MODULE IV

CELL, MOLECULAR BIOLOGY AND BIOTECHNOLOGY

CELL

Cell membrane- Structure and function

- The cell membrane (plasma membrane) is a thin semi-permeable membrane that surrounds the cytoplasm of a cell.
- Its function is to protect the integrity of the interior of the cell by allowing certain substances into the cell while keeping other substances out.
- It also serves as a base of attachment for the cytoskeleton in some organisms and the cell wall in others.
- Thus the cell membrane also serves to help support the cell and help maintain its shape.
- Another function of the membrane is to regulate cell growth through the balance of endocytosis and exocytosis.
- In endocytosis, lipids and proteins are removed from the cell membrane as substances are internalized.
- In exocytosis, vesicles containing lipids and proteins fuse with the cell membrane increasing cell size.
- Animal cells, plant cells, prokaryotic cells, and fungal cells have plasma membranes. Internal organelles are also encased by membranes.

Cell Membrane Structure

- The cell membrane is primarily composed of a mix of proteins and lipids.
- Depending on the membrane's location and role in the body, lipids can make up anywhere from 20 to 80 percent of the membrane, with the remainder being proteins.

- While lipids help to give membranes their flexibility, proteins monitor and maintain the cell's chemical climate and assist in the transfer of molecules across the membrane.

Dry Weight
40% lipid – E.g. phospholipid molecules and cholesterol.
60% protein – E.g. channel proteins and carrier proteins.
1-10% carbohydrate – Often found attached to proteins/lipids on the outside of the cell membrane – a coat of carbohydrate surrounding a cell is often called the glycocalyx.

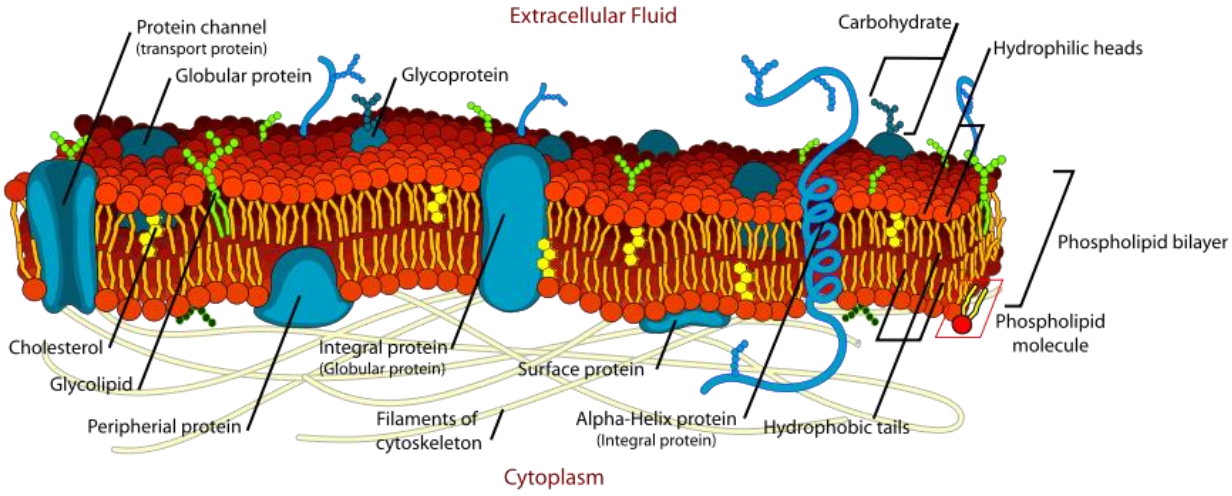


Fig: Structure of Cell membrane

Phospholipids

The membrane bilayer contains many kinds of phospholipid molecules, with different sized head and tail molecules.

These consist of a head molecule, a phosphate molecule, a glycerol and two fatty acid chains.

- **Head group-** This is a polar group e.g. a sugar or choline – meaning that the head end of the phospholipid is hydrophilic.
- **Tail of 2 fatty acid chains** – normally consisting of between 14-24 carbons (but the most common carbon lengths are 16 and 18). If the chain contains a cis double bond then the chain is kinked – therefore reducing the tight packing of the membrane and so increasing its movement. As the tail is made of fatty acids, it does not form hydrogen bonds with water and therefore is hydrophobic and non-polar.

Phospholipid molecules are therefore amphipathic – being both hydrophilic and hydrophobic. They spontaneously form bilayers in the water with the head groups facing out and the tail groups facing in.

In the bilayer, there are van der Waal forces between the fatty acid tails of the phospholipid, with electrostatic and hydrogen bonds between the hydrophilic groups and water.

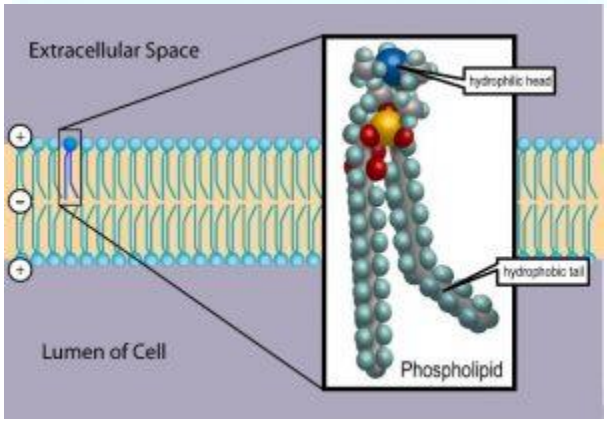


Fig: Structure of Phospholipid

Cholesterol

- Cholesterol is vital for many functions in a cell, including very importantly, a major constituent of the cell membrane.
- Cholesterol itself consists of a polar head, a **planar steroid ring** and a non-polar hydrocarbon tail.
- Cholesterol is important in the membrane as it helps to maintain cell membrane stability and fluidity at varying temperatures.
- Cholesterol is bound to neighbouring phospholipid molecules via hydrogen bonds and therefore at low temperatures, reduces their packing.
- Overall this means at low temperatures, when rate of movement is lowest, a fluid phase is maintained.
- At high temperatures, cholesterol helps to stop the formation of crystalline structures and the rigid planar steroid ring prevents intrachain vibration and therefore making the membrane less fluid.

Cell Membrane Proteins

- The cell membrane contains two types of associated proteins.
- **Peripheral membrane proteins** are exterior to and connected to the membrane by interactions with other proteins.
- **Integral membrane proteins** are inserted into the membrane and most pass through the membrane.
- Portions of these transmembrane proteins are exposed on both sides of the membrane. Cell membrane proteins have a number of different functions.
- **Structural proteins** help to give the cell support and shape.
- Cell membrane **receptor proteins** help cells communicate with their external environment through the use of hormones, neurotransmitters, and other signaling molecules.
- **Transport proteins**, such as globular proteins, transport molecules across cell membranes through facilitated diffusion.

- **Glycoproteins** have a carbohydrate chain attached to them.
- They are embedded in the cell membrane and help in cell to cell communications and molecule transport across the membrane.

Functions of the Cell Membrane

Cell membranes are vital for the normal functioning of all the cells in our bodies. Their main functions consist of:

- Forming a continuous, highly selectively permeable barrier – both around cells and intracellular compartments.
- Allowing the control of an enclosed chemical environment – important to maintain ion gradients.
- Communication – both with the extracellular and extra-organelle space.
- Recognition – including recognition of signaling molecules, adhesion proteins and other host cells (very important in the immune system).
- Signal generation – in response to a stimulus creating a change in membrane potential.

In a cell, different parts of the membrane have different functions and therefore their structure is specialised for this. An example of this specialisation can be seen in the different parts of a nerve; the cell membrane in the axon is specialised for electrical conduction whereas the end of the nerve is specialised for synapsing, meaning the composition of the membrane is different.

Cell organelles with special reference to Mitochondria and Ribosomes

Important Cell organelles

Within the cytoplasm, the major organelles and cellular structures include:

(1) Nucleolus
(2) Nucleus
(3) Ribosome
(5) Endoplasmic reticulum
(6) Golgi apparatus
(7) Cytoplasm
(8) Mitochondria
(9) Vacuole
(10) Lysosome
(12) Centriole

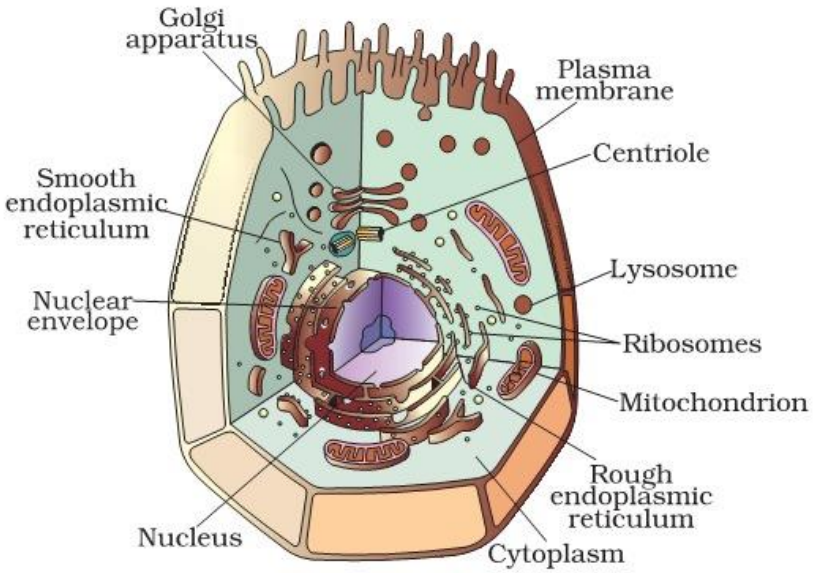
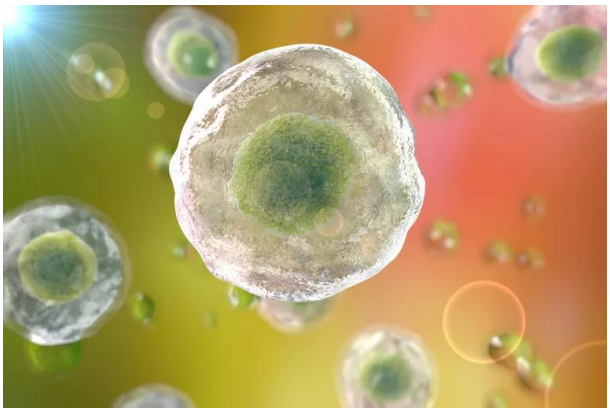


Fig. 5.5: *Animal cell*

1. **<u>Nucleus</u>**



The cell nucleus is a membrane-bound structure that contains a cell's hereditary information and controls its growth and reproduction. It is the command center of a eukaryotic cell and is usually the most notable cell organelle in both size and function.

Function

- The key function of the nucleus is to control cell growth and multiplication.
- This involves regulating gene expression, initiating cellular reproduction, and storing genetic material necessary for all of these tasks.
- In order for a nucleus to carry out important reproductive roles and other cell activities, it needs proteins and ribosomes.

Protein and Ribosome Synthesis

- The nucleus regulates the synthesis of proteins in the cytoplasm through the use of messenger RNA (mRNA).
- Messenger RNA is a transcribed DNA segment that serves as a template for protein production.
- It is produced in the nucleus and travels to the cytoplasm through the nuclear pores of the nuclear envelope, which you'll read about below. Once in the cytoplasm, ribosomes and another RNA molecule called transfer RNA work together to translate mRNA in order to produce proteins.

Physical Characteristics

The shape of a nucleus varies from cell to cell but is often depicted as spherical.

To understand more about the role of the nucleus, read about the structure and function of each of its parts.

- **Nuclear Envelope and Nuclear Pores**
  - The cell nucleus is bound by a double membrane called the **nuclear envelope**.
  - This membrane separates the contents of the nucleus from the cytoplasm, the gel-like substance containing all other organelles.
  - The nuclear envelope consists of phospholipids that form a lipid bilayer much like that of the cell membrane.
  - This lipid bilayer has nuclear pores that allow substances to enter and exit the nucleus, or transfer from the cytoplasm to the nucleoplasm.
  - The nuclear envelope helps to maintain the shape of the nucleus. It is connected to the **endoplasmic reticulum (ER)** in such a way that the internal chamber of the nuclear envelope is continuous with the lumen, or inside, of the ER. This also allows the transfer of materials as well.

- Chromatin
  - The nucleus houses chromosomes containing DNA.
  - DNA holds heredity information and instructions for cell growth, development, and reproduction. When a cell is "resting", or not dividing, its chromosomes are organized into long entangled structures called chromatin.

- Nucleoplasm

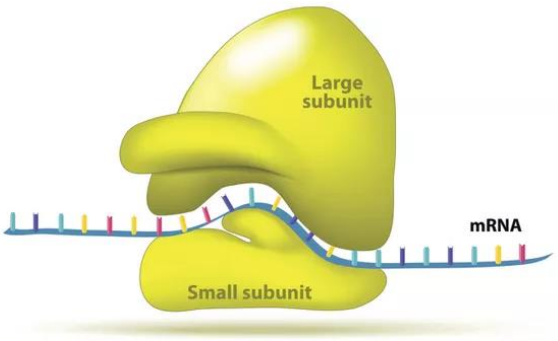
- Nucleoplasm is the gelatinous substance within the nuclear envelope. Also called karyoplasm, this semi-aqueous material is similar to cytoplasm in that it is composed mainly of water with dissolved salts, enzymes, and organic molecules suspended within.
- The nucleolus and chromosomes are surrounded by nucleoplasm, which cushions and protects nuclear contents.
- Like the nuclear envelope, the nucleoplasm supports the nucleus to hold its shape. It also provides a medium by which materials, such as enzymes and nucleotides (DNA and RNA subunits), can be transported throughout the nucleus to its various parts.

- Nucleolus
  - Contained within the nucleus is a dense, membrane-less structure composed of RNA and proteins called the nucleolus.
  - The nucleolus contains nucleolar organizers, the parts of chromosomes carrying the genes for ribosome synthesis.
  - The nucleolus helps to synthesize ribosomes by transcribing and assembling ribosomal RNA subunits.
  - These subunits join together to form ribosomes during protein synthesis.

2. Ribosomes

- Ribosomes are cell organelles that consist of RNA and proteins. They are responsible for assembling the proteins of the cell.
- Depending on the protein production level of a particular cell, ribosomes may number in the millions.

RIBOSOME

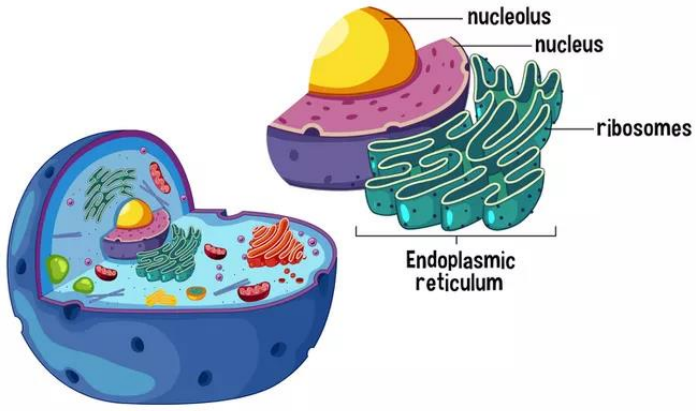


- Ribosomes are typically composed of two subunits: a large subunit and a small subunit.
- Eukarotic ribosomes (80S), such as those in plant cells and animal cells, are larger in size than prokaryotic ribosomes (70S), such as those in bacteria.
- Ribosomal subunits are synthesized in the nucleolus and cross over the nuclear membrane to the cytoplasm through nuclear pores.
- Both ribosomal subunits join together when the ribosome attaches to messenger RNA (mRNA) during protein synthesis.
- Ribosomes along with another RNA molecule, transfer RNA (tRNA), help to translate the protein-coding genes in mRNA into proteins.
- Ribosomes link amino acids together to form polypeptide chains, which are further modified before becoming functional proteins.
- Ribosomes are typically composed of two subunits: a large subunit and a small subunit.
- Eukarotic ribosomes (80S), such as those in plant cells and animal cells, are larger in size than prokaryotic ribosomes (70S), such as those in bacteria.
- Ribosomal subunits are synthesized in the nucleolus and cross over the nuclear membrane to the cytoplasm through nuclear pores.
- Both ribosomal subunits join together when the ribosome attaches to messenger RNA (mRNA) during protein synthesis.

- Ribosomes along with another RNA molecule, transfer RNA (tRNA), help to translate the protein-coding genes in mRNA into proteins.
- Ribosomes link amino acids together to form polypeptide chains, which are further modified before becoming functional proteins.

Location in the Cell

- There are two places where ribosomes commonly exist within a eukaryotic cell: suspended in the cytosol and bound to the endoplasmic reticulum. These ribosomes are called **free ribosomes and bound ribosomes** respectively.
- In both cases, the ribosomes usually form aggregates called polysomes or polyribosomes during protein synthesis.
- Polyribosomes are clusters of ribosomes that attach to a mRNA molecule during protein synthesis.
- This allows for multiple copies of a protein to be synthesized at once from a single mRNA molecule.

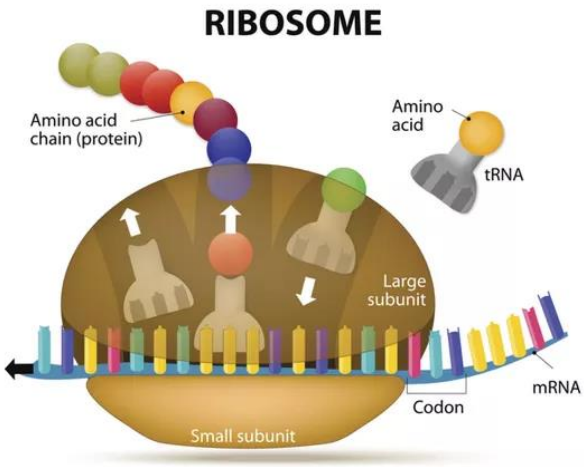


- Free ribosomes usually make proteins that will function in the cytosol (fluid component of the cytoplasm), while bound ribosomes usually make proteins that are exported from the cell or included in the cell's membranes.

- Interestingly enough, free ribosomes and bound ribosomes are interchangeable and the cell can change their numbers according to metabolic needs.
- Organelles such as mitochondria and chloroplasts in eukaryotic organisms have their own ribosomes. Ribosomes in these organelles are more like ribosomes found in bacteria with regard to size.
- The subunits comprising ribosomes in mitochondria and chloroplasts are smaller (30S to 50S) than the subunits of ribosomes found throughout the rest of the cell (40S to 60S).

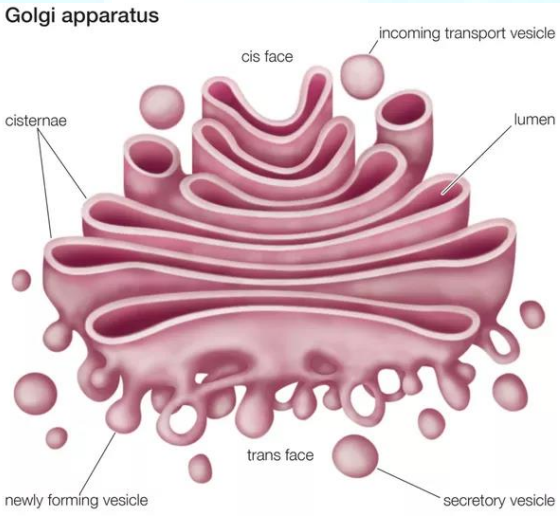
Ribosomes and Protein Assembly

- Protein synthesis occurs by the processes of **transcription and translation.**
- In transcription, the genetic code contained within DNA is transcribed into an RNA version of the code known as messenger RNA (mRNA).
- The mRNA transcript is transported from the nucleus to the cytoplasm where it undergoes translation.
- In translation, a growing amino acid chain, also called a polypeptide chain, is produced.
- Ribosomes help to translate mRNA by binding to the molecule and linking amino acids together to produce a polypeptide chain.
- The polypeptide chain eventually becomes a fully functioning protein. Proteins are very important biological polymers in our cells as they are involved in virtually all cell functions.



There are some differences between protein synthesis in eukaryotes and prokaryotes. Since eukaryotic ribosomes are larger than those in prokaryotes, they require more protein components. Other differences include different initiator amino acid sequences to start protein synthesis as well as different elongation and termination factors.

3. Golgi Apparatus



- The golgi apparatus consists of a system of membrane-bound vesicles arranged approximately parallel to each other in stacks called cisterns.
- These membranes often have connections with the membranes of ER and therefore constitute another portion of a complex cellular membrane system.

- The material synthesized near the ER is packaged and dispatched to various targets inside and outside the cell through the golgi apparatus.
- Its functions include the storage, modification and packaging of products in vesicles.
- In some cases, complex sugars may be made from simple sugars in the golgi apparatus.
- The golgi apparatus is also involved in the formation of lysosomes.

4. Mitochondria

- Mitochondria are considered the "powerhouses" of eukaryotic cells.
- These organelles generate power by converting energy into forms that are usable by the cell.
- Located in the cytoplasm, mitochondria are the sites of cellular respiration.
- Cellular respiration is a process that ultimately generates fuel for the cell's activities from the foods we eat.
- Mitochondria produce the energy required to perform processes such as cell division, growth, and cell death.
- Mitochondria have a distinctive oblong or oval shape and are bounded by a double membrane.
- The inner membrane is folded creating structures known as cristae.
- Mitochondria are found in both animal and plant cells.
- They are found in all body cell types, except for mature red blood cells.
- The number of mitochondria within a cell varies depending on the type and function of the cell.
- As mentioned, red blood cells do not contain mitochondria at all.
- The absence of mitochondria and other organelles in red blood cells leaves room for the millions of hemoglobin molecules needed in order to transport oxygen throughout the body.

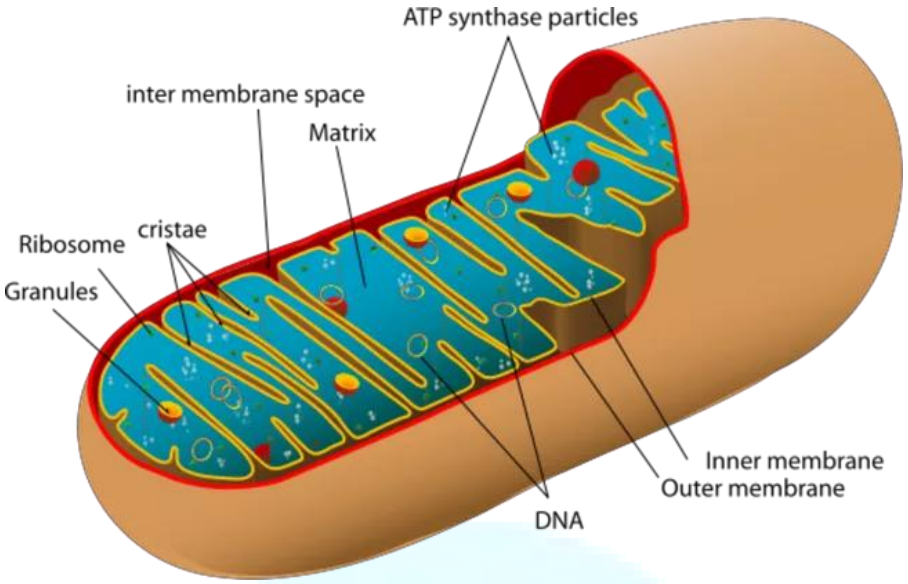
- Muscle cells, on the other hand, may contain thousands of mitochondria needed to provide the energy required for muscle activity. Mitochondria are also abundant in fat cells and liver cells.

Mitochondrial DNA

- Mitochondria have their own DNA, ribosomes and can make their own proteins.
- Mitochondrial DNA (mtDNA) encodes for proteins that are involved in electron transport and oxidative phosphorylation, which occur in cellular respiration.
- In oxidative phosphorylation, energy in the form of ATP is generated within the mitochondrial matrix.
- Proteins synthesized from mtDNA also encode for the production of the RNA molecules transfer RNA and ribosomal RNA.
- Mitochondrial DNA differs from DNA found in the cell nucleus in that it does not possess the DNA repair mechanisms that help prevent mutations in nuclear DNA.
- As a result, mtDNA has a much higher mutation rate than nuclear DNA.
- Exposure to reactive oxygen produced during oxidative phosphorylation also damages mtDNA.

Mitochondria Anatomy and Reproduction

- **Mitochondrial Membranes**
    - Mitochondria are bounded by a double membrane.
    - Each of these membranes is a phospholipid bilayer with embedded proteins.
    -



- The outermost membrane is smooth while the inner membrane has many folds.
- These folds are called cristae. The folds enhance the "productivity" of cellular respiration by increasing the available surface area.
- Within the inner mitochondrial membrane are a series of protein complexes and electron carrier molecules, which form the electron transport chain (ETC).
- The ETC represents the third stage of aerobic cellular respiration and the stage where the vast majority of ATP molecules are generated.
- ATP is the body's main source of energy and is used by cells to perform important functions, such as muscle contraction and cell division.

- **Mitochondrial Spaces**
  - The double membranes divide the mitochondrion into two distinct parts: the intermembrane space and the mitochondrial matrix.
  - The intermembrane space is the narrow space between the outer membrane and the inner membrane, while the mitochondrial

matrix is the area that is completely enclosed by the innermost membrane.
    - The mitochondrial matrix contains mitochondrial DNA (mtDNA), ribosomes, and enzymes.
    - Several of the steps in cellular respiration, including the Citric Acid Cycle and oxidative phosphorylation occur in the matrix due to its high concentration of enzymes.

- **Mitochondrial Reproduction**
    - Mitochondria are semi-autonomous in that they are only partially dependent on the cell to replicate and grow.
    - They have their own DNA, ribosomes, make their own proteins, and have some control over their reproduction.
    - Similar to bacteria, mitochondria have circular DNA and replicate by a reproductive process called binary fission.
    - Prior to replication, mitochondria merge together in a process called fusion.
    - Fusion is needed in order to maintain stability as, without it, mitochondria will get smaller as they divide.
    - These smaller mitochondria are not able to produce sufficient amounts of energy needed for proper cell function.

5. **<u>Endoplasmic reticulum</u>**

The endoplasmic reticulum (ER) is an important organelle in eukaryotic cells. It plays a major role in the production, processing, and transport of proteins and lipids. The ER produces transmembrane proteins and lipids for its membrane and many other cell components including lysosomes, secretory vesicles, the Golgi apparatus, the cell membrane, and plant cell vacuoles.

- The endoplasmic reticulum is a network of tubules and flattened sacs that serve a variety of functions in plant and animal cells.

- The two regions of the ER differ in both structure and function.
- Rough ER has ribosomes attached to the cytoplasmic side of the membrane. Smooth ER lacks attached ribosomes.
- Typically, the smooth ER is a tubule network and the rough ER is a series of flattened sacs.
- The space inside of the ER is called the lumen.
- The ER is very extensive, extending from the cell membrane through the cytoplasm and forming a continuous connection with the nuclear envelope.
- Since the ER is connected with the nuclear envelope, the lumen of the ER and the space inside the nuclear envelope are part of the same compartment.

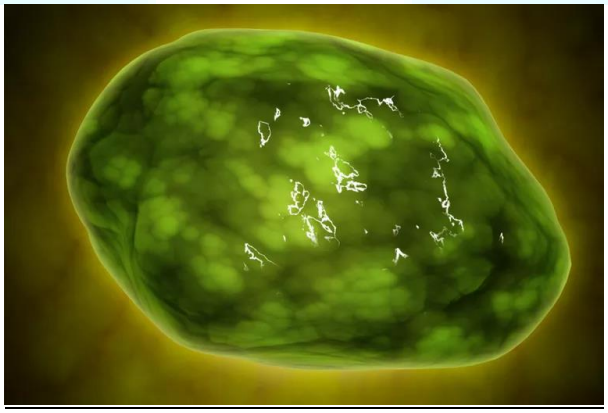
Rough Endoplasmic Reticulum

- The rough endoplasmic reticulum manufactures membranes and secretory proteins.
- The ribosomes attached to the rough ER synthesize proteins by the process of translation.
- In certain leukocytes (white blood cells), the rough ER produces antibodies.
- In pancreatic cells, the rough ER produces insulin.
- The rough and smooth ER are usually interconnected and the proteins and membranes made by the rough ER move into the smooth ER to be transferred to other locations.
- Some proteins are sent to the Golgi apparatus by special transport vesicles.
- After the proteins have been modified in the Golgi, they are transported to their proper destinations within the cell or exported from the cell by exocytosis.

Smooth Endoplasmic Reticulum

- The smooth ER has a wide range of functions including carbohydrate and lipid synthesis.
- Lipids such as phospholipids and cholesterol are necessary for the construction of cell membranes.
- Smooth ER also serves as a transitional area for vesicles that transport ER products to various destinations.
- In liver cells the smooth ER produces enzymes that help to detoxify certain compounds.
- In muscles the smooth ER assists in the contraction of muscle cells, and in brain cells it synthesizes male and female hormones.

6. Lysosomes

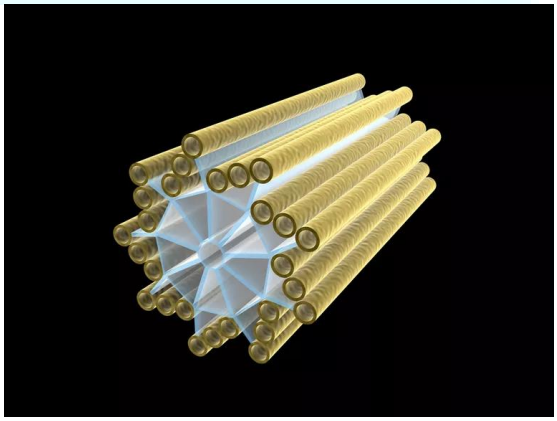


- Lysosomes are a kind of waste disposal system of the cell.
- Lysosomes help to keep the cell clean by digesting any foreign material as well as worn-out cell organelles.
- Foreign materials entering the cell, such as bacteria or food, as well as old organelles end up in the lysosomes, which break them up into small pieces.
- Lysosomes are able to do this because they contain powerful digestive enzymes capable of breaking down all organic material.

- During the disturbance in cellular metabolism, for example, when the cell gets damaged, lysosomes may burst and the enzymes digest their own cell.
- Therefore, lysosomes are also known as the ‘suicide bags’ of a cell.
- Structurally, lysosomes are membrane-bound sacs filled with digestive enzymes. These enzymes are made by RER.

7. **Centrioles**

Centrioles are cylindrical cell structures that are composed of groupings of microtubules, which are tube-shaped molecules or strands of protein.



- Without centrioles, chromosomes would not be able to move during the formation of new cells.
- Centrioles help to organize the assembly of microtubules during cell division. To put it simply, chromosomes use the centriole's microtubules as a highway during the cell division process.

Where Centrioles Are Found

- Centrioles are found in all animal cells and only a few species of lower plant cells. Two centrioles—a mother centriole and a daughter centriole—are found within the cell in a structure called a centrosome.

Composition

- Most centrioles are made up of nine sets of microtubule triplets, with the exception of some species, such as crabs which have nine sets of microtubule doublets.
- There are a few other species that deviate from the standard centriole structure. Microtubules are composed of a single type of globular protein called tubulin.

Two Main Functions

- During mitosis or cell division, the centrosome and centrioles replicate and migrate to opposite ends of the cell.
- Centrioles help to arrange the microtubules that move chromosomes during cell division to ensure each daughter cell receives the appropriate number of chromosomes.
- Centrioles are also important for the formation of cell structures known as cilia and flagella.
- Cilia and flagella, found on the outside surface of cells, aid in cellular movement.
- A centriole combined with several additional protein structures is modified to become a basal body.
- Basal bodies are the anchoring sites for moving cilia and flagella.

8. **Plasma Membrane or Cell Membrane**

- Cell membrane is also called the plasma membrane.
- It can be observed only through an electron microscope.
- Plasma membrane is the outermost covering of the cell that separates the contents of the cell from its external environment.

9. **Cytoplasm**

- It is the jelly-like substance present between the cell membrane and the nucleus.

- The cytoplasm is the fluid content inside the plasma membrane.
- It also contains many specialized cell organelles [mitochondria, golgi bodies, ribosomes, etc].
- Each of these organelles performs a specific function for the cell.
- Cell organelles are enclosed by membranes.
- The significance of membranes can be illustrated with the example of viruses.
- Viruses lack any membranes and hence do not show characteristics of life until they enter a living body and use its cell machinery to multiply.

MOLECULAR BIOLOGY

Organization of eukaryotic genome

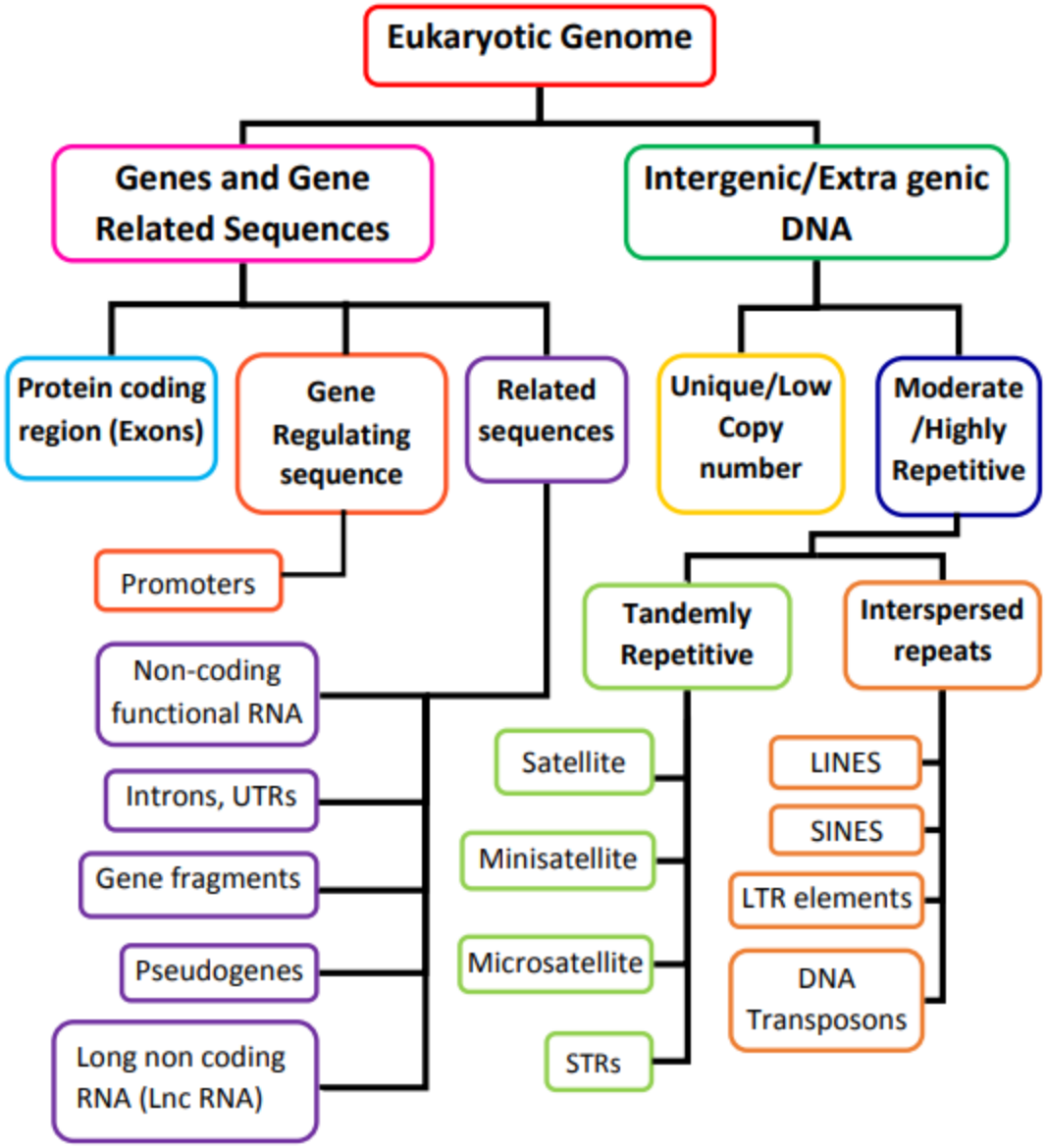
A genome is an organism's complete set of DNA, comprising of nuclear and mitochondrial DNA. Each genome contains all of the information needed to build and maintain that organism. A human haploid cell, consist of 23 nuclear chromosomes and one mitochondrial chromosome, contains more than 3.2 billion DNA base pairs.

Eukaryotic genome is linear and conforms the Watson-Crick Double Helix structural model. Embedded in Nucleosome-complex DNA & Protein (Histone) structure that pack together to form chromosomes. Eukaryotic genome have unique features of Exon - Intron organization of protein coding genes, representing coding sequence and intervening sequence that represents the functionality of RNA part inside the genome.

Configuration of Eukaryotic genome:

The configuration of eukaryotic genome includes protein coding region, gene regulating region, gene related sequence and intergenic DNA or extra genic DNA which includes low copy number and moderate or high copy

number repetitive sequence, the flow chart representation of configuration is given below:



1. GENES AND GENE RELATED SEQUENCES

A. PROTEIN CODING REGION (EXONS)

- Protein coding sequences are the DNA sequences that are transcribed into mRNA later translated to proteins.

- The complete protein coding genes capacity of the genome is contained within the exomes (the part of the genome formed by exons, the sequences which when transcribed remain within the mature RNA sequence after introns are removed using RNA splicing)and consists of DNA sequences encoded by exons that can be later translated into proteins.
- It consists of ORF (Open reading frame). These are the reading frame that has the potential to code for the proteins/peptide.
- It is stretch of codons that do not contain a stop codon (UAA, UAG, and UGA).
- An AUG with the ORF may indicate where translation starts.

B. GENE REGULATING SEQUENCES

Promoters are combinations of short sequence elements (usually located in the immediate upstream region of the gene often within 200 bp of the transcription start site) which serve to initiate transcription. They can be subdivided into different components.

- The **Core promoter** directs the basal transcription complex to initiate transcription of the gene. In the absence of additional regulatory elements it permits constitutive expression of the gene, but at very low (basal) levels. Core promoter elements are typically located very close to the transcription initiation site, at about nucleotide position -45 to +40. They include: the TATA box located at position ca. -25, surrounded by GC-rich sequences and recognized by the TATA- binding protein subunit of TFIID; the BRE sequence located immediately upstream of the TATA elements at around -35 and recognized by the TFIIB component; the Inr (initiator) sequence located at the start site of transcription and bound by TFIID; the DPE or Downstream Promoter Element, located at about position +30 relative to transcription and recognized by TFIID.

- The **proximal promoter** region is the sequence located immediately upstream of the core promoter, usually from -50 to -200 bp (promoter elements found further upstream would be said to map to the distal promoter region). They include: GC boxes (also called Sp1 boxes, the consensus sequence is GGGCGG which is often found in multiple copies within 100 bp of the transcription initiation site); CCAAT boxes typically located at position -75.
- **ENHANCERS** are positive transcriptional control elements which are particularly prevalent in the cells of complex eukaryotes such as mammals but which are absent or very poorly represented in simple eukaryotes such as yeast. They serve to increase the basal level of transcription which is initiated through the core promoter elements. Their function, unlike those of the core promoter, are independent of both their orientation and, to some extent, their distance from the genes they regulate. Enhancers often contain within a span of only 200-300 bp.
- **SILENCERS** serve to reduce transcription levels.
- **BOUNDARY ELEMENTS (INSULATORS)** are regions of DNA, often spanning from 0.5 kb to 3 kb, which function to block the spreading of the influence of agents that have a positive effect on transcription (enhancers) or negative one (silencers, hetero-chromatinlike repressive effects).
- **RESPONSE ELEMENTS** modulate transcription in response to specific external stimuli. They are usually located a short distance upstream of the promoter elements (often within 1kb of the transcription start site). A variety of such elements respond to specific hormones or to intracellular second messengers such as cyclic AMP.

C. RELATED SEQUENCES

1) NON CODING FUNCTIONAL RNA:

A non-coding RNA (ncRNA) is an RNA molecule that is not translated into a protein. Less frequently used synonyms are non-protein-coding RNA (npcRNA), non-messenger RNA (nmRNA) and functional RNA (fRNA). The DNA sequence from which a functional non-coding RNA is transcribed is often called an RNA gene.

2) INTRONS, UTRS

The term intron refers to both the DNA sequence within a gene and the corresponding sequence in RNA transcripts. At least four distinct classes of introns have been identified.

- Introns in nuclear protein-coding genes that are removed by spliceosomes (spliceosomal introns)
- Introns in nuclear and archaeal transfer RNA genes that are removed by proteins (tRNA introns)
- Self-splicing group I introns that are removed by RNA catalysis.
- Self-splicing group II introns that are removed by RNA catalysis.
- Group III introns are proposed to be a fifth family, but little is known about the biochemical apparatus that mediates their splicing. They appear to be related to group II introns, and possibly to spliceosomal intron.

In molecular genetics, an untranslated region (or UTR) refers to either of two sections, one on each side of a coding sequence on a strand of mRNA. If it is found on the 5' side, it is called the 5' UTR (or leader sequence), or if it is found on the 3' side, it is called the 3' UTR (or trailer sequence).

3) GENE FRAGMENTS

Gene fragments are pieces of genes containing only the exons (those parts of the gene which actually encode the protein sequence). They are composed of cDNA.

4) PSEUDOGENES:

Pseudogenes are dysfunctional relatives of genes that have lost their gene expression in the cell or their ability to code protein. Pseudogenes often result from the accumulation of multiple mutations within a gene whose product is not required for the survival of the organism. Depending on their DNA sequence characteristics pseudogenes are mainly of two types:

- **Processed pseudogene:** They have all the normal parts of a protein-coding gene, but was originally thought to be incapacitated based on presumed DNA code errors.
- **Unprocessed pseudogene:** They lacks the intervening non-protein coding sequences called introns, which are typically spliced out when a messenger RNA (mRNA transcript) is produced from a gene.

5) LONG NON CODING RNA (LNC RNA)

Long non-coding RNAs (lncRNA) are a type of non-coding RNAs (ncRNAs) that exceed 200 nucleotides in length. lncRNAs are a relatively abundant component of the mammalian transcriptome and have been implicated in several cellular functions, including the regulation of gene transcription through the recruitment of chromatin-modifying enzymes.

Human genome contains many thousands of Lnc-RNA these is again divided into following categories:

a. MACRO Lnc-RNA: Lnc-RNA particularly lying in imprinted clusters, predominantly unspliced “macro” transcripts that can also show a low level of splicing, and both the unspliced and spliced transcript have the potential to be functional. Three known imprinted macro lncRNAs have been shown to silence from three to ten genes in cis in imprinted gene clusters.

b. BIDIRECTIONAL Lnc-RNA: It is oriented head to head with a protein coding gene within 1 kb. It exhibits a similar expression patterns to its protein coding counterpart that suggests that may be subject to share

a regulatory pressure. Non coding locus that originates from within the promoter region of a protein-coding gene, with transcription proceeding in the opposite direction on the other strand.

c. SENSE- OVERLAPPING LncRNA: These can be considered transcript variants of protein - coding RNA as they overlap with a known annotated gene on the same genomic strand. Majority of these lack substantial orf for protein translation while others have orfs that shares the same start codon as protein coding transcript for that gene, but encode protein for several reasons including non-sense mediated decay issues that limit the translation of mRNAs with premature termination stop codons and triggered NMD mediated destruction of the m RNA or an upstream alternative open reading frame which inhibits the translation of the predicted ORF.

d. SENSE- INTRONIC LncRNA: Sense intronic reside within of a coding gene, but do not intersect any exons.

e. RETAINED- INTRON: An Alternative spliced transcript that contain transcribed intronic sequence with respect to isoform of the locus, coding or other variants.

f. NC_RNA HOST: It has a separate non coding transcript that host non coding micro RNA. Such as MIRs.

g. LINC RNA (LONGINTERGENIC_RNA): "Intergenic" refers to long non-coding RNAs that are transcribed from non-coding DNA sequences between protein-coding genes. Requires lack of coding potential and may not be conserved between species. These have length >200bp. They are named according to the 3- protein coding gene nearby.

h. ANTISENSE: RNA molecules that are transcribed from the antisense strand and overlap in part well defined spliced sense or intron less sense RNA. Antisense -overlapping lncRNA have the tendency to undergo fewer splicing events and typically show lower abundance than sense transcript.

i. AMBIGUOUS _ORF: Has transcripts that are believed to be protein coding, but have more than one possible open reading frame.

j. 3 PRIME_OVERLAPPING_NCRNA: Has transcripts where ditag and/or published experimental data strongly supports the existence of long (>200bp) non-coding transcripts that overlap the 3'UTR of a protein-coding locus on the same strand.

k. NONCODING: Contains transcripts which are known from the literature to not be protein coding.

2. INTERGENIC /EXTRAGENIC DNA

An Intergenic region (IGR) is a stretch of DNA sequences located between genes. Intergenic regions are a subset of Noncoding DNA. Occasionally some intergenic DNA acts to control genes nearby, but most of it has no currently known function. It is sometimes referred to as junk DNA.

In humans, intergenic regions comprise about 75% of the genome, whereas this number is much less in bacteria (15%) and yeast (30%) Intergenic regions are different from intragenic regions (or introns), which are short, noncoding regions that are found within genes, especially within the genes of eukaryotic organisms.

Do contain functionally important elements such as promoters and enhancers. Also intergenic regions may contain as yet unidentified genes such as noncoding RNAs. They are thought to have regulatory functions.

A. UNIQUE/LOW COPY NUMBER

Low Copy Number (LCN) is a DNA profiling or DNA testing technique developed by the Forensic Science Service (FSS) which has been in use since 1999.

LCN is an extension of Second Generation Multiplex Plus (SGM Plus) profiling technique. It is a more sensitive technique because it involves a greater amount of copying via polymerase chain reaction (PCR) from a smaller amount of starting material, meaning that a profile can be obtained from only a few cells, which may be as small as a millionth the size of a grain of salt, and amount to a just few cells of skin or sweat left from a fingerprint.

B. MODERATE/HIGHLY REPETITIVE

Genome contain some repetitive DNA sequences, including repetitive coding DNA. However, the majority of highly repetitive DNA sequences occur outside genes. Some of the sequences are present at certain sub chromosomal regions as large arrays of tandem repeats. This type of DNA, known as heterochromatin, remains highly condensed throughout the cell cycle and does not generally contain genes.

1) TANDEMLY REPETITIVE REGION

Highly repeated noncoding human DNA often occurs in arrays (or blocks) of tandem repeats of sequence which may be a simple one (1-10 nucleotides), or a moderately complex one (tens to hundreds of nucleotides). Individual arrays can occur at a few or many different chromosomal locations.

1(a) Satellite DNA: Satellite DNA is transcriptionally inactive as is the vast majority of minisatellite DNA, but in the case of microsatellite DNA a significant percentage is located in coding DNA.

- Tandem repeats occur in DNA when a pattern of one or more nucleotides is repeated and the repetitions are directly adjacent to each other.

1(b) Minisatellites DNA: It comprises of a collection of moderately sized arrays of tandemly repeated DNA sequence which are dispersed over considerable portions of nuclear genome.

- A minisatellite, is a tract of repetitive DNA in which certain DNA motifs (ranging in length from 10–60 base pairs) are typically repeated 5-50 times.
- Minisatellites occur at more than 1,000 locations in the human genome and they are notable for their high mutation rate and high diversity in the population.
- Minisatellites are prominent in the centromeres and telomeres of chromosomes, the latter protecting the chromosomes from damage.
- Hypervariable minisatellite DNA sequences are highly polymorphic and are organized in over 1000 arrays (from 0.1 to 20 kb long) of short tandem repeats. its significance is not clear, although it has been reported to be a 'hotspot' for homologous recombination in human cells.

1(c) Microsatellites DNA: Microsatellite DNA, also called simple sequence repeats (SSR), are small arrays of tandem repeats of a simple sequence (usually less than 10 bp). They are interspersed throughout the genome, accounting for over 60 Mb (2% of genome), and are thought to have arisen mostly by replication slippage.

A microsatellite is a tract of repetitive DNA in which certain DNA motifs (ranging in length from 2–5 base pairs) are repeated, typically 5-50 times. Microsatellites occur at thousands of locations in the human genome and they are notable for their high mutation rate and high diversity in the population. Microsatellites

are often referred to as short tandem repeats (STRs) by forensic geneticists, or as simple sequence repeats (SSRs) For example, the sequence TATATATATA is a dinucleotide microsatellite.

1(d) STRS: Short tandem repeat is a microsatellite, consisting of a unit of two to thirteen nucleotides repeated hundreds of times in a row on the DNA strand.

2) INTERSPERSED REPEAT

Interspersed repeats mainly come from transposable elements (TEs), but they also include some protein coding gene families and pseudogenes. Transposable elements are able to integrate into the genome at another site within the cell. It is believed that TEs are an important driving force on genome evolution of higher eukaryotes. TEs can be classified into two categories, Class 1 (retrotransposons) and Class 2 (DNA transposons).

2(a) LINES:

- LINEs (long interspersed nuclear elements) have been very successful transposons.
- They have a comparatively long evolutionary history, occurring in other mammals, including mice.
- As autonomous transposons, they can make all the products needed for retro transposition, including the essential reverse transcriptase.
- Human LINEs consist of three distantly related families: LINE-1, LINE2, and LINE-3, collectively comprising about 20% of the genome.
- They are located primarily in euchromatic regions and are located preferentially in the dark AT-rich G bands (Giemsa-positive) of metaphase chromosomes.

2(b) SINES:

- SINEs (short interspersed nuclear elements) are retrotransposons about 100–400 bp in length.

- They have been very successful in colonizing mammalian genomes, resulting in various interspersed DNA families, some with extremely high copy numbers.
- Unlike LINEs, SINEs do not encode any proteins and they cannot transpose independently. However, SINEs and LINEs
- share sequences at their 3' end, and SINEs have been shown to be mobilized by neighboring LINEs.
- By parasitizing on the LINE element transposition machinery, SINEs can attain very high copy numbers.
- The human Alu family is the most prominent SINE family in terms of copy number, and is the most abundant sequence in the human genome, occurring on average more than once every 3 kb.

2(c) LTR ELEMENTS:

- LTR transposons include autonomous and non-autonomous retrovirus-like elements that are flanked by long terminal repeats (LTRs) containing necessary transcriptional regulatory elements.
- Endogenous retroviral sequences contain gag and pol genes, which encode a protease, reverse transcriptase, RNAse H, and integrase.
- They are thus able to transpose independently. There are three major classes of human endogenous retroviral sequence (HERV), with a cumulative copy number of about 240,000, accounting for a total of about 4.6% of the human genome.

2(d) DNA TRANSPOSONS:

- The cut-and-paste transposition mechanism of class II TEs does not involve an RNA intermediate. The transpositions are catalyzed by several transposase enzymes.
- Some transposases non-specifically bind to any target site in DNA, whereas others bind to specific DNA sequence targets.

- DNA transposons have terminal inverted repeats and encode a transposase that regulates transposition.
- They account for close to 3% of the human genome.
- Ritually all the resident human DNA transposon sequences are no longer active; they are therefore transposon fossils.
- DNA transposons tend to have short lifespans within a species, unlike some of the other transposable elements such as LINEs.
- However, quite a few functional human genes seem to have originated from DNA transposons, notably genes encoding the RAG1 and RAG2 recombinases and the major centromere binding protein CENPB.

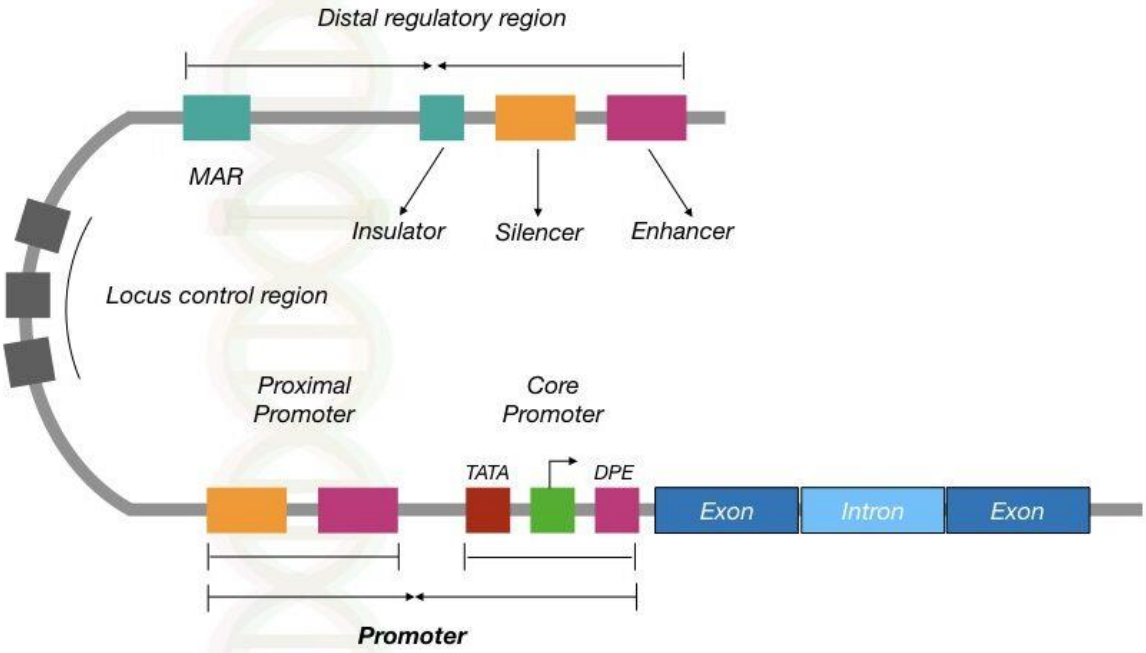
Gene content and genomic size complexity of eukaryotic genome

A functional segment of DNA which manufacture protein, regulate gene expression and renowned as a hereditary unit is known as a gene.

Structure of gene:

- Genes are actually DNA strands thus are made up of the nucleotide chain. The chemical structure of a gene comprises nucleotides.
- A part of DNA- genes are made up of **A, T, G and C nucleotides.**
- With the nucleotides of the opposite strand, it binds with hydrogen bonds and with the adjacent nucleotide, it binds with phosphodiester bonds.
- The nucleotides are the combination of **nitrogenous bases (A, T, G and C), phosphate and pentose sugar.**
- In general, the gene structure consists of two types of elements: **core elements and regulatory elements**.
    - **Core element:** The core elements or sequences actually take parts in protein formation.
    - **Regulatory elements:** maintain gene expression.

    - **Exons** are core elements. Sequences on the other side like promoters, enhancers and silencers are regulatory elements of a gene.
- The third type of element called **maintenance elements** possesses information for DNA repair, modification and replication. The functional or physical structure of a gene comprises introns, exons, promotes, enhancers and UTRs.
- **Introns** are intervening non-coding sequences removed from the final transcript.
- **Exons** are coding parts of a gene which are joined after splicing and constructs the final transcript.
- **Regulatory elements** are located on the extreme ends of a gene.
- **Promotes** are non-coding sequences but facilitates binding sites for enzymes and transcriptional factors to work. The promoter consists of TATA box and CCAAT sequences for enzyme binding.
- The entire promoter region is located on the 5’ end and made up of core promoter and proximal promoter sequences.
- Here, the core promoter facilitates RNA polymerase bindings (and other proteins) to start transcription. While the proximal promoter provides bindings for transcriptional factors.
- The **enhancer** induces transcription while the silencer represses it. Collectively, enhancers and silencers located far away from the exon, regulate gene expression.
-



- The 3' untranslated regions are non-coding regions of gene help in aborting the process of transcription and to form the final transcript.
- Once the RNA polymerase reaches the untranslated region it stops synthesizing RNA and detached from the strand.
- The eukaryotic gene structure consists of more regulatory sequences than prokaryotic genes. In addition to this, the entire machinery of transcription and translation is different in both.
- The operon concept of prokaryotic genes consists of a gene cluster of similar functions. Introns are not a part of an operon.
- Contrary, the eukaryotic genes consist of introns (non-coding DNA) at regular intervals. Each and every gene have their own promoter region to facilitate transcription.
- All the non-coding elements that help in gene regulation are divided into two categories viz cis-acting elements and trans-acting elements.
- Promoters, enhancers, silencers, activators, insulators, locus control regions and MARs- matrix attachment regions are categorized into cis-elements.

- While other transcriptional proteins which are formed from some genes are categorized into **trans-elements**.

Conserved exons and recombination

- The evolution of new genes is critical for speciation. Exon recombination, also known as exon shuffling or domain shuffling, is an important means of new gene formation.
- It is observed across vertebrates, invertebrates, and in some plants such as potatoes and sunflowers.
- During exon recombination, exons from the same or different genes recombine and produce new exon-intron combinations, which might evolve into new genes.
- Exon shuffling follows "splice frame rules." Each exon has three reading frames.
- The incoming exons can recombine and join at any one of the three reading frames and cause frameshift mutations.
- Therefore, not all recombination events are useful; some can even result in a premature stop codon and immature protein.

Exon shuffling in the human genome

- Along with gene duplication and divergence, the exon shuffling is attributed to the evolution of several human-specific genes. For example, around 25 million years ago, a gene called MCH (Melanin-concentrating hormone) underwent exon recombination by retrotransposition in the early primates.
- It created de novo intron-exon boundaries, which later evolved into the Hominidae specific conserved gene PMCHL1 - although this is a pseudogene, the antisense RNA is expressed in the human brain.
- Since the original MCH gene encoded a neuropeptide involved in balancing energy requirements and body weight in rodents, the PMCHL-1 is expected to have similar functions.

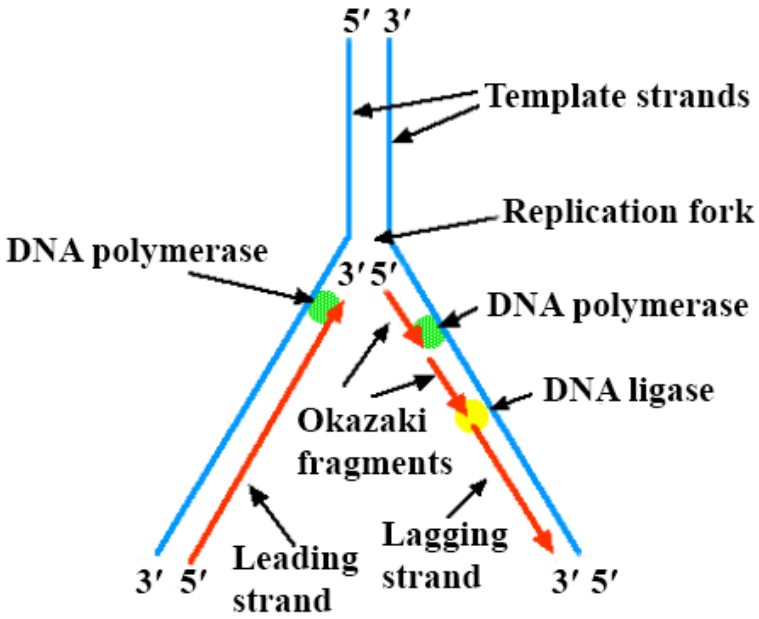
- Illegitimate recombination (IR) is one of the most commonly observed mechanisms of exon recombination or exon shuffling.
- IR leads to duplication of exons; this has been observed in several human diseases such as Duchenne and Becker muscular dystrophy, familial hypercholesterolemia, Lesch-Nyhan syndrome, hemophilia, and lipoprotein lipase deficiency.

DNA replication , Repair and Recombination

The **process of DNA duplication** is called DNA replication. Replication follows several steps that involve multiple proteins called replication enzymes and RNA. In eukaryotic cells, such as animal cells and plant cells, DNA replication occurs in the S phase of interphase during the cell cycle. The process of DNA replication is vital for cell growth, repair, and reproduction in organisms.

Preparation for Replication

Step 1: Replication Fork Formation



- Before DNA can be replicated, the double stranded molecule must be “unzipped” into two single strands.
- DNA has four bases called adenine (A), thymine (T), cytosine (C) and guanine (G) that form pairs between the two strands.
- Adenine only pairs with thymine and cytosine only binds with guanine.
- In order to unwind DNA, these interactions between base pairs must be broken.
- This is performed by an enzyme known as DNA helicase. DNA helicase disrupts the hydrogen bonding between base pairs to separate the strands into a Y shape known as the replication fork. This area will be the template for replication to begin.
- DNA is directional in both strands, signified by a 5' and 3' end.
- This notation signifies which side group is attached the DNA backbone. The 5' end has a phosphate (P) group attached, while the 3' end has a hydroxyl (OH) group attached.
- This directionality is important for replication as it only progresses in the 5' to 3' direction.
- However, the replication fork is bi-directional; one strand is oriented in the 3' to 5' direction (leading strand) while the other is oriented 5' to 3' (lagging strand).
- The two sides are therefore replicated with two different processes to accommodate the directional difference.

Replication Begins

Step 2: Primer Binding

- The leading strand is the simplest to replicate.
- Once the DNA strands have been separated, a short piece of RNA called a primer binds to the 3' end of the strand.
- The primer always binds as the starting point for replication. Primers are generated by the enzyme DNA primase.

ENTRI

DNA Elongation

Step 3: Elongation

- Enzymes known as DNA polymerases are responsible creating the new strand by a process called elongation.
- There are five different known types of DNA polymerases in bacteria and human cells. In bacteria such as E. coli, polymerase III is the main replication enzyme, while polymerase I, II, IV and V are responsible for error checking and repair.
- DNA polymerase III binds to the strand at the site of the primer and begins adding new base pairs complementary to the strand during replication.
- In eukaryotic cells, polymerases alpha, delta, and epsilon are the primary polymerases involved in DNA replication.
- Because replication proceeds in the 5' to 3' direction on the leading strand, the newly formed strand is continuous.
- The lagging strand begins replication by binding with multiple **primers**.
- Each primer is only several bases apart.
- DNA polymerase then adds pieces of DNA, called **Okazaki fragments**, to the strand between primers.
- This process of replication is discontinuous as the newly created fragments are disjointed.

Step 4: Termination

- Once both the continuous and discontinuous strands are formed, an enzyme called **exonuclease** removes all RNA primers from the original strands.
- These primers are then replaced with appropriate bases.
- Another exonuclease “proofreads” the newly formed DNA to check, remove and replace any errors.

- Another enzyme called **DNA ligase** joins Okazaki fragments together forming a single unified strand.
- The ends of the linear DNA present a problem as DNA polymerase can only add nucleotides in the 5′ to 3′ direction.
- The ends of the parent strands consist of repeated DNA sequences called **telomeres.**
- Telomeres act as protective caps at the end of chromosomes to prevent nearby chromosomes from fusing.
- A special type of DNA polymerase enzyme called **telomerase** catalyzes the synthesis of telomere sequences at the ends of the DNA.
- Once completed, the parent strand and its complementary DNA strand coils into the familiar double helix shape.
- In the end, replication produces two DNA molecules, each with one strand from the parent molecule and one new strand.

Enzyme involved in replication

DNA replication would not occur without enzymes that catalyze various steps in the process. Enzymes that participate in the eukaryotic DNA replication process include:

- **DNA helicase** - unwinds and separates double stranded DNA as it moves along the DNA. It forms the replication fork by breaking hydrogen bonds between nucleotide pairs in DNA.
- **DNA primase** - a type of RNA polymerase that generates RNA primers. Primers are short RNA molecules that act as templates for the starting point of DNA replication.
- **DNA polymerases** - synthesize new DNA molecules by adding nucleotides to leading and lagging DNA strands.
- **Topoisomerase or DNA Gyrase** - unwinds and rewinds DNA strands to prevent the DNA from becoming tangled or supercoiled.
- **Exonucleases** - group of enzymes that remove nucleotide bases from the end of a DNA chain.

- **DNA ligase** - joins DNA fragments together by forming phosphodiester bonds between nucleotides.

Prokaryotic and eukaryotic DNA replication

DNA Replication in Prokaryotes

Centring on the general principle of DNA replication, the prokaryotic DNA replication in prokaryotic cells takes place just before a cell divides in an organism and ensures both daughter cells receive an exact copy of the parent's genetic material. The process uses the semiconservative model of replication which results in a double-stranded DNA with one parental and one daughter strand.

The **Steps of Prokaryotic DNA Replication** are as follows:

- The DNA replication process is bi-directional begins at a spot on the DNA molecule called the origin of replication.
- At this spot, enzymes unwind the double helix structure of the DNA which makes its components accessible for replication.
- The helix is unwound by the helicase enzyme to form a pair of replication forks, and the unwound helix is stabilised by SSB proteins and DNA isomerases.
- Primase forms 10 base RNA primers which initiate the synthesis of the leading and the lagging strand.
- The leading continues to synthesise in the 5' to 3' direction by DNAP III (DNA Polymerase III)
- The lagging strand is also synthesised in the 5' to 3' direction but it is discontinued through the formation of Okazaki fragments.
- DNA polymerase I removes the 10 base RNA primers and replaces the gap with deoxynucleotides.

- Then DNA ligase seals the breaks between Okazaki fragments as well as around the primers to form continuous strands.
- The entire process of replication takes place in the cell cytoplasm.

DNA Replication in Eukaryotes

The eukaryotic DNA replication takes place in the cell nucleus and only occurs in the S phase at many chromosomal origins. Similar to prokaryotic DNA replication, eukaryotic cells also use the semi-conservative process of replication but there are multiple origins of replication.

The **Steps of the Eukaryotic DNA Replication** are as follow:

- The replication process starts in a chromosome at multiple origins, with one origin being at 30-300 kb of DNA depending on the tissue and species.
- A replication bubble of two forks forms at each origin. The DNA replicated under the control of a single origin is called a replicon. The synthesis proceeds until all bubbles merge together.
- The process starts with the unwinding of DNA with the help of enzymes, which makes its components accessible for replication.
- The unwound helix forms a pair of replication forks and is stabilised by DNA topoisomerases and SSB proteins.
- The RNA primers required for the process are made by DNA polymerases α which initiates the synthesis of the lagging strand and makes the first primer. It then extends it with a short region of DNA.
- The Okazaki fragments and the leading strand are synthesised by DNA polymerase δ.
- The leading strand is synthesised continuously whilst the lagging strand is synthesised discontinuously. Both strands are synthesised in the 5′to 3′ direction.
- At completion, DNA ligase seals the breaks around the primers and between the Okazaki fragments.

Although there are some similarities between DNA replication in prokaryotes and eukaryotes, the differences are many. Here we will discuss the differences between prokaryotes' and eukaryotes' DNA replication process.

DNA damage and repair

Damage to cellular DNA is involved in mutagenesis and the development of cancer. The DNA in a human cell undergoes several thousand to a million damaging events per day, generated by both external (exogenous) and internal metabolic (endogenous) processes. Changes to the cellular genome can generate errors in the transcription of DNA and ensuing translation into proteins necessary for signaling and cellular function. Genomic mutations can also be carried over into daughter generations of cells if the mutation is not repaired prior to mitosis. Once cells lose their ability to effectively repair damaged DNA, there are three possible responses.

- The cell may become senescent, i.e., irreversibly dormant. In 2005, multiple laboratories reported that senescence could occur in cancer cells in vivo as well as in vitro, stopping mitosis and preventing the cell from evolving further.
- The cell may become apoptotic. Sufficient DNA damage may trigger an apoptotic signaling cascade, forcing the cell into programmed cell death.
- The cell may become malignant, i.e., develop immortal characteristics and begin uncontrolled division.

To compensate for the degree and types of DNA damage that occur, cells have developed multiple repair processes including mismatch, base excision, and nucleotide excision repair mechanisms, with little process redundancy. Cells may have evolved to proceed into apoptosis or senescence if overwhelming damage occurs rather than expend energy to effectively

repair the damage. The rate at which a cell is able to make repairs is contingent on factors including cell type and cell age.

Source of DNA Damage

The four types of factors that cause DNA damage are:

1. **Hydrolysis**
2. **Deamination**
3. **Alkylation and**
4. **Oxidation**

1. Hydrolysis:

- DNA consists of long strands of sugar molecules called deoxyribose that are linked together by phosphate groups.
- Each sugar molecule carries one of the four natural DNA bases: adenine, guanine, cytosine, or thymine (A, G, C, or T).
- The chemical bond between a DNA base and its respective deoxyribose, although relatively stable, is nonetheless subject to chance cleavage by a water molecule in a process known as spontaneous hydrolysis.
- Loss of the "purine" bases (guanine and adenine) is referred to as depurination, whereas loss of the "pyrimidine" bases (cytosine and thymine) is called depyrimidination. In mammalian cells, it is estimated that depurination occurs at the rate of about 10,000 purine bases lost per cell generation.
- The rate of depyrimidination is considerably slower, resulting in the loss of about 500 pyrimidine bases per cell generation.
- The baseless sugars that result from these processes are commonly referred to as AP-sites (apurinic/apyrimidinic).
- They are potentially lethal to the cell, as they act to block the progress of DNA replication, but are efficiently repaired in a series of enzyme-catalyzed reactions collectively referred to as the base excision repair (BER) pathway.

- In fact, AP-sites are intentionally created during the course of BER.

2. Deamination:

- The bases that make up DNA are also vulnerable to modification of their chemical structure.
- One form of modification, called spontaneous deamination, is the loss of an amino group (-NH2). For example, cytosine (C), which is paired with guanine (G) in normal, double-stranded DNA, has an amino group attached to the fourth carbon (C4) of the base.
- When that amino group is lost, either through spontaneous, chemical, or enzymatic hydrolysis, a uracil (U) base is formed, and a normal C-G DNA base pair is changed to a pre-mutagenic U-G base pair (uracil is not a normal part of DNA).
- The U-G base pair is called pre-mutagenic because if it is not repaired before DNA replication, a mutation will result.
- During DNA replication, the DNA strands separate, and each strand is copied by a DNA polymerase protein complex.
- On one strand, the uracil (U) will pair with a new adenine (A), while on the other strand the guanine (G) will pair with a new cytosine(C).
- Thus, one DNA double-strand contains a normal C-G base pair, but the other double-strand has a mutant U-A base pair.
- This process is called mutation fixation, and the mutation of the G to an A is said to be fixed (meaning “fixed in place,” not “repaired”).
- In other words, the cell now accepts the new mutant base pair as normal.
- It is estimated that approximately 400 cytosine deamination events per genome occur every day.
- Clearly, it is very important for the cell to repair DNA damage before DNA replication commences, in order to avoid mutation fixation.
- One cause of normal human aging is the gradual accumulation over time of mutations in our cellular DNA.

3. Alkylation:

- Another type of base modification is alkylation.
- Alkylation occurs when a reactive mutagen transfers an alkyl group (typically a small hydrocarbon side chain such as a methyl or ethyl group, denoted as -CH3 and –C2H5, respectively) to a DNA base.
- The nitrogen atoms of the purine bases (N3 of adenine and N7 of guanine) and the oxygen atom of guanine (O6) are particularly susceptible to alkylation in the form of methylation.
- Methylation of DNA bases can occur through the action of exogenous (environmental) and endogenous (intracellular) agents. For example, exogenous chemicals such as dimethyl sulfate, used in many industrial processes and formed during the combustion of sulfur-containing fossil and N-methyl-N-nitrosoamine, a component of tobacco smoke, are powerful alkylating agents.
- These chemicals are known to greatly elevate mutation rates in cultured cells and cause cancer in rodents.
- Inside every cell is a small molecule known as S-adenosylmethionine or “SAM” SAM, which is required for normal cellular metabolism, is an endogenous methyl donor.
- The function of SAM is to provide an activated methyl group for virtually every normal biological methylation reaction.
- SAM helps to make important molecules such as adrenaline, a hormone secreted in times of stress; creatine, which provides energy for muscle contraction; and phosphatidylcholine, an important component of cell membranes.
- However, SAM can also methylate inappropriate targets, such as adenine and guanine. Such endogenous DNA-alkylation damage must be continually repaired; otherwise, mutation fixation can occur.

4. Oxidation:

- Oxidative damage to DNA bases occurs when an oxygen atom binds to a carbon atom in the DNA base.
- High-energy radiation, like X-rays and gamma radiation, causes exogenous oxidative DNA base damage by interacting with water molecules to create highly reactive oxygen species, which then attack DNA bases at susceptible carbon atoms.
- Oxidative base damage is also endogenously produced by reactive oxygen species released during normal respiration in mitochondria, the cell's "energy factories."
- Humans enjoy a long life span; thus, it would seem that healthy, DNA repair-proficient cells could correct most of the naturally occurring endogenous DNA damage.
- Unfortunately, when levels of endogenous DNA damage are high, which might occur as the result of an inactivating mutation in a DNA repair gene, or when we are exposed to harmful exogenous agents like radiation or dangerous chemicals, the cell's DNA repair systems become overwhelmed.
- Lack of DNA repair results in a high mutation rate, which in turn may lead to cell death, cancer, and other diseases.
- Also, if the level of DNA repair activity declines with age, then the mutational burden of the cell will increase as we grow older.

Methods for Repairing DNA Damages

(a) Direct Reversal of Base Damage:

- Spontaneous addition of a methyl group (CH3-) (an example of alkylation) to Cs followed by domination to a T is the most frequent cause of point mutations in humans.
- Fortunately, most of these changes are repaired by enzymes which are known as glycosylases that remove the mismatched T

restoring the correct C. The DNA backbone need not be broken for this.

- DNA used to get damaged by alkylation in cancer chemotherapy ("chemo") due to some of the drugs used also damage DNA by alkylation. Some of the methyl groups can be removed by a protein encoded by our MGMT gene. The removal of each methyl group requires another molecule of protein as the protein can only do it once.
- Each of the myriad types of chemical alterations to bases requires its own mechanism to correct. The cell needs are more general mechanisms capable of correcting all sorts of chemical damage with a limited toolbox. The mechanisms of excision repair this requirement.

b) Excision Repair:

- In this process the damaged base or bases are removed and then replaced with the correct ones in a localized burst of DNA synthesis.
- There are **three modes of excision repair**, each of which employs specialized sets of enzymes namely.
    - **Base Excision Repair (BER)**
    - **Nucleotide Excision Repair (NER)**
    - **Mismatch Repair (MMR)**

Base Excision Repair (BER)

The **steps** and by players of BER are:

(i) Removal of the damaged base by a DNA glycosylase.
(ii) Removal of its deoxyribose phosphate in the backbone which produces a gap.

(iii) Replacement with the correct nucleotide. This relies on DNA polymerase beta, one of at least 11 DNA polymerases encoded by our genes.

(iv) Ligation of the break in the strand. Two enzymes are known that can do this; both require ATP to provide the needed energy.

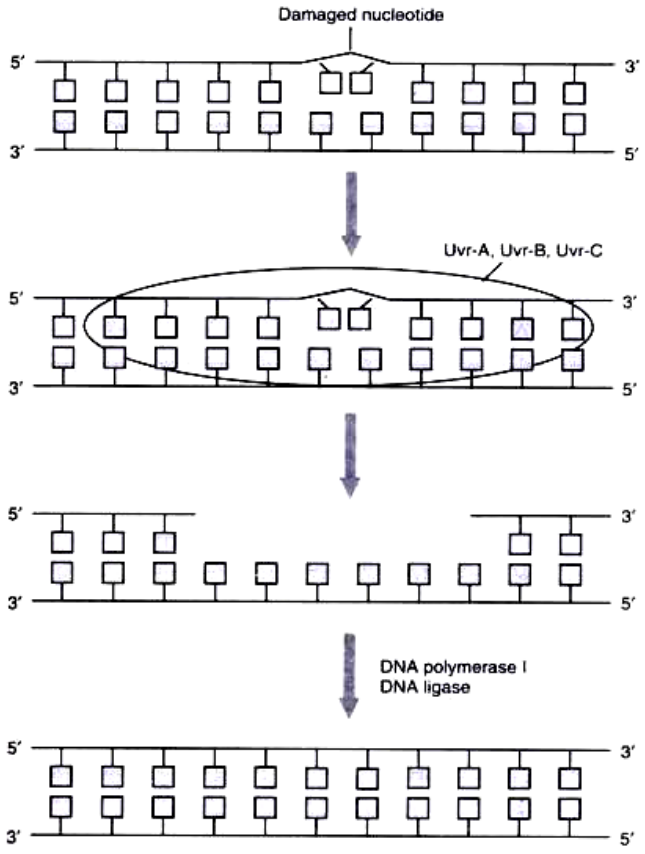


Fig. 4.5. Nucleotide Excision Repair (NER) steps

Nucleotide Excision Repair (NER)

Nucleotide Excision Repair (NER) differs from BER in several ways. It uses different enzymes. NER removes a large "patch" around the damage.

The **steps** and key players of NER are:

(i) The damage one or more protein factors recognize. These assemble at the location.
(ii) The DNA is unwound which produces a "bubble". The enzyme system that does this is Transcription Factor IIH, TFIIH, (which also functions in normal transcription).
(iii) Cuts are made on both the 3′ side and the 5′ side of the damaged area so the tract containing the damage can be removed.
(iv) A fresh burst of DNA synthesis – using the intact (opposite) strand as a template – fills in the correct nucleotides. The DNA polymerases responsible are designated polymerase delta and epsilon.
(v) A DNA ligase covalently inserts the fresh piece into the backbone.

Mismatch Repair (MMR)

Mismatch Repair (MMR) deals with correcting mismatches of the normal bases.

1. It can enlist the aid of enzymes involved in both base-excision repair (BER) and nucleotide-excision repair (NER) as well as using enzymes specialized for this function.
2. Recognition of a mismatch requires several different proteins including one encoded by MSH2.
3. Cutting the mismatch out also requires several proteins, including one encoded by MLH1.
4. The process of repairing starts with the protein MutS which binds to mismatched base pairs.
5. MutL is recruited to the complex and activates MutH which binds to GATC sequences.

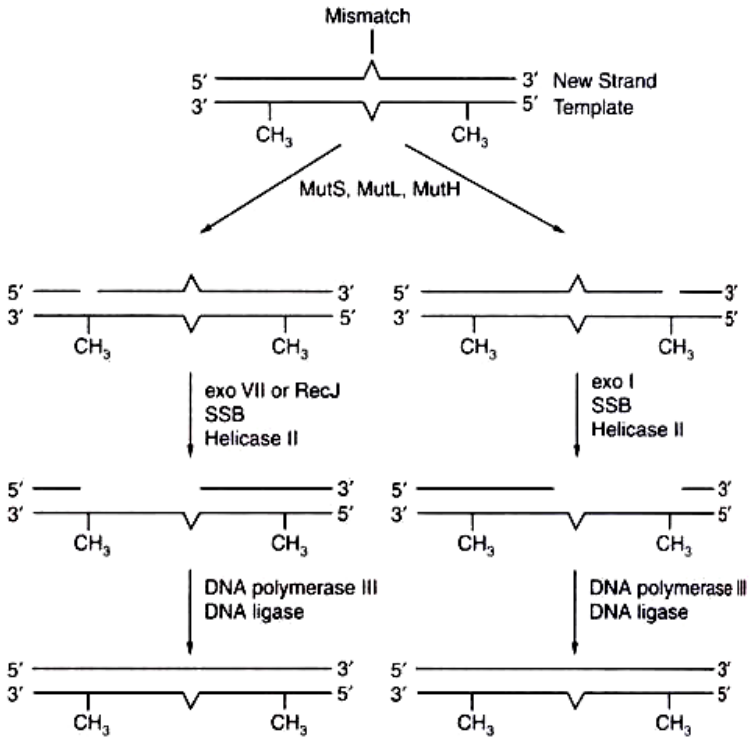


Fig. 4.6. Mismatch Repair (MMR) steps

6. Activation of MutH cleaves the unmethylated strand at the GATC site.
7. Then, the segment from the cleavage site to the mismatch is removed by exonuclease (with assistance from helices II and SSB proteins).
8. If the cleavage occurs on the 3′ side of the mismatch, this step is carried out by exonuclease I.
9. It degrades a single strand only in the 3′ to 5′ direction.
10. If the cleavage occurs on the 5′ side of the mismatch, exonuclease VII or RecJ is used to degrade the single stranded DNA.
11. Mismatch repair is very expensive and inefficient as the distance between the GATC site and the mismatch could be as long as 1,000 base pairs.
12. Homologs of MutS and MutL have been found in yeast, mammals, and other eukaryotes. MSH1 to MSH5 are

homologous to MutS; MLH1, PMS1 and PMS2 are homologous to MutL. Colon cancer relate to mutations of MSH2, PMS1 and PMS2.

DNA Recombination

At least four types of naturally occurring recombination have been identified in living organisms.

- **General or homologous recombination** occurs between DNA molecules of very similar sequence, such as homologous chromosomes in diploid organisms. General recombination can occur throughout the genome of diploid organisms, using one or a small number of common enzymatic pathways. This chapter will be concerned almost entirely with general recombination.
- **Illegitimate or nonhomologous recombination** occurs in regions where no large-scale sequence similarity is apparent, e.g. translocations between different chromosomes or deletions that remove several genes along a chromosome. However, when the DNA sequence at the breakpoints for these events is analyzed, short regions of sequence similarity are found in some cases. For instance, recombination between two similar genes that are several million bp apart can lead to deletion of the intervening genes in somatic cells.
- **Site-specific recombination** occurs between particular short sequences (about 12 to 24 bp) present on otherwise dissimilar parental molecules. Site-specific recombination requires a special enzymatic machinery, basically one enzyme or enzyme system for each particular site. Good examples are the systems for integration of some bacteriophage, such as l, into a bacterial chromosome and the rearrangement of immunoglobulin genes in vertebrate animals.
- The third type is **replicative recombination**, which generates a new copy of a segment of DNA. Many transposable elements use a process of

replicative recombination to generate a new copy of the transposable element at a new location.

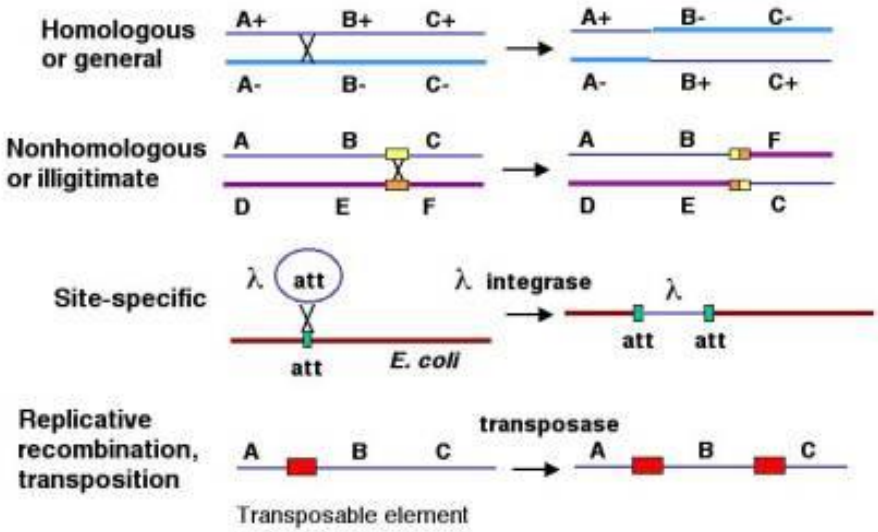
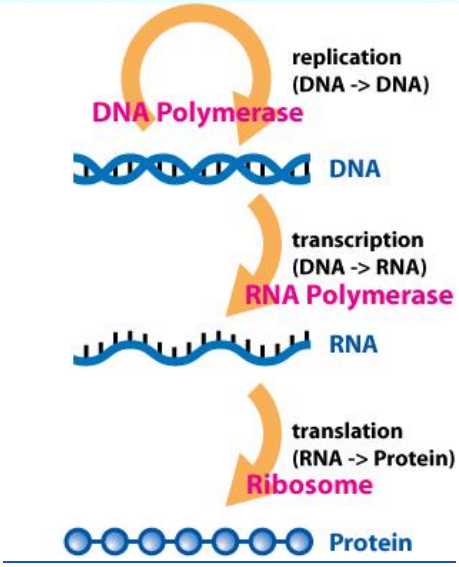


Fig: Types of natural recombination

Transcription and RNA processing

Transcription

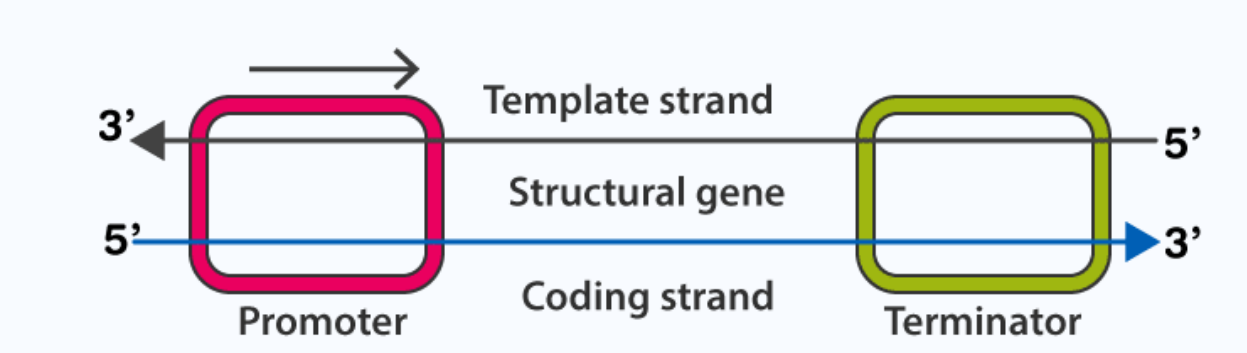


- It is one of the first processes in gene expression.

- The genetic information flows from DNA to protein and this flow of information takes place in a sequential process of transcription and translation.
- Only one strand of DNA is copied during the process of transcription known as the template strand and the RNA formed is called the mRNA.
- The main motive of transcription is to make a copy of RNA from the DNA sequence.
- The RNA transcript carries the information used to encode a protein.

RNA Polymerase

- The RNA polymerase is the main enzyme involved in transcription.
- It uses single-strand DNA to synthesize a complementary RNA strand.
- The DNA-dependent RNA polymerase binds to the promoter and catalyses the polymerization in the 5′ to 3′ direction on the template strand.
- Once it reaches the terminator sequence, the process terminates and the newly synthesised RNA strand is released.
- Transcription Unit is a stretch of a DNA transcribed into an RNA molecule.
- Its function is to encode at least one gene.
- Suppose if gene encodes protein than mRNA is produced by transcription.
- A protein encoded by the DNA transcription unit may comprise a coding sequence.
- Compared to DNA replication, transcription has a lower copying fidelity.



Stages of Transcription

Transcription proceeds in enzymatically catalysed steps i.e.

1. Initiation
2. Elongation
3. Termination

Initiation

- RNA polymerase attaches to the DNA molecule and moves along the DNA strand until it recognises a promoter sequence.
- These are known as the transcription start sites.
- The DNA double helix then unwinds and all the bases on each of the DNA strands are exposed.
- This acts as a template for a new mRNA strand.

Elongation

- Ribonucleotides are added to the template strand that enables the growth of mRNA growth.

Termination

- RNA polymerase encounters a terminator sequence and the transcription stops.
- RNA polymerase then releases the DNA template.

RNA Processing

The transcribed RNA is known as the pre-mRNA. It is processed further to convert it into mature RNA. RNA processing include:

1. Capping
2. Polyadenylation
3. Splicing

Capping

A methylated guanine cap is added to protect the mRNA. It involves:

- Addition of methylated guanine
- It occurs at 5′ end of mRNA transcript
- It protects the mRNA from degradation

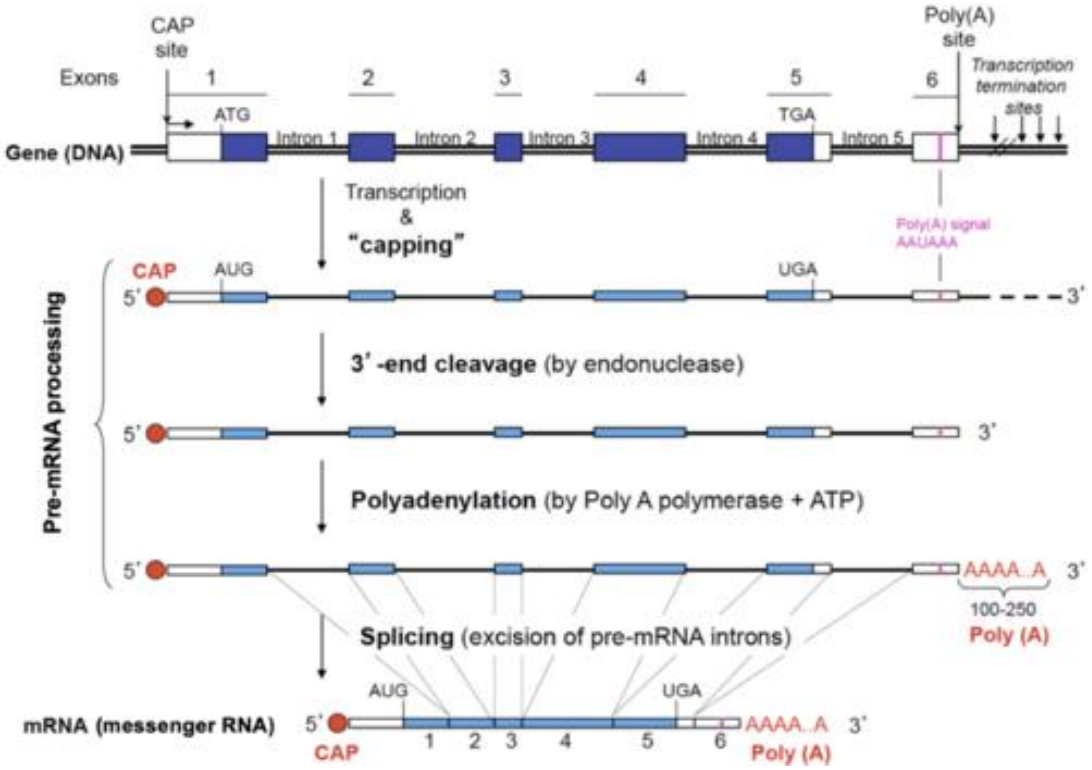
Polyadenylation

The poly-A tail also protects the mRNA from degradation. It involves:

- The endonucleases cleave the mRNA at a specific sequence.
- The enzyme polyA polymerase facilitates the addition of several adenine nucleotides.

Splicing

- The non-coding sequences, i.e., the introns are removed by spliceosome excision.
- The coding sequences or the exons join together by ligation.



Thus several proteins can be made from a single pre-mRNA. A mature mRNA is obtained at the end of transcription.

Prokaryotic and Eukaryotic transcription

Similarities Between Prokaryotic and Eukaryotic Transcription

- In both kinds of transcriptions, the RNA provides the template for the synthesis.
- One strand of DNA duplex acts as the template in both transcriptions.
- Both Prokaryotic and Eukaryotic transcriptions produce RNA molecules.
- The chemical composition of both transcriptions is similar.
- The enzyme RNA polymerase facilitates both kinds of transcriptions.

ENTRI

Differences Between Prokaryotic and Eukaryotic Transcription

Process Timing

In the case of prokaryotic transcription, both the processes of transcription and translation occur simultaneously and continuously in the cytoplasm. These processes do not occur simultaneously in eukaryotic transcription.

Process Location

The transcription and translation both occur in the cytoplasm in prokaryotic transcription. However, in eukaryotic transcription, the transcription takes place in the nucleus and the translation occurs in the cytoplasm.

Genetic Association

The prokaryotic transcription initiation is simple as the DNA is not associated with the histone protein. In eukaryotic transcription, with the DNA being associated with the protein, the process becomes complex.

RNA Processing

The RNA processing takes place in the cytoplasm for prokaryotic transcription and in the nucleus for eukaryotic transcription.

Types of RNA

There is only one type of RNA polymerase enzyme in prokaryotic transcription and it helps to synthesise all the other types of RNA in the cells (mRNA, tRNA, and rRNA). Eukaryotic transcription involves three types of RNA. There is RNA Polymerase I that helps in the rRNA synthesis, RNA Polymerase II for mRNA, and RNA Polymerase III that aids in the synthesis of tRNA and 5S rRNA.

RNA Polymerase Composition

RNA polymerase in prokaryotic transcription has 5 polypeptides. In eukaryotic transcription, RNA polymerase I have 14 subunits, and RNA polymerase II has 10-12 subunits.

Location of the Promoter Region

The promoter region is located upstream to the start site in both kinds of transcriptions but in eukaryotic transcription, sometimes, the promoter region is located downstream to the start site in RNA Polymerase III (present only in eukaryotic transcription).

Presence of σ Factor

One of the critical prokaryotic and eukaryotic transcription differences lies in the presence of the σ factor. Prokaryotic transcription initiation requires the presence of σ factor which is not present in eukaryotic transcription which requires initiation factors.

Binding of the RNA with the Promoter Region

In prokaryotic transcription, the RNA polymerase recognizes and binds with the promoter region with the help of the σ factor. This is possible in eukaryotic transcription only when the initiation factors are present in the promoter region.

Presence of TATA Box, CAT Box, and Pribnow Box

While the Pribnow boxes are present at 10 locations in the case of prokaryotic transcription, they are absent in eukaryotic transcription. TATA boxes and CAT boxes are not present in the promoter region in case of prokaryotic transcription and the Pribnow box is the sequence that is considered functionally equivalent to the TATA box. In eukaryotic transcription, TATA boxes are present 25-35 base pairs before the start of the transcription initiation site of a gene.

Presence of Introns

Introns are absent in prokaryotic transcription and thus there is no splicing of mRNA. As they are present in eukaryotic transcription, splicing is also present.

Modification of the Primary Transcript

An essential difference between prokaryotic and eukaryotic transcription is that the primary transcript does not undergo any post-transcriptional modification in prokaryotic transcription but it happens in the case of eukaryotic transcription.

RNA Capping

When we differentiate between prokaryotic and eukaryotic transcription, one of the essential points to consider is the RNA capping. It is absent in prokaryotic transcription and the mRNA does not have a 5' guanosine cap. On the other hand, eukaryotic transcription includes RNA capping that takes place at the 5' position mRNA.

Prokaryotic Transcription	Eukaryotic Transcription
Transcription and translation occur simultaneously.	Transcription and translation don't occur simultaneously.
Prokaryotic transcription occurs in the cytoplasm	Eukaryotic transcription occurs in the nucleus and translation occurs in the cytoplasm.
RNAs are released and processed in the cytoplasm	RNAs are released and processed in the nucleus
RNA polymerases are a complex of five polypeptides.	RNA polymerases are a complex of 10 -15 polypeptides.

ENTRI

Binding of transcription complexes

RNA polymerase II (the polymerase that transcribes eukaryotic mRNAs) cannot bind to a promoter by itself. Rather, it requires other protein factors that are added in a specific order to the promoter and interact with RNA polymerase II to efficiently transcribe the mRNA. An important question to be resolved is whether trans-acting factors must bind to the promoter before this complex is built. The following table describes features of the transcription factors (TF).

TFIID

- **Transcription factor II D (TFIID)** is one of several general transcription factors that make up the RNA polymerase II preinitiation complex.
- RNA polymerase II holoenzyme is a form of eukaryotic RNA polymerase II that is recruited to the promoters of protein-coding genes in living cells.
- It consists of RNA polymerase II, a subset of general transcription factors, and regulatory proteins known as SRB proteins.
- Before the start of transcription, the transcription Factor II D (TFIID) complex binds to the core promoter DNA of the gene through specific recognition of promoter sequence motifs, including the TATA box, Initiator, Downstream Promoter, Motif Ten, or Downstream Regulatory elements.
- Functions
    - Coordinates the activities of more than 70 polypeptides required for initiation of transcription by RNA polymerase II.
    - Binds to the core promoter to position the polymerase properly.
    - Serves as the scaffold for assembly of the remainder of the transcription complex.
    - Acts as a channel for regulatory signals.

TFIIA

- Transcription factor TFIIA is a nuclear protein involved in the RNA polymerase II-dependent transcription of DNA.
- TFIIA is one of several general (basal) transcription factors (GTFs) that are required for all transcription events that use RNA polymerase II.
- Other GTFs include TFIID, a complex composed of the TATA binding protein TBP and TBP-associated factors (TAFs), as well as the factors TFIIB, TFIIE, TFIIF, and TFIIH.
- Together, these factors are responsible for promoter recognition and the formation of a transcription preinitiation complex (PIC) capable of initiating RNA synthesis from a DNA template.
- Functions
    - TFIIA interacts with the TBP subunit of TFIID and aids in the binding of TBP to TATA-box containing promoter DNA.
    - Interaction of TFIIA with TBP facilitates formation of and stabilizes the preinitiation complex.
    - Interaction of TFIIA with TBP also results in the exclusion of negative (repressive) factors that might otherwise bind to TBP and interfere with PIC formation.
    - TFIIA also acts as a coactivator for some transcriptional activators, assisting with their ability to increase, or activate, transcription.
    - The requirement for TFIIA in vitro transcription systems has been variable, and it can be considered either as a GTF and/or a loosely associated TAF-like coactivator.
    - Genetic analysis in yeast has shown that TFIIA is essential for viability.

TFIIB

- Transcription factor II B (TFIIB) is a general transcription factor that is involved in the formation of the RNA polymerase II preinitiation complex (PIC) and aids in stimulating transcription initiation.

- TFIIB is localised to the nucleus and provides a platform for PIC formation by binding and stabilising the DNA-TBP (TATA-binding protein) complex and by recruiting RNA polymerase II and other transcription factors.
- It is encoded by the TFIIB gene, and is homologous to archaeal transcription factor B and analogous to bacterial sigma factors.

TFIIF- RNA Polymerase II Complex

- Transcription factor II F (TFIIF) is one of several general transcription factors that make up the RNA polymerase II preinitiation complex.
- TFIIF is encoded by the **GTF2F1, GTF2F2, and GTF2F2L genes.**
- TFIIF binds to RNA polymerase II when the enzyme is already unbound to any other transcription factor, thus preventing it from contacting DNA outside the promoter.
- Furthermore, TFIIF stabilizes the RNA polymerase II while it's contacting TBP and TFIIB.

Translation

- The second major step in gene expression is called **translation**.
- After the messenger RNA makes a complementary strand to a single strand of DNA in transcription, it then gets processed during RNA splicing and is then ready for translation.
- Since the process of translation occurs in the cytoplasm of the cell, it has to first move out of the nucleus through the nuclear pores and out into the cytoplasm where it will encounter the ribosomes needed for translation.
- In the process of translation, messenger RNA works together with the transfer RNA i.e. tRNA and ribosome for the synthesis of proteins.
- The whole process of transcription and translation is called **gene expression.**

- Protein synthesis can be defined as the process in which the molecules of amino acids are arranged as a single line into proteins by involving **ribosomal RNA, transfer RNA, messenger RNA,** and other enzymes.

Prokaryotic and eukaryotic gene expression

Prokaryotic gene expression

- Because prokaryotic organisms lack a cell nucleus, the processes of transcription and translation occur almost simultaneously.
- When the protein is no longer needed, transcription stops.
- As a result, the primary method to control what type and how much protein is expressed in a prokaryotic cell is through the regulation of DNA transcription into RNA.
- All the subsequent steps happen automatically.
- When more protein is required, more transcription occurs.
- Therefore, in prokaryotic cells, the control of gene expression is almost entirely at the transcriptional level.

Eukaryotic gene expression

- Eukaryotic cells, in contrast, have intracellular organelles and are much more complex.
- Recall that in eukaryotic cells, the DNA is contained inside the cell's nucleus and it is transcribed into mRNA there.
- The newly synthesized mRNA is then transported out of the nucleus into the cytoplasm, where ribosomes translate the mRNA into protein.
- The processes of transcription and translation are physically separated by the nuclear membrane; transcription occurs only within the nucleus, and translation only occurs outside the nucleus in the cytoplasm.
- The regulation of gene expression can occur at all stages of the process

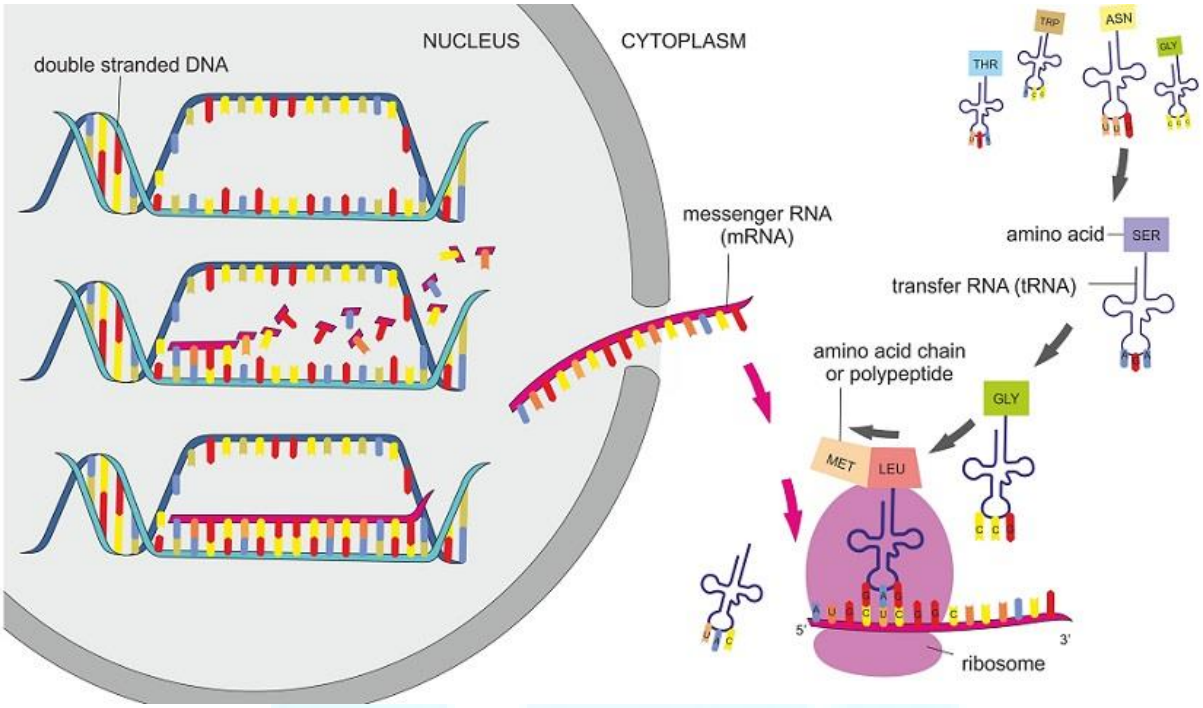
- **Epigenetic level:** regulates how tightly the DNA is wound around histone proteins to package it into chromosomes
- **Transcriptional level:** regulates how much transcription takes place
- **Post-transcriptional level:** regulates aspects of RNA processing (such as splicing) and transport out of the nucleus
- **Translational level:** regulates how much of the RNA is translated into protein
- **Post-translational level:** regulates how long the protein lasts after it has been made and whether the protein is processed into an active form.

Prokaryotic organisms	Eukaryotic organisms
• Lack nucleus • RNA transcription and protein translation occur almost simultaneously • Gene expression is regulated primarily at the transcriptional level	• Contain nucleus • RNA transcription occurs prior to protein translation, and it takes place in the nucleus. RNA translation to protein occurs in the cytoplasm. • RNA post-processing includes addition of a 5′ cap, poly-A tail, and excision of introns and splicing of exons. • Gene expression is regulated at many levels (epigenetic, transcriptional, post-transcriptional, translational, and posttranslational)

ENTRI

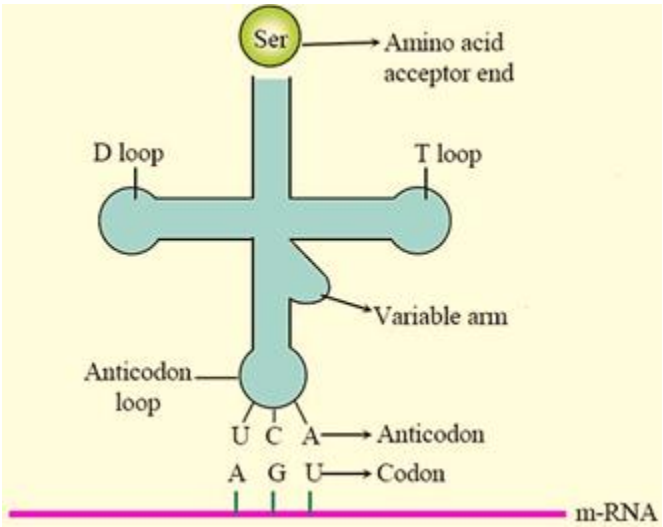
Translational machinery, mechanism of initiation, elongation and termination

Translation Process in Protein Synthesis



- The translation is a process of protein synthesis from mRNA with the help of ribosomes.
- Translational unit of mRNA from 5′ to 3` includes start codon, region coded polypeptide, a stop codon, and untranslated regions (UTRs) at 5`end & 3`end both for more efficiency of the process.
- The ribosome is the place where the whole machinery of translation is present.
- Each eukaryotic ribosome has two parts: a smaller 40S subunit and a larger 60S subunit.
- The smallest unit has a point for attachment of mRNA.
- Along with the largest subunit, it forms a P-site or peptidyl transfer (Donor site).
- There are binding sites for initiation factors, elongation factors, translocation, etc.

Structure and Role of tRNA in Protein Synthesis

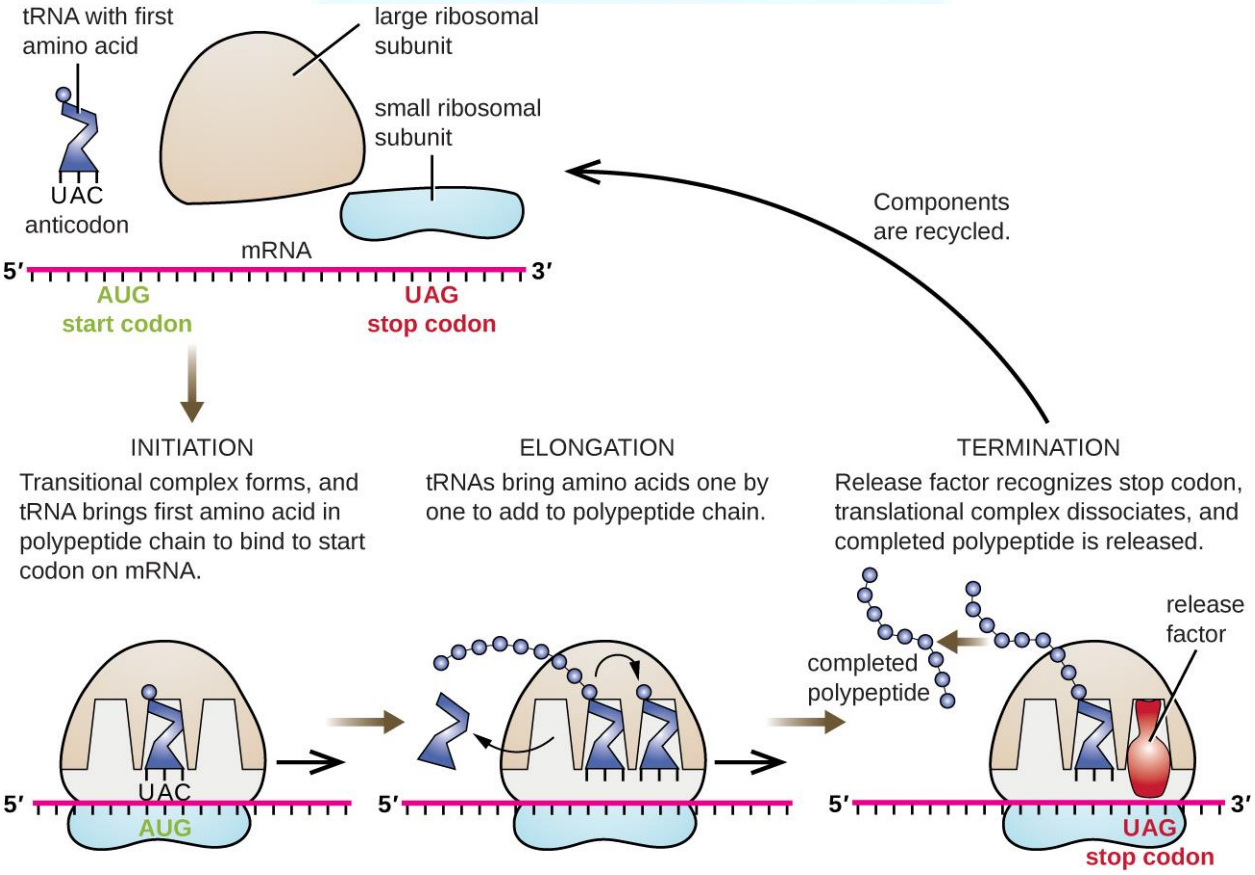


- The transfer RNA(tRNA) is a family of about 60 small sized ribonucleic acids that can recognize the codon of mRNA and exhibit a higher affinity for 21 activated amino acids which combine with them and carry them to the site of protein synthesis. tRNA molecules have been variously termed as soluble RNA or supernatant RNA or adapted RNA of the cell.
- Structurally, tRNA looks like a cloverleaf or inverted L shaped molecule which on one end has an amino acid receptor end and on the other end has an anticodon loop.
- The L shape results due to the modification in the nucleotides of tRNA such as pseudouridine, dihydrouridine(DHU), inosine and ribothymidine.
- The bent in the chain of each tRNA molecule contains a definite sequence of three nitrogenous bases that constitute the anticodon.
- It recognizes the codon on mRNA. The main constituents of tRNA are-
    - **Anticodon Loop:** It contains 7 bases out of which three bases form the anticodon loop and attaches to the codon of mRNA.
    - **DHU Loop:** This loop serves as the binding site for aminoacyl synthetase enzyme and it contains around 8-12 bases. The D arm contains the modified nucleotide called dihydrouridine.

- **T Ψ C Loop:** This loop contains two modified nucleotides- pseudouridine and ribothymidine. This loop serves as the attachment site for ribosomes.
- **AA Binding Site:** This site serves as the binding site for amino acid. It contains a CCA- OH group.
- **Variable Loop:** It is generally present between the TΨC loop and anticodon loop.

The function of tRNA is specific in protein synthesis as they pick up specific amino acids from the amino acid pool and carry over the mRNA strand.

Protein Synthesis Steps Involved



The three stages of translation are-

- Initiation involves assembling ribosomes around mRNA and activating amino acid and delivering it to the transfer RNA.
- Elongation is the process in which the RNA strand gets longer by adding amino acids.
- The termination process only involves releasing a polypeptide chain.

Explanation of Steps of Translation

1.Initiation

- Initiation in prokaryotes requires large and small ribosome subunits, the mRNA, initiating transfer RNA, and 3 initiation factors (IFs).
- Amino acids are activated by binding with the enzyme called aminoacyl tRNA synthetase in presence of ATP forming an enzyme complex and P site.

Amino acid and ATP in the presence of aminoacyl transfer RNA synthetase Pi + AA-AMP-Enzyme complex

- Transfer of amino acid to tRNA -

AA-AMP-Enzyme complex + transfer RNA Amino Acid- tRNA + AMP + Enzyme.

- Two sites at ribosome are present that are called A-site and P-site where units of ribosome bind to the cap region of messenger RNA and comparatively smaller units bind to mRNA followed by binding of them with the larger subunits. It makes AUG lie on P-site and methionyl tRNA binds to P-site.

2.Elongation of the Polypeptide Chain

- At the 2nd codon, other aminoacyl transfer RNA complexes that are charged initiate binding at A-site.

- At P-site- peptide bond between the carboxyl molecule and the amino molecule is observed whereas at A-site bond between amino molecule and amino acid is formed through the enzyme named as a peptidyl transferase.
- Sliding of ribosome over messenger RNA from one codon to its alternate codon in the direction of 5' to 3`.
- A polypeptide chain is formed by the attachment of amino acids to one alternate to another in a chain formed by the peptide bond, and the attachment is based in accordance with the sequence of codons resulting in elongation of the protein chain.

3. Termination of Polypeptide

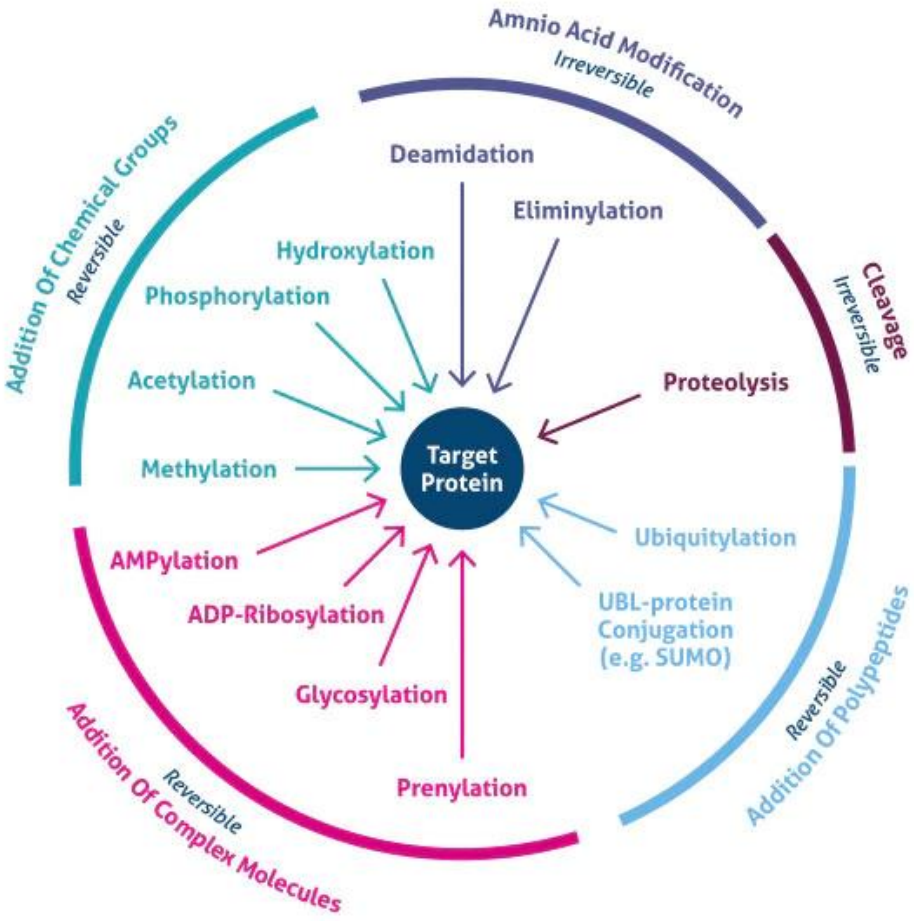
- Reaching the A-site of the ribosome at a termination codon which is present, not coding for any amino acid, no charged transfer RNA binds to the A-site of ribosome.
- A polypeptide is now not associated with the ribosome and dissociates and is catalyzed by a "release factor", a factor that releases 3 termination codons called UGA, UAG, and UAA.

Post-translational modification(PTM) of protein

- PTMs can happen at any step of the protein lifespan.
- Many proteins are modified shortly after translation is completed to mediate proper folding or to direct the nascent protein to distinct cellular locations (such as the nucleus or membrane).
- Other modifications occur after folding and localization are completed to activate or inactivate catalytic activity.
- Proteins are also covalently linked to tags that target a protein for degradation.

- They are modified through a combination of post-translational cleavage and the addition of functional groups through a step-wise mechanism of protein maturation or activation.
- PTMs occur at distinct amino acid side chains or peptide linkages and are most often mediated by enzymatic activity.
- Indeed, 5% of the proteome comprises enzymes that perform more than 200 types of PTMs (4).
- These enzymes include kinases, phosphatases, transferases, and ligases, which add or remove functional groups, proteins, lipids, or sugars to or from amino acid side chains, and proteases, which cleave peptide bonds to remove specific sequences or regulatory subunits.
- Many proteins can also modify themselves using autocatalytic domains, such as autokinase and autoprotolytic domains.
- PTMs can also be reversible based on the nature of the modification.
- As an example, phosphatases hydrolyze the phosphate group to remove it from the protein and reverse its biological activity.

Types of post-translational modifications (PTMs)



1. **Protein phosphorylation**

- Protein phosphorylation is the one of the most commonly occurring and most-studied post-translational modifications.
- It entails the phosphorylation of a specific amino acid residue through the addition of a phosphate group to a polar group R via a kinase, most commonly occurring at serine, tyrosine or threonine residues.
- The addition of the phosphate group results in a modification of the protein, whereby it transitions from being hydrophobic apolar to hydrophilic polar, enabling its interaction with other molecules – essentially "activating" it.
- A reversible post-translational modification, protein phosphorylation is important for cell regulation and the activation and deactivation of

enzymes and receptors, which can be implicated in disease processes such as cancer.

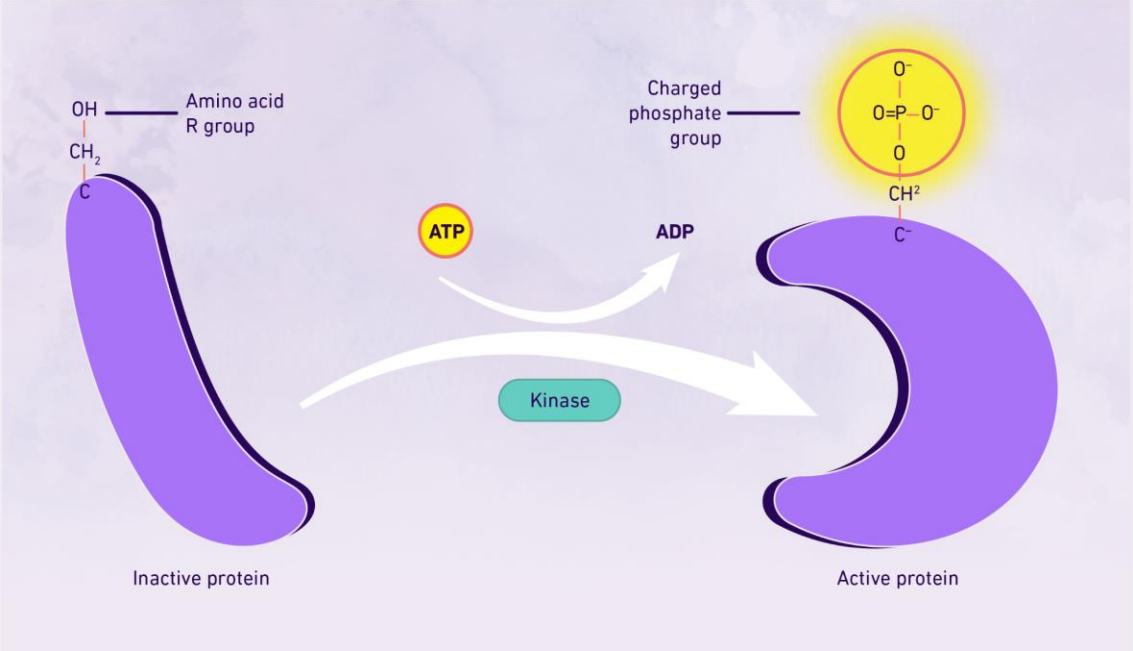


Fig: The key steps in protein phosphorylation

2. **Protein glycosylation**

- Protein glycosylation is recognized as one of the most “complicated" yet most commonly occurring post-translational modifications.
- It involves the covalent addition of a carbohydrate moiety to an amino acid, forming a glycoprotein.
- Glycosylation reactions are diverse and catalyzed by various different enzymes, which attach specific glycans to specific amino acids.
- Glycoproteins are estimated to make up ~50% of the proteome; however, the study of the glycoproteome is challenging because of the large number and diversity of glycoprotein isoforms.
- Glycosylation of eukaryotic proteins is usually categorized into two major types; N-linked, whereby a sugar molecule is attached to the amide nitrogen of asparagine, and O-linked, where a sugar molecule is attached to the oxygen atom of serine or threonine.

- There are an array of applications from glycoproteome research; many glycoproteins serve structural functions, whereas immunoglobulins are central to immunity and surface-presenting glycoproteins and glycolipids determine human blood group type.

3. Protein ubiquitination

- Ubiquitin is a small protein – approximately 8kDa in size – that can bind to a substrate protein in a process known as ubiquitination, a type of post-translation modification that serves to regulate a protein's function or mark it for degradation.

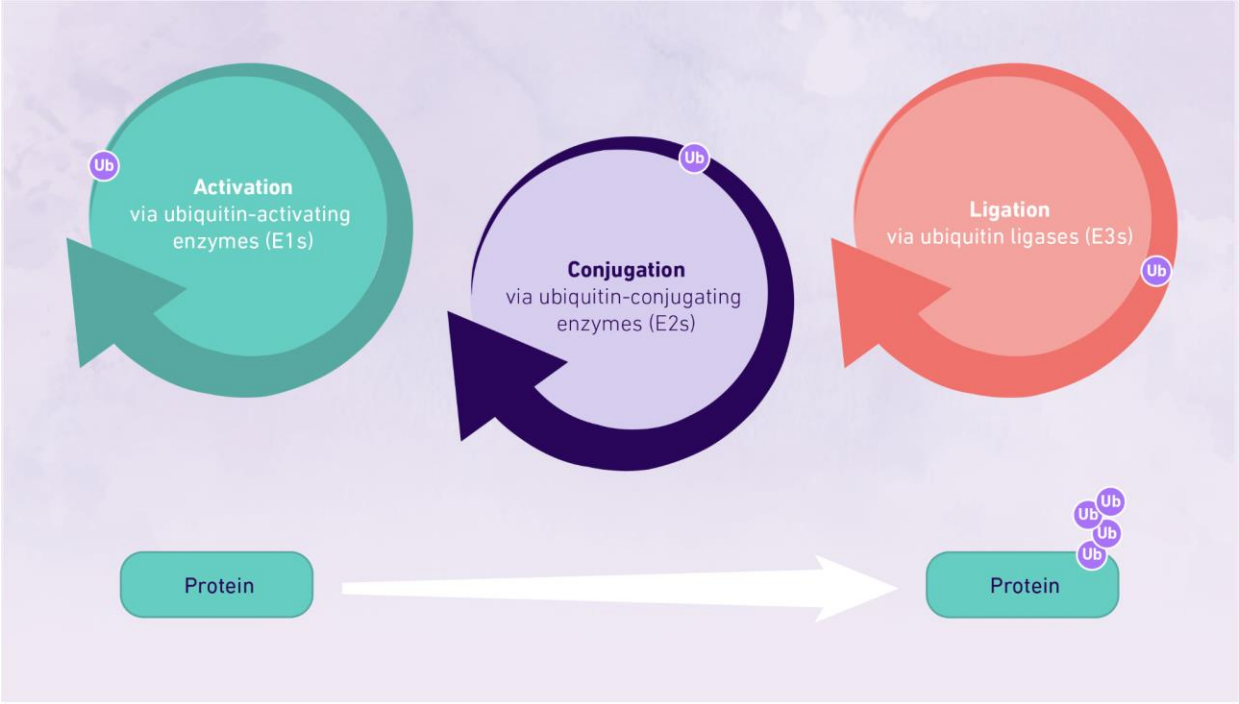


Fig: The key steps involved in protein ubiquitination

- Ubiquitination occurs in three sequential steps that are catalyzed by three groups of enzymes.
- This process generally culminates with an isopeptide bond forming between ubiquitin and the lysine residue of the protein substrate.
- Monoubiquitination refers to the addition of one ubiquitin molecule, whereas the addition of several ubiquitin proteins is known as polyubiquitination.

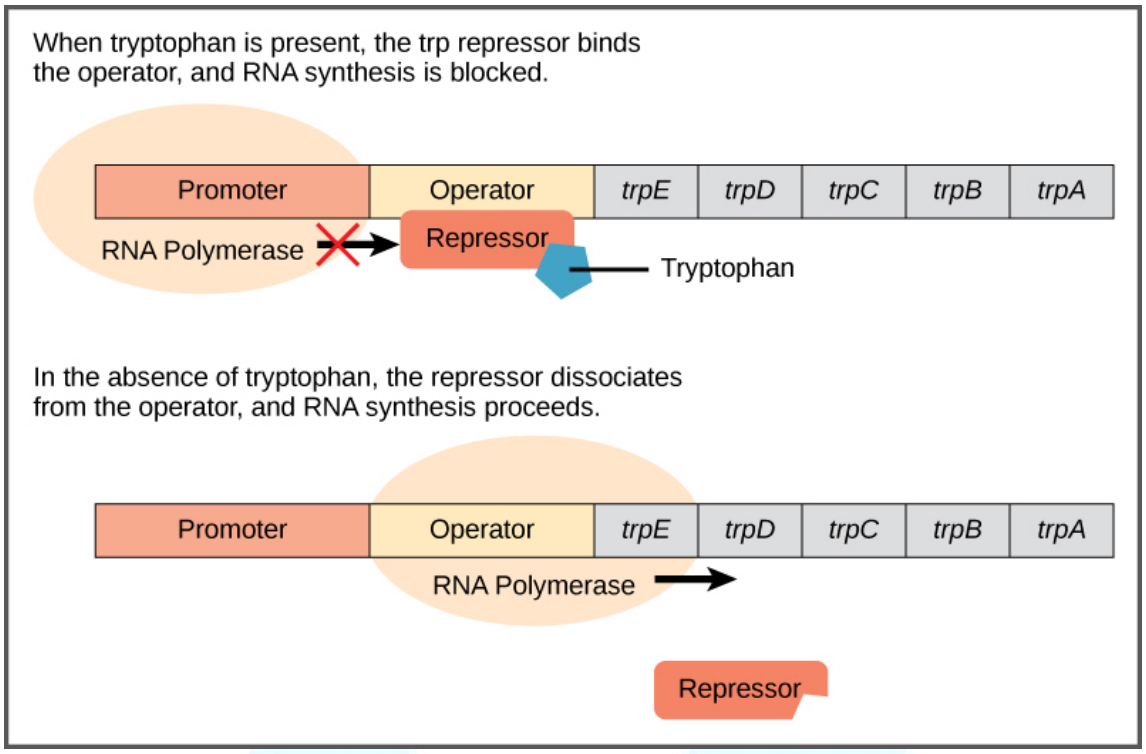
- Ubiquitination serves several functions, the most common being to flag proteins for degradation by the proteasome, but there are others including: immune and inflammatory response, organelle biogenesis and signaling roles in DNA repair.

Gene regulation mechanism in Prokaryotes

- The DNA of prokaryotes is organized into a circular chromosome supercoiled in the nucleoid region of the cell cytoplasm.
- Proteins that are needed for a specific function, or that are involved in the same biochemical pathway, are encoded together in blocks called **operons.**
    - For example, all of the genes needed to use lactose as an energy source are coded next to each other in the lactose (or lac) operon.
- In prokaryotic cells, there are three types of regulatory molecules that can affect the expression of operons: repressors, activators, and inducers.
- **Repressors** are proteins that suppress transcription of a gene in response to an external stimulus, whereas **activators** are proteins that increase the transcription of a gene in response to an external stimulus.
- Finally, inducers are small molecules that either activate or repress transcription depending on the needs of the cell and the availability of substrate.

The trp Operon: A Repressor Operon

- Bacteria such as E. coli need amino acids to survive.
- Tryptophan is one such amino acid that E. coli can ingest from the environment.

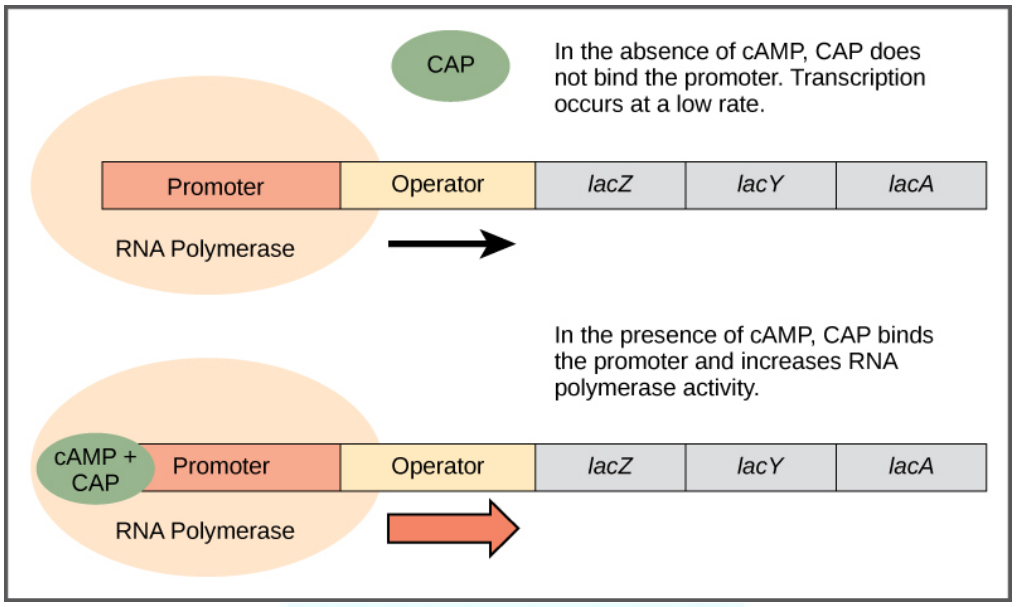


- E. coli can also synthesize tryptophan using enzymes that are encoded by five genes.
- These five genes are next to each other in what is called the **tryptophan (trp) operon.**
- If tryptophan is present in the environment, then E. coli does not need to synthesize it and the switch controlling the activation of the genes in the trp operon is switched off.
- However, when tryptophan availability is low, the switch controlling the operon is turned on, transcription is initiated, the genes are expressed, and tryptophan is synthesized.
- The five genes that are needed to synthesize tryptophan in E. coli are located next to each other in the trp operon.
- When tryptophan is plentiful, two tryptophan molecules bind the repressor protein at the operator sequence.
- This physically blocks the RNA polymerase from transcribing the tryptophan genes.

- When tryptophan is absent, the repressor protein does not bind to the operator and the genes are transcribed.
- A DNA sequence that codes for proteins is referred to as the coding region.
- The five coding regions for the tryptophan biosynthesis enzymes are arranged sequentially on the chromosome in the operon.
- Just before the coding region is the transcriptional start site.
- This is the region of DNA to which RNA polymerase binds to initiate transcription.
- The promoter sequence is upstream of the transcriptional start site; each operon has a sequence within or near the promoter to which proteins (activators or repressors) can bind and regulate transcription.
- A DNA sequence called the operator sequence is encoded between the promoter region and the first trp coding gene.
- This operator contains the DNA code to which the repressor protein can bind.
- When tryptophan is present in the cell, two tryptophan molecules bind to the trp repressor, which changes shape to bind to the trp operator.
- Binding of the tryptophan–repressor complex at the operator physically prevents the RNA polymerase from binding, and transcribing the downstream genes.
- When tryptophan is not present in the cell, the repressor by itself does not bind to the operator; therefore, the operon is active and tryptophan is synthesized.
- Because the repressor protein actively binds to the operator to keep the genes turned off, the trp operon is negatively regulated and the proteins that bind to the operator to silence trp expression are negative regulators.

Catabolite Activator Protein (CAP): An Activator Regulator

- Just as the trp operon is negatively regulated by tryptophan molecules, there are proteins that bind to the operator sequences that act as a positive regulator to turn genes on and activate them.
- For example, when glucose is scarce, E. coli bacteria can turn to other sugar sources for fuel.
- To do this, new genes to process these alternate genes must be transcribed.
- When glucose levels drop, cyclic AMP (cAMP) begins to accumulate in the cell.
- The cAMP molecule is a signaling molecule that is involved in glucose and energy metabolism in E. coli.
- When glucose levels decline in the cell, accumulating cAMP binds to the positive regulator catabolite activator protein (CAP), a protein that binds to the promoters of operons that control the processing of alternative sugars.
- When cAMP binds to CAP, the complex binds to the promoter region of the genes that are needed to use the alternate sugar sources.
- In these operons, a CAP binding site is located upstream of the RNA polymerase binding site in the promoter.
- This increases the binding ability of RNA polymerase to the promoter region and the transcription of the genes.
- When glucose levels fall, E. coli may use other sugars for fuel but must transcribe new genes to do so. As glucose supplies become limited, cAMP levels increase.
- This cAMP binds to the CAP protein, a positive regulator that binds to an operator region upstream of the genes required to use other sugar sources.



The lac Operon: An Inducer Operon

- The third type of gene regulation in prokaryotic cells occurs through inducible operons, which have proteins that bind to activate or repress transcription depending on the local environment and the needs of the cell.
- The lac operon is a typical inducible operon.
- As mentioned previously, E. coli is able to use other sugars as energy sources when glucose concentrations are low.
- To do so, the cAMP–CAP protein complex serves as a positive regulator to induce transcription.
- One such sugar source is lactose. The lac operon encodes the genes necessary to acquire and process the lactose from the local environment.
- CAP binds to the operator sequence upstream of the promoter that initiates transcription of the lac operon.
- However, for the lac operon to be activated, two conditions must be met.

- First, the level of glucose must be very low or non-existent. Second, lactose must be present.
- Only when glucose is absent and lactose is present will the lac operon be transcribed.
- This makes sense for the cell, because it would be energetically wasteful to create the proteins to process lactose if glucose was plentiful or lactose was not available.

Gene regulation mechanism in Eukaryotes

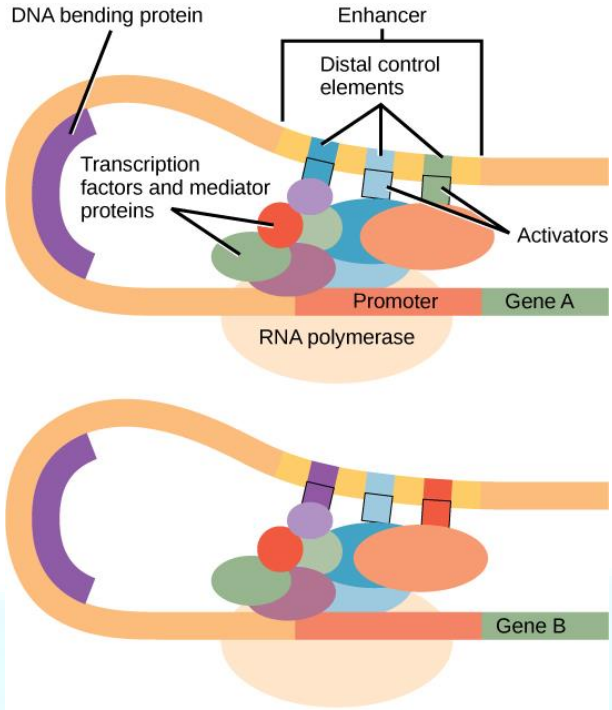
- Like prokaryotic cells, the transcription of genes in eukaryotes requires the actions of an RNA polymerase to bind to a sequence upstream of a gene to initiate transcription.
- However, unlike prokaryotic cells, the eukaryotic RNA polymerase requires other proteins, or transcription factors, to facilitate transcription initiation.
- Transcription factors are proteins that bind to the promoter sequence and other regulatory sequences to control the transcription of the target gene.
- RNA polymerase by itself cannot initiate transcription in eukaryotic cells.
- Transcription factors must bind to the promoter region first and recruit RNA polymerase to the site for transcription to be established.

The Promoter and the Transcription Machinery

- Genes are organized to make the control of gene expression easier.
- The promoter region is immediately upstream of the coding sequence.
- This region can be short (only a few nucleotides in length) or quite long (hundreds of nucleotides long).
- The longer the promoter, the more available space for proteins to bind.
- This also adds more control to the transcription process.

- The length of the promoter is gene-specific and can differ dramatically between genes.
- Consequently, the level of control of gene expression can also differ quite dramatically between genes.
- The purpose of the promoter is to bind transcription factors that control the initiation of transcription.
- Within the promoter region, just upstream of the transcriptional start site, resides the TATA box.
- This box is simply a repeat of thymine and adenine dinucleotides (literally, TATA repeats).
- RNA polymerase binds to the transcription initiation complex, allowing transcription to occur.
- To initiate transcription, a transcription factor (TFIID) is the first to bind to the TATA box.
- Binding of TFIID recruits other transcription factors, including TFIIB, TFIIE, TFIIF, and TFIIH to the TATA box.
- Once this complex is assembled, RNA polymerase can bind to its upstream sequence.
- When bound along with the transcription factors, RNA polymerase is phosphorylated.
- This releases part of the protein from the DNA to activate the transcription initiation complex and places RNA polymerase in the correct orientation to begin transcription; DNA-bending protein brings the enhancer, which can be quite a distance from the gene, in contact with transcription factors and mediator proteins.
- An enhancer is a DNA sequence that promotes transcription.
- Each enhancer is made up of short DNA sequences called distal control elements.
- Activators bound to the distal control elements interact with mediator proteins and transcription factors.

- Two different genes may have the same promoter but different distal control elements, enabling differential gene expression.



- In addition to the general transcription factors, other transcription factors can bind to the promoter to regulate gene transcription.
- These transcription factors bind to the promoters of a specific set of genes.
- They are not general transcription factors that bind to every promoter complex, but are recruited to a specific sequence on the promoter of a specific gene.
- There are hundreds of transcription factors in a cell that each bind specifically to a particular DNA sequence motif.
- When transcription factors bind to the promoter just upstream of the encoded gene, it is referred to as a **cis-acting element**, because it is on the same chromosome just next to the gene.
- The region that a particular transcription factor binds to is called the **transcription factor binding site**.

- Transcription factors respond to environmental stimuli that cause the proteins to find their binding sites and initiate transcription of the gene that is needed.

Enhancers and Transcription

- In some eukaryotic genes, there are regions that help increase or enhance transcription.
- These regions, called enhancers, are not necessarily close to the genes they enhance.
- They can be located upstream of a gene, within the coding region of the gene, downstream of a gene, or may be thousands of nucleotides away.
- Enhancer regions are binding sequences, or sites, for transcription factors.
- When a DNA-bending protein binds, the shape of the DNA changes.
- This shape change allows for the interaction of the activators bound to the enhancers with the transcription factors bound to the promoter region and the RNA polymerase.
- Whereas DNA is generally depicted as a straight line in two dimensions, it is actually a three-dimensional object.
- Therefore, a nucleotide sequence thousands of nucleotides away can fold over and interact with a specific promoter.

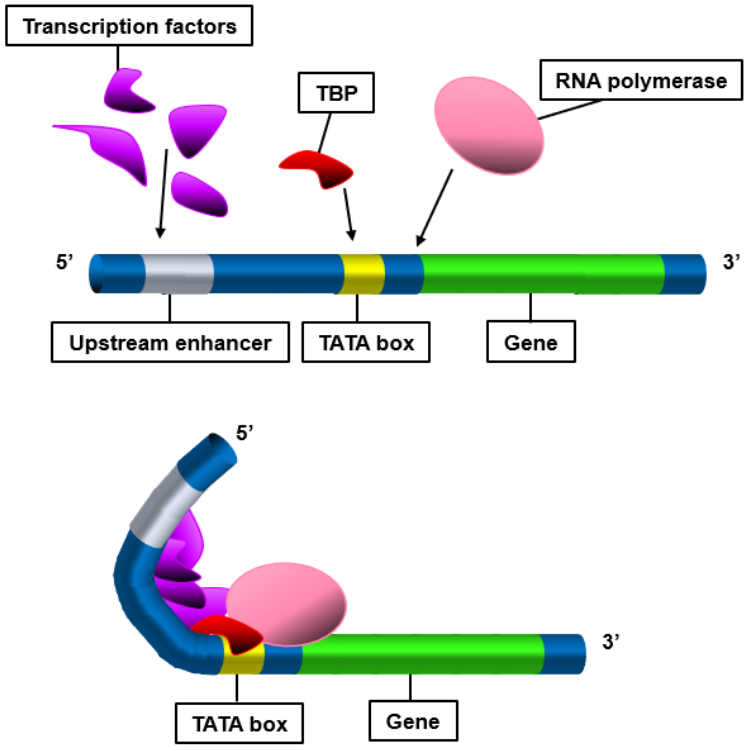
Turning Genes Off: Transcriptional Repressors

- Like prokaryotic cells, eukaryotic cells also have mechanisms to prevent transcription.
- Transcriptional repressors can bind to promoter or enhancer regions and block transcription.
- Like the transcriptional activators, repressors respond to external stimuli to prevent the binding of activating transcription factors.

Transcriptional signals-TATA, CAAT box, Enhancers

TATA

- TATA box or Goldberg-Hogness box is a consensus sequence found in the promoter region of eukaryotic genes.
- TATA box is a core element of the promoter.
- It is located 25 base pairs upstream of the transcription initiation site.
- It is conserved throughout the evolutionary process.
- The consensus sequence of TATA box is TATAWAW, where W is either A or T.
- It is a non-coding DNA sequence.
- However, TATA box is important for the transcription of the gene.
- It is the binding site of TATA-binding protein (TBP) and other transcription factors. Once they bind into the TATA box, the recruitment of RNA polymerase II and proper regulation of transcription takes place in eukaryotic genes.



- Since TATA box is involved in the regulation of transcription, mutations of TATA box can result in phenotypic changes leading to diseases such as gastric cancer, spinocerebellar ataxia, Huntington's disease, blindness, β-thalassemia, immunosuppression, Gilbert's syndrome, and HIV-1, etc.

CAAT box

- CAAT box is a region of nucleotides with the consensus sequence of GGCCAATCT.
- Similar to TATA box, CAAT box is also located in the promoter region of the gene.
- Therefore, it is located about 75-80 base pairs upstream to the transcription site.

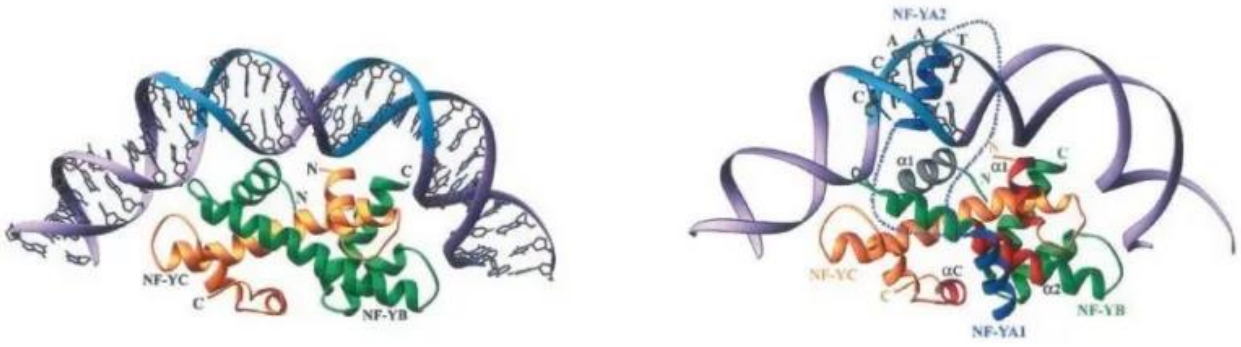


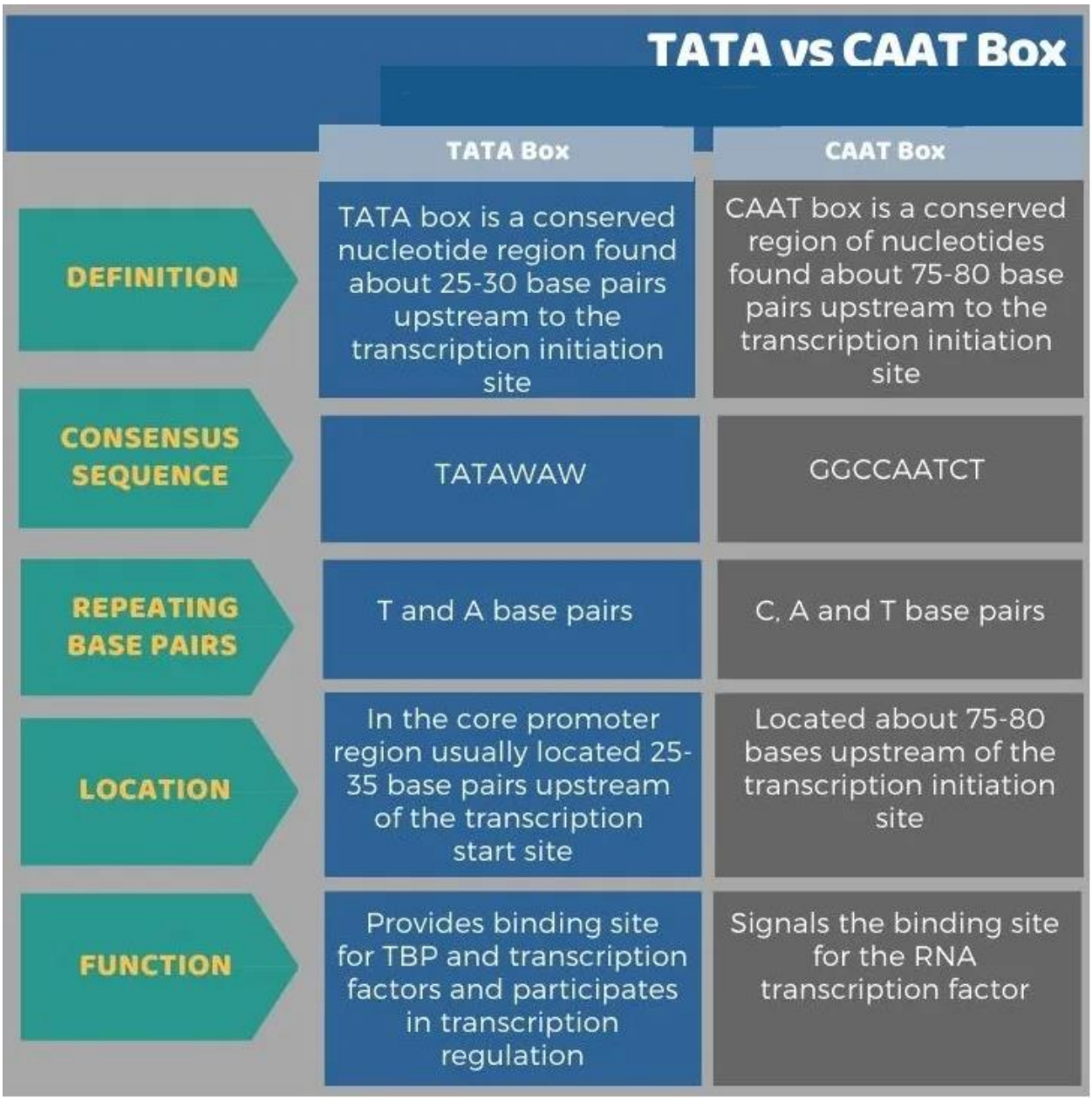
Fig: CAAT Box

- Specific transcription factors bind to the CAAT box.
- This binding stabilizes the preinitiation complex for easier binding of the enzyme RNA polymerase.
- Likewise, the CAAT box works as a regulatory sequence.
- Mutations in the CAAT box region highly affect the promoter response and regulation of transcription.

What are the Similarities Between TATA and CAAT Box?

- TATA and CAAT are two regions of nucleotides found in the regulatory sequence of eukaryotic genes.
- They are found upstream to the transcription initiation site.
- Both are consensus DNA sequences.
- They are essential for the regulation of transcription, so they are regulatory sequences.

What is the Difference Between TATA and CAAT Box?



Enhancers

- An enhancer is a DNA sequence that promotes transcription. Each enhancer is made up of short DNA sequences called distal control elements. Activators bound to the distal control elements interact with mediator proteins and transcription factors.
- In some eukaryotic genes, there are regions that help increase or enhance transcription.

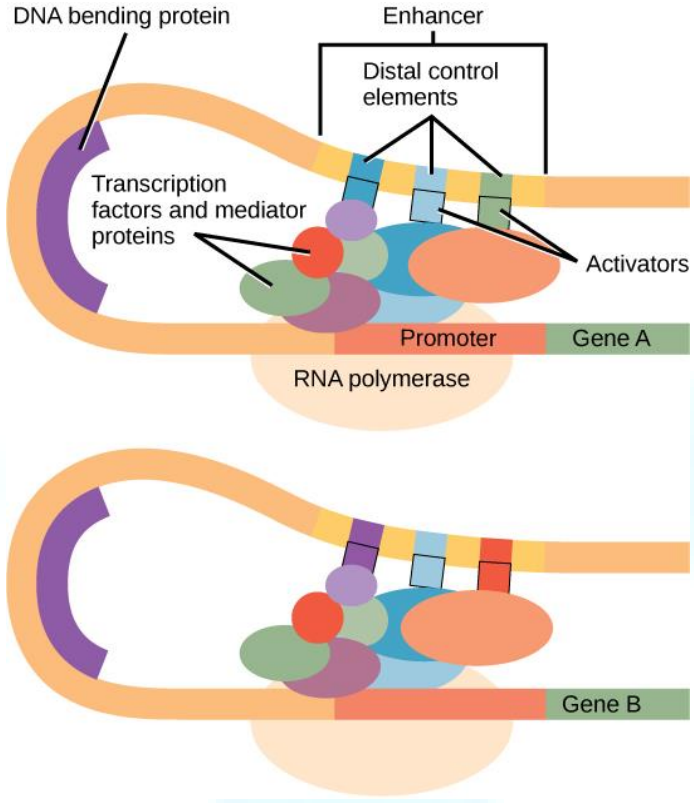


Fig: Enhancers

- These regions, called **enhancers**, are not necessarily close to the genes they enhance.
- They can be located upstream of a gene, within the coding region of the gene, downstream of a gene, or may be thousands of nucleotides away.
- Enhancer regions are binding sequences, or sites, for transcription factors.

- When a DNA-bending protein binds to an enhancer, the shape of the DNA changes.
- This shape change allows the interaction between the activators bound to the enhancers and the transcription factors bound to the promoter region and the RNA polymerase to occur.
- Whereas DNA is generally depicted as a straight line in two dimensions, it is actually a three-dimensional object.
- Therefore, a nucleotide sequence thousands of nucleotides away can fold over and interact with a specific promoter.