in 1938 by Warren Weaver, who was the director of natural sciences department at Rokefeller foundation. Though DNA was identified to be the genetic material as early as 1869, the exact nature of its function was understood only after the discovery of the molecular structure of DNA by J.D Watson and F.H.C Crick in 1952. The concept of base pairing introduced by them directly lead to the understanding DNA replication, RNA synthesis etc. Following these discoveries, the new field of molecular biology progressed rapidly allowing scientists to map out the fundamental biological principles behind macromolecular synthesis and function inside the cell. Today molecular biology has evolved into a very important field of biology with important applications like disease diagnosis and therapy, industrial applications like biotechnology products and to improve agricultural yields.

6.1 THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

The central dogma of molecular biology tries to explain how the genetic information contained in the DNA is conveyed or expressed in the biological system. This dogma was first stated by Crick in 1958 and later revised by Watson in 1965. This describes the detailed residue by residue transfer of the genetic information from DNA to Protein. This dogma is the frame work for understanding the information transfer between the important biological macromolecules: the DNA, RNA and proteins. When cells divide, a new copy of DNA has to be synthesised from the existing DNA. This process where a new DNA molecule is synthesised with the existing DNA as a template is known as replication. The information in the DNA can be copied to RNA by a process called transcription. RNA molecules can act as templates to produce protein molecules by a process called translation.

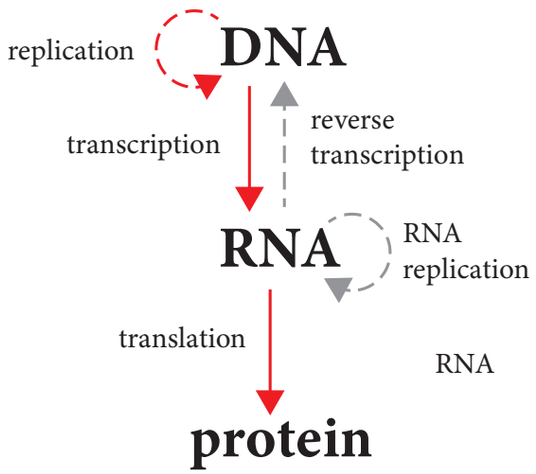


Figure 1: The central dogma of molecular biology

This simple information flow from DNA to RNA to Protein is the common mechanism in most organisms. But rarely, as in the case of retroviruses DNA molecule can be synthesised from RNA molecules (reverse transcription). Similarly RNA molecules are also known to replicate ie synthesise new RNA molecules in RNA viruses such as polio virus and mengoviruses.

6.2 DNA replication

Genetic information has to flow from the parental cells or organisms to progeny by a very faithful mechanism. This copy of DNA from DNA known as replication is the first and fundamental step in the central dogma. Any DNA replication mechanisms should consider and overcome the following complexities

1. As we have studied already the strands of DNA run in opposite directions (5'->3') on both strands. So the new nucleotides added should maintain that anti parallel directionality.
2. DNA is helical and a supply of energy is required to unwind it.
3. Such unwound single strands should be prevented from base pairing once again (re-annealing)
4. The DNA molecule is large and so the replication process should start and proceed at several points.
5. There should be a special mechanism to identify and correct errors during replication.

The replication of DNA molecule is an enzymatic reaction catalyzed by the enzyme DNA polymerase. However several other protein molecules are also associated with the replication process considering the above mentioned complexities.

6.2.1 The models of DNA replication

After the discovery of Watson and Crick double helical DNA structure, three different types of models for DNA replication were proposed, which were the conservative, semi conservative and dispersal models.

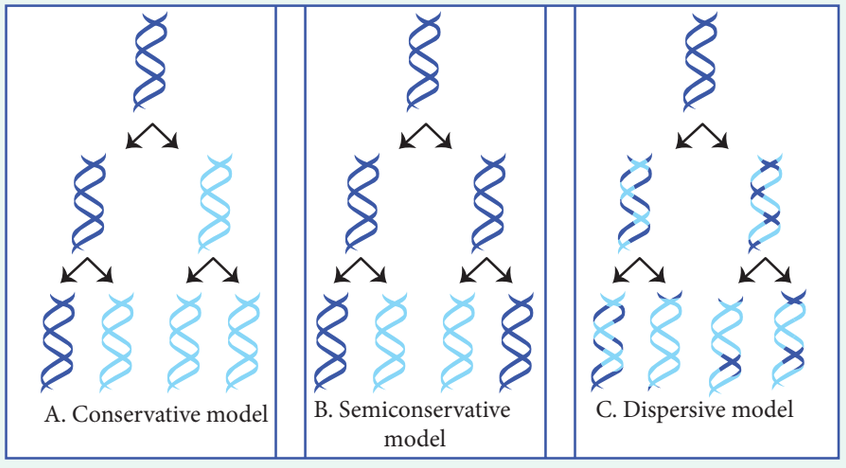


Figure 2 Different models of DNA replication

6.2.2 The conservative model

In this model the two parental strands base pair once again after forming a new copy. The newly synthesized strands combine together to form another daughter molecule. So in effect at the end of one cycle of replication an intact parental DNA and a new daughter DNA are obtained.

6.2.3 The Semiconservative model

According to this model the two strands separate in parental DNA and each of these strands act as a template to form a new DNA molecule. Thereby, after a cycle of replication

two DNA molecules are formed and each of these two molecules will have one parental strand and one newly synthesized strand.

6.2.4 The dispersive model

This model states that the parental DNA is randomly segmented into multiple double stranded DNA fragments. These fragments undergo replication similar to conservative model. The fragments later reassemble into complete DNA molecules. This model of replication will result in synthesis of two DNA molecules with interspersed parental and newly synthesized double helix.

6.2.5 The Meselson-Stahl experiment and the conformation of semiconservative model

The semiconservative model of DNA replication was confirmed by Matthew Meselson and Franklin Stahl in 1958. They took advantage of the fact that though ^{14}N isotope of nitrogen is normally present in DNA, the presence of the heavier ^{15}N isotope doesn't affect the function of DNA.

In this experiment, three different batches of the bacterium *E. coli* were grown. The first batch was grown with the medium containing ^{15}N as the only source of nitrogen for several generations. The bacteria took up this ^{15}N for growth and eventually all the nitrogen in the DNA were ^{15}N isotope. The *E. coli* were then switched suddenly to the medium containing the lighter ^{14}N as the only source of nitrogen, so as the bacterium divides these lighter nitrogen gets incorporated into the DNA. After switching to the light ^{14}N medium, sample DNA were isolated for every generation of *E. coli*, starting from generation zero (ie., every 20 mins). The DNA extracted subjected to caesium chloride (CsCl) density gradient centrifugation. This method of centrifugation separates the DNA into bands according to their density, with heavier DNA band at the bottom and progressively lighter DNA bands at the top.

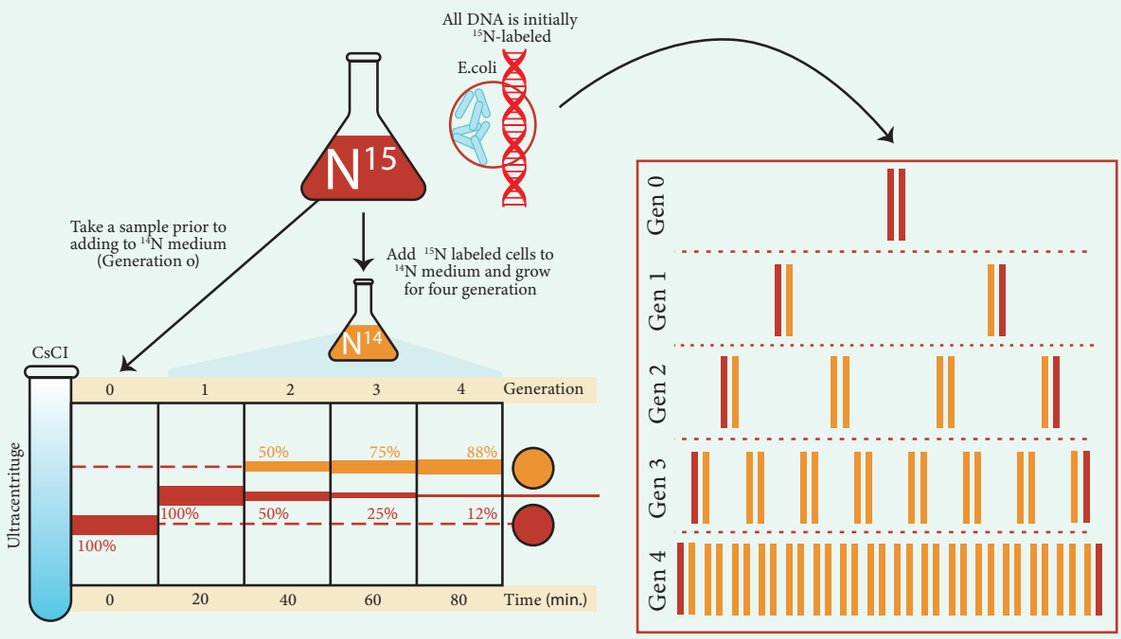


Figure 3 The schematic representation of Meselson-Stahl experiment

The experiment and the results are shown in figure 3. It could be seen that at generation zero, a single band at the bottom of the tube was produced. This is due to the fact that all the nitrogen in the DNA were the heavy ^{15}N isotope. At generation one, a single band was produced but in the intermediary position between where ^{15}N and ^{14}N isotopes were expected. This directly proves semi conservative replication as one strand is made of lighter isotope of nitrogen and the other strand is made of heavy isotope of nitrogen. Similarly the DNA at generation two showed two bands one at the intermediary region and another band at the lighter ^{14}N region. Subsequent generations of DNA showed these two bands, however the intermediary band was shown to become thinner and the ^{14}N band becoming broader. This is due to the fact that the strands containing ^{14}N isotope increases as the DNA replicates in ^{14}N medium.

6.2.6 Overview of DNA Replication

1. DNA replication is the process by which DNA makes a copy of itself during cell division.
2. The first step in DNA replication is to denature or 'unzip' the double helix structure of the DNA molecule, where the two strands of DNA are seperated.
3. DNA denaturation is carried out by an enzyme called helicase which breaks the hydrogen bonds holding the complementary bases of DNA together. Single strand binding (SSB) proteins bind to these strands to prevent them from annealing together.
4. Since the DNA is a very large molecule, the separation of the two single strands of DNA creates a 'Y' shape called a replication 'fork'.
5. The two separated strands will act as templates for synthesis of the new strands of DNA.
6. One of the strands is oriented in the 3' to 5' direction (towards the replication fork), this is the leading strand.
7. The other strand is oriented in the 5' to 3' direction (away from the replication fork), this is the lagging strand. As a result of their different orientations, the two strands are replicated differently.
8. Leading Strand replication: A short piece of RNA called a primer (produced by an enzyme called primase) comes along and binds to the end of the leading strand. The primer acts as the starting point for DNA synthesis. An enzyme called DNA polymerase binds to the leading strand and then 'walks' along it, adding new complementary nucleotide bases (A, C, G and T) to the strand of DNA in the 5' to 3' direction in a continuous fashion.
9. Lagging strand replication: Numerous RNA primers are made by the primase enzyme and bind at various points along the lagging strand. Chunks of DNA, called Okazaki fragments, are then added to the lagging strand also in the 5′ to 3′ direction, by DNA polymerase. This type of replication is called discontinuous as the Okazaki fragments will need to be joined up later.
10. Once all of the bases are matched up (A with T, C with G), an enzyme called exonuclease strips away the primer(s). The gaps where the primer(s) were are then filled by yet more complementary nucleotides.

6.2.7 The DNA polymerase

The DNA polymerase or the DNA dependent DNA synthetase is a large multi subunit enzyme which adds deoxyribonucleotides to the 3' end of the growing strand one by one. Hence it could be thought of an enzyme that elongates the existing nucleic acid (DNA or RNA primer) in 5'->3' direction. The DNA polymerase enzyme can perform two other activities which are the 3'→5' exonuclease activity and 5'→3' exonuclease activity. Exonucleases are enzymes which can remove bases one by one from the ends of a DNA molecule. The 3'→5' exonuclease activity of DNA polymerases provides as an error correction or proof reading action by removing incorrectly incorporated bases during replication. There are different types of DNA polymerases adapted to different types of cells/organisms and the features described above are of the most predominate enzyme called DNA polymerase I. There are other DNA polymerases like DNA polymerase II, III, alpha, beta, delta etc.,

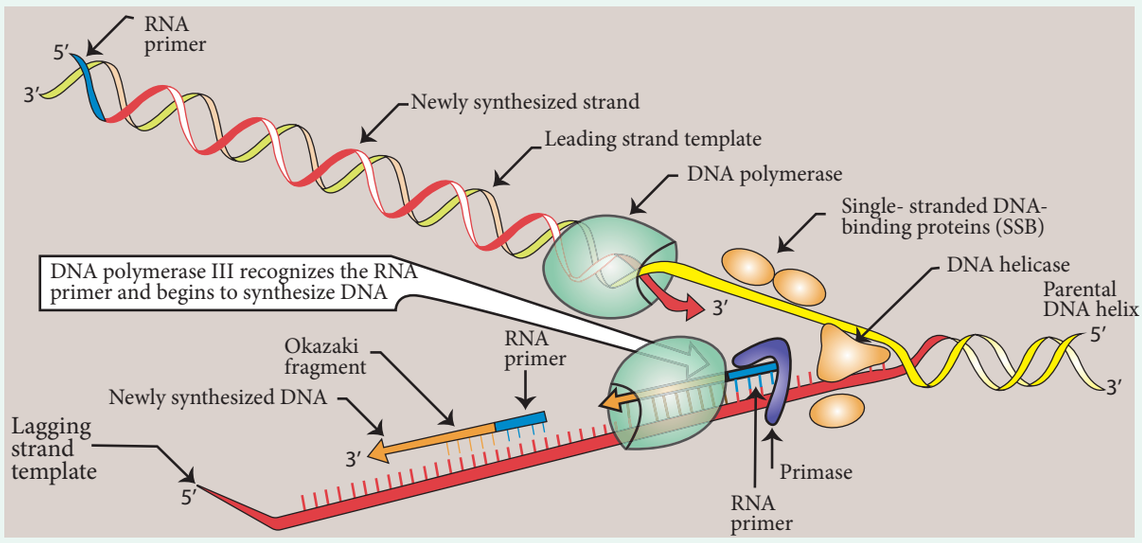


Figure 4 Overview of the DNA replication process

6.2.8 Difference between prokaryotes and eukaryotes in DNA replication

S. No.	Prokaryotes	Eukaryotes
1	It occurs inside the cytoplasm	It occurs inside the nucleus
2	There is a single origin of replication	Origin of replication are numerous
3	DNA polymerase III carries out both initiation and elongation	Initiation is carried out by DNA polymerase alpha while elongation by DNA polymerase by delta and epsilon
4	DNA repair and gap filling are done by polymerase I	The same are performed by DNA polymerase beta

5	RNA primer is removed by DNA polymerase I (5'->3' exonuclease activity)	RNA primer is removed by DNA polymerase beta
6	Okazaki fragments are long	Okazaki fragments are short
7	Replication is very rapid	Replication is very slow
8	DNA gyrase is needed to remove the supercoils in the circular bacterial DNA	DNA gyrase is not needed

6.2.9 The polymerase chain reaction – an essential tool for molecular biology

Polymerase chain reaction (PCR) is a method used to make billions of copies of a small region of DNA. This process is also known as PCR amplification. Depending on the application the DNA region being amplified can vary. Some examples include

1. A gene whose function a researcher wants to understand
2. A genetic marker used by forensic scientists to match crime scene DNA with suspects.
3. A genetic marker region in DNA which could be used for diagnosis of a disease like cancer

The goal of PCR amplification is to make enough quantity of the target DNA so that it could be processed for further analysis including sequencing, electrophoresis or plasmid cloning. PCR is routinely used in many areas of biology and medicine, including molecular biology research, medical diagnostics, ecology etc.,

6.2.9.1 The steps involved in PCR amplification

This reaction is essentially a DNA replication carried out in a tube. The key ingredients of a PCR reaction are the enzyme Taq polymerase, primers, template DNA, and deoxy ribo nucleotides (DNA building blocks). The Taq Polymerase is the DNA polymerase enzyme of the heat tolerant bacterium *Thermus aquaticus*. This enzyme can be functional even in temperatures around 70°C. The ingredients are assembled in a tube, along with cofactors needed by the enzyme and are subjected to repeated cycles of heating and cooling that allow DNA to be synthesized. Each cycle of amplification consists of three steps, which are

i. Denaturation (96°C): Heat the reaction strongly to separate or denature the DNA strands. This provides single-stranded template for the next step.

ii. Annealing (55 - 65°C): Cool the reaction so the primers can bind to their complementary sequences on the single-stranded template DNA.

iii. Extension (72°C): Raise the reaction temperatures so Taq polymerase extends the primers, synthesizing new strands of DNA.

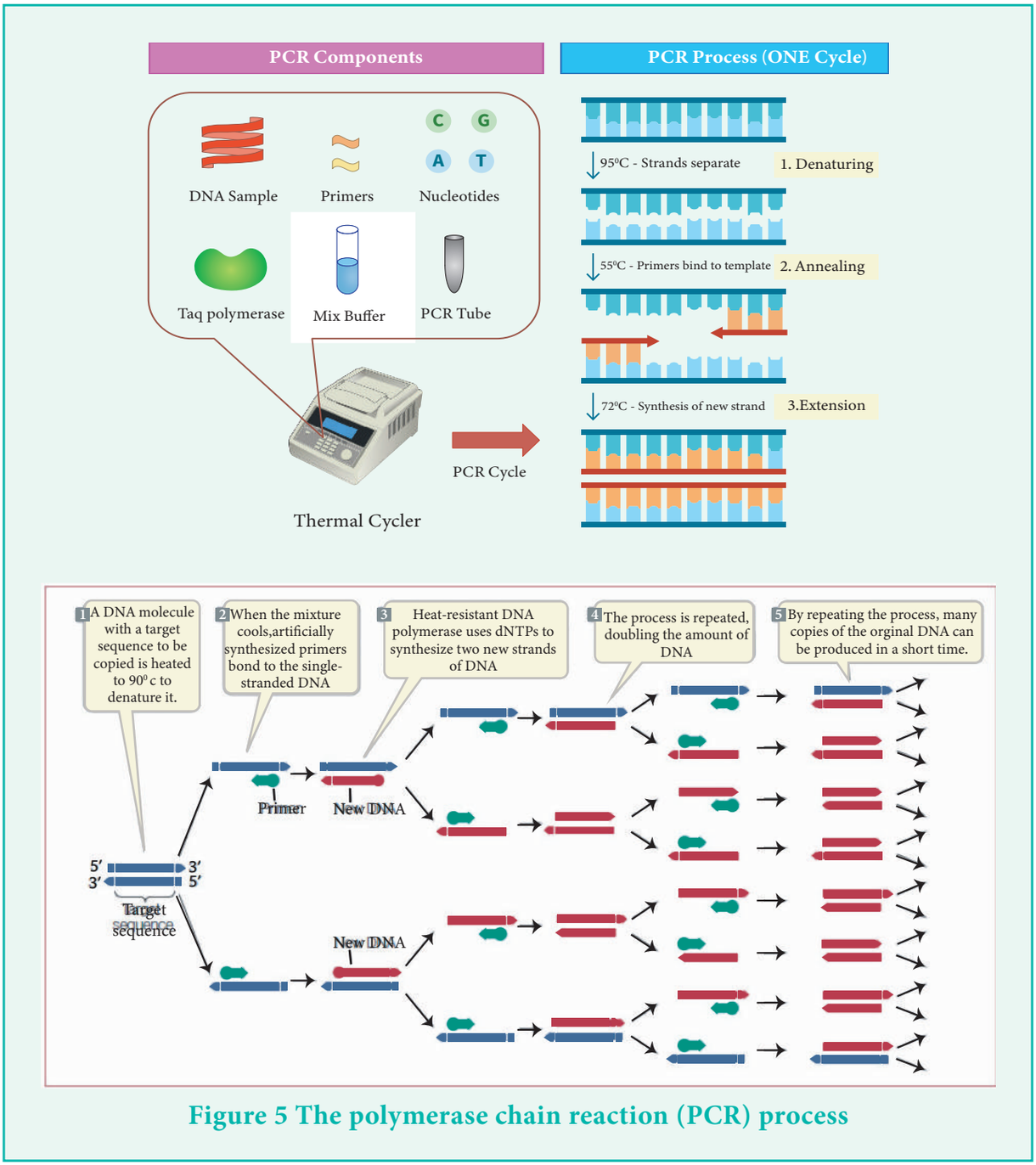


Figure 5 The polymerase chain reaction (PCR) process

6.3 TRANSCRIPTION

As described earlier transcription is the process by which the information in a strand of DNA (deoxyribonucleic acid) is copied into a new molecule of RNA (ribonucleic acid) with the help of the enzyme called RNA polymerase and a few other proteins.

6.3.1 Genes and Gene expression

Genes are described as units of heredity in classical genetics. However in molecular biology terms, genes can be considered as a region in DNA which acts as an information

template for a protein or RNA molecule (in case of tRNA or rRNA genes). Each gene contain a distinct sequence of nucleotides which determines the aminoacid sequence or nucleotide sequence a cell might synthesize. If a gene i.e, a segment of DNA can act as a template to synthesize an RNA transcript by the process of translation, that gene is considered to be expressed.

6.3.2 Overview of transcription

1. It is a process where a DNA sequence is read by an enzyme called RNA polymerase and a complementary, antiparallel RNA strand is synthesized.
2. Apart from RNA polymerase many proteins called transcription factors control the rate of transcription by binding to specific DNA sequence.
3. Transcription between eukaryotes and prokaryotes will vary slightly.
4. Some viruses (eg. HIV) have the ability to transcribe RNA into DNA. HIV has an RNA genome that is reverse transcribed into DNA.
5. RNA polymerase, together with one or more general transcription factors, binds to promoter DNA.
6. RNA polymerase creates a transcription bubble, which separates the two strands of the DNA helix. This is done by breaking the hydrogen bonds between complementary DNA nucleotides.

There are three stages in transcription which are initiation, elongation and termination.

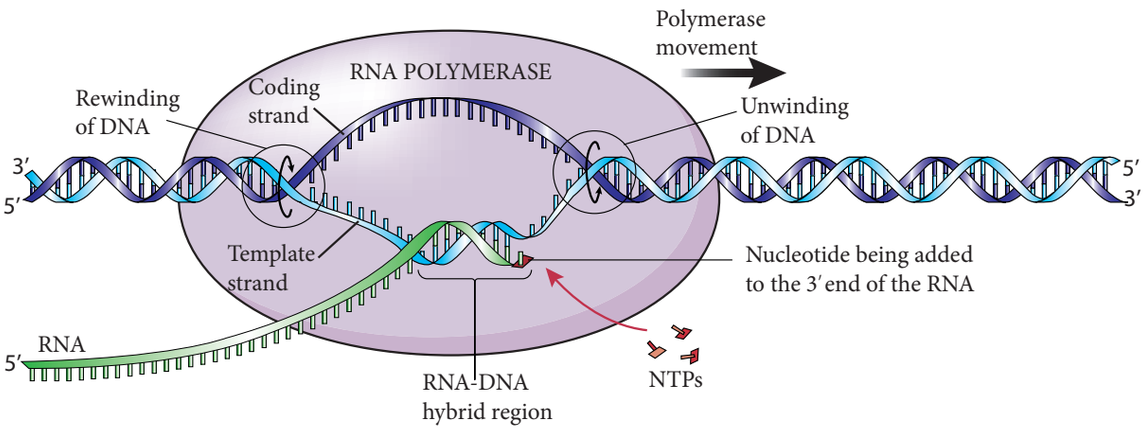


Figure 6. Representation of the transcription process

6.3.2.1 Initiation of transcription

1. RNA polymerase, together with one or more general transcription factors, binds to promoter DNA.
2. RNA polymerase, along with transcription factor sigma, binds to a sequence of DNA called the promoter, found near the beginning of a gene.

3. Each gene (or group of co-transcribed genes, in bacteria) has its own promoter.
4. Once bound, RNA polymerase separates the DNA strands, providing the single-stranded template needed for transcription.
5. RNA polymerase creates a transcription bubble, which separates the two strands of the DNA helix. This is done by breaking the hydrogen bonds between complementary DNA nucleotides.

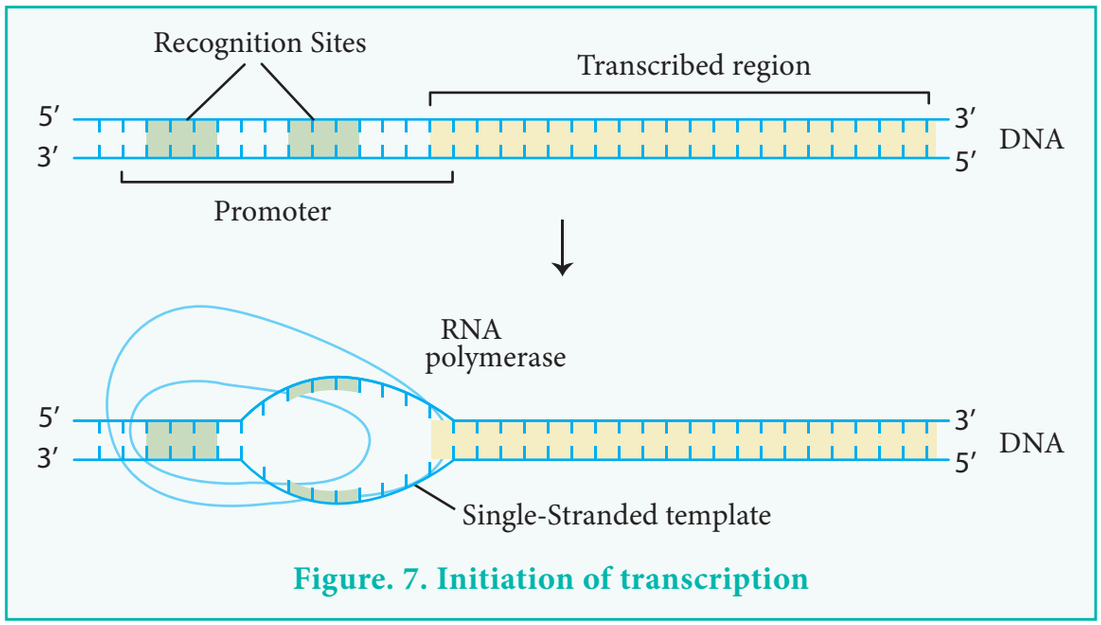


Figure. 7. Initiation of transcription

6.3.2.2 Elongation of transcription

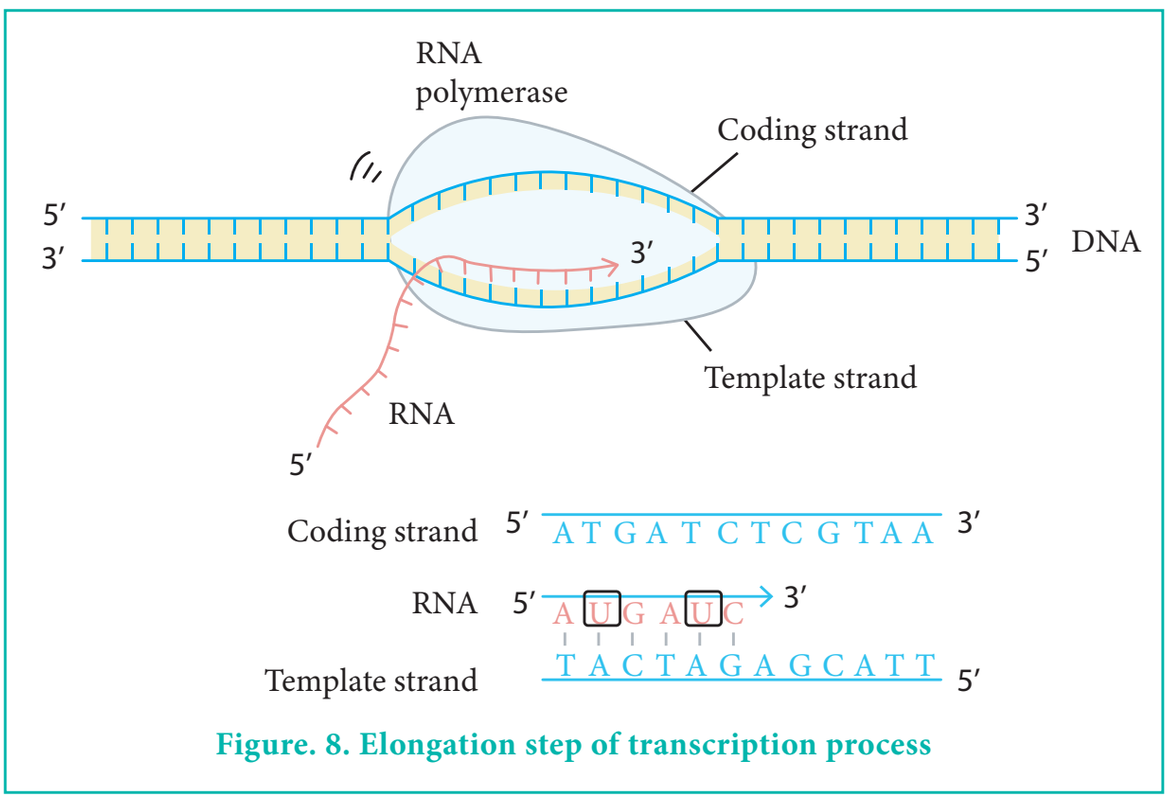


Figure. 8. Elongation step of transcription process

1. After separation of the DNA strands, one strand of DNA, the template strand, acts as a template for RNA polymerase.
2. The enzyme RNA polymerase reads this template strand one base at a time and synthesizes an RNA molecule out of complementary nucleotides always in 5' to 3' direction.
3. The RNA transcript carries the same information as the non-template (coding) strand of DNA the base thymine (T) is replaced by uracil (U).

6.3.2.3 Termination of transcription

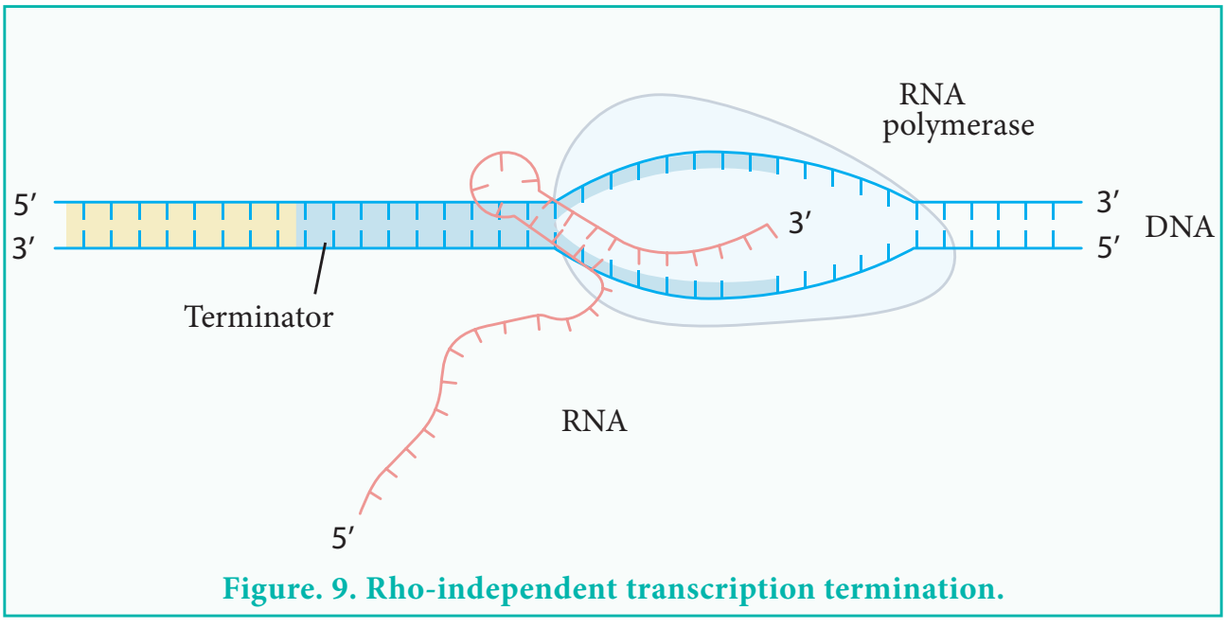


Figure. 9. Rho-independent transcription termination.

1. Two different types of transcription termination are observed in bacteria: the Rho-independent termination and Rho-dependent termination.
2. In Rho-independent termination, the RNA transcript forms a hairpin loop structure rich in GC base pairs, followed by a poly U region.
3. The hair pin structure of the newly synthesized RNA transcript produces a mechanical stress in the RNA polymerase and stalls it from moving. At the same time the relatively weak hydrogen bonds made by the poly U region with template DNA pulls out of the RNA polymerase, terminating transcription.
4. In the "Rho-dependent" type of termination, a protein called Rho destabilizes the interaction between the template and the mRNA and releases the newly synthesized mRNA from the elongation complex.

6.3.3 Types of RNA, synthesis and processing

There are three types of RNA which play a major role in protein synthesis, which are the ribosomal RNA or rRNA, the transfer RNA ot tRNA and the messenger RNA or mRNA

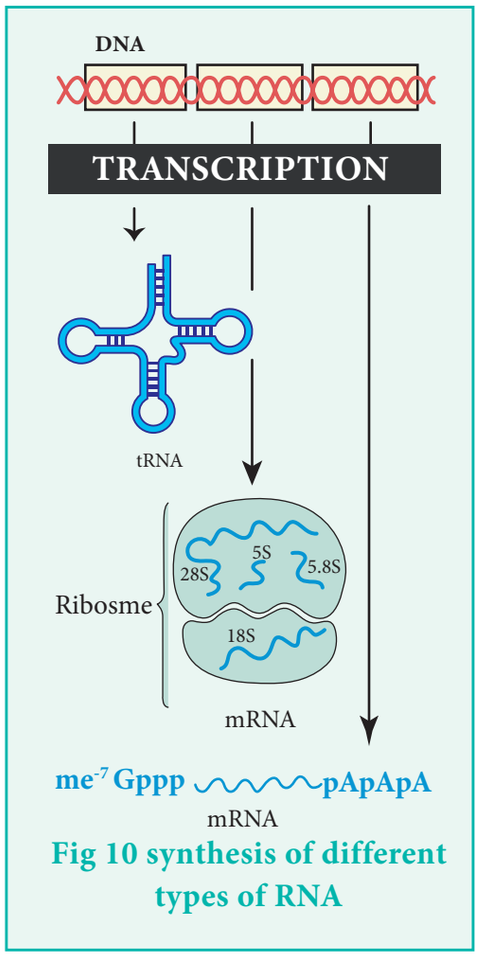


Fig 10 synthesis of different types of RNA

6.3.3.1 The Ribosomal RNA

These molecules are components of ribosomes. They form complexes with ribosomal proteins and act as important machinery for protein synthesis. There are three different species of rRNA, the 5S, 16S and 23S rRNA in prokaryotes and four in eukaryotes, the 5S, 5.8S, 18S and 28S rRNA. Here S represents the svedbergs unit or the sedimentation rate during centrifugation.

6.3.3.2 The Transfer RNA

These are the smallest among the three RNA types. Their primary function is to transfer aminoacids during protein synthesis. So there is atleast one tRNA molecule for every aminoacid. Figure 11 shows the secondary and tertiary structure of a typical tRNA molecule. The secondary structure is like a clover leaf and aminoacids are covalently linked to the 5'end. It could also be noted that there are many unusual bases in the tRNA like dihydrouracil. They also form extensive intra chain base pairing forming different loops in the RNA structure like the TψC loop, anticodon loop and D loop.

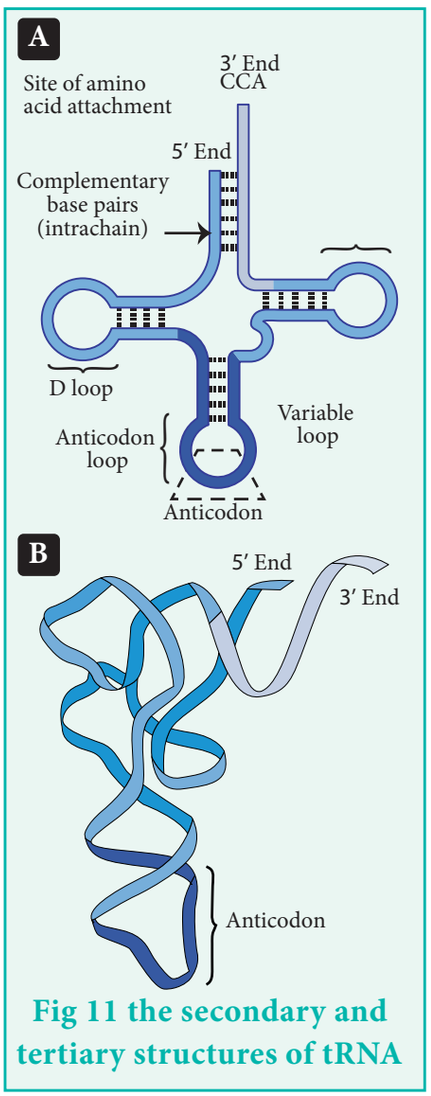


Fig 11 the secondary and tertiary structures of tRNA

6.3.3.3 The messenger RNA

This carries the genetic information from the DNA to the cytoplasam and hence the name messenger RNA. They are highly heterogeneous varying in size and sequence content. They act as template for protein synthesis. If the mRNA contains information for only one protein it is called monocistronic. The eukaryotic mRNAs are mostly monocistronic. However prokaryotic mRNA can contain information for encoding more than one mRNA and are called polycistronic.

6.3.4 Post transcriptional modification of RNA

In eukaryotes the RNA molecules are further processed. The most important modificatios are

1. A 5-methyl guanosine base is added to the 5' end of the transcript
2. The 3' end is added with poly Adenine bases known as the poly A tail.
3. The eukaryotic genes and there by their corresponding RNA transcripts can have multiple non coding regions in between known as *introns*. The coding regions are known as exons. The process by which the noncoding *intron* regions are removed from the RNA transcript is known as RNA splicing. This processed RNA molecule which is ready for translation is known as an mRNA or a messenger RNA.

6.3.5 Difference between Prokaryotic and Eukaryotic in Transcription

Sl. No	Prokaryotic Transcription	Eukaryotic Transcription
1	Coupled transcription-translation is the rule	Coupled transcription translation is not possible.
2	Occurs in the cytoplasm	Occurs in the nucleus
3	There is no definite phase for occurrence	Majority of transcription takes place in the G1 and G2 phase of the cell cycle
4	A single RNA polymerase synthesis all the three types of RNA(mRNA,tRNA,rRNA)	Different types RNA polymerases (I,II,III etc) synthesize different RNA molecules such as rRNA,mRNA,tRNA etc...
5	Initiation of transcription factor does not need any proteins or initiation factors	Initiation of transcription proteins requires transcription factors.These are TFIIA,TFIIB,TFIID,TFIIE,TFIIH.These recognises the TATA box.
6	Transcription usually involves more than one gene (polycistronic)	Transcription normally involves only one gene (monocistronic)
7	Usually there is no post transcriptional modification of the primary transcript.	Primary transcript undergo post translational modification(RNA editing)
8	Introns absent in the primary transcript.	Introns present in the primary transcript.

6.4 TRANSLATION

As we have already studied, Proteins are the key molecules which literally carry out every single molecular process inside our cell. The instructions about how these proteins should be made is encoded in the DNA and is transcribed from the chromosome in the form RNA by a process which we saw as transcription. Similalry the process by which a new protein molecule is synthesized using this mRNA as a template is known as translation.

6.4.1 The genetic code

During translation, a cellular organelle called ribosome "reads" the information in a messenger RNA (mRNA) and uses it to build a polypeptide or chain of amino acids. In an mRNA, three consecutive nucleotides are bases known as codons provide the instructions for building a polypeptide. Each codon codes for a single aminoacid. There are totally 61 codons which codes for amino acids and there are three more codons UAA, UAG and UGA which are known as stop codons as they are signals for termination of translation. The codon AUG codes for the aminoacid methionine and it also acts as a start codon. The genetic code is universal i.e, baring a few rare instances it is followed by all organisms. Similalry the genetic code is degenerate, meaning that one codon can only represent one aminoacid but one aminoacid can be represented by more than one codon. For example the codon UUU represents only phenylalanine. However phenylalanine can be represented by two codons UUU and UUC.

Second letter

First Letter	U	C	A	G	Third Letter
U	UUU, UUC: Phe UUA, UUG: Leu	UCU, UCC, UCA, UCG: Ser	UAU, UAC: Tyr **UAA STOP** **UAG STOP**	UGU, UGC: Cys **UGA STOP** UGG Trp	U C A G
C	CUU, CUC, CUA, CUG: Leu	CCU, CCC, CCA, CCG: Pro	CAU, CAU: His CAA, CAG: Gln	CGU, CGC, CGA, CGG: Arg	U C A G
A	AUU, AUC, AUA: Ile AUG Met	ACU, ACC, ACA, ACG: Thr	AAU, AAC: Asn AAA, AAG: Lys	AGU, AGC: Ser AGA, AGG: Arg	U C A G
G	GUU, GUC, GUA, GUG: Val	GCU, GCC, GCA, GCG: Ala	GAU, GAC: Asp GAA, GAG: Glu	GGU, GGC, GGA, GGG: Gly	U C A G

Figure 12 The universal genetic code

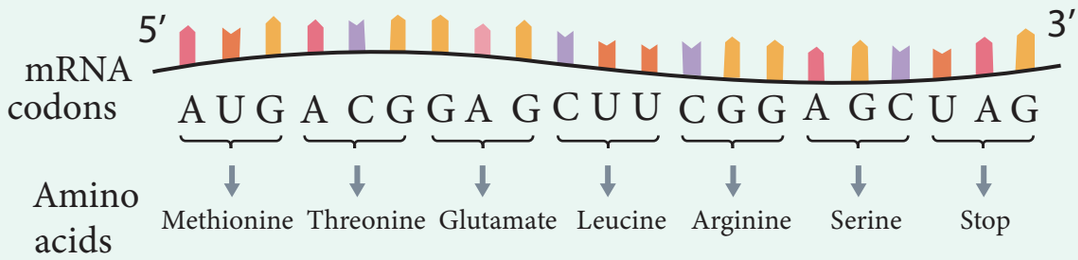


Figure 13 Representational decoding of codons in a sample mRNA to corresponding amino acid sequence.

6.4.2 Overview of translation

How is an mRNA molecule "read" to make a polypeptide? Two important components act as a machinery for this which are the tRNAs and ribosomes. As we have seen already

the tRNAs paly a very important role in protein synthesis. They act as molecular "bridges" that connect mRNA codons to the amino acids they encode. Each tRNA has an anticodon loop, which has a sequence of three nucleotides called an anticodon. They can bind to specific codon in the mRNA. The other end of the tRNA carries the amino acid specified by the codons.

6.4.3 Ribosomes

Ribosomes are the cellular organelles where polypeptides (proteins) are built. They themselves are made up of different proteins and rRNA. Each ribosome has two subunits, a large one and a small one, which sandwiches the mRNA as shown in figure 14

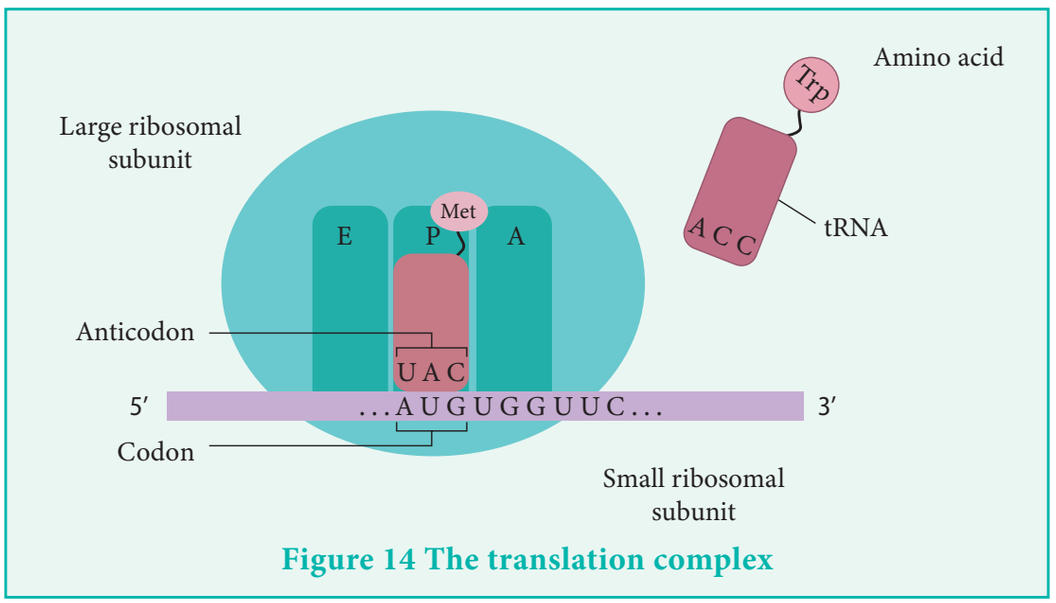


Figure 14 The translation complex

The ribosome provides three different structural packets where tRNA can enter and transfer its amnoacid to the growing polypeptide chain. These slots are called the A, P, and E sites. The ribosome itself can act as an enzyme, catalyzing the chemical reaction that links amino acids together to make a chain.

6.4.4 Molecular events in translation

The translation process can be divided into three different stages: initiation, elongation, and termination.

6.4.4.1 Translation Initiation

Inititaion require three important components

- A ribosome – the large subunit and the small subunit
- An mRNA with instructions for the encoding protein.
- An "initiator" tRNA carrying the first amino acid in the protein, which is almost always methionine (Met)

The methonine tRNA attaches to the small ribosomal subunit. Together, they bind to the 5' end of the mRNA by recognizing the 5' cap in eukaryotes and then they move along the mRNA in the 3' direction, stopping when they reach the start codon AUG. This is later

joined by the large subunit of the ribosome and this complex is known as the translation initiation complex.

In bacteria the small ribosomal subunit attaches directly to a pattern of sequence in the mRNA known as Shine-Dalgarno sequences which are present just before start codons and "point them out" to the ribosome.

Since the bacterial mRNA are polycistronic, they are transcribed in groups (called operons), so one bacterial mRNA can contain the coding sequences for several genes. A Shine-Dalgarno sequence marks the start of each coding sequence, letting the ribosome find the right start codon for each gene.

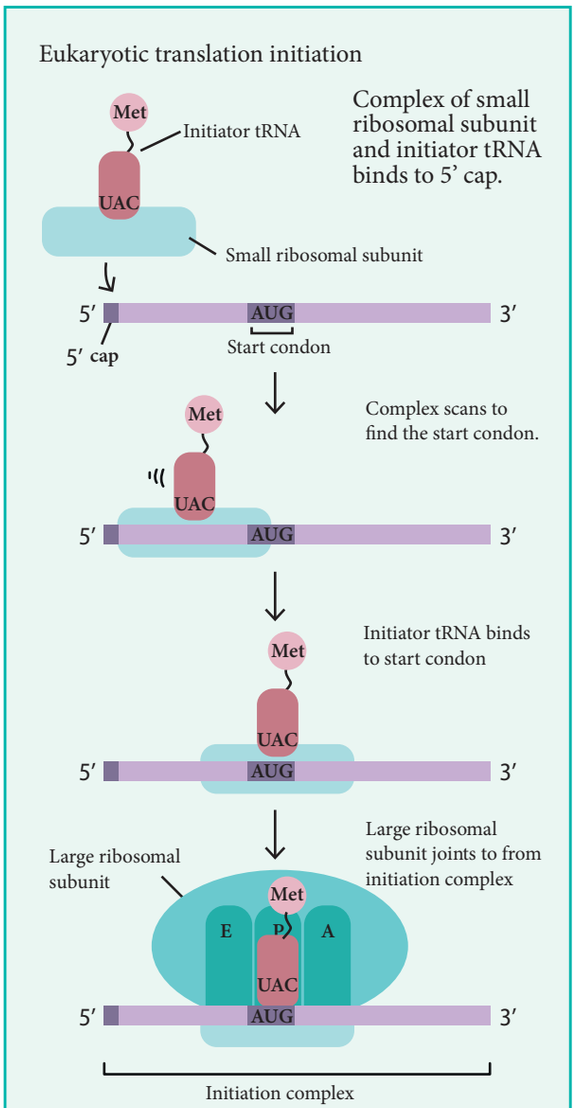


Figure 15 The translation initation process in eukaryotes

6.4.4.2 Translation Elongation

1. The first methionine-carrying tRNA starts in the middle slot or the P site of the ribosome. The subsequent codon is open to the A site.
2. The next tRNA with the anticodon which is complementary to this codon occupies the A site of the ribosome.
3. Once the matching tRNA has landed in the A site the first amino acid methionine from the first tRNA is covalently bonded (peptide bond) to the amino acid of the second tRNA in the A site.
4. Thus a dipeptide is formed with the methionine as N-terminus and the other amino acid is the C-terminus.

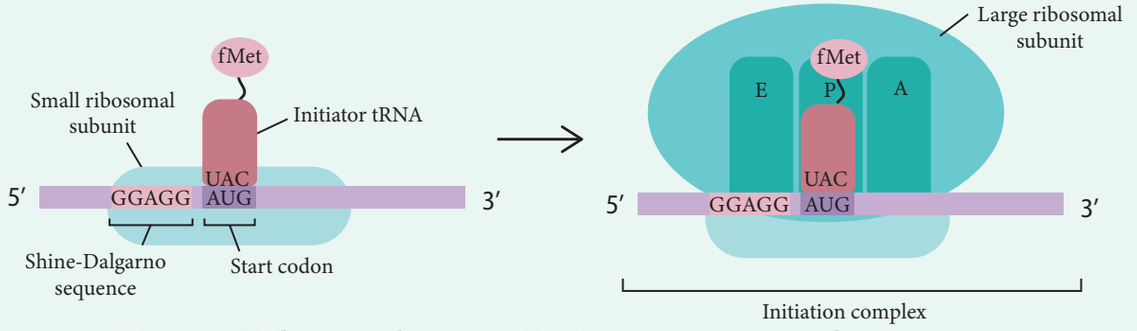


Figure 16 The translation initiation process in prokaryotes

5. Once the peptide bond is formed, the mRNA is pulled onward through the ribosome by exactly one codon. This shift allows the first, empty tRNA to drift out via the E site at the same time exposing the a new and subsequent codon to the A site.

6. The entire cycle repeats and every time a new aminoacid is added as per the information in the codons being exposed to the A site. Thus a polypeptide stats growing in the ribosome.

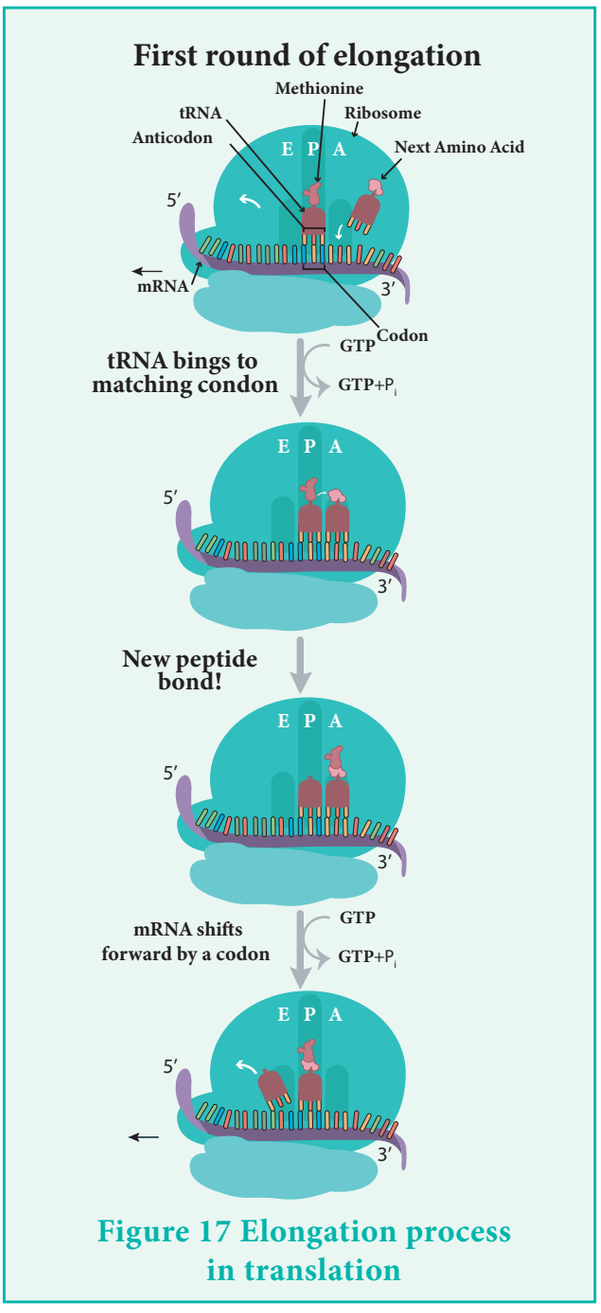


Figure 17 Elongation process in translation

6.4.4.3 Termination of translation.

The growing polypeptide is stoped when the A site of the ribosome encounters one of the stop codons in the mRNA which are UAA, UAG, or UGA.

Stop codons are recognized by proteins called release factors, which fit neatly into the P site helping the synthesized polypeptide to be released.

The small and large ribosomal subunits separate from the mRNA and from each other and are free to take part in another different translation process.

6.4.5 Post-Translational Modifications

1. After translation, amino acids may be chemically altered or removed.

2. The chemical modifications include phosphorylation, glycosylation, glycation, nitrosylation, methylation, acetylation, lipidation etc

3. The new polypeptide will also fold into a distinct 3D structure and may join with other polypeptides to make a multi-subunit protein.

6.4.6 Difference between prokaryotic and eukaryotic translation

S. No	Prokaryotic translation	Eukaryotic translation
1	Transcription and translation are part of a continuous process and occurs simultaneously	Transcription and translation are separate process
2	The 5' end of the transcribed mRNA is immediately available for translation without any modification	The primary transcript is subjected to post transcriptional modifications and has to cross the nuclear membrane to reach the cytoplasam for translation
3	70S ribosome is involved, with a 50S larger subunit and a 30S smaller sub-unit	80S ribosome is involved with a 60S larger subunit and a 40S smaller subunit
4	Larger subunit of ribosome is made of 23S, 5S rRNA and 36 different proteins	Larger subunit is made of 28S, 5S, 5.8S rRNA and 40 different proteins
5	Smaller subunit of ribosome is composed of 16S rRNA and 21 different proteins	Smaller subunit of ribosome is made of 18S rRNA and 33 proteins
6	mRNA is polycistronic	mRNA is monocystronic

6.5 RECOMBINANT DNA TECHNOLOGY

This technology is also known as DNA cloning allows us to insert a DNA fragment of interest such as a new gene into an organism, which doesn't contain that fragment of DNA. Though there are a variety of strategies to achieve this, the simplest of it is introducing a new gene into bacteria through plasmid vectors. Plasmids are small circular DNA molecules present in bacteria. They can be used to carry a foreign gene into the bacteria and hence the name vectors. The entire process is explained in figure 18. and consists of the following important steps

1. A bacterial plasmid designed with an antibiotic resistant marker gene is selected as a vector.
2. The circular plasmid vector is cleaved at specific sites using special enzymes called restriction endonuclease which are enzymes that cleave DNA at specific sequence patterns. For example EcoRI is a restriction enzyme which cleaves at the sequence GAATC.
3. The gene to be cloned is inserted in the gap produced by the endonuclease and an enzyme called DNA ligase is used to join the DNA pieces. This results in a circular plasmid vector containing the gene of interest or the recombinant plasmid.
4. This recombinant plasmid vector is now transferred into the cell by means of a process called transformation.

5. These bacteria can be grown in a media containing specific antibiotics as per the resistant gene present in the plasmid vector. The bacterial cells which have taken up the recombinant plasmid vector can only grow in the presence of antibiotics (positive clones).
6. These positive clones can then be subjected to further applications like production of recombinant proteins, studying expression etc.,
7. This method has been successfully used for commercial production of many proteins like Human Insulin, antibodies, enzymes of medical and industrial interests etc.,

6.6 GENOME SEQUENCING AND THE BIRTH OF GENOMICS ERA

The birth of the 21st century has seen many revolutions in the field of molecular biology. Today molecular biologists are more interested in studying the collective behaviour of the genome rather than studying one individual gene. The entire information present in the DNA of an organism can be considered as its genome. Genomics can be defined as a study of whole genomes of organisms and uses a combination of techniques such as recombinant DNA technology, PCR amplification, Advanced DNA sequencing methods and bioinformatics analysis. Genomics experiments tries to harness the availability of entire sequence of genome which is possible today by the pioneering sequencing experiment by Fredrick Sanger and the recent next generation sequencing technologies most of which are build upon Sanger sequencing procedure

6.6.1 The Sanger sequencing procedure

1. It is also known as dideoxy chain termination method.
2. It sequence is generated by replicating a template DNA strand (the region to be sequenced) and interrupting the replication process at one of the four bases.
3. Four different reaction mixtures are produced such that the growing chain terminates in A,T, G or C
4. The reaction mixture also contains dideoxynucleoside triphosphates (ddNTPs) along with usual dNTPs.
5. If during replication ddNTPs is incorporated instead of usual dNTPs in the growing DNA strand then the replication stops at that nucleotide.
6. The ddNTPs are radio labelled (with P_{32})or fluorescently labelled so the chain can be visualized in a gel electrophoresis procedure.

The details of Sanger sequencing procedure is depicted in figure 19 This sequencing procedure was performed in a massively parallel fashion and the human genome was sequenced in the year 2000.

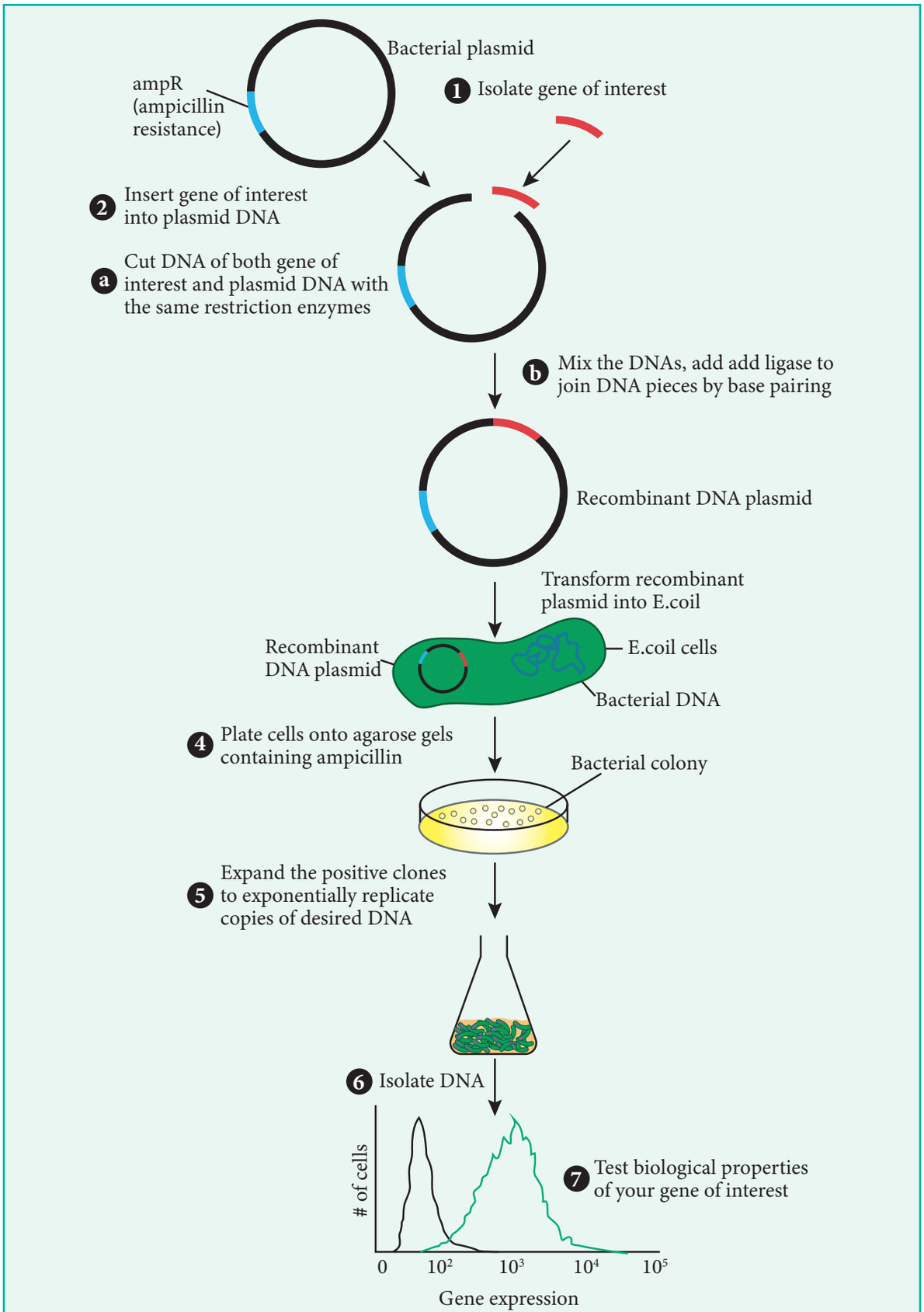


Figure 18 The steps in recombinant DNA technology (bacterial gene cloning)

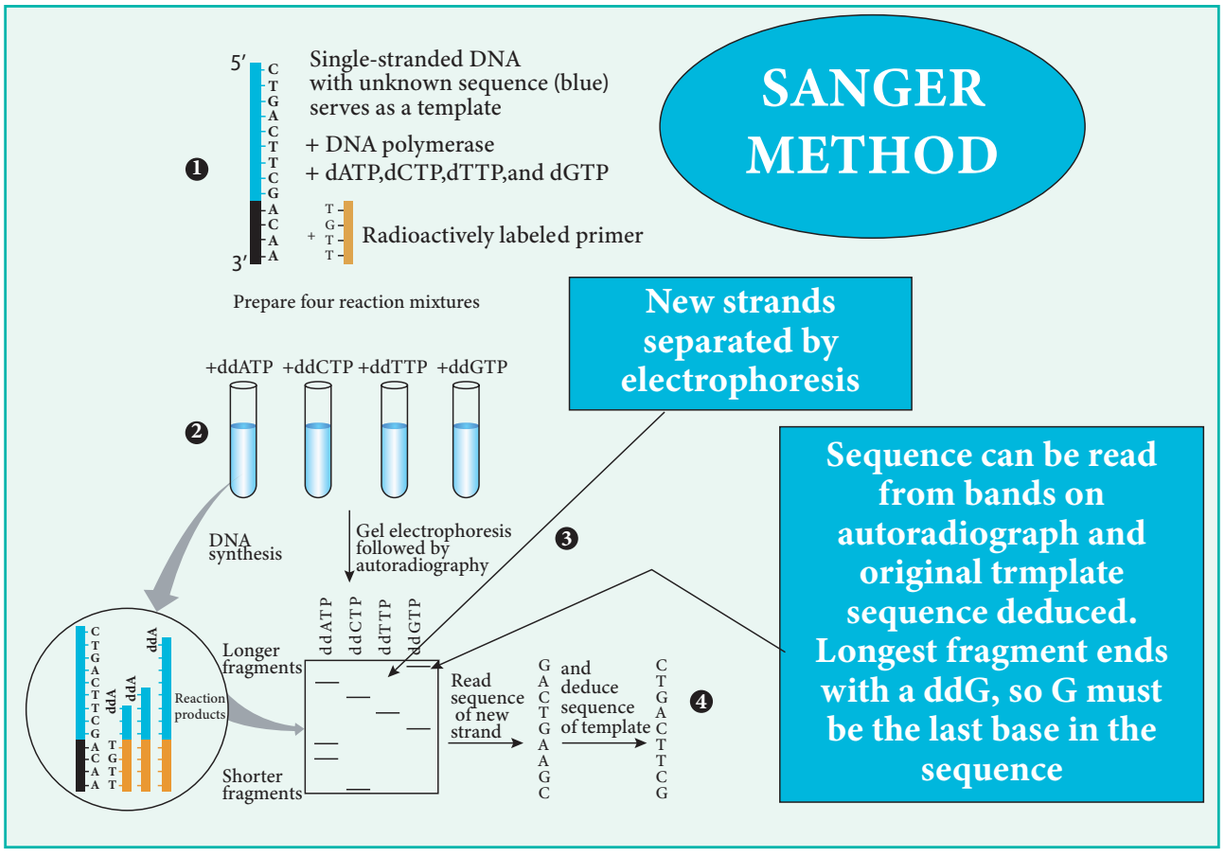


Figure 19 Schematic representation of Sanger sequencing procedure

6.6.2 The Next Generation Sequencing (NGS) Technologies and their role in genomics.

All the Next Generation Sequencing (NGS) methods adopt the principle of Sanger's method i.e., sequencing-by-synthesis. However they all use technologies that do not permenantly terminate the growing chain, so that the synthesis can be monitored as each base is incorporated. This improvement allows us to conduct sequencing of millions of DNA fragments simultaneously and hence achieving high throughput, which was not possible in Sanger's method. Moreover the sequencing costs has become much lower using these technologies and today a human genome can be sequenced at around $200 and within 48 hours. Some of the popular methods are the Illumina reversible dye terminator method, ion torrent method and nano pore DNA sequencing methods.

The advancements in genome sequencing has led to more ambitious projects like the 1000 genomes project or 1KGP. It was an international collaboratory effort to sequence 1000 different human genomes and to catalogue their genetic variations. This project was started in 2008 and was completed in 2012 with the complete genome sequence of 1098 humans. With the help of this data, so far about 100 regions in human genome has been identified to have association with diseases like coronary heart disease, diabetes mellitus etc.,

The Cancer Genome Atlas program (TCGA) is another ambitious project which was started in 2007 and so far has collected genomics data from 20,000 cancer patients and includes 33 types of cancer. This data is freely available for researchers around the world

and act as a potential resource for combating caner.

6.6.3 Applications of NGS technologies

Apart from obtaining the sequence of DNA or genomes, the current sequencing strategies can be tailored for different applications. A few important applications are discussed below.

Transcriptomic Profiling (RNA-Seq)

In this method the RNA content in a cell can be isolated and sequenced in a quantitative fashion. As a result this method can profile the genes that are expressed in a cell in a given condition.

Polymorphism and Variation Discovery

This experiment identifies the mutations or variations in a genome sequence and associates them with phenotypes such as diseases.

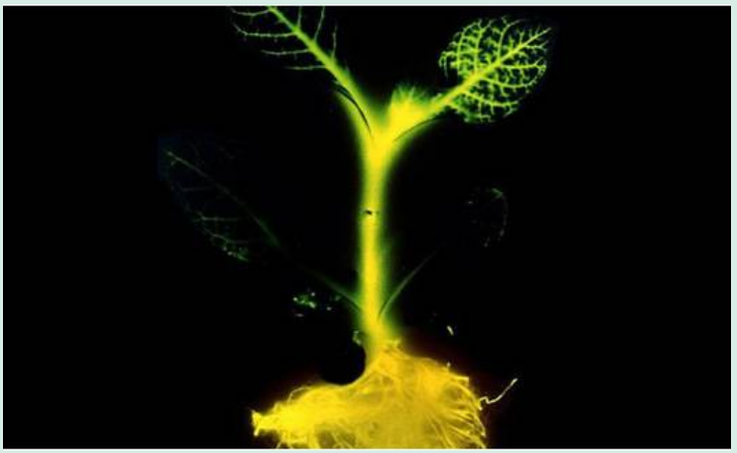
Protein-DNA Interaction Analysis (ChIP-Seq)

This method is also known as Chromatin Immuno Precipitation (ChIP), which precipitates the protein (histone) DNA complexes alone, by means of specific antibodies and then that region of DNA alone is sequenced. This method helps us to identify the genome regions which can interact with proteins like histones.

Metagenomics

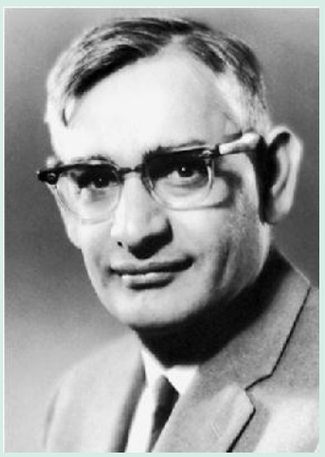
Metagenomics can be defined as the study of genetic material recovered directly from environmental samples. For example, the study of the microbiome in the gut of an individual. These samples contain extremely large but unknown numbers of unculturable microbial species. So such samples are directly subjected to DNA isolation and sequencing (without culturing) and considered as the genome of an entire community.

DO YOU KNOW?



Scientist Steve Howell in 1986 created a tobacco plant that can glow in dark like a firefly. He isolated a gene called luciferase which are present in fireflies and are the reason for the fireflies to glow in dark. He introduced this gene into toboco plants and showed that they can also glow in dark. These type of plants which contains foreign genes are known as transgenic plants or genetically modified (GM) plants. The methods used for producing such organisms are collectively known as genetic engineering. These methods promise increased agricultural yields by producing transgenic plants which can be pest resistant, drought resistant, saline resistant etc., However extreme caution has to be employed when introducing these plants to main stream agriculture as several factors has to be considered such as the balance in eco system, cost and effectiveness of the GM variety while cultivating in the field etc., Some countries have also banned GM crops and many countries follow the policies of strictly regulated use.

DO YOU KNOW?



Har Gobind Khorana was an Indian-American biochemist who did pioneering experiments to reveal the genetic code. He was born to Krishna Devi Khorana and Ghanpat Rai Khorana in Punjab during 1922. He finished his Bachelaor's and Master's degrees in Punjab University (Lahore). He lived in pre independent India until 1945, after which he moved to England and obtained his PhD in organic chemistry from The University of Liverpool. In 1960 he was appointed as a professor in University of Wisconsin (USA) where he did a serried of experiments to understand how RNA codes for proteins. He synthesized repeating units of nucleotides 'UCU CUC UCU ...' codes for the amino acids serine and leucine repeatedly. By this way he was able to show that three RNA nucleotides were responsible for an amino acid. He also identified stop codons. In short he was able to synthesize a gene or in other words 'a nucleotide sequences which could code for a desired aminoacid sequence' He was awarded Nobel prize in 1968 as recognition for his works in understanding the molecular biology of protein synthesis.

Summary

- DNA Replication : Each strand of the DNA double helix serves as a template for the synthesis of a complementary daughter strand by the semiconservative mode of replication which was confirmed by the Meselson Stahl experiment.
- The two strands of the double helix unwind with the help of helicase. DNA, being a large molecule, the separation of the two single strands creates a 'Y' shaped structure called a 'replication fork'. DNA polymerases synthesize the new DNA strands only in the 5' to 3' direction.Therefore, one of the newly synthesized strands grows in the 5' to 3' direction towards the replication fork (leading strand) and and one in the 5' to 3' direction away from the replication fork (lagging strand).
- A short stretch of RNA acts as a primer for DNA polymerase and it is synthesized by RNA primase.The leading strand needs only one primer, but the lagging strand needs many and short fragments of DNA called Okazaki fragments are synthesized, which are eventually ligated.
- In *E.coli*, DNA polymerase III elongates a DNA chain and proofreads mismatched bases with its 3' to 5' exonuclease activity. DNA polymerase I removes RNA primers using its 5' to 3' exonuclease activity and fills the gaps.
- Transcription: The process of RNA synthesis directed by a DNA template is called transcription and it is performed by RNA polymerase. Transcription occurs in 3 steps: initiation, elongation and termination.
- RNA polymerase binds to a sequence of DNA called the promoter, found near the start of a gene. Once bound, RNA polymerase separates the DNA strands, exposing the single-stranded DNA template. Complementary bases are added (instead of T, U is added) one at a time by RNA polymerase.The newly synthesized RNA strand grows from 5' to 3'. Termination occurs with the help of a rho factor or with specific terminator sequences.
- In eukaryotes post transcriptional modifications occur by which, a methyl guanosine cap and a poly A tail are added to 5' and 3' ends, respectively, to mRNA. Introns are spliced out of mRNA, tRNA and rRNA primary transcripts.
- Translation: The three major components in translation are mRNA, ribosomes and tRNA. The mRNA code is read in the 5' to 3' direction, and the protein is synthesized from the amino terminus to the carboxyl terminus.Three steps occur during translation: Initiation, elongation and termination.
- There is at least one specific tRNA and one aminoacyl tRNA synthetase for each amino acid. An anticodon in tRNA recognizes the codon present in mRNA.
- Initiation of translation: mRNA associates with the small ribosomal subunit and this process requires a sequence called Shine-Dalgarno sequence on mRNA.

The initiating codon is AUG and the initiating amino acid is formyl methionine.

- Elongation is facilitated by factors that facilitate the binding of aminoacyl tRNA to the A site on the ribosome as well as the movement of the ribosome along the mRNA.
- Peptide bond formation is catalyzed by peptidyl transferase and after peptide bond formation, it moves forward along the mRNA in 5' to 3' direction to the next codon.
- Termination begins when one of the stop codons moves into the A site. The recognition of these codons by release factors releases the newly formed protein from the m-RNA-ribosome complex.
- Post-translational modifications like removal of amino acids, phosphorylation, glycosylation, hydroxylation and protein folding also occur after translation in eukaryotes.
- Recombinant DNA technology: This technology, is also known as DNA cloning,allows insertion of a DNA fragment of interest, such as a new gene, into an organism. The simplest strategy is introducing a new gene into bacteria through plasmid vectors. An outline of gene cloning using plasmid vectors is given.
- PCR amplification of DNA, sequencing of DNA using Sanger's dideoxy chain termination method and the latest Next Generation Sequencing (NGS) methods are explained along with their applications.

EVALUATION

I Fill in the Blanks

1. The information in the DNA can be copied to RNA by a process called _____
2. Name the three different models for DNA replication.
3. A short piece of RNA produced by primase enzyme is called as _____
4. The fragments of DNA in the lagging strand during DNA replication are called _____
5. _____ is an enzyme that strips away the primer(s).

II State True or False

6. In Prokaryotes, DNA replication occurs inside the nucleus.
7. Replication is very slow in Eukaryotes.
8. The eukaryotic mRNAs are mostly monocistronic.

9. In Prokaryotes, introns are absent in the primary transcript.

10. The reaction that makes billions of copies of a small region of DNA is known as _____

11. Transcription is a process by which DNA sequence is read by an enzyme called _____ and a complementary, antiparallel RNA strand is synthesized.

12. RNA polymerase, along with transcription factor sigma, binds to a sequence of DNA called the _____

13. While comparing the coding strand of DNA and a transcribed RNA the base _____is replaced by _____.

14. Name the three types of RNA.

15. _____is type of RNA carries the genetic information from the DNA to the cytoplasm.

16. An mRNA that can contain information for encoding more than one mRNA and are called _____.

17. The cytoplasmic cellular organelle involved in translation process is _____.

18. _____ play a very important role in protein synthesis and acts as molecular "bridges" connecting the mRNA and the growing polypeptide chain.

19. State the different stages of translation process.

20. _____ is also known as DNA cloning which allows us to insert a new gene into an organism.

III Match the following

21. Sanger's Method	-	Bacterial Translation
22. Shine Dalgarno sequence	-	UAG
23. Stop codon	-	AUG
24. Taq polymerase	-	DNA sequencing
25. Start Codon	-	Polymerase Chain Reaction

IV Answer the following questions

1. Write in brief about the central dogma of molecular biology.
2. Describe the Meselson-Stahl experiment and the conformation of semi-conservative model.
3. Explain the events involved in DNA Replication briefly.
4. Compare the different models of DNA replication.
5. Differentiate between the process of DNA replication in prokaryotes and eukaryotes.
6. Explain the role of the DNA polymerase in DNA replication.
7. Describe the steps involved in PCR amplification.
8. Give an Overview of transcription.
9. Compare Prokaryotes and Eukaryotes with respect to Transcription
10. Write briefly about the Molecular events in translation.
11. Write short notes on Recombinant DNA technology.
12. Explain the Sanger sequencing procedure in detail.
13. Discuss about the Post transcriptional modification and Post-translational modifications events.
14. Compare the three main types of RNA in living organisms.
15. Discuss about Next Generation Sequencing (NGS) Technologies and their role in genomics.
16. Explain the different applications of NGS technologies in genomics.

CONCEPT MAP