UNIT-1: INTRODUCTION TO MOLECULAR BIOLOGY

SCOPE AND HISTORY

One of the most outstanding contributions in the field of biological science is the discovery of DNA structure and the role of RNA displayed in the synthesis of protein. The spectacular success in the analysis and comprehensive understanding of genetic code and the nature of gene regulation laid foundation in the field of molecular biology.(This molecular approaches in turn have shown greater diagram of positive influence in understanding traditional disciplines such as genetics, cytology, embryology and evolutionary biology,

Molecular biology has study in any subject of biology is considered incomplete, if it is not studied at the molecular level using techniques of recombinant DNA. For instance, using the techniques of reverse genetics, genes for a variety of morphological traits are being identified and the mechanisms underlying the development of these traits are being uncovered.

Structures and functions of different cell constituents are being resolved at the atomic level, giving a new dimension to the study of cytology, which is more appropriately described as cell biology now. For example cell wall, cell membrane, cytoskeleton, nuclear membrane, chromosomes, nucleolus, different cell organelles (chloroplast, mitochondria, Endoplasmic Reticulum etc.), have been studied achieving very high resolution sometimes approaching atomic level. These developments gave birth to new areas of biology like 'genomics', '*in-silico* biology', 'systems biology', 'integrative biology', 'bioinformatics', etc.

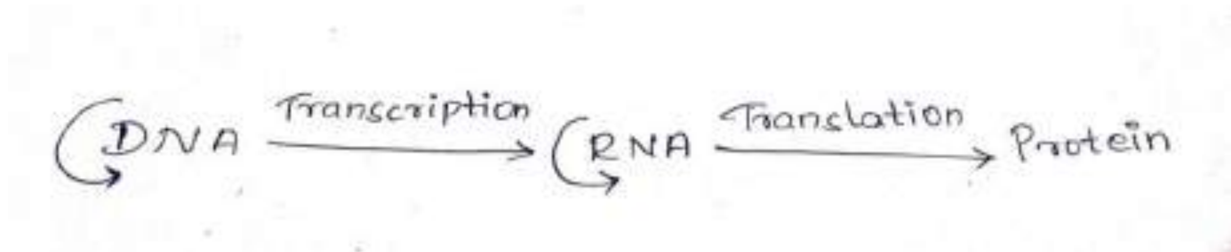
DNA was discovered in 1869 by Johann Friedrich Miescher, a Swiss researcher. The demonstration that DNA contained genetic information was first made in 1944, by Avery, Macleod and McCarty.

CENTRAL DOGMA OF MOLECULAR BIOLOGY

DNA itself cannot directly order the production of amino acids but forms its transcripts first and then is translated into proteins. This unidirectional flow of information is given by central dogma.

DNA is the chemical basis of heredity and may be regarded as the reserve bank of genetic information.

DNA is exclusively responsible for maintaining the identity of different species of organisms over millions of years. Further, every aspect of cellular function is under the control of DNA. The DNA is organized into genes, the fundamental units of genetic information. The genes control the protein synthesis through the mediation of RNA.



The interrelationship of these three classes of biomolecules (DNA, RNA and proteins) constitutes the *Central Dogma of Molecular Biology* or more commonly the *Central Dogma of Life*.

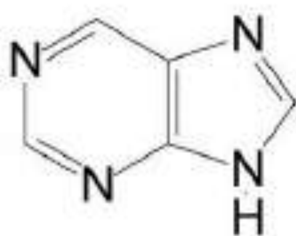
STRUCTURE OF DNA

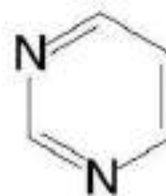
NUCLEOSIDE

Nucleosides are the nitrogen base combined with pentose sugar. In DNA it is known as deoxyribo nucleosides. Free nucleosides occur in small amount in most of the cell.

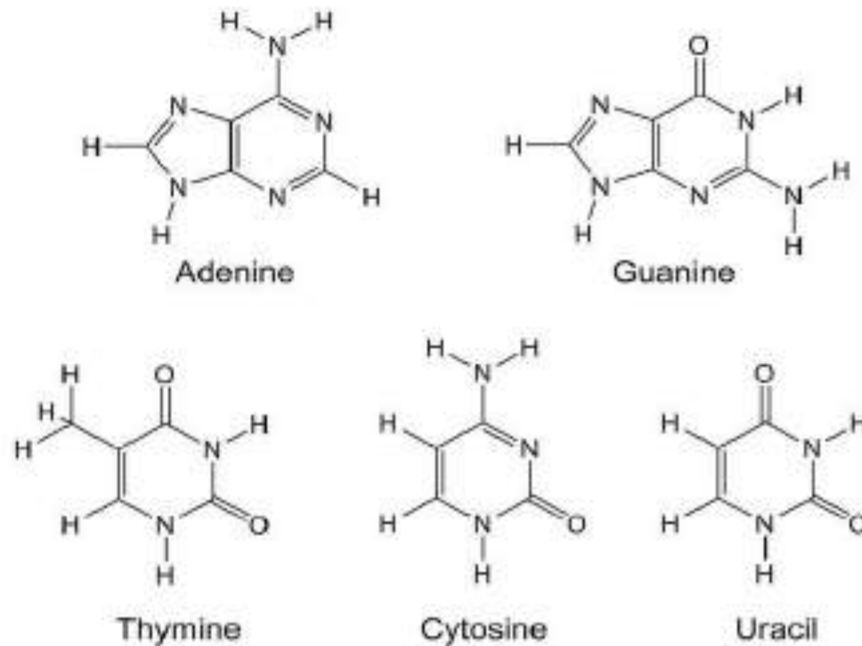
Nucleoside, a structural subunit of nucleic acids, the heredity-controlling components of all living cells, consisting of nucleobase (nitrogenous base) and a five-carbon sugar (either ribose or deoxyribose) and the nitrogen- containing compound is either a purine (adenine or guanine) or a pyrimidine (cytosine, thymine or uracil).

Nucleosides = pentose sugar + nitrogen base



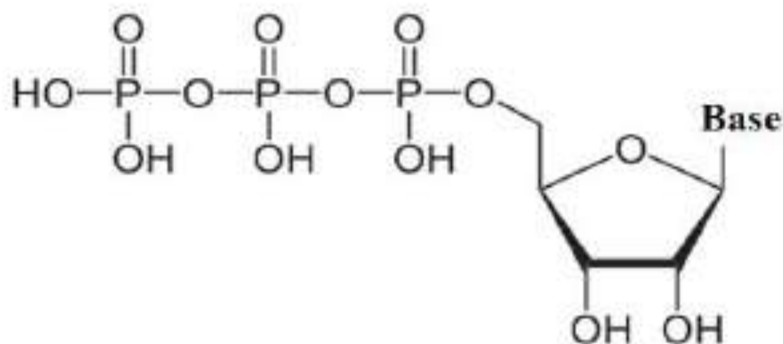


Purine Pyrimidine



NUCLEOTIDE

Nucleotides are composed of a nitrogenous base, a pentose sugar and a phosphate. Nucleotides perform a wide variety of functions in the living cells, besides being the building blocks or monomeric units in the nucleic acid (DNA or RNA) structure. These include their role as structural components of some coenzymes of B-complex vitamins (eg. FAD, NAD+), in the energy reactions of cells (ATP is the energy currency), and in the control of metabolic reactions. [Nucleotide is nucleoside + phosphate].



Structure of Nucleotide

BASE PAIRING

Base pair is nothing but the complementary nitrogenous molecules that are connected by hydrogen bonds. Base pairs are found in double- stranded DNA and RNA, where the bonds between them connect the two strands, making the double-stranded structures possible. Base pairs themselves are formed from bases, which are complementary nitrogen-rich compounds known as purines and pyrimidine.

According to Watson-Crick base-pairing, which forms the basis for the helical configuration of double-stranded DNA, DNA contains four bases: two purines, adenine (A) and guanine (G) and the two pyrimidine, cytosine (C) and thymine (T). Within the DNA molecule 'A' bonds only with 'T' by two hydrogen bonds and 'C' bonds only with 'G' with three hydrogen bonds. In RNA, thymine is replaced by uracil (U).

Each step of the DNA ladder is made up of a purine and pyrimidine base pair i.e. a double ring and a single ring compound held together by hydrogen bonding. 2 purines occupy too much space while 2 pyrimidine will occupy too much little, because of the purine- pyrimidine pairing the total number of purines in double stranded DNA molecule is always equal to the total number of pyrimidines, thus A+G = C+T then, A/T =1 and G/C= 1.

BASE STACKING

Base stacking is a common arrangement of nucleobases found in the three dimensional structure of nucleic acids. Bases (or base pairs) are planar, and these planes stack at constant distance (about 3.4 Å), excluding water and maximizing Van der Waals interactions. In terms of structural stability of nucleic acids in aqueous solution, the stacking interactions of bases play a larger role than the hydrogen bonds formed by the bases.

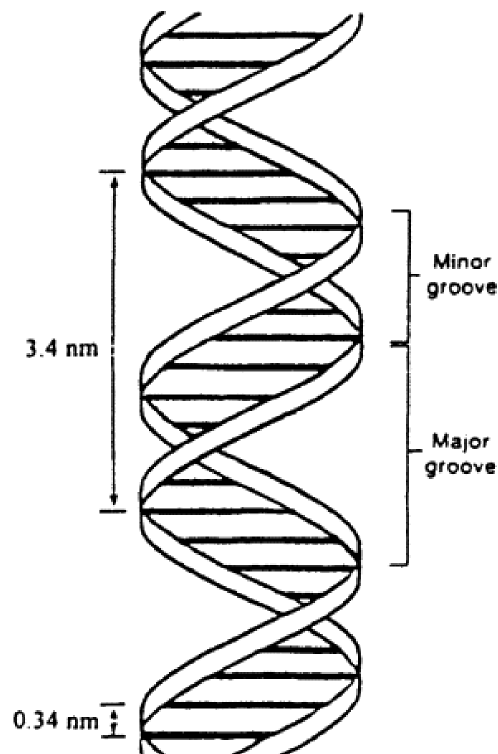
DOUBLE HELIX

The double helical structure of DNA was proposed by James Watson and Francis Crick in 1953 (Nobel Prize, 1962). The elucidation of DNA structure is considered as a milestone in the era of modern biology. The structure of DNA double helix is comparable to a twisted ladder.

FEATURES OF WATSON AND CRICK MODEL

The salient features of Watson - Crick Model of DNA are given below:

1. The DNA is a right handed double helix. It consists of two poly deoxy ribo nucleotide chains (strands) twisted around each other on a common axis.
2. The two strands are antiparallel, i.e. one strand runs in the 5' to 3' direction while the other in 3' to 5' direction. This is comparable to two parallel adjacent roads carrying traffic in opposite direction.
3. The width or diameter of a double helix is 20 Å (2 nm).
4. Each turn (pitch) of the helix is 34 Å.
5. Each strand of DNA has a hydrophilic deoxyribose phosphate backbone (3'-5' Phospho diester bonds) on the outside of the molecule, while the hydrophobic bases are stacked inside.
6. The two polynucleotide chains are not identical but complementary to each other due to base pairing.
7. The two strands are held together by hydrogen bonds formed by complementary base pairs. 'A' bonds only with 'T' by two hydrogen bonds and 'C' bonds only with 'G' with three hydrogen bonds.
8. The hydrogen bonds are formed between a purine and a pyrimidine only. If two purines face each other, they would not fit into the allowable space. And two pyrimidines would be too far to form hydrogen bonds. The only base arrangement possible in DNA structure, from special considerations is A-T, T-A, G-C AND C-G.
9. The complementary base pairing in DNA helix proves Chargaff's rule. The content of adenine equals to that of thymine (A=T) and guanine equals to that of cytosine (G=C0.
10. The genetic information resides on one of the two strands known as template strand or sense strand. The opposite strand is antisense strand. The double helix has (wide) major grooves and (narrow) minor grooves along the Phosphodiester backbone. Proteins interact with DNA at these grooves, without disrupting the base pairs and double helix.



Structure of DNA

MAJOR AND MINOR GROOVES

Due to the double helical structure of two strands the DNA is long extended polymer with two grooves of different size. It is due to change in geometry of the base pair. i.e., when the angle between glycosidic bonds is 120° it results in minor groove. When the angle between glycosidic bonds is 240° it results in major groove. Thus, more and more base pairs stack on top of each other.

SUPERCOILING

- The term 'supercoiling' means literally the coiling of a coil.
- DNA supercoiling is generally a manifestation of structural strain.
- Supercoiling occurs when the molecule relieves the helical stress by twisting around itself. Over-twisting leads to positive, while under-twisting leads to negative supercoiling.
- If DNA is in the form of a circular molecule, or if the ends are rigidly held so that it forms a loop, then over-twisting or under-twisting leads to the supercoiled state.
- Supercoiling is important in a number of biological processes, such as compacting DNA, and by regulating access to the genetic code.

- DNA supercoiling strongly affects DNA metabolism and possible gene expression.

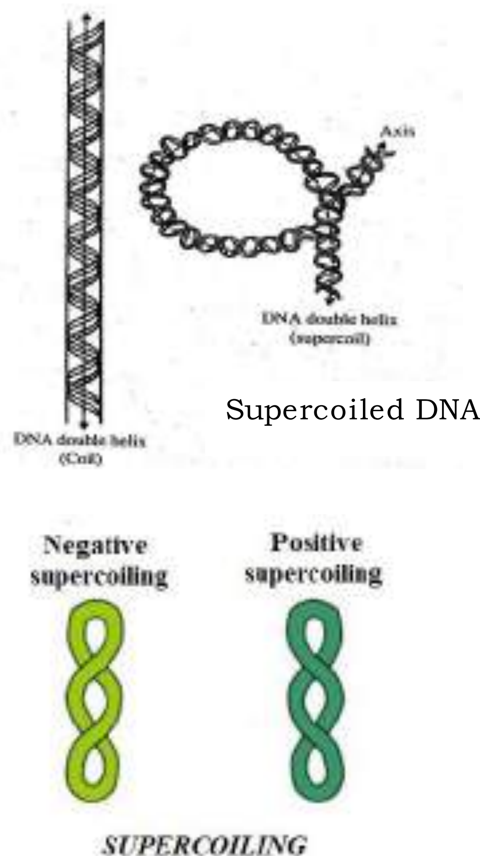
Based on the nature of twisting two types of super coils can result viz.

Negative Super Coils:

- Negative supercoiling is the left-handed, double helical form of DNA.
- Prokaryotes and eukaryotes usually have negative supercoiled DNA.
- It is naturally prevalent as it prepares the molecule for processes that require separation of the DNA strands without the need of additional energy.

Positive Supercoils:

- Positive supercoiling is the right-handed, double helical form of DNA.
- It is twisted tightly in a right handed direction until the helix creates knot.
- Positive supercoiling is more condensed as the supercoil forms at the direction of DNA helix.



Supercoiled DNA

Twist:

- It is the total number of turns of duplex.

Writh:

- It is the turning of the axis of the helix in the space.

Linking number:

- Which is defined as the "number of times that the two strands of double helix of a closed molecule cross over each other". It has two terms; namely twist (T) and writh (W).

$$L = T + W$$

- The numbers of times that two strands are interwined.
- It is also the number of cleaves necessary for separating 2 DNA strands.

Super helical density can be assayed from the topoisomerase activity.

FORMS OF DNA

DNA can exist in 4 forms on the basis of sugar puckering. They are A, B, C and D forms.

- A-DNA- in this type 11 bp are present in one full turn of the helix.
- B-DNA- in this type 10 bp are present in one full turn of the helix.
- C-DNA- in this type 9 bp are present in one full turn of the helix.
- D-DNA- in this type 8 bp are present in one full turn of the helix.
- Z-DNA- in this type 12 bp are present in one full turn of the helix.

Among the DNA mentioned, Z-DNA shows left handed helix and remaining shows right handed helix.

- **A-Form DNA**

ĐNA can undergo reversible conformational change and give rise to A-form up on dehydration. It predominates if relative humidity is 75%. The diameter of the helix is 25.5 Å. There are 11 base pairs per turn. The rise per base pair (28Å/11) is less than that in B-DNA.

Salient features of A-DNA

- It has very deep major groove and very shallow minor groove.
- The extra width of the A-DNA goes in to creation of a 6 Å cylindrical hole at the centre of the helix.
- The diameter of helix is 25.5 Å and can accommodate 11 base pairs per turn of the helix.
- The pitch per turn of helix is 24.6 Å.

➢ **B-Form of DNA**

It is the most stable configuration. It is a right handed double helix and can be purified when the relative humidity is 92% in the medium. The B-form of DNA was explained by Watson and Crick. It is long thinner with 3.4 Å thick. The purines and pyrimidines are stalked perpendicular to the helical axis. The diameter of the helix is 20 Å. There are 10 bases per turn of the helix. The pitch of the helix is 34 Å.

Salient features

- It is regarded as genetic material, where it carries heredity characters from one generation to another.
- It takes part in expression of phenotypic character of organism.
- It takes part in biosynthesis of different kinds of RNA and protein.
- Progeny rod undergoes circularization to form double stranded circular DNA. Thus at the end 2 double stranded daughter circular DNA molecule are formed. Each with one parental strand and one new strand.
- It undergoes mutation, responsible for the evolution of new varieties and species.

➢ **Z- form of DNA [Zig-Zag DNA]**

This was discovered by A. Wang and Alexander Rich, 25 years after the discovery of the Watson-Crick structure. It is a left handed helix of a bizarre confirmation. This is due to zig-zag arrangement of phosphate ribose backbone. This is due to presence of alternative purine and pyrimidine residues. The glycosidic bond can be in one of the two conformations called ‘syn’ and ‘anti’.

Salient features of Z- DNA

- 12 base pairs (bp) can be seen with in each helical turn of 37 Å.
- It is found in natural samples when a dinucleotide repeats are present such as d (CGCGCG).
- The Z-DNA formation is extremely sequence and composition dependent.
- The formation of Z-DNA is solvent dependent. The DNA can acquire Z-DNA conformation in presence of alcohol, along with very high ionic strength.

STRUCTURE AND FUNCTION

mRNA (messenger RNA)

- Is transcribed on the DNA template and therefore carries the genetic information of the DNA.
- It was Jacob and Monad (1961) proposed the name m-RNA for bearing the transcripts of the DNA for protein synthesis.
- The total population of the mRNA is about 5-10% of the total RNA present.
- The mRNA is short lived, since they are easily broken down into the ribonucleotides by the enzyme ribonuclease.
- In the *E-coli* the mRNA are short lived and are alive only for 2 min, therefore the cell does not contain high concentration of the mRNA at a time. In contrast the mRNA of the eukaryotes is much stable.
- The mRNA is single stranded and sometimes are coiled but the coils lack the complementary bases.

The structure of the mRNA of prokaryotes and the eukaryotes contain:

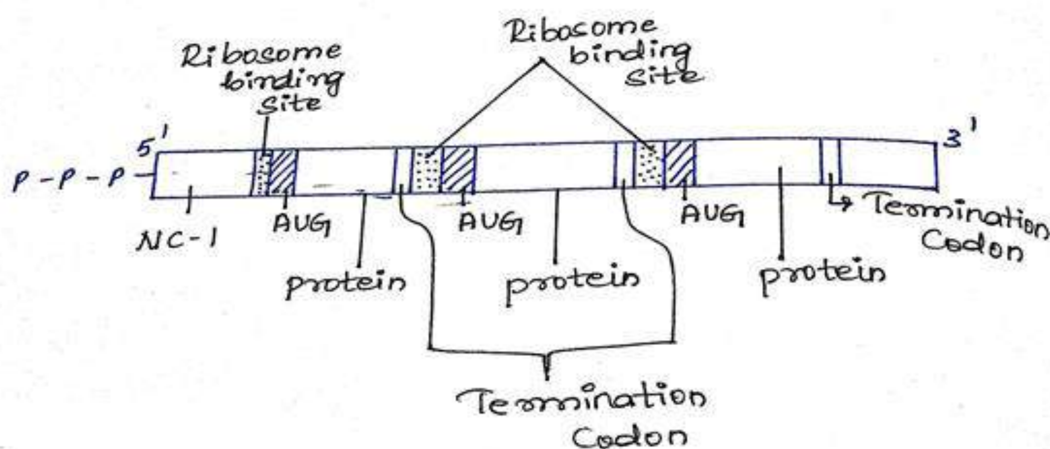
1. ***5' cap***- this is seen and is of importance in the eukaryotes and the animal viruses. The 5' end of the mRNA contains a cap which is formed after the methylation of any of the four nucleotides.
2. ***Non-coding regions***- there are 2 non-coding regions first followed by the cap and the second followed by the termination codon.
3. ***Initiation codon***- both the prokaryotes and eukaryotes contain the initiation codon AUG/ GUG. The binding and the start of the protein synthesis takes place in the bacterial cell. The ribosomes directly bind to the initiation codons and the protein synthesis begins.

4. ***Coding region***- it is the most important of mRNA which is about 1500 nucleotides long and help in the translation and production of long chains of amino acids (proteins). After attaching with several ribosomes the combination of this mRNA strand with several ribosomes is called the polyzones.

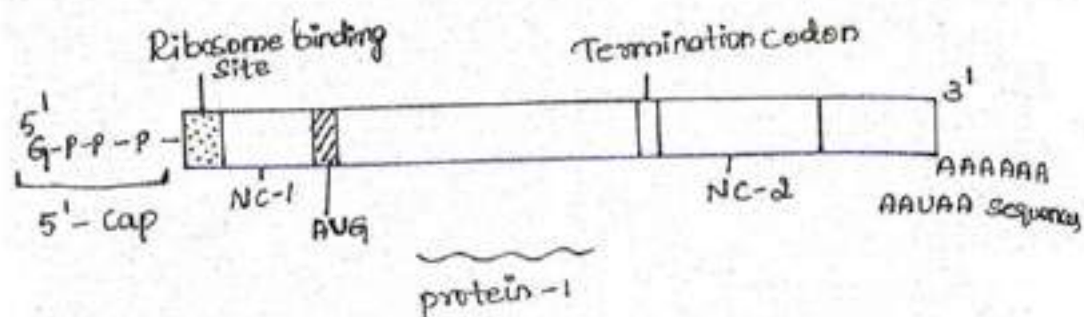
   The prokaryotic mRNA is polycistronic as it produces multiple proteins that as separately translated from the same mRNA molecule.

   Whereas the eukaryotic mRNA is typically monocistronic as it produces only one polypeptide chain per mRNA molecule.

5. ***Termination codon***- the termination codon is required to give the stop signal to the protein synthesis. In the eukaryotes there are 3 termination codons namely UAA, UGA, UAG, these codons bring about the termination of translation process.



Structure of Prokaryotic mRNA



Structure of Eukaryotic mRNA

Functions of mRNA

- The mRNA carries genetic information from DNA to ribosomes.
- These ribosomes then use the mRNA as template to translate the genetic information into amino acid sequence of protein.

rRNA (ribosomal RNA)

rRNA forms about 80% of the total cellular RNA and is quite stable, rRNA is associated with ribosomes which are factories of the protein synthesis. 4 kinds of rRNA are found in animals, plants and prokaryotes which are distinguish on the basis of their sedimentation coefficients.

In prokaryotes the molecular species of rRNA include 16s, 5s and 23s types, whereas eukaryotes have 18.5s, 5.8s and 28s species. In ribosomes, rRNA is found to be complexed with 50 different protein molecules.

Function of rRNA

- It makes complex with proteins and form ribosomal subunits which provide space for protein synthesis.
- mRNA also contains multiple regulatory regions that can determine the timing and rate of translation.

tRNA (transfer RNA)

20 different amino acids required for the protein synthesis are present in the cytoplasm before joining an appropriate amino acid together to form a protein they are activated by attaching to the RNA. The catalyzation of these attachments takes place in the presence of amino acyl tRNA synthetase.

Structure of tRNA/ Clover Leaf Model

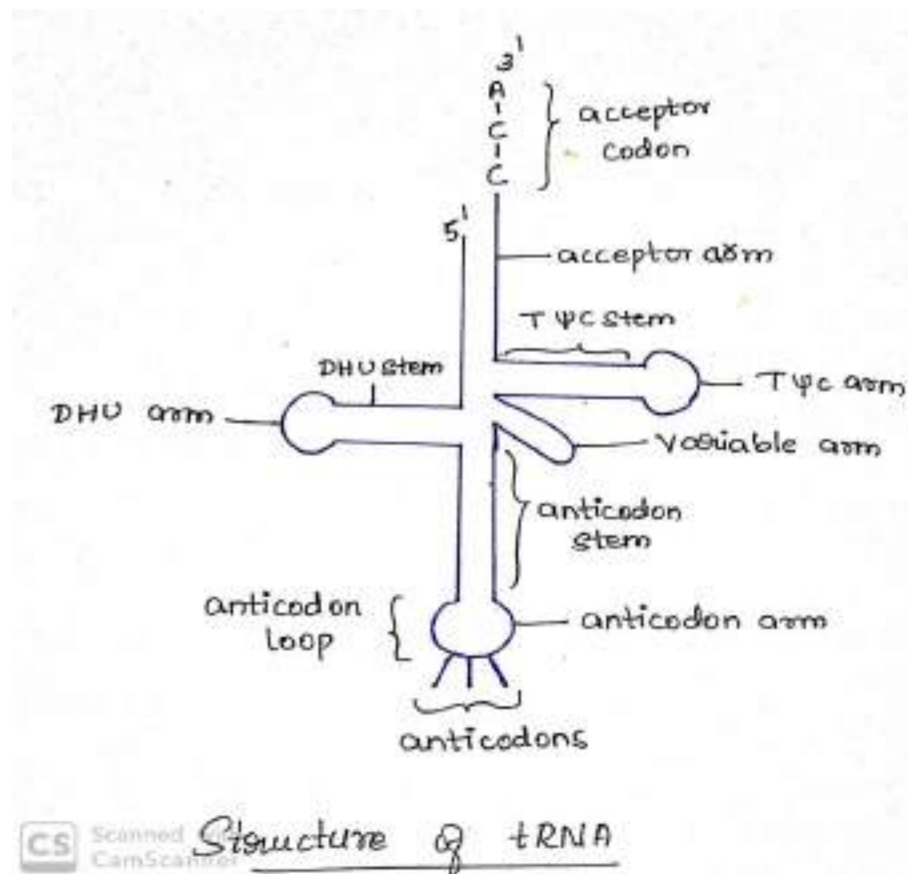
For the first time, R Holley (1968) isolated the tRNA from the yeast cells and proposed the ‘Clover Leaf Model’ to explain the structure of tRNA. The typical Clover Leaf structure of the tRNA has the following features:

1. A single polynucleotide chain of tRNA is folded to form 5 arms, namely
    a. Acceptor arm
    b. DHU arm

c. Anticodon arm
d. Variable arm
e. TΨC arm

An arm consists of a stem and a loop. Except the acceptor arm the rest 4 arms consists of their respective stems and loops.

- The acceptor arm consists of 7 basepairs and 4 unpaired bases containing a 3-CCA sequence and the 4th variable purine (A or G) at the 3' end of the poly nucleotide chain. The last residue acts as the amino acid attachment site.
- The DHU arm namely Dihydro uridine arm consists of 7-12 unpaired and acts as site of recognition of the amino acid activating enzyme amino acyl tRNA synthetase.
- All the tRNA molecules contain different nucleotide triplet codon on the anticodon loop, it is also called the anticodon o the codon recognising site. It is complementary to the codon of the mRNA molecule.
- The tRNA also possesses the TΨC arm, which consists of stem of 5 base pairs and a loop of 7 unpaired bases, including pseudouridine. The TΨC arm has a ribosome recognising site and helps in the binding of the tRNA molecule to the ribosomes.
- In some of the tRNA especially those with the long chains of variable arm is present between the anticodon arm and TΨC arm.
- The variable arm may or may not contain the stem, the tRNA molecule that initiates the protein synthesis is called the initiator tRNA.
- The initiator tRNA of the eukaryotes differs from the prokaryotes. The tRNA specifies methionine as the starting amino acid in the eukaryotic protein synthesis and N-formyl methionine in the prokaryotes.



Structure of tRNA

Functions of tRNA

- Help in the recognition of Aminoacyl tRNA synthetase enzyme.
- Picks up specific amino acid from cytoplasm and carries to site of protein synthesis.
- Attaches itself to ribosome in accordance with sequence specified by mRNA.
- Transmits amino acid to polypeptide chain.
- Participate in non-protein synthetic processes such as a primer during reverse transcription in retrovirus life cycle.

SECONDARY STRUCTURES IN RNA

Secondary structure is the set of interactions between bases, i.e., which parts of strands are bound to each other. The secondary structure of RNA consists of a single polynucleotide. Base pairing in RNA occurs when RNA folds between complementarity regions. Both single- and double-stranded regions are often found in RNA molecules. The antiparallel strands form a helical shape. The four basic elements in the secondary structure of RNA are helices, loops, bulges, and junctions. Stem-loop or hairpin loop is the most common element of RNA secondary structure. *[Clover Leaf Model].*

UNIT-2: DNA REPLICATION AND REPAIR

INTRODUCTION

Ability of DNA molecule to make exact copy of its own is called DNA replication.

DNA replication takes place during S-phase of the interphase of the cell cycle.

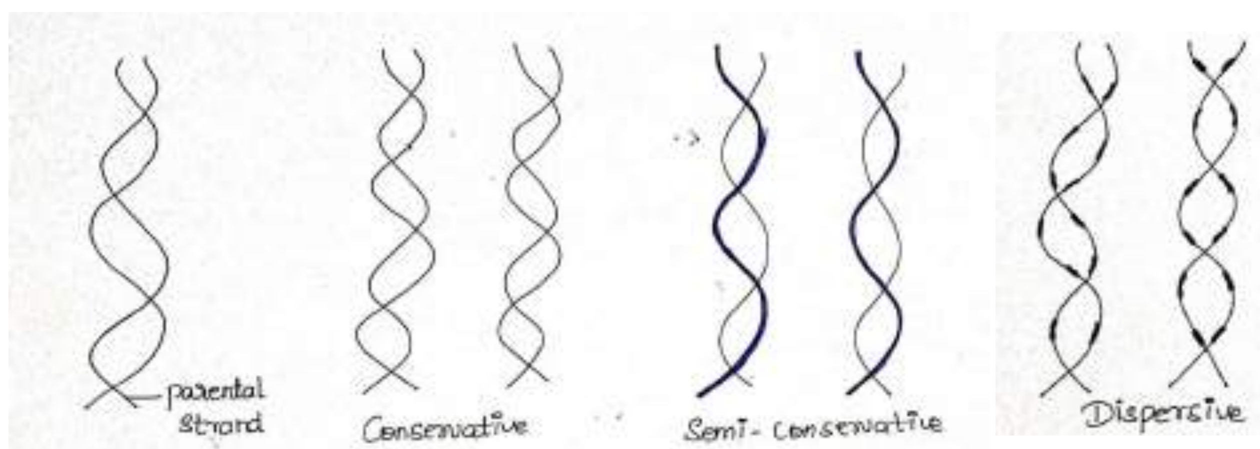
DNA replication takes place in the following 3 ways.

1. Conservative type
2. Semi-conservative type
3. Dispersive type

Conservative type: according to it the original DNA molecule (parental DNA) remains unchanged and the other is formed entirely new.

Semi-conservative type: according to it two daughter DNA molecules formed would possess one strand completely new and one is parental.

Dispersive type: according to it the parental DNA breaks at several points, each piece would replicate and the pieces would join randomly. Thus the 2 strands consists of some portions are new and some portions are old pieces dispersed through the molecules.



Among these semi-conservative type is the most common type of DNA replication.

TYPES AND FUNCTION OF DNA POLYMERASES IN PROKARYOTIC AND EUKARYOTIC REPLICATION

In 1957, Korenberg discovered an enzyme in *E.coli* which is known as DNA polymerase.

The following enzyme involves in DNA replication in Eukaryotes:

- **DNA polymerase α:** this enzyme initiates DNA replication. In DNA template, it functions at both leading and lagging strands to initiates the replication process. Since it has primase actively (RNA polymerase). It takes part in the initiation of DNA replication in both leading and lagging strands.it has 4 subunits. It is present in the nucleus.

  Functions:
  - Polymerisation activity: it helps in polymerisation of nucleotide units in both leading and lagging strands of DNA template.it has primase activity, hence it requires to initiate DNA replication in both leading and lagging strand.

- **DNA pol β:** it is a single polypeptide chain with molecular weight is 30,000 to 50,000. It is present in the nucleus.

  Functions:
  - It involves in DNA repair and polymerisation.
- **DNA pol γ:** it is a single polypeptide chain with molecular weight ranging from 150,000 to 300,000. It is present in only in mitochondria.

  Functions:
  - It involves in polymerisation of mitochondria/ DNA during replication.
- **DNA pol Σ (sigma):** this enzyme takes part in replication of DNA results in the formation of continuous strand called leading strand on DNA template.

  Functions:
  - It involves in polymerisation of nucleotide units in 5’ to 3' direction on leading strand.
- **DNA pol δ (delta):** this enzyme takes part on replication of DNA which results in the formation of discontinuous strand called lagging strand. It is composed of 2 sub-units and present in nucleus.
  - ***Functions:*** It involves in polymerisation of nucleotide units in 5' to 3' direction on lagging strand.

The following enzymes are involved in DNA replication in Prokaryotes.

There are 3 types of DNA polymerases namely,

1. DNA polymerase I
2. DNA polymerase II
3. DNA polymerase III

- **DNA polymerase I**: it is regarded as a DNA repair enzyme. It is a single polypeptide chain, its molecular weight is about 109,000 and it consists of a zinc atom per chain. Hence, it is also regarded as a 'metallo-enzyme'. It is more or less spherical in shape with diameter 65Å. It is attached at regular intervals to the DNA chain.

  Further the DNA pol I consist of following 3 types of enzymes.
  - **5' - 3' Polymerase**- this enzyme catalyses the synthesis of new DNA (polymer) from its nucleotides (monomers) only in 5'- 3' direction is called polymerisation.
  - **3'- 5' exonuclease**- is an enzyme that degrades nucleic acids from one end and is unable to make internal cuts in a polynucleotide chain.
  - **5' - 3' exonuclease**- this enzyme functions in removal of the DNA segment damaged by UV light, irradiation and other agents in 5' - 3' direction. This enzyme plays an important role in the removal of thymine- dimer there by continues the process of replication.

  Functions
  - Polymerisation activity- it can synthesise only short DNA segment by removing short RNA primer on DNA template. It polymerises about 100 nucleotide molecules per minute on DNA template.
  - Exonuclease activity- it removes nucleotide units in both 3' ⟶ 5' and 5' ⟶ 3' strand.

- **DNA pol II**: it is a single polynucleotide chain with a molecular weight of about 90-120 KDa.

  Functions
  - *Polymerisation activity*- just like polymerase it also synthesise long DNA segment by polymerisation activity on DNA template.
    It polymerises about 40 nucleotide molecules per minute on DNA template.
  - *Exonuclease activity*- like DNA pol-I, it also helps to remove nucleotide units of DNA strand in 5' ⟶ 3' strand.

    - It acts as editing on DNA repairing factor in DNA repair and replication.
    - In the absence of polymerase I, this enzyme can take up the responsibility of elongating Okazaki fragments.

- **DNA pol III:** it consists of 2 polypeptide chain with a molecular weight of about 140 KDa. It is unlike DNA pol-I and DNA pol-II multimeric in nature, where it is composed of 10 subunits remaining structural features of this enzymes are similar to that of DNA pol-I.

  Functions

    - *Polymerisation activity*: it can synthesise long DNA strand on DNA template in 5'-3' direction only. Where it polymerises DNA nucleotide units 15000 per minute on DNA template.
    - *Exonuclease activity:* it removes nucleotide units only in 3' ⟶ 5' direction of the DNA strand.

PROOF READING ACTIVITY

During the replication of DNA the correct base sequence is maintained this is possible due to the presence of DNA polymerase with exonuclease activity. This capability of DNA polymerase to maintain the correct base sequence in DNA is known as **fidelity**.

In *E.coli* for every 10^4 nucleotides a wrong base is incorporated into daughter DNA during replication. If this is not repaired it could potentially damage the cell when DNA undergoes replication. If this is not repaired it could potentially damage the cell when DNA undergoes further replication.

The presence of such mechanism in the living cells were discovered from the studies based on the *E.coli*. It has been revealed that the 3'- 5' exonuclease activity of the polymerase-I is capable of rectifying such mis- incorporations.

Types of DNA damages

The DNA damages may be due to any of the following reasons:

1. Substitution of a wrong base in one of the strand during replication. It results from the replication error where it is not corrected by editing function.
   Ex: Replacement of uracil by thiamine by spontaneous deamination.
   It can be corrected by specific N-Glycosylase enzyme.

2. Depurination in DNA strands in the presence of chemicals like alkylating agents [ex: Ethyl ethane sulphonate, methyl-bis-amine] that breaks the N-Glycosidic bond of a purine nucleotide.
3. Variety of ionizing radiations and some chemicals can change the bases in to entirety different compounds. Sometimes the free-radicals produced during metabolism can alter the bases. Ex: Thymine dimers Due to the formation of dimers the DNA get distorted that Inhibits advancement Replication fork during DNA replication.
4. Single and double strand Nicks A large number of agents [like peroxides, metal ions, sulphydroxy compounds] can cause cleavage of phosphodiester bonds. Some tines the ionizing radiations can also cause such nicks in DNA.
5. Sometimes certain antibiotics like mitomycin can cause cross-linking of the bases on one strand with opposite base on the complementary strand.

Depending on the basis of mechanism of action, the DNA repair systems are of following types.

1. **Direct repair system:** It involves removal of a single mismatched base pair from the DNA. Ex: Photo reactivation of thyamine dimers.
2. **Excision repair system:** It involves the removal of a stretch of DNA that contain the dam- age during which normal DNA on either side of damaged DNA is also removed. The cleave site in such system may be either general or specific (where cleavage occurs as apurinic site).
3. **SOS repair:** It helps the DNA to tolerate the damage which is achieved by passing problem of DNA replicating.
4. **Recombination repair system:** In this system there is reconstruction of functional DNA molecule from undamaged fragments. Thus the information lost during damage from the gene can be recombined by this mechanism.
5. **Mismatch repair system:** It involves the removal of mismatched base from one of the strand and joins corresponding correct base pairs in to the DNA which is complementary to the base present on the opposite strand.

5' - 3' EXONUCLEASE ACTIVITY: is a protein encoded by a gene. It removes nucleotides from 5' end of primer by hydrolysis. This enzyme functions in removal of the DNA segment damaged by UV light, irradiation and other agents in 5' - 3' direction.

This enzyme plays an important role in the removal of thymine- dimer there by continues the process of replication.

ENDONUCLEASE ACTIVITY: these are enzymes capable of making an incision within the inner portion (middle) of the double stranded DNA molecule in which the original strand acts as DNA template for the synthesis of new strand based on base complementary rule.

There are 2 types:

- 5' - 3' endonuclease activity
- 3' - 5' endonuclease activity

TOPOISOMERASE ACTIVITY

These are the class of enzymes that helps to relax the torsional strains generated by unwinding duplex DNA. It catalyses the inter conversion of different topological isomers of DNA. They are of two types; namely topoisomerase I and topoisomerase II.

- **Topoisomerase-I:** Cleaves only one strand of duplex DNA.
- **Topoisomerase-II:** It cleaves both the strands of duplex DNA.

The cuts usually product 5' - OH group and 3' - Phosphate which is covalently linked to a tyrosine in the enzyme. Generally these enzymes are involved in relaxation of super coils except DNA gyrase of *E.coli* (topoisomerase-II) that generate negative supercoils.

TELOMERIC DNA REPLICATION

As DNA polymerase alone cannot replicate the ends of chromosomes, telomere aids in their replication and prevents chromosome degradation.

The telomeric sequence is maintained by a special in rhyme called telomerase enzyme.

Telomerase is an enzyme in eukaryotic cells that adds a specific sequence of DNA to the telomeres of chromosomes after they divide giving the chromosomes stability over time. While replicating DNA, The Eukaryotic DNA replication in Science cannot replicate the sequences present at the ends of the chromosomes.

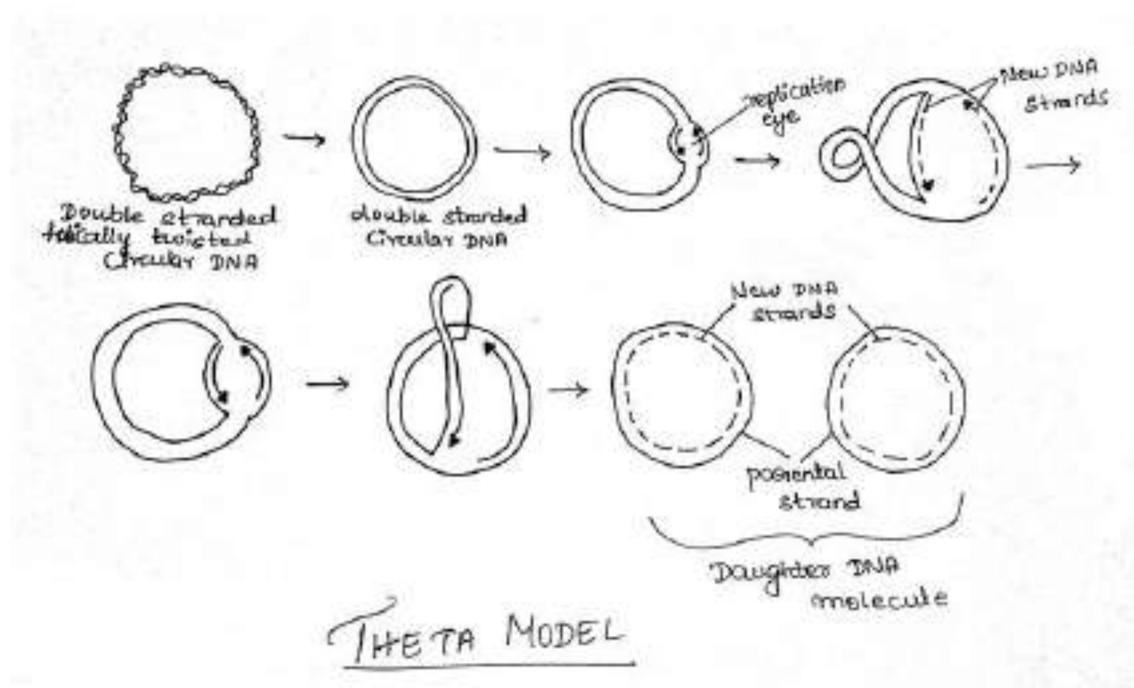
The ends of the linear chromosomes are known as telomeres; repetitive sequences that code for no particular gene. Hence, these sequences and the information they carry may get lost. Telomeres 'cap' the end sequences and themselves get lost in the process of DNA replication.

Telomeres are repetitive nucleotide sequences located at the terminal of linear chromosomes of most Eukaryotic organisms. Most prokaryotes do not have telomeres.

PLASMID REPLICATION

THETA MODEL/ CAIRN'S MODEL

This model was put forward by Cairn's. This type of replication is usually found in *E.coli* and *Bacillus subtilis*.



According to this model DNA replication takes place in following manner or steps:

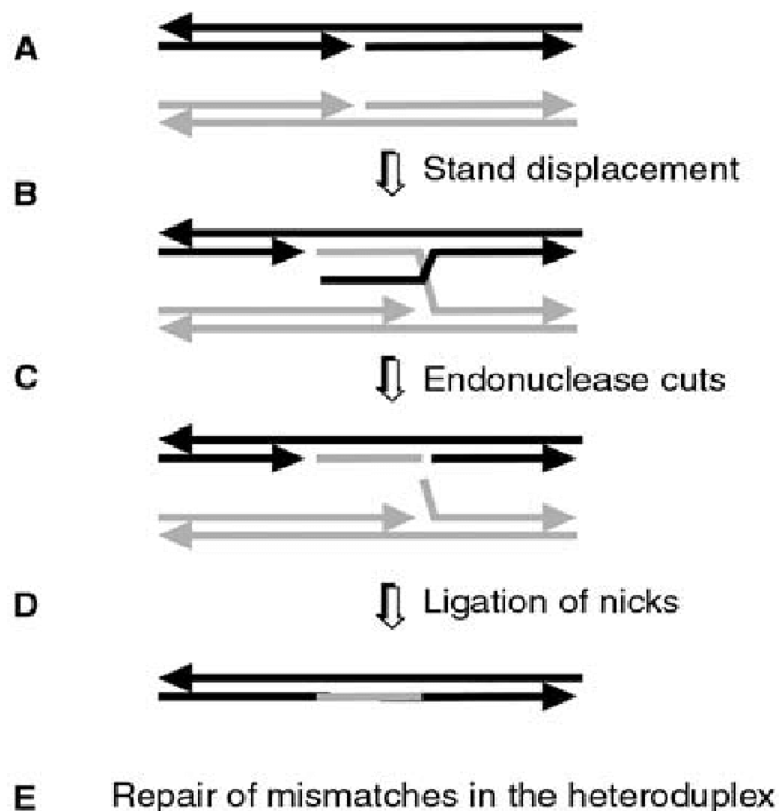
1. Unwinding of double stranded circular helically coiled DNA takes place at specific point called origin. This results in the formation of bi-directional replication eye.
2. As the unwinding of double stranded DNA helix proceeds from the point of origin replication of DNA continues simultaneously.
3. Unwinding creates super twisting of 2 strands. This leads to tension of the strands.
4. To release the tension temporary break and reunion of the strand occurs.
5. This results in free rotation of the 2 strands and replication process continues further.

Thus at the end of replication 2 daughter DNA molecules are formed each with one parental strand and other new strand.

STRAND DISPLACEMENT MODEL

Strand displacement replication has been associated with broad host range plasmids. The mechanism involves the parents duplex strand have two single stranded replication initiation sites.

Vertical lines shows hybridization between DNA strands. Duplex melting is through the Helicase enzyme allowing to sites to form hair pin. The hair pin bas are recognised by RepB, which initiates the synthesis of RNA primer. The extension takes place at 3'- OH of Primer. 2 daughter loops are formed as parents strands are displaced and dissociated from each other, leaving single stranded DNA intermediates. Synthesis continuous in both the direction extending area of the loop formation. Elongation is completed and termination occurs at SSI sides in which replication began. At this point, the newly synthesized daughter strands are restored. Segregation of two daughter strands are ligated, resulting into DNA duplexes each containing a parental strand and a daughter strand.



ROLLING CIRCLE MODEL OR σ MODEL

This model was proposed by Gilbert and Dressler in 1968. This type of replication is found in bacteria that carry out conjugation, a genetic recombination process.

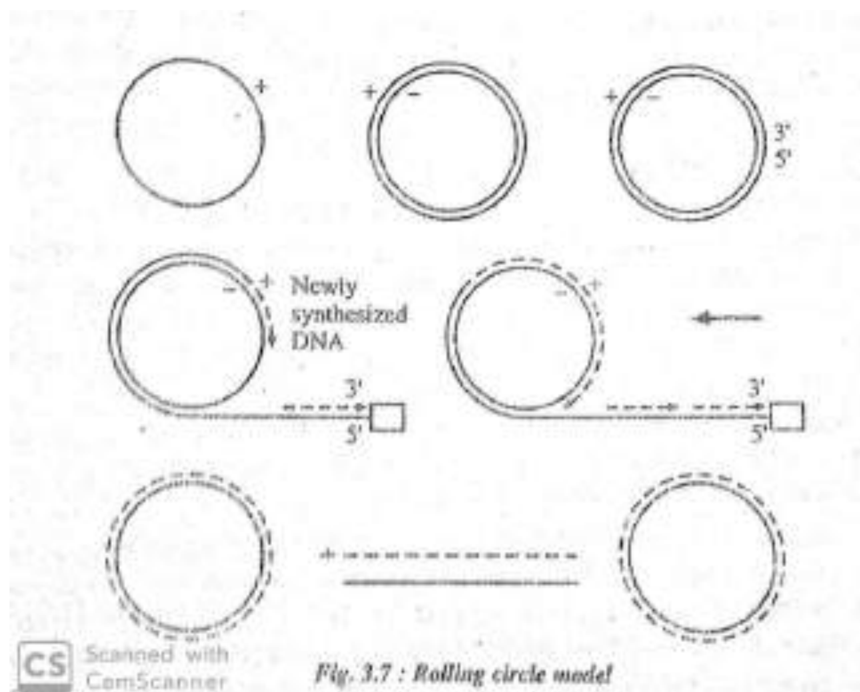


Fig. 3.7 : Rolling circle model

1. Unwinding of double stranded helically coiled circular DNA takes place, this results in a separation of 2 rings.
2. One of the parental DNA ring breaks at a specific point by an endonuclease which results in the formation of a linear strand with 5' - 3' end attached to host membrane and other remains as an intact strand.
   The 3' end serves as a primer for the synthesis of new DNA strand under the catalytic action of DNA polymerase and intact DNA strand acts as a DNA template, where it rolls up 360° and synthesis new strand based on base complimentary rule.
3. As replication proceed on DNA template broken ring separates as tail segment.
4. Tail segment also acts as a DNA template and synthesises new DNA strand to become double stranded progeny rod.
5. Then progeny rod undergoes circularisation to form new double stranded circular DNA.

6. Thus at the end of replication 2 daughter double stranded circular DNA molecules are formed, each with one parental strand and other new strand.

DNA REPAIR

For an organism to survive and reproduce; it has to guard its genetic information i.e, to address the challenges possessed by the physical and chemical agents, the cell has evolved several DNA repair mechanisms. A wide variety of repair systems have been observed in both prokaryotes and eukaryotes.

❖ NUCLEOTIDE EXCISION REPAIR

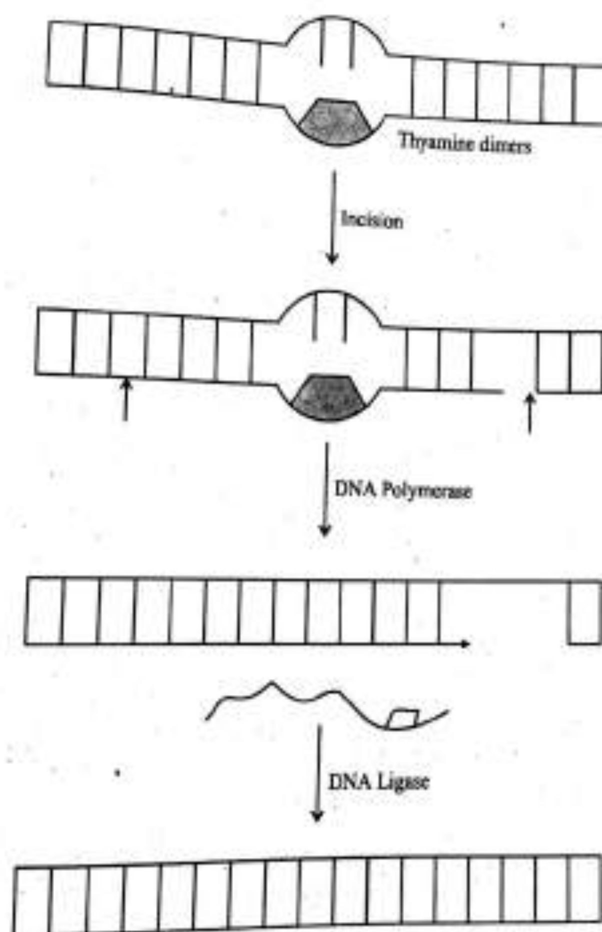
Large distortions in the helical structure of DNA are repaired by Nuclear excision repair. Here the enzyme nicks the DNA on either side of damage, this results in a fragment of 12 to 13 nucleotides in *E.coli* where the fifth Phospho diester bond on the 3' end and eight Phospho diester bond on the 5' end are hydrolysed. The dual incision is followed by the removal of the oligonucleotides and the gap is filled by DNA polymerase I and followed by ligation by DNA ligase.

Incision activity in *E.coli* can be attributed to complex products of 3 different genes uvr-A, uvr-B and uvr-C. The nicked DNA fragment is removed by the Helicase-II (encoded by uvr-D).

Following Table gives a list of different gene products and their functions that are involved in Nucleotide excision repair.

Proteins of Nucleotide excision repair	
Gene product	**Function**
Uvr A	It act as DNA binding protein
Uvr B	Binds to the Uvr A DNA complex and nicks DNA at 3' end of Damage
Uvr C	Binds Uvr B - DNA complex and nicks DNA at 5' end of damage.
Uvr D	[Helicase-II] unwinds DNA and removes damaged DNA
Pol-A	Functions as polymerase-I and helps in Gap filling
Lig	Helps in Gap sealing

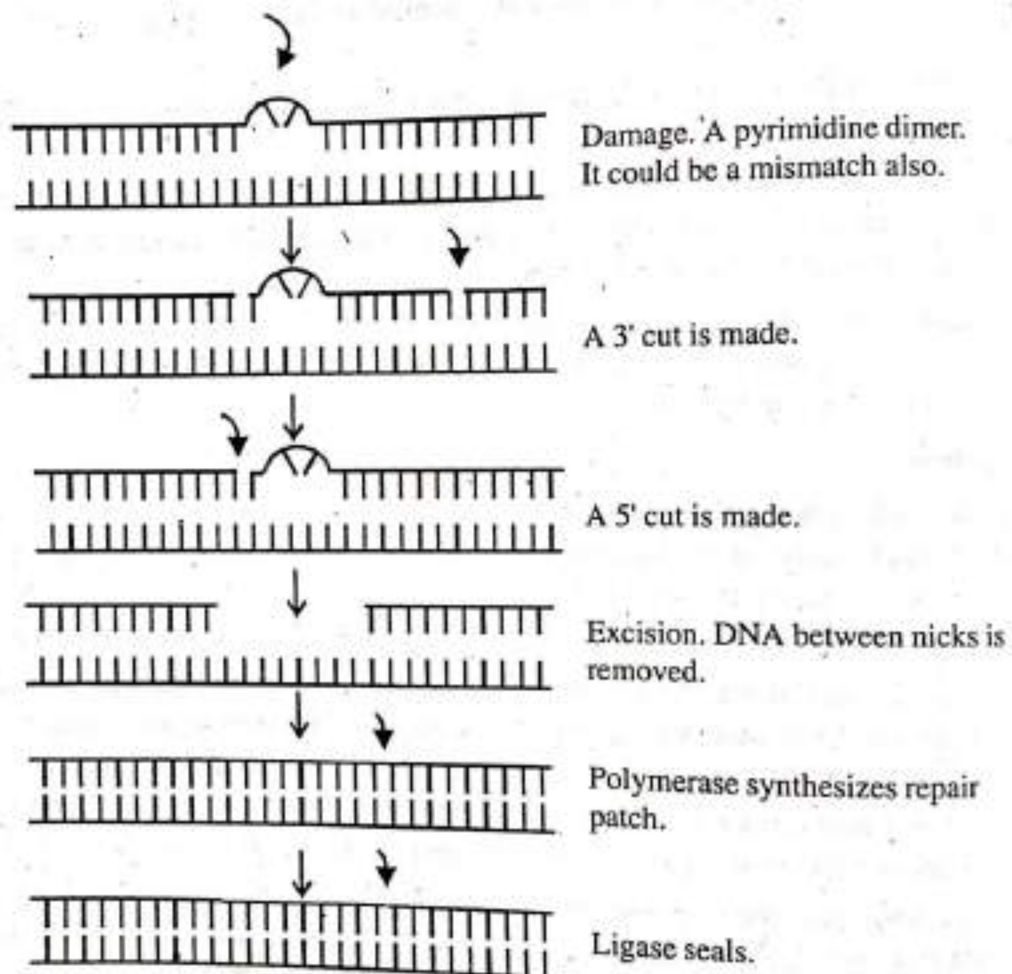
Eukaryotic system contains most complex incision-Excision activities and require more enzymes and is least understood.



Mechanism of Nucleotide Excision Repair in E.coli

❖ BASE EXCISION REPAIR

It is the most versatile, efficient non-specific and common to all organisms. It is a multistep enzymatic process. It is employed to remove very small DNA lesions such as products of deamination of cytosine, thyamine (uracil) or alkylated bases like 3-methyl adenine or other modified bases like 8-oxo guanine. The lesions are removed by cleaving glycosyl bond catalysed by "DNA Glycosylase". This gives rise to a basic sugar which is often known as apurinic or apyrimidinic site or **APsite**. Once the APsite is formed another group of Enzymes known as "AP-endonucleases" which nicks the DNA strand at upstream of the APsite, to create a free 3' – OH terminus adjacent to the APsite. The DNA polymerase-I replaces DNA and fills the gap. The removal of displaced strand is carried out by 5'→3' exonuclease activity of polymerase-I. The gap is sealed by DNA ligase.



Base Excision Repair

❖ MISMATCH REPAIR

The overall fidelity of replication in *E.coli* is improved by correcting mismatches after replication. Since base excision repair is involved in removal of altered bases, whereas the mismatch repair is normally concerned with the removal of normal bases involved in mismatch. For this, repair system should show some discrimination between template and daughter strand. This is achieved by methylating the template DNA in presence of Dam-methylase that methylates N6 position of all Adenines within GATC sequences.

The mechanism of mismatch repair has been thoroughly studied in *E.coli*. From such studies it has been observed that other proteins like Mut L, Mut H and Mut S are also involved in repair mechanism. The Mut S recognises the mismatches and binds to them. Then Mut L protein forms a complex with Mut S. Mut H binds Mut L at GATC sequences. Due to this the

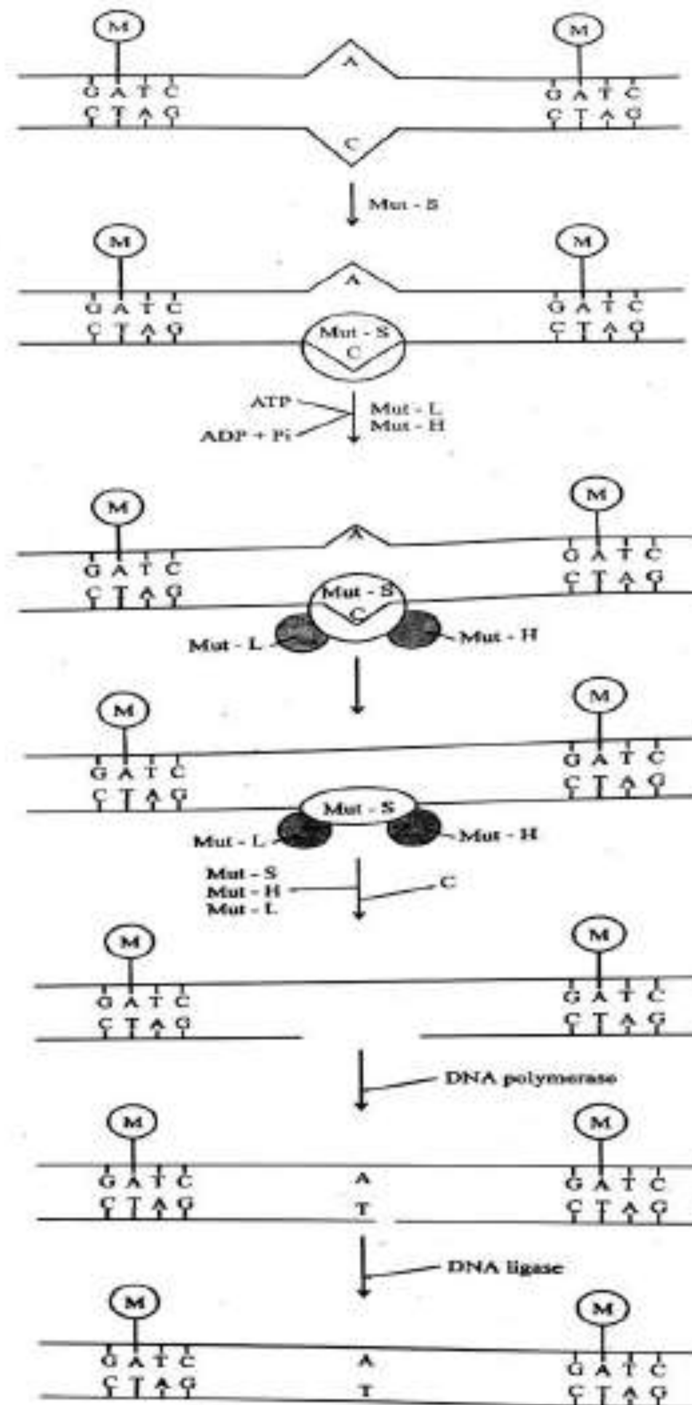
DNA mismatch will be looped. The Mut H bound has site specific endonuclease activity, that catalyses cleavage of unmethylated strand on the 5' end of the Guanine in GATC sequence. Further steps of repairmmechanism are dependent on location of mismatch relative to the cleavage site.

Once the nick is made the excision takes place. This process requires cooperation of Helicase-II, single strand DNA Binding Proteins (SSB) and specific exonuclease. The excision is followed by the DNA synthesis in presence of DNA-polymerase - III and ligation through DNA ligase.

Proteins involved in mismatch repair	
Proteins/Enzymes	**Biological Function**
Dam methylase	Methylates Adenine residue for self-recognition
DNA Helicase-II	Unwinding of duplex DNA
SSB [Single strand binding proteins]	Prevents the base pairing of separated strands
DNA polymerase - III	Synthesises leading and lagging strands
Exonuclease	Removes mismatch base pairs
DNA ligase	Sealing of Gaps/Nicks
Mut S	Recognises mismatch base pair
Mut H	Nicks at GATC site of unmethylated strand
Mut L	Facilitates binding of Mut H

The mismatch repair system is quite expensive in *E.coli* with reference to the ATP utilisation. Because the mismatch may be 1,000 base pairs away from the GATC sequence hence degradation and replacement of such a lengthy sequences demands enormous investment of energy.

The eukaryotic cells also contain the same version of mismatch repair system. There are several proteins which are structurally and functionally similar to the bacterial Mut S and Mut L proteins. The detailed mechanism of DNA mismatch repair in Eukaryotes is not well understood.



Mismatch Repair

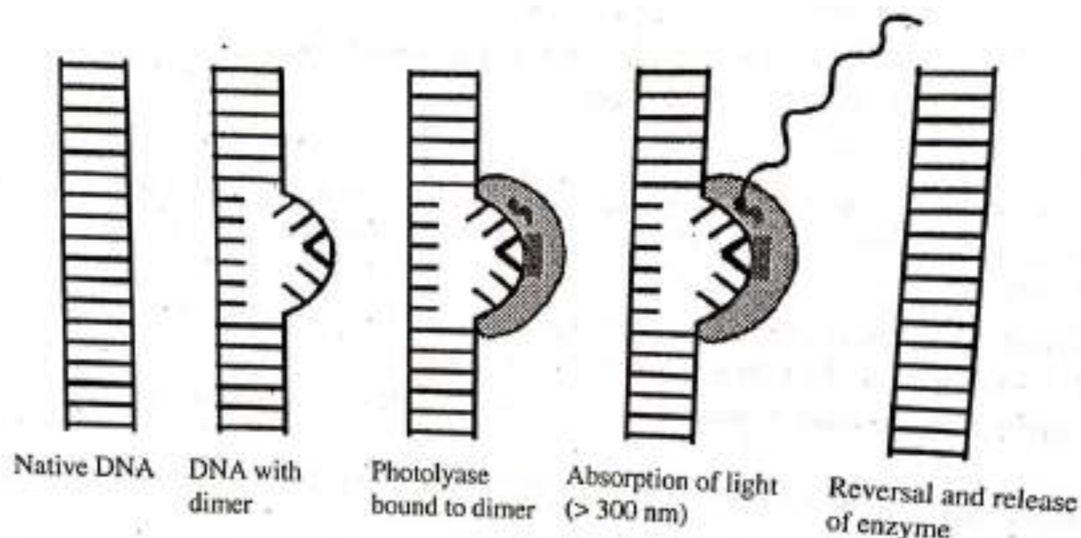
❖ PHOTO-REACTIVATION

Some of the damages are repaired without removing nucleotide/base. The best characterized example is the photo reactivation of pyrimidine dimers. Photoreactivation is an enzymatic cleavage of Thyamine dimers activated by the visible light.

DNA up on exposure to UV light can give rise to cyclobutane thymine dimers. This causes distortion in the DNA. These dimers can be split by the photolyase enzyme that absorb visible light [λ = 300 - 600 nm] and use light energy to split thymine dimers. The enzyme is generally monomer with molecular weight ranging from 55-65 kDa. These enzymes generally contain two cofactors [chromophores] that are non-covalently linked to the protein. The first chromophore is N^5, N^{10}, methenyl-tetrahydro foliate [pterene] and the second chromophore is $FADH^-$.

When the enzyme binds to the damaged site the pterene absorbs light in the visible region and transfers it to the $FADH^-$ that in turn transfers an electron to the pyrimidine dimer. The dimer gets split this results in a free radical which later converted to thyamine residues.

The enzyme is erratically distributed in the biological world. It is widely distributed in bacteria [except *Bacillus subtilis*] plants and animals [absent in placental mammals].

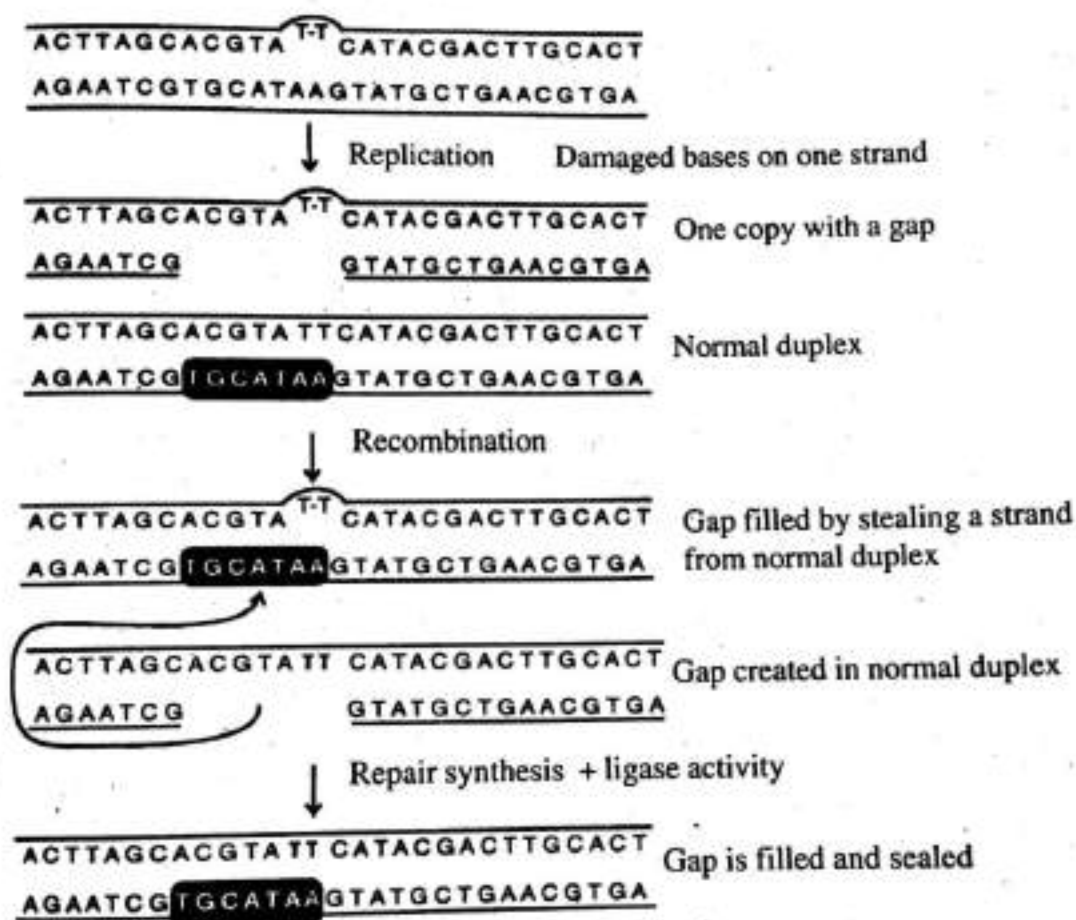


Enzymatic Photo-reactivation of thyamine dimer.

- **RECOMBINATION REPAIR SYSTEM**

It functions after replication. It is a post replicative repair system. This system is effective in dealing with the damages produced in daughter DNA by replication of a damaged DNA. Consider a situation of structural distortion due to thyamine dimers on one strand of the duplex, then the enzyme synthesises a daughter DNA. Replication restarts some distance farther along, leaving behind a gap in newly synthesized strand.

The newly synthesized duplex has different structure. One of the strand has damaged adduct that facing newly synthesized strand with lengthy gap, while the other duplicate remain unchanged. The gap opposite to the damaged site on first duplex is filled by taking homologous single strand DNA from normal duplex and pulls it in to the gap of the abnormal one. (i.e., single strand exchange). Following this the recipient duplex has parental damage strand facing a wild type daughter strand. The gap generated in normal strand is filled by the polymerase I generating normal duplex. So, the damage is confined to the original distortion and is not passed along.

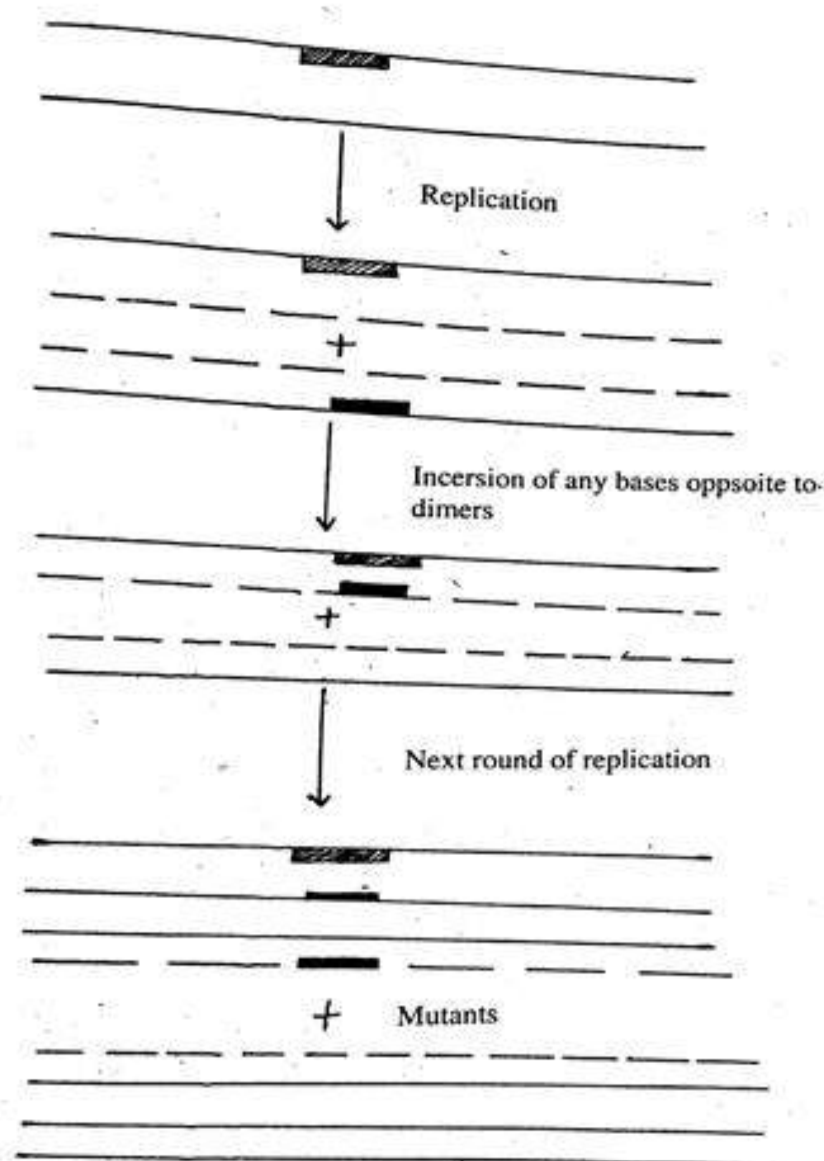


Recombination Repair System

SOS REPAIR

It includes bypass system that allows the growth of DNA chain across damaged segments in an error prone manner. Jean Weigle conducted experiments with *E.coli* to predict, the existence of SOS repair system. When normal DNA pol-III reaches the lesion on DNA during replication, before it could be repaired, the **rec A** gene is induced to produce the

Rec A protein which binds to the single stranded DNA at the site of distortion (due to pyrimidine dimers) and interacts with the 'ε' sub unit of polymerase and suppresses its editing function. Hence the polymerase-III can proceed across the lesions. The SOS repair controls all other types of repair mechanisms like excision repair, Recombination repair etc. The presence of Rec A at the distortion site inhibits the editing and thus allows mispaired base to remain in the daughter DNA as a mutant.



Mechanism of SOS Repair

Transcription in Prokaryotes:

In prokaryotic organisms transcription occurs in three phases known as initiation, elongation and termination.

RNA is synthesized by a single RNA polymerase enzyme which contains multiple polypeptide subunits. In E. coli, the RNA polymerase has five subunits: two α, one β, one β' and one σ subunit ($\alpha_2\beta\beta'\sigma$). This form is called the holoenzyme. The σ subunit may dissociate from the other subunits to leave a form known as the core enzyme.

These two forms of the RNA polymerase have different roles in transcription:

(i) Initiation:

Transcription cannot start randomly but must begin specifically at the start of a gene. Signals for the initiation of transcription occur in the promoter sequence which lies directly upstream of the transcribed sequence of the gene. The promoter contains specific DNA sequences that act as points of attachment for the RNA polymerase.

In E. coli, two sequence elements recognized by the RNA polymerase known as the -10 sequence and the -35 sequence arc present. The exact sequences can vary between promoters but all conform to an overall pattern known as the consensus sequence. The σ subunit of the RNA polymerase is responsible for recognizing and binding the promoter, probably at the -35 Box.

In the absence of the σ subunit the enzyme can still bind to DNA but binding is more random. When the enzyme binds to the promoter it initially forms a closed promoter complex in which the promoter DNA remains as a double helix. T he enzyme covers about 60 base pairs of the promoter including the -10 and -35 boxes. To allow transcription to begin, the double helix partially dissociates at the – 10 box, which is rich in weak A-1 bonds to give an open promoter complex.

The σ subunit then dissociates from the open promoter complex leaving the core enzyme. At the same time the first two ribonucleotides bind to the DNA, the first phosphodiester bond is formed and transcription is initiated (Fig. 7.7).

(ii) Elongation:

During elongation the RNA polymerase moves along the DNA molecule melting and unwinding the double helix as it progresses. The enzyme adds ribonucleotides to the 3' end of the growing RNA molecule with the order of addition determined by the order of the bases on the template strand.

In most cases, a leader sequence of variable length is transcribed before the coding sequence of the gene is reached. Similarly, at the end of the coding sequence a noncoding trailer sequence is transcribed before transcription ends.

During transcription only a small portion of the double helix is unwound at any one time. The unwound area contains the newly synthesized RNA base-paired with the template DNA strand and extends over 12-17 bases.

The unwound area needs to remain small because unwinding in one region necessitates over-winding in adjacent regions and this imposes strain on the DNA molecule. To overcome this problem, the RNA is released from the template DNA as it is synthesized allowing the DNA double helix to reform (Fig. 7.8).

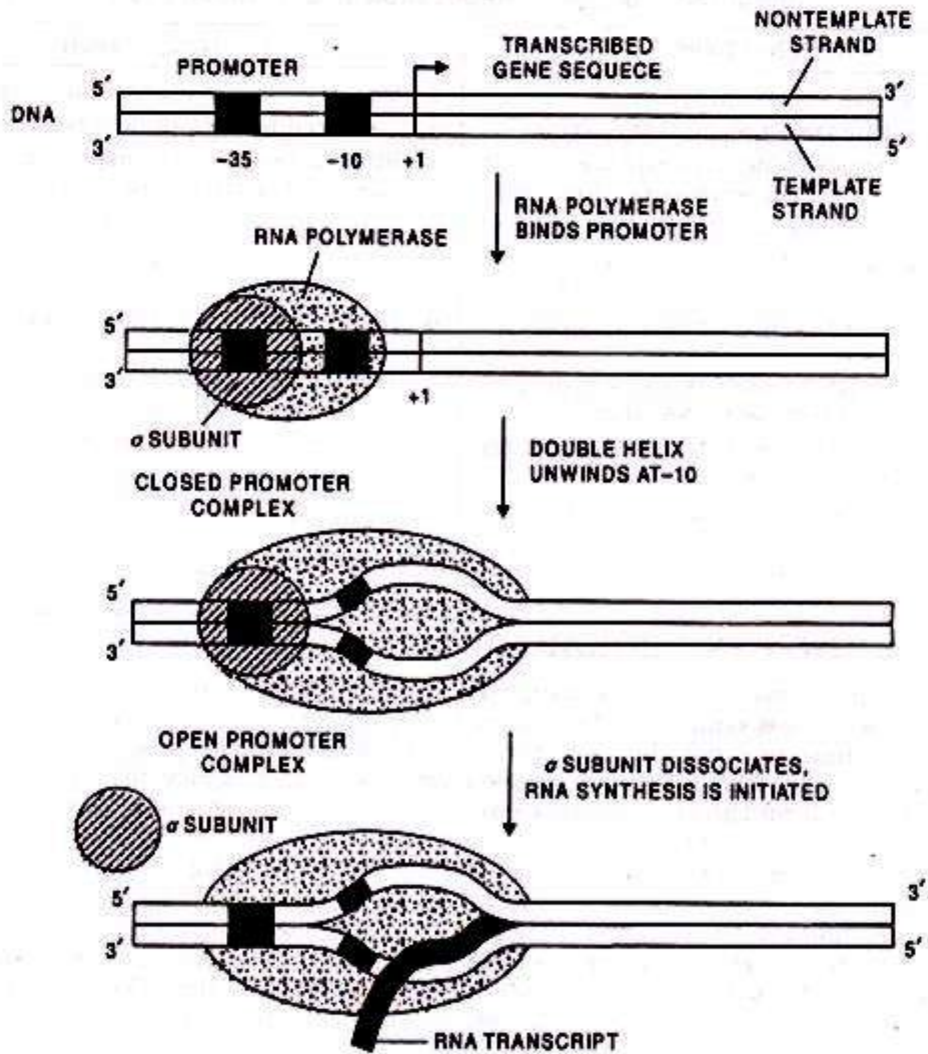


Fig. 7.7. Initiation of transcription in prokaryotes.

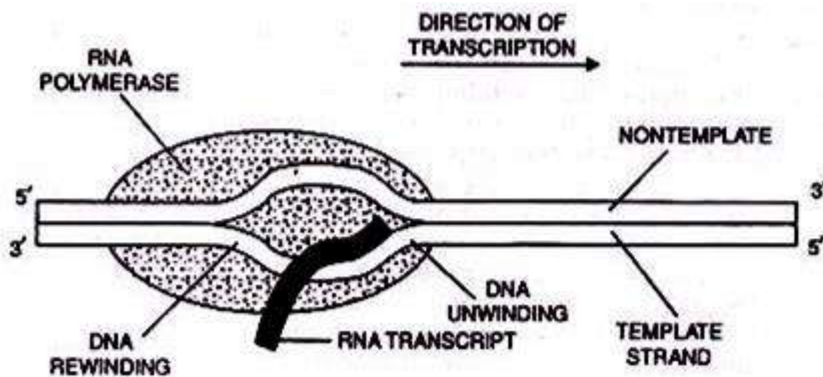


Fig. 7.8 Transcription elongation in prokaryotes.

(iii) Termination:

The termination of transcription occurs non-randomly and takes place at specific points after the end of the coding sequence. In E. coli, termination occurs at sequences known as palindromes. These are symmetrical about their middle such that the first half of the sequence is followed by its exact complement in the second half.

In single-stranded RNA molecules this feature allows the first half of the sequence to base pair with the second half to form what is known as a stem-loop structure (Fig. 7.9). These appear to act as signals for termination. In some cases the stem-loop sequence is followed by a run of 5-10 As in the DNA which form weak A-U base pairs with the newly synthesized RNA.

It is thought that the RNA polymerase pauses just after the stem-loop and that the weak A-U base pairs break causing the transcript to detach from the template. In other cases the run of As is absent and a different mechanism occurs based on binding of a protein called Rho (ρ) which disrupts base-pairing between the template and the transcript when the polymerase pauses after the stem-loop. The termination of transcription involves the release of the transcript and the core enzyme which may then re-associate with the σ subunit and go on to another round of transcription (Fig. 7.9 & 7.10).

In many bacteria, genes of related functions are grouped together in operons. An operon acts as a single transcription unit and thus produces polycistronic mRNA. In eukaryotes, only monocistronic mRNAs are generally produced.

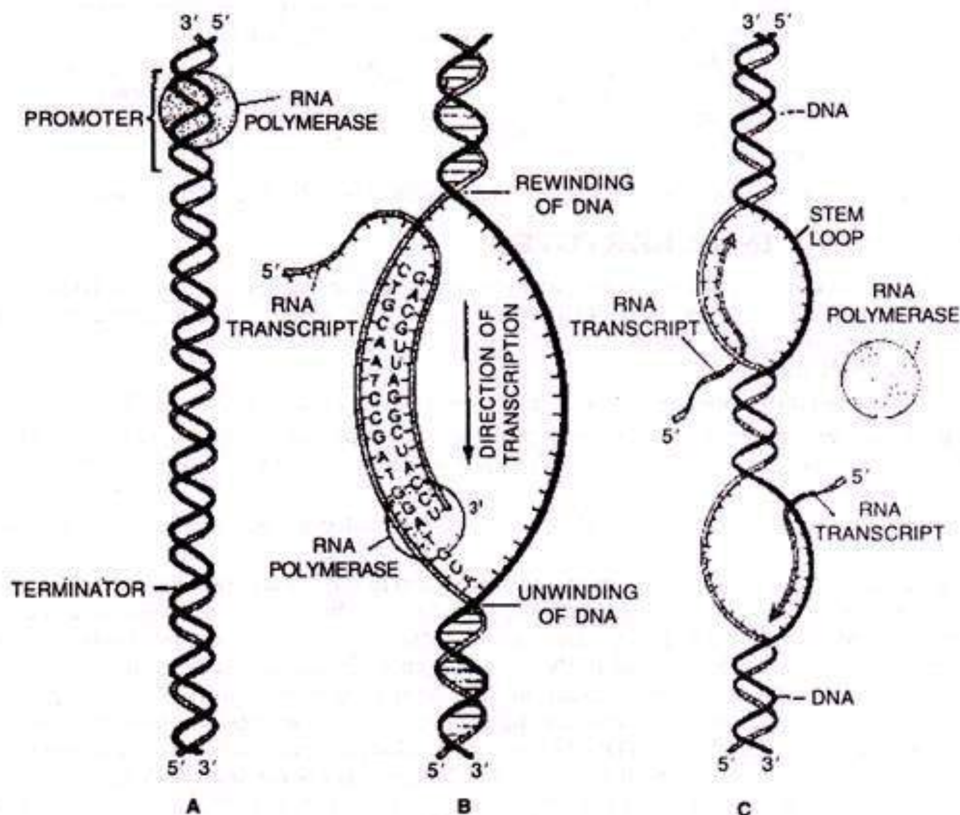


Fig. 7.9 A–C. A summarized diagram of general process of transcription in prokaryotes : **A.** Binding to promoter and RNA chain initiation : **B.** Elongation of RNA polypeptide chain, **C.** Termination of transcription.

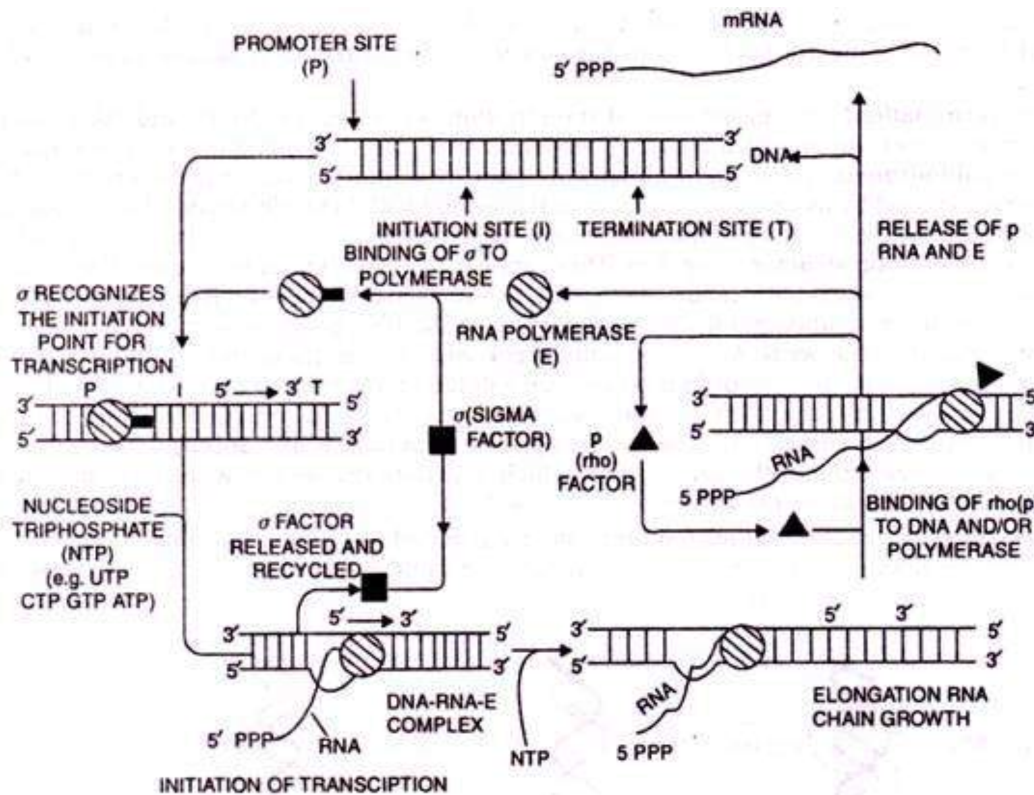


Fig 7.10. Diagrammatic presentation of the synthesis of RNA by E.coli polymerase

Transcription in Eukaryotes:

Transcription occurs in eukaryotes in a way similar to prokaryotes. However, initiation is more complex, termination does not involve stem-loop structures and transcription is carried out by three enzymes (RNA polymerases I, II and III) each of which transcribes a specific set of genes and functions in a slightly different way.

RNA polymerase I transcribes genes encoding three of the four ribosomal RNAs (18S, 28S and 5.8S). RNA polymerase II enzyme transcribes genes that encode proteins. Binding of RNA polymerase II to its promoter involves several different DNA sequence elements and a number of proteins called transcription factors. RNA polymerase III transcribes a set of short genes that encode transfer RNAs and the 5S ribosomal RNA.

Unlike the situation in prokaryotic genes, transcription in eukaryotes occurs within the nucleus and mRNA moves out of the nucleus into the cytoplasm for translation. The initiation and regulation of transcription is more extensive than prokaryotes. Another major difference between prokaryotes and eukaryotes lies in the fact that the mRNA in eukaryotes is processed from the primary RNA transcript, a process called maturation.

Initially at the 5′ end a cap (consisting of 7-methyl guanosine or 7 mG) and a tail of poly A at the 3′ end are added (Fig. 7.11) The cap is a chemically modified molecule of guanosine triphosphate (GTP). The primary eukaryotic mRNA transcript is much longer and localised into the nucleus, when it is also called heterogenous nuclear RNA (hnRNA) or pre- mRNA.

The eukaryotic primary mRNAs are made up of two types of segments; non-coding introns and the coding exons. The introns are removed by a process called RNA splicing. Of a pair of small nuclear ribonucleoprotein (SnRNPs pronounced "snurps"), one binds to 5' splice site and the other to 3' splice site.

A spliceosome forms because of interaction between SnRNPs and other proteins. This spliceosome uses energy of ATP to cut the RNA, releases the introns and joins two adjacent exons to produce mature mRNA. Besides, these two post-transcriptional modifications, RNA editing may also take place before translation begins.

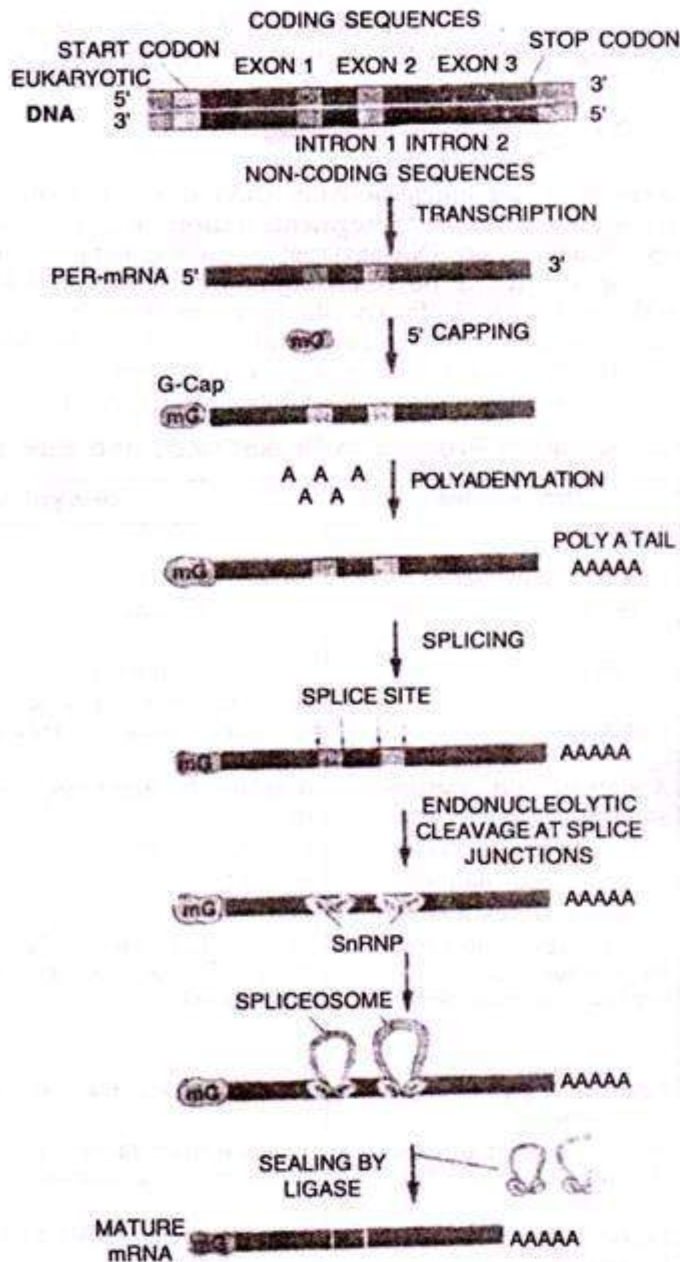


Fig. 7.11. Transcription in eukaryotes

Post-Transcription Processing:

Primary transcript is often larger than the functional RNAs. It is called heterogeneous or hnRNA especially in case of mRNA.

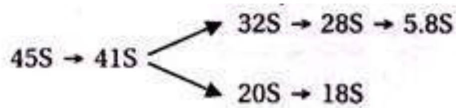
Post-transcription processing is required to convert primary transcript into functional RNAs.

It is of four types:

(i) Cleavage:

Larger RNA precursors are cleaved to form smaller RNAs. Primary transcript of rRNA is 45S in eukaryotes.

It is cleaved to form the following:



Primary transcript is cleaved by ribonuclease-P (an RNA enzyme) to form 5-7 tRNA precursors,

(ii) Splicing:

Eukaryotic transcripts possess extra segments (introns or intervening sequences). They are removed by nucleases. Ribozyme (an-RNA enzyme) is a self splicing intron involved in some of these reactions as well as catalysing polymerisation.

(iii) Terminal Additions:

Additional nucleotides are added to the ends of RNAs for specific functions, e.g., CCA segment in tRNA, cap nucleotides at 5' end of mRNA or poly-A segments at 3' end of mRNA.

(iv) Nucleotide Modifications:

They are most common in tRNA-methylation (e.g., methyl cytosine, methyl guanosine), deamination (e.g., inosine from adenine), dihydrouracil, pseudouracil, etc.

TRANSLATION IN PROKARYOTES:

Translation process consists of three major phases or stages, viz:

(1) Initiation,

(2) Elongation and

(3) Termination.

These are briefly discussed below:

1. Initiation:

This is the first phase of translation. Start or initiation codon [AUG] is responsible for initiation of translation process.

Initiation of translation in prokaryotes involves the assembly of the components of the translation system which are: the two ribosomal subunits (small and large), the mRNA to be translated, the first (formyl) aminoacyl tRNA (the tRNA charged with the first amino acid), GTP (as a source of energy), and three initiation factors (IF 1, IF 2 and IF 3) which help the assembly of the initiation complex.

The ribosome consists of three sites, the A site, the P site, and the E site. The A site is the point of entry for the aminoacyl tRNA (except for the first aminoacyl tRNA, fMet-$tRNA_f^{Met}$, which enters at the P site). The P site is where the peptidyl tRNA is formed in the ribosome. And the E site which is the exit site of the now uncharged tRNA after it gives its amino acid to the growing peptide chain.

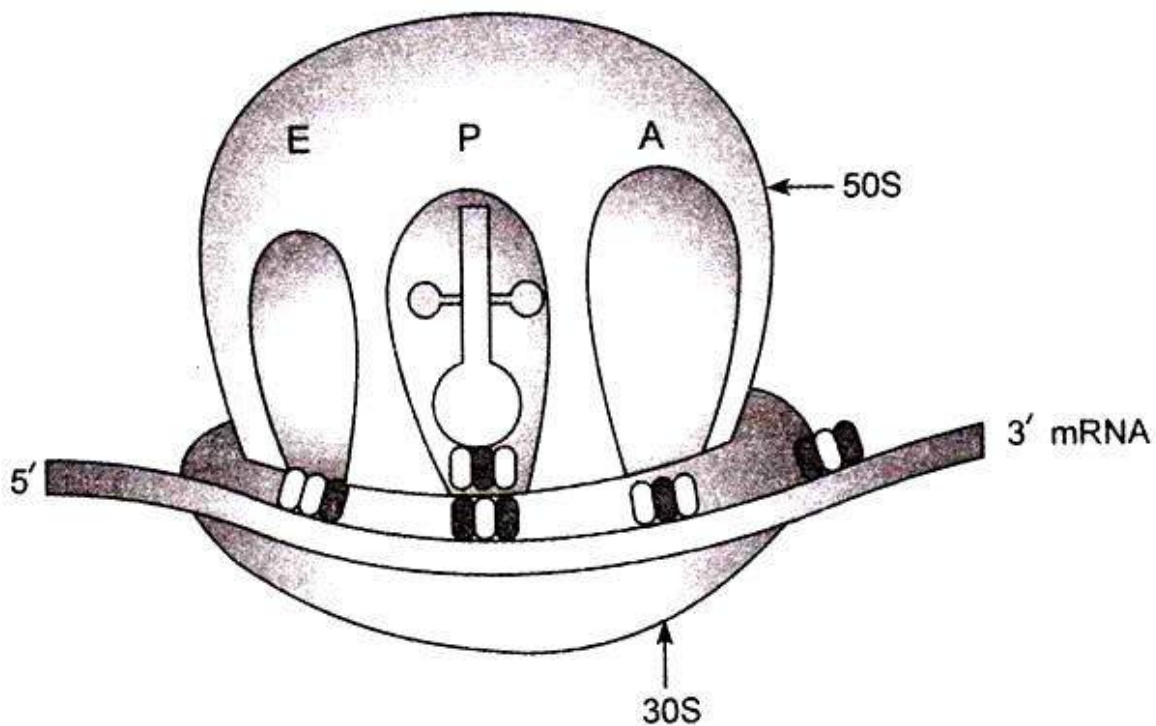


Fig. 23.1. Initiation

Translation begins with the binding of the small ribosomal subunit to a specific sequence on the mRNA chain. Initiation of translation begins with the 50S and 30S ribosomal subunits. IF1 (initiation factor 1) blocks the A site to ensure that the IMet-tRNA can bind only to the P site and that no other aminoacyl-tRNA can bind in the A site during initiation, while IF3 blocks the E site and prevents the two subunits from associating.

IF2 is a small GTPase which binds fmet-tRNA$_f^{Met}$ and helps its binding with the small ribosomal subunit. The 3′ end of the 16S rRNA of the small 30S ribosomal subunit recognizes the ribosomal binding site on the mRNA (Shine-Dalgarno sequence or SD), through its anti-SD sequence, 5-10 base pairs upstream of the start codon. The Shine-Dalgarno sequence is found only in prokaryotes.

This helps to correctly position the ribosome onto the mRNA so that the P site is directly on the AUG initiation codon. IF3 helps to position fMet-tRNA$_f^{met}$ into the P site, such that fMet-tRNA$_f$ ^{met} interacts via base pairing with the mRNA initiation codon (AUG). Initiation ends as the large ribosomal subunit joins the complex causing the dissociation of initiation factors.

The small subunit binds via complementary base pairing between one of its internal subunits and the ribosome binding site. This site a sequence of about ten nucleotides on the mRNA. It is located anywhere from 5 and 11 nucleotides from the initiating codon [AUG],

After binding of the small subunit, a special tRNA molecule, called N-formyl methionine, or fMet, recognizes and binds to the initiator codon. Then the large subunit binds resulting in the formation of the initiation complex. As soon as the initiation complex is formed, the fMet-tRNA occupies the P site of the ribosome and the A site is left empty.

This entire initiation process is facilitated by extra proteins, called initiation factors that help with the binding of ribosomal subunits and tRNA to the mRNA chain.

2. Elongation:

This is the second phase or middle phase of translation. Elongation begins after the formation of the initiation complex. Elongation of the polypeptide chain involves addition of amino acids to the carboxyl end of the growing chain. The growing protein exits the ribosome through the polypeptide exit tunnel in the large subunit.

Elongation starts when the fmet-tRNA enters the P site, causing a conformational change which opens the A site for the new aminoacyl-tRNA to bind. This binding is facilitated by elongation factor-T4 (EF-T4), a small GTPase. Now the P site contains the beginning of the peptide chain of the protein to be encoded and the A site has the next amino acid to be added to the peptide chain.

The growing polypeptide connected to the tRNA in the P site is detached from the tRNA in the P site and a peptide bond is formed between the last amino acids of the polypeptide and the amino acid still attached to the tRNA in the A site.

This process, known as peptide bond formation, is catalyzed by a ribozyme, peptidyltransferase, an activity intrinsic to the 23S ribosomal RNA in the 50S ribosomal subunit. Now, the A site has newly formed peptide, while the P site has an unloaded tRNA (tRNA with no amino acids). In the final stage of elongation, translocation, the ribosome moves 3 nucleotides towards the 3′ end of mRNA.

Since tRNAs are linked to mRNA by codon-anticodon base-pairing, tRNAs move relative to the ribosome taking the nascent polypeptide from the A site to the P site and moving the uncharged tRNA to the E exit site.

This process is catalyzed by elongation factor G (EF-G). The ribosome continues to translate the remaining codons on the mRNA as more aminoacyl-tRNA binds to the A site, until the ribosome reaches a stop codon on mRNA(UAA, UGA, or UAG).

When the A site opens again, the next appropriate aminoacyl tRNA can bind there and the same reaction takes place, yielding a three-amino acid peptide chain. This process repeats, creating a polypeptide chain in the P site of the ribosome. A single ribosome can translate 60 nucleotides per second. This speed can be vastly augmented when ribosomes unite together to form polyribosomes.

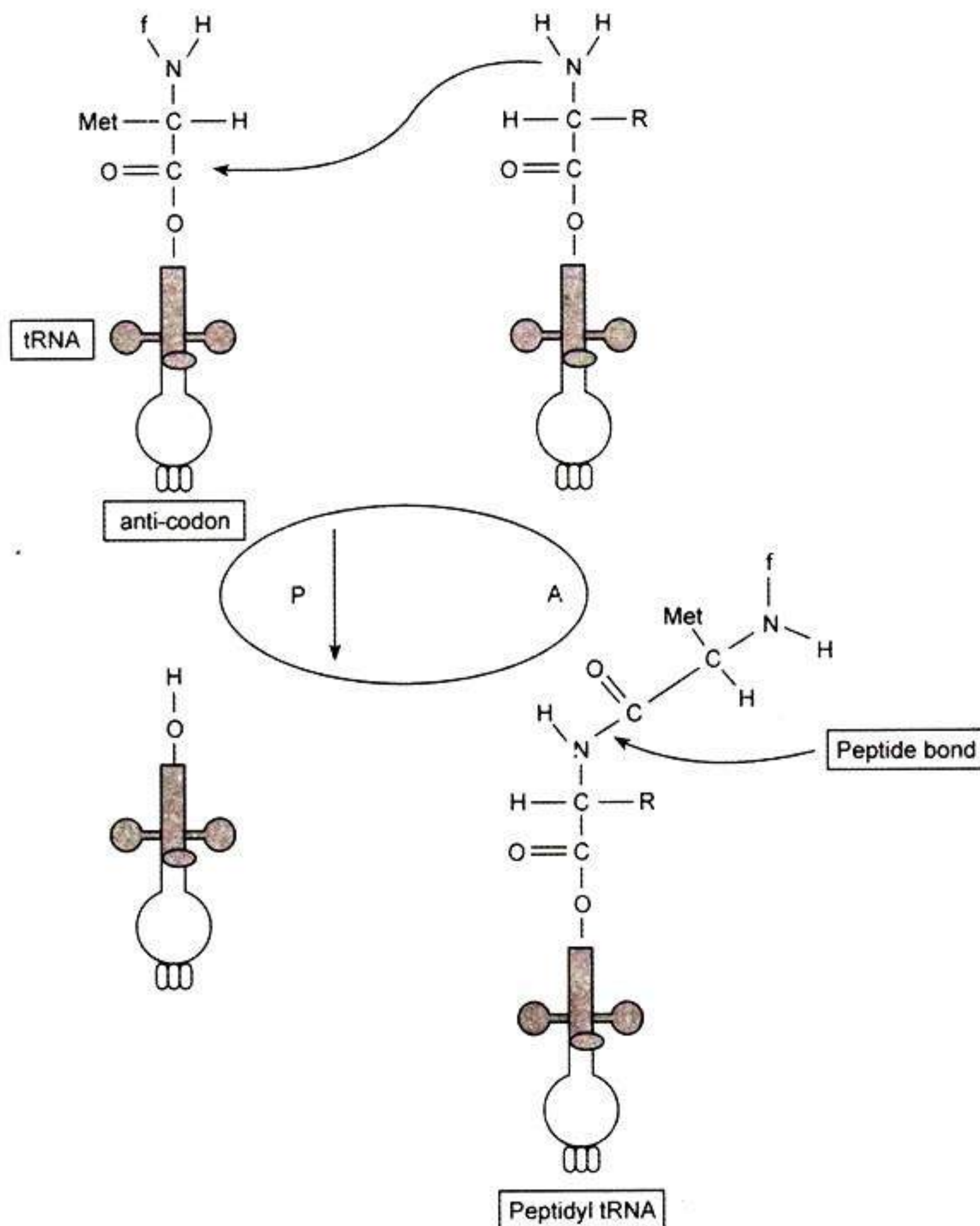


Fig. 23.2. Peptide Formation.

3. Termination:

This is the last phase of translation. Termination occurs when one of the three termination codons moves into the A site. These codons are not recognized by any tRNAs. Instead, they are recognized by proteins called release factors, namely RF1 (recognizing the UAA and UAG stop codons) or RF2 (recognizing the UAA and UGA stop codons).

These factors trigger the hydrolysis of the ester bond in peptidyl-tRNA and the release of the newly synthesized protein from the ribosome. A third release factor RF-3 catalyzes the release of RF-1 and RF-2 at the end of the termination process.

Translation in Eukaryotes:

The mechanism of translation in eukaryotes is similar to that of prokaryotes in several aspects.

Translation process consists of three phases or stages, viz:

(i) Initiation,

(ii) Elongation and

(iii) Termination.

These are discussed as follows:

1. Initiation:

The process of initiation of translation in eukaryotes is of two types, viz:

(i) Cap-dependent initiation, and

(ii) Cap-independent initiation.

i. Cap-Dependent Initiation:

Initiation of translation usually involves the interaction of certain key proteins with a special tag bound to the 5′-end of an mRNA molecule, the 5′ cap. The protein factors bind the small ribosomal subunit (also referred to as the 40S subunit), and these initiation factors hold the mRNA in place.

The eukaryotic Initiation Factor 3 (eIF3) is associated with the small ribosomal subunit, and plays a role in keeping the large ribosomal subunit from prematurely binding.

The factor eIF3 also interacts with the eIF4F complex which consists of three other initiation factors [eIF4A, eIF4E and eIF4G]. The factor eIF4G is a protein which directly associates with both eIF3 and the other two components.

The eIF4E is the cap-binding protein. It is the rate-limiting step of capdependent initiation, and is often cleaved from the complex by some viral proteases to limit the cell's ability to translate its own transcripts.

The eIF4A is an ATP-dependent RNA helicase, which aids the ribosome in resolving certain secondary structures formed by the mRNA transcript. There is another protein associated with the eIF4F complex called the Poly-A Binding Protein (PABP), which binds the poly-A tail of most eukaryotic mRNA molecules. This protein is considered to play a role in circularization of the mRNA during translation.

This pre-initiation complex (43S subunit, or the 40S and mRNA) along with protein factors move along the mRNA chain towards its 3'-end. It scans for the 'start' codon (typically AUG) on the mRNA. The start codon indicates the site where the mRNA will begin coding for the protein. In eukaryotes and archaea, the amino acid encoded by the start codon is methionine.

The initiator tRNA charged with Met forms pan of the ribosomal complex and thus all proteins start with this amino acid. The Met-charged initiator tRNA is brought to the P-site of the small ribosomal subunit by eukaryotic Initiation Factor 2 (eIF2). It hydrolyzes GTP, and signals for the dissociation of several factors from the small ribosomal subunit which results in the association of the large subunit (or the 60S subunit).

The complete ribosome (80S) then commences translation elongation, during which the sequence between the 'start' and 'stop' codons is translated from mRNA into an amino acid sequence. In this way a protein is synthesized.

ii. The Cap-Independent Initiation:

This is lesser known method of translation in eukaryotes. This method of translation has been recently discovered. It has been found to be important in conditions that require the translation of specific mRNAs. It works despite cellular stress or the inability to translate most mRNAs. Examples of such type of translation are factors responding to apoptosis and stress-induced responses.

The best studied example of the cap-independent mode of translation initiation in eukaryotes is the Internal Ribosome Entry Site (IRES) approach. The main difference between cap-independent translation and cap-dependent translation is that the former does not require the ribosome to start scanning from the 5' end of the mRNA cap until the start codon.

The ribosome can be trafficked to the start site by ITAFs (IRES trans-acting factors) bypassing the need to scan from the 5' end of the un-translated region of the mRNA.

2. Elongation:

Elongation is dependent on eukaryotic elongation factors At the end of the initiation step, the mRNA is positioned so that the next codon can be translated during the elongation stage of protein synthesis.

The initiator tRNA occupies the P site in the ribosome; and the A site is ready to receive an aminoacyl-tRNA. During chain elongation, each additional amino acid is added to the nascent polypeptide chain in a three-step micro-cycle.

The steps in this micro-cycle are:

(i) Positioning the correct aminoacyl-tRNA in the A site of the ribosome;

(ii) Forming the peptide bond and

(iii) Shifting the mRNA by one codon relative to the ribosome.

The translation machinery works relatively slowly compared to the enzyme systems that catalyze DNA replication. Proteins are synthesised at a rate of only 18 amino acid residues per second, whereas bacterial replisomes synthesize DNA at a rate of 1,000 nucleotides per second.

This difference in rate reflects, in part, the difference between polymerizing four types of nucleotides to make nucleic acids and polymerizing 20 types of amino acids to make proteins. Testing and rejecting incorrect aminoacyl- tRNA molecules takes time and slows protein synthesis.

The rate of transcription in prokaryotes is approximately 55 nucleotides per second, which corresponds to about 18 codons per second, or the same rate at which the mRNA is translated.

In bacteria, translation initiation occurs as soon as the 5′ end of an mRNA is synthesized, and translation and transcription are coupled. This tight coupling is not possible in eukaryotes because transcription and translation are carried out in separate compartments of the cell (the nucleus and cytoplasm).

Eukaryotic mRNA precursors must be processed in the nucleus [e.g., capping, polyadenylation, splicing) before they are exported to the cytoplasm for translation.

3. Termination:

This is the last phase of translation. Termination occurs when one of the three termination codons moves into the A site. These codons are not recognized by any tRNAs.

Termination of elongation is dependent on eukaryotic release factors In eukaryotes, there is only one release factor that is eRF, which recognizes all three stop codons [in place of RF1, RF2, or RF3 factors in prokaryotes]. However, the overall process of termination is similar to that of prokaryotes.

Protein Targeting:

A typical mammalian cell may contain numerous kinds of proteins and numerous individual protein molecules. The eukaryotic cell is a multi-compartmental structure. Its many organelles each requires different proteins. Except a few of them which are synthesized in mitochondria and chloroplasts all other proteins necessary for the cell and the ones to be secreted by the cell are synthesized in the cytosol on free ribosomes and on ribosomes bound to the endoplasmic reticulum.

Most proteins are coded by the nuclear genome and synthesized in the cytoplasm. The proteins are present in the ER, mitochondria, chloroplasts, Golgi, peroxisomes, nucleus, in the cytosol and in the membranes of all these organelles. They are selectively transported into their appropriate organelles inside the cell and across the plasma membrane to be secreted outside the cell.

Some of them are carried into membrane bound vesicles which bud off from one organelle and transported in definite pathways. Different destinations of different proteins require sophisticated system for labelling and sorting newly synthesized proteins and ensuring that they reach their proper places. This transportation of proteins to their final destinations is called protein targeting.

Proteins destined for cytoplasm and those to be incorporated into mitochondria, chloroplasts and nuclei are synthesized on free ribosomes in the cytoplasm. Proteins destined for cellular membranes, lysosomes and extracellular transport, use a special distribution system. The main structures in this system are the rough endoplasmic reticulum (RER) and Golgi complex.

The RER is a network of interconnected membrane enclosed vesicles or vacuoles. The endoplasmic reticulum is coated with polyribosomes to give it a rough appearance. The golgi complex is also a stack of membrane bound sacs but they are not interconnected. The golgi complex acts as a switching center for proteins to various destinations.

Proteins to be directed to their destinations via Golgi complex are synthesized by ribosomes associated with endoplasmic reticulum.

Signal Sequence:

Protein sorting requires proper address labels which are in the form of peptide signal sequences. A signal sequence that directs the protein to its target is present in the form of 13-35 amino acids in the newly synthesized protein itself. It is the first to be synthesized and is mostly present at the amino N-terminal, sometimes at the carboxyl C- terminal.

It is known as signal sequence or leader sequence. Some proteins are further sorted to a sub-compartment within the target organelle. For this purpose, a second signal sequence is present behind the first signal sequence which is cleaved.

Proteins carried inside the membrane bound vesicles are called cargo proteins. An embedded or integrated protein is carried in the membrane of the vesicle, while secretory protein is carried within the lumen of the vesicle. The vesicle buds off from the donor surface and fuses with the target surface releasing its contents into the target organelle and the membrane protein is incorporated into the membrane of the target organelle. The process is repeated during the passage of protein from ER to Golgi to lysosomes and from Golgi to plasma membrane.

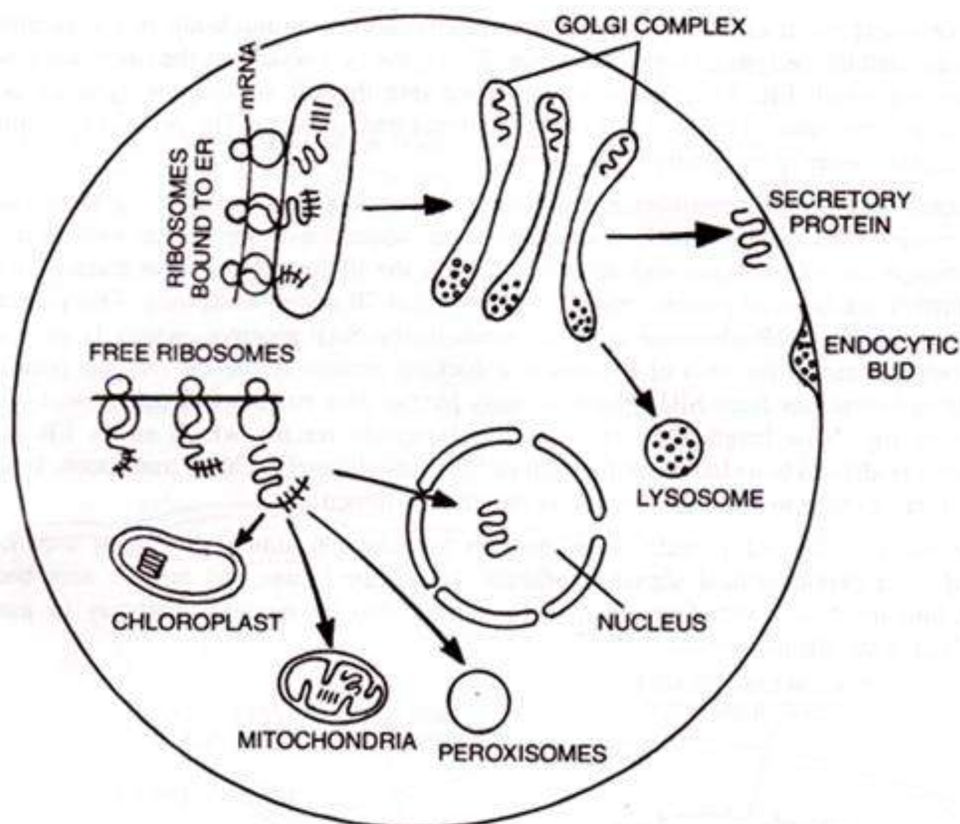


Fig. 21.1. Nascent proteins targeting to different organelles of the cell and cell secretion.

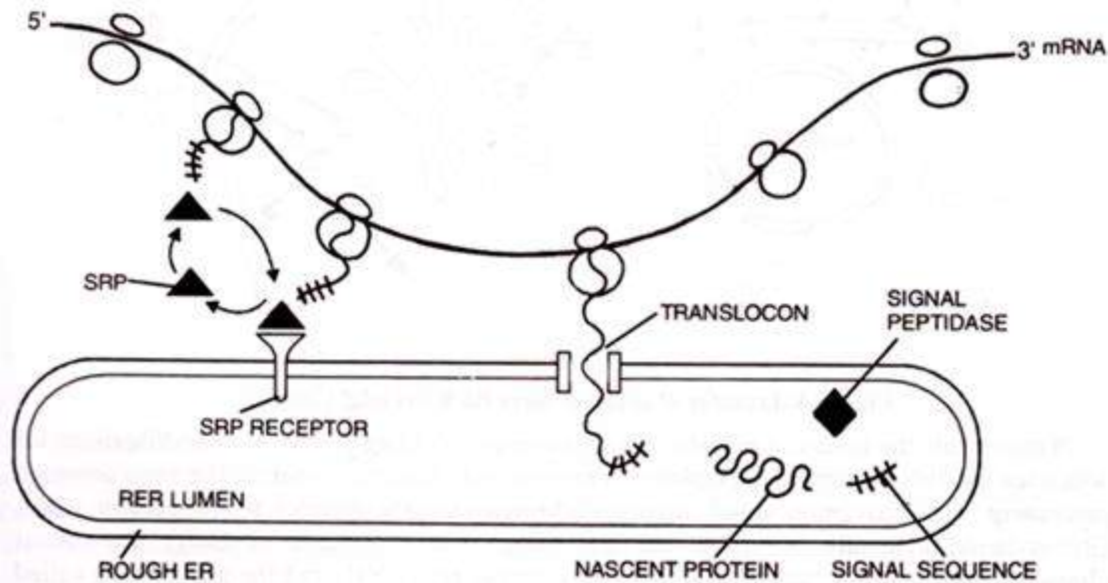


Fig. 21.2. Transport of proteins into ER

Transport of Proteins into ER:

A short N-terminus signal sequence at the beginning of the growing nascent protein chain' determines whether a ribosome synthesizing the proteins binds to ER or not. The protein synthesis always begins on free ribosomes. As the signal sequence emerges out of the ribosome, the large ribosomal sub-unit binds to ER membrane.

This is decided by the type of signal sequence. This is the first sorting as the ribosome binds to ER, forming rough ER. Translocation takes place into the ER while growing chain is still bound to the ribosome. This is called co-translational translocation. The process is facilitated by the signal sequence recognition mechanism.

Signal Sequence Recognition Mechanism:

It consists of a signal recognition particle (SRP) present in the cytosol. SRP binds to the signal sequence of the nascent protein as soon as it emerges out of ribosome and directs it towards the ER membrane. The binding of SRP stops further synthesis of protein chain when it is about 70 amino acids long.

This prevents it from folding. The SRP-ribosome complex binds to the SAP receptor, which is an integral membrane protein in the wall of ER and is a docking protein of the ER. At this point GTP hydrolysis hydrolyses frees SRP which is ready for the next round of directing next nascent protein of ER.

Now lengthening of nascent polypeptide restarts which enters ER lumen. Ribosome is aligned to a channel in the wall of ER. This channel is called translocon. It allows the elongating chain to enter the translocon into the ER lumen.

As the growing polypeptide chain emerges into the ER lumen, the signal sequence is cleaved by a peptide called signal peptidase. Inside the lumen, the protein may become folded into its final active form or may be carried into its secretary pathway or may be embedded in the ER membrane.

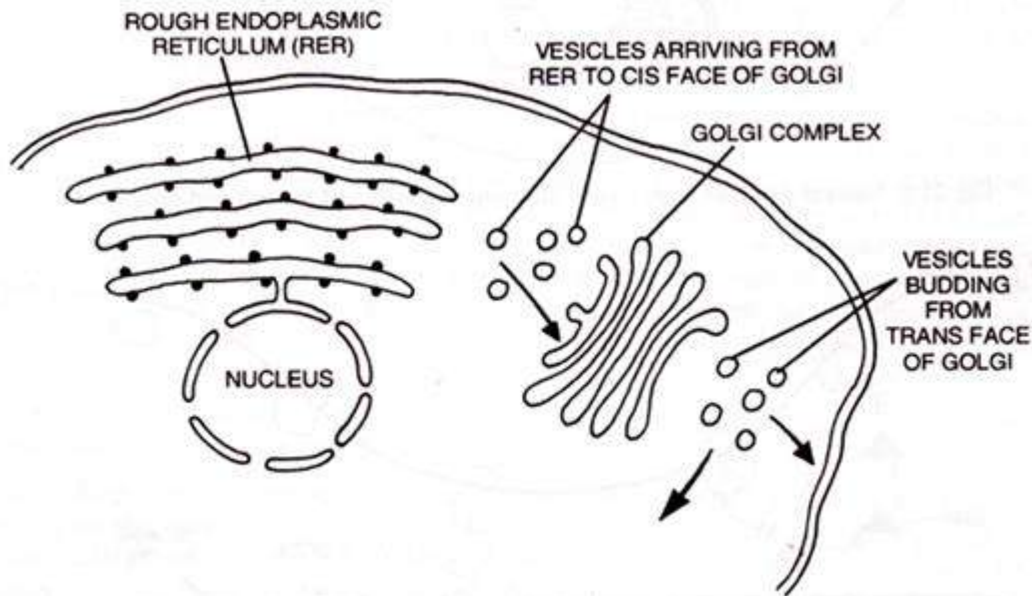


Fig. 21.3. Transfer of proteins form RER to Golgi Complex

Once inside the lumen of ER, the protein undergoes folding and several modifications for which the ER lumen contains a number of enzymes and chaprone proteins. The most common processing is glycosylation which involves addition of carbohydrates to the protein chain. Glycosylation generally occurs in the ER lumen but sometimes in Golgi also.

Most oligosachharides or glycons are attached to the amino group NH_3 and the proteins are called N-linked glycoprotiens e.g. oligosachharide attached to aspargine. A preformed oligosachharide is added to the proteins. This structure is Man 9 $(Glc\ NAC)_2$ called high mannose structure.

This contains mannose, glucose and N-acetylglucosamine). All nascent proteins start the sorting pathway by addition of the same pre-formed oligosachharide in plants and animals. Almost all proteins that enter the secretary pathway are glycosylated.

In ER lumen, after glycosylation, many protiens are folded and stabilized by disulphide proteins bonds (-S-S-). This reaction is catalyzed by an enzyme, protein disulphide isomerase (PDI). Most of human proteins are stabilized by disulphide bonds.

Role of Golgi Complex in Protein Transportation:

The role of Golgi complex is to act as a switching center for proteins to various destinations. Both ER and Golgi apparatus are flattened cisternae. Transport of proteins from one compartment (donor) to the next one (target) is carried out in transport vesicles. The vesicles contain cargo proteins in their lumen and integral membrane proteins in their membranes.

The vesicles bud off from ER and fuse with the cis-compartment or receiving compartment of Golgi. In this process cargo proteins are delivered into the lumen of Golgi and membrane proteins become part of the membrane of the target vesicles. The proteins are glycosylated, folded, modified and sorted in ER. This process of glycosylation, modification and sorting of proteins continues in successive Golgi cisternae.

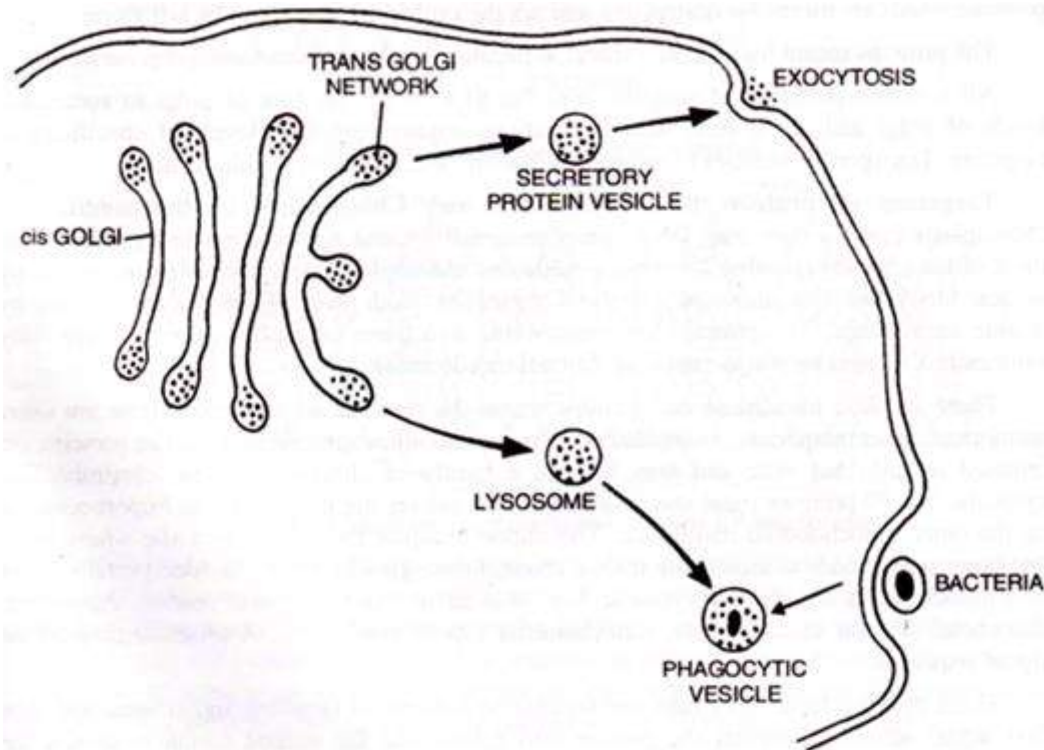


Fig. 21.4. Transport of proteins from Golgi to Lysosomes

Starting from the cis-compartment to medial compartment and lastly to trans-Golgi network proteins are exported to the end target. In trans-golgi network (TGN) proteins are further sorted to be delivered to lysosomes, for secretion outside the cell and to plasma membrane according to signals present in the nascent proteins.

Transport of Proteins from Golgi to Lysosomes:

The lysosomal enzymes and lysosomal membrane proteins are synthesized in rough ER and transported to Golgi cisternae and ultimately to lysosomes. The sorting signal that directs the lysosomal enzymes from the trans- Golgi network (TGN) to lysosomes is mannose 6-phosphate (M6P). The attachment of M6P to lysosomal enzymes prevents their further modification.

Separation of M6P bearing lysosomal enzymes from other proteins takes place in TGN. The wall of TGN contains M6P receptors. These M6P receptors bind to lysosomal proteins. The vesicles containing these receptor bearing proteins bud off from TGN. These vesicles are called lysosomes. Later these vesicles fuse with vesicles which have arisen by pinacocytosis and phagocysis to form secondary lysosomes. Low pH of Lysosomes triggers the dissociation of enzymes from the receptors.

The M6P receptors are recycled back to trans-golgi network in vesicles. Lysosomes contain hydrolyzing proteolytic enzyme, which digests proteins meant for degradation. A protein named ubiquitin marks the proteins meant for destruction. Ubiquitin is present in all eukaryotic cells. This mechanism degrades only those proteins which are meant for destruction and not the proteins which are to be left alone.

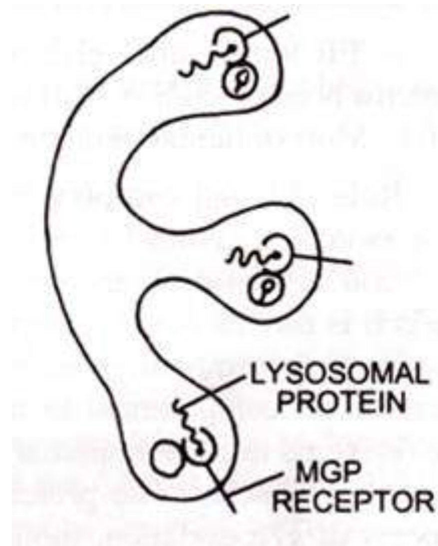


Fig. 5. Trans-Gologi Network

The proteins meant for secretion travel to plasma membrane from trans-golgi network.

All this transportation of vesicles from the RER to the cis face of golgi to successive levels of golgi and on to their final destinations requires the high levels of specificity in targeting. Transport of vesicles to wrong destinations would lead to cellular chaos.

Targeting of Proteins to Mitochondria and Chloroplasts:

Mitochondria and chloroplasts possess their own DNA, ribosomes, mRNA and synthesize a few proteins. But most of the proteins required for mitochondria and chloroplasts are synthesized in cytosol by nuclear DNA and then imported into these organelles. Both these organelles are covered by double membranes. The proteins are translocated into these organelles after they are fully synthesized. This is known as post-translational translocation.

There are four mitochondrial locations where the proteins are targeted. These are outer membrane, inner membrane, intermembranal space and mitochondrial matrix. The proteins are released in unfolded state and they bind to a family of chaprones. These chaprones are cytosolic hsp 70 proteins (heat shock proteins) that deliver the proteins to an import receptor on the outer mitochondrial membrane.

The import receptor then slides to a site where inner and otuer mitochondrial membrane form a channel through which the unfolded protein enters into mitochondria leaving out cytosolic hsp 70 protiens. As the protein reaches the matrix, mitochondrial heat shock protein, mitochondrial hsp 70 binds to it. A protease cleaves the signal sequence.

These proteins have more than one successive N-terminal targeting signal sequence. The first signal sequence imports the protein into matrix and the second signal re-directs the protein into membranes or inter membranal space.

Mitochondria processes machinery for cellular respiration. Each membrane and each compartment of mitochondria has its unique proteins. Enzymes of electron transport chain lie in the inner membrane while most enzymes of citric acid cycle are found in the matrix.

Protein Targeting to Chloroplasts:

The newly synthesized proteins by free ribosomes are impored into chloroplasts as in mitochondria. Calvin cycle enzymes fix atmospheric CO2 into carbohydrates during photosynthesis.

Protein Targeting into Nucleus:

The nuclear envelope consists of outer and inner membranes and has inter membranous space between them. The outer membrane is continuous with ER and has ribosomes on it. Proteins for the nucleus are synthesized on free ribosomes in the cytosol and imported into nucleus through 3000-4000 nuclear pores known as nuclear pore complexes which are special gates.

The proteins that are imported into nucleus are in fully folded state and do not require any chaprones. Protiens imported into nucleus have targeting signal sequences on them which are called nuclear localization signals (NLS). Each one has 4-8 amino acids and they are internal sequences and not terminal. NLS is not cleaved from the protein. Due to this feature proteins can re-enter the nucleus whenever the nuclear envelope is lost during cell division.

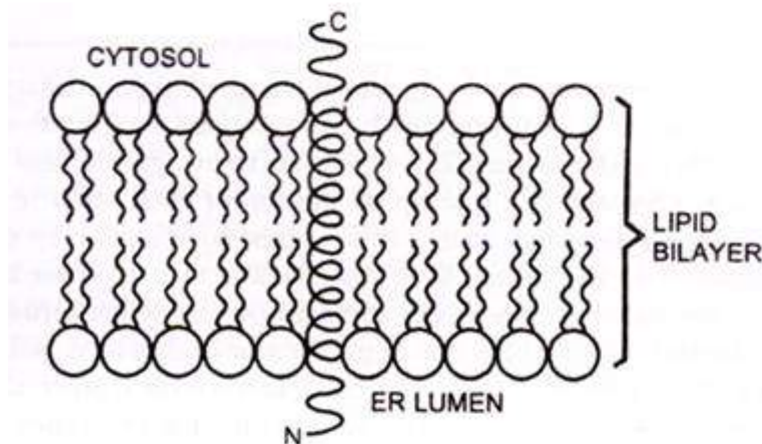


Fig. 21.6. Single pass topogenic sequence

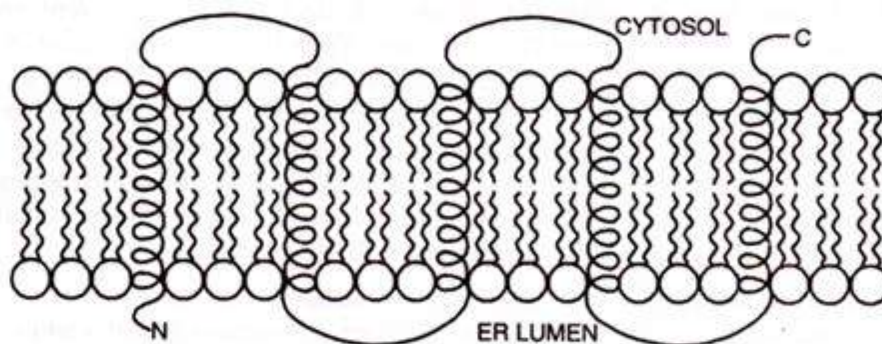


Fig. 21.7. Multi pass protein domains across ER membrane.

Membrane Proteins:

The proteins embedded in different membranes may have single trans-membrane domain which is a segment of 20-25 amino acids. Other proteins may have many trans-membrane domains connected by loops on both sides of the membrane. These proteins are called multi-pass orientation proteins. In photosynthetic bacteria a protein called bacterio-rodospin spans 12-14 times across the lipid bilayer membrane of bacteria. It traps energy from sunlight and uses it to pump protons across the bacterial membrane.