CHAPTER-5

MOLECULAR BASIS OF INHERITANCE

Topic-1 Nucleic Acid – DNA and RNA

Concepts Covered • *Nucleic acids, packaging of DNA helix, nucleosome, experiments to show DNA as a genetic material, RNA, process of protein synthesis.*

Revision Notes

Genetic Material

- **Nucleic Acids**
  - DNA and RNA are the two types of nucleic acids.
  - DNA is the genetic material in all organisms except some viruses.
  - RNA is the genetic material in some viruses.
  - RNA mostly functions as messenger.
- **Structure of Polynucleotide Chain**
  - Polynucleotides are the polymers of nucleotides.
  - DNA and RNA are examples of polynucleotides.
  - **A nucleotide has 3 components :**
    1. A nitrogenous base
    2. A pentose sugar (ribose in RNA and deoxyribose in DNA)
    3. A phosphate group
  - Nitrogen bases are of 2 types :
    (a) **Purines :** It includes Adenine (A) and Guanine (G).
    (b) **Pyrimidines :** It includes Cytosine (C), Thymine (T) and Uracil (U). Thymine (5-methyl Uracil) present only in DNA and Uracil only in RNA (In place of thymine).
  - A nitrogenous base is linked to the pentose sugar through an N-glycosidic linkage to form nucleoside.

Nucleosides in RNA	Nucleosides in DNA
Adenosine	Deoxyadenosine
Guanosine	Deoxyguanosine
Cytidine	Deoxycytidine
Uridine	Deoxythymidine

  - Nitrogen base + sugar + phosphate group = Nucleotide (deoxyribonucleotide). In RNA, every nucleotide residue has an additional – OH group present at 2'-position in the ribose.
  - 2 nucleotides are linked through **3' → 5' phosphodiester** bond to form dinucleotide.
  - When series of nucleotides are linked together, it forms polynucleotide.
- **Structure of DNA**
  - **Johann Friedrich Miescher (1869) :** Identified DNA and named it as 'Nuclein'.
  - **James Watson & Francis Crick** proposed the double helix model of DNA. It was based on the X-ray diffraction data produced by **Maurice Wilkins & Rosalind Franklin**.
  - DNA is made of two polynucleotide chains coiled in a right-handed fashion. Its backbone is formed of sugar and phosphates. The bases project inside.
  - The two chains have anti-parallel polarity i.e., one chain has the polarity 5' → 3' and the other has 3' → 5'.
  - Nitrogen bases of opposite chains are held together by hydrogen bonds forming base pairs (bp).
  - There are two hydrogen bonds between A and T (A = T) and three H-bonds between C and G (C ≡ G).
  - Purine comes opposite to a pyrimidine. This generates a uniform distance between the two strands.
- **Erwin Chargaff's Rule**
  - Purines and pyrimidines are always in equal amounts i.e., A + G = T + C.
  - In DNA, the proportion of A is equal to T and the proportion of G is equal to C i.e., A = T and G = C.
  - The base ratio A + T/G + C may vary from species to species but constant for a given species.
  - Length of DNA = number of base pairs × distance between two adjacent base pairs.

- ϕ 174 (a bacteriophage) has 5386 nucleotides.
- Bacteriophage lambda has 48502 base pairs (bp).
- *E. coli* has 4.6×10^6 bp.
- Haploid content of human DNA = 3.3×10^9 bp.
- Number of base pairs in human = 6.6×10^9
- Length of DNA in humans = 6.6×10^9 bp $\times$ 0.34×10^{-9} m/bp = 2.2 m
- Length of DNA in *E. coli* =1.36 mm (1.36×10^{-3} m).
  $\therefore$ The number of base pairs = 1.36×10^{-3} m/0.34×10^{-9} m/bp = 4×10^6 bp.

Packaging of DNA Helix

- In prokaryotes (e.g., *E. coli*), the DNA molecule is held with some positively charged non-histone basic proteins like negatively charged polyamines and form 'nucleoid'.
- In eukaryotes, there is a set of positively charged basic proteins called histones.
- Histone proteins are rich in positively charged basic amino acid residues lysine and arginine.
- There are five types of histone proteins-H1, H2A, H2B, H3 and H4.
- Two molecules each of H2A, H2B, H3 and H4 organize to form a unit of eight molecules called as histone octamer.
- Negatively charged DNA is wrapped around positively charged histone octamer to form a structure called a nucleosome.
- Nucleosomes are connected with the help of linker DNA on which H1 Histone is present.

Nucleosome

- A typical nucleosome contains 200 bp of DNA helix.
- Therefore, the total number of nucleosomes in human = 6.6×10^9 bp/200 bp = 3.3×10^7.
- Nucleosomes constitute the repeated unit to form chromatin.
- Chromatin is the thread-like stained bodies.
- Nucleosomes in chromatin appears as "beads-on-string" when it is viewed under the electron microscope.
- Chromatin is packaged to form a solenoid or a zig-zag structure.
- Further supercoiling constitute a looped structure called chromatin fibre.
  These chromatin fibres further coil and condense at the metaphase stage of cell division to form chromosomes.
- Chromatin is packaged $\rightarrow$ solenoid $\rightarrow$ chromatin fibres $\rightarrow$ coiled and condensed at metaphase stage $\rightarrow$ chromosomes.
- Higher level packaging of chromatin requires non-histone chromosomal (NHC) proteins.
- Two types of chromatin are :
  (a) **Euchromatin :** Loosely packed and transcriptionally active chromatin and is light-stained.
  (b) **Heterochromatin :** Densely packed and inactive region of chromatin and stains dark.

Key Facts

In Eukaryotes, DNA polymerases are of 5 types, these are DNA polymerase α, β, γ, δ and ε.

Nucleotide arrangement in DNA can be seen by X-ray Crystallography.

Example 1

Q. Calculate the number of beaded structures (nucleo somes) present in the nucleus of diploid eukaryotic cell which possess 2.2×10^6 bp.

Sol. One nucleosome has 200 bp.

The number of beaded structures (nucleosomes) present in the nucleus of diploid eukaryotic cell which possess 2.2×10^6 bp.

$\therefore \frac{2.2 \times 10^6}{200} = 1.1 \times 10^4$ or 11×10^3 nucleosomes

The Search for Genetic Material

Griffith's Experiment - Transforming Principle

- **Griffith** (1928) used mice and a bacterial strain, *Streptococcus pneumoniae*.
- *Streptococcus pneumoniae* has two strains :
  (a) **Smooth (S) strain (Virulent) :** Has polysaccharide mucous coat. Causes pneumonia.
  (b) **Rough (R) strain (Non-virulent) :** No mucous coat. Does not cause pneumonia.

Experiment

- S-strain $\rightarrow$ Inject into mice $\rightarrow$ Mice die
- R-strain $\rightarrow$ Inject into mice $\rightarrow$ Mice live
- S-strain (Hk) $\rightarrow$ Inject into mice $\rightarrow$ Mice live
- S-strain (Hk) + R-strain (live) $\rightarrow$ Inject into mice $\rightarrow$ Mice die
- He concluded that there exists some 'transforming principle', that is transferred from heat-killed S-strain to R-strain. It enabled R-strain to synthesize smooth polysaccharide coat and become virulent. This must be due to the transfer of genetic material.

Biochemical Characterization of Transforming Principle

- Oswald Avery, Colin MacLeod & Maclyn McCarty in 1944 worked to determine the biochemical nature of 'transforming principle' in Griffith's experiment.
- They purified biochemicals (proteins, DNA, RNA, etc.) from heat-killed S cells using suitable enzymes.
- They discovered that :
  - **(a)** Digestion of protein and RNA (using Proteases and RNases) did not affect transformation. So, the transforming substance was not a protein or RNA.
  - **(b)** Digestion of DNA with DNase inhibited transformation. It means that DNA caused the transformation of R cells to S cells i.e., DNA was the transforming substance.

The Genetic Material is DNA

- The fact that DNA is the genetic material also came from the experiments of **Alfred Hershey** and **Martha Chase (1952)**.
- They worked with viruses that infect bacteria and are called bacteriophages.

Hershey-Chase Experiment—Blender Experiment

- Hershey and Chase made two preparations of bacteriophage - In one, proteins were labelled with ^{35}S by putting in a medium containing radioactive sulphur (^{35}S). In the second, DNA was labelled with ^{32}P by putting in a medium containing radioactive Phosphorous (^{32}P).
- These preparations were used separately to infect *E. coli*.
- After infection, the *E. coli* cells were gently agitated in a blender to separate the phage particles from the bacteria.
- Then the culture was centrifuged. Heavier bacterial cells were formed as a pellet at the bottom. Lighter viral components outside the bacterial cells remained in the supernatant.
- They found that,
  - **(a)** Supernatant contains viral protein labelled with 35S, i.e., the viral protein had not entered the bacterial cells.
  - **(b)** The bacterial pellet contains radioactive ^{32}P. This shows that viral DNA labelled with ^{32}P had entered the bacterial cells. This proves that DNA is the genetic material.

Properties of Genetic Material

- A molecule that can act as a genetic material must fulfill the following criteria :
  - **(a)** Be able to generate its replica by the process of replication.
  - **(b)** Chemically and structurally be stable.
  - **(c)** Allow slow changes, the mutations that are required for evolution.
  - **(d)** It should be able to store genetic information which can be inherited.
  - **(e)** Be able to express itself as 'Mendelian Characters'.

Key Words

Euchromatin : The region of chromatin which is loosely packed and genetically active.

Heterochromatin : The chromatin that is more densely packed, stains dark and is genetically inactive.

DNA is a better Genetic Material than RNA due to the following reasons :

- DNA is chemically less reactive and structurally more stable. It can to undergo repair.
- Due to the unstable nature of RNA, RNA viruses (e.g., *Q*β bacteriophage, Tobacco Mosaic Virus, etc.) mutate and evolve faster.
- For the storage of genetic information, DNA is better due to its stability. But for the transmission of genetic information, RNA is better.
- RNA can directly code for protein synthesis, hence can easily express the characters. DNA is dependent on RNA for protein synthesis.

Reasons for stability (less reactivity) of DNA	Reasons for mutability (high reactivity) of RNA
Double-stranded	Single-stranded
Presence of thymine	Presence of Uracil
Absence of 2'-OH	Presence of 2'-OH

- The two DNA strands are complementary. On heating, they separate. When appropriate conditions are provided they come together. (In Griffith's experiment, when the bacteria were heat-killed, some properties of DNA did not destroy).

RNA World

- RNA is a single-stranded structure but it is often folded back upon itself forming helices. Nitrogenous bases are like those of DNA except that there is **uracil** in place of **thymine**.

- RNA was the first regulatory chemical and genetic material in early life forms.
- It acts as genetic material and biocatalyst.
- Essential life processes (metabolism, translation, splicing, etc) evolved around RNA.
- DNA has evolved from RNA with chemical modifications that made it more stable.

Central Dogma of Molecular Biology

- It was proposed by **Francis Crick** (1958). It states that the genetic information flows unidirectionally from DNA $\rightarrow$ RNA $\rightarrow$ Protein.

Reverse Transcription : H. Temin and Baltimore in 1978 gave the concept of reverse flow of genetic information i.e., the formation of DNA from RNA. This is called Reverse Central Dogma or Teminism or reverse transcription. This takes place in some of the viruses in the presence of an enzyme called reverse transcriptase.

Types of RNA

- RNA is of 3 types –mRNA, tRNA and rRNA.
- mRNA constitutes 2–5% of the total cellular RNA, tRNA is about 15% and rRNA is about 70–80%.
- **mRNA (messenger RNA) :** Provides a template for translation (protein synthesis) and is transcribed from DNA.
- **rRNA (ribosomal RNA) :** Structural and catalytic role during translation. e.g., ^{23}S rRNA in bacteria acts as ribozyme.
  It is the component of ribosome and is the most stable type of RNA.
- **tRNA (transfer RNA or sRNA or soluble RNA or adaptor RNA) :** Brings amino acids for protein synthesis and reads the genetic code.
- tRNA is the smallest amongst all the RNA and is made up of 70–80 nucleotides only.

DNA Replication

- Replication is the copying of DNA from parental DNA.
- **Watson** & **Crick** proposed a semi-conservative mode of replication.
- It suggests that the parental DNA strands act as a template for the synthesis of new complementary strands. After the completion of replication, each DNA molecule would have one parental and one new strand.

Experimental Proof

- **Mathew Meselson** & **Franklin Stahl (1958)** experimentally proved semi-conservative mode.
- **Meselson & Stahl's Experiment :** They cultured *E. coli* in a medium containing $N^{15}H_4Cl$ (N^{15} : heavy isotope of N). N^{15} was incorporated into both strands of bacterial DNA and the DNA became heavier.
- Another preparation containing N salts labelled with N^{14} was also made. N^{14} was also incorporated in both strands of DNA and became lighter.
- These two types of DNA can be separated by centrifugation in a CsCl density gradient.
- They took *E. coli* cells from the N^{15} medium and transferred them to the N^{14} medium.
- After one generation (i.e., after 20 minutes), they isolated and centrifuged the DNA. Its density was intermediate (hybrid) between ^{15}N DNA and ^{14}N DNA. This showed that in the newly formed DNA, one strand is old (N^{15} type) and one strand is new (N^{14} type). This confirms the semi-conservative mode of replication.
- After II generations (i.e., after 40 minutes), there were equal amounts of hybrid DNA and light DNA.
- **Taylor et. al (1958)** performed similar experiments on *Vicia faba* (faba beans) using radioactive thymidine to detect distribution of newly synthesized DNA in the chromosomes. It proved that the DNA in chromosomes also replicate semi-conservatively.

The Machinery and Enzymes for Replication

- DNA replication starts at a point called *origin* (*ori*).
- A unit of replication with one origin is called a *replicon*.
- During replication, the two strands unwind and separate by breaking H-bonds in the presence of an enzyme, Helicase.
- Unwinding of the DNA molecule at a point forms a 'Y'-shaped structure called replication fork.
- The separated strands act as templates for the synthesis of new strands.
- DNA replicates in the 5'$\rightarrow$3' direction.
- Deoxyribonucleoside triphosphates (dATP, dGTP, dCTP & dTTP) act as substrate and also provide energy for polymerization.
- Firstly, a small RNA primer is synthesized in presence of an enzyme, primase.
- In the presence of an enzyme, DNA dependent *DNA polymerase*, many nucleotides join with one another to primer strand and form a polynucleotide chain (new strand).
- The DNA polymerase forms one new strand (leading strand) on a continuous stretch in the 3'$\rightarrow$5' direction (Continuous synthesis).
- The other new strand is formed in small stretches (Okazaki fragments) in the 5'$\rightarrow$3' direction (Discontinuous synthesis).
- The Okazaki fragments are then joined together to form a new strand by an enzyme, DNA ligase. This new strand is called lagging strand.
- If a wrong base is introduced in the new strand, DNA polymerase can do proofreading.
- *E. coli* completes replication within 38 minutes i.e., 2000 bp per second.
- In eukaryotes, the replication of DNA takes place at the S-phase of the cell cycle. Failure in cell division after DNA replication results in polyploidy.

- **Transcription**
  - It is the process of copying genetic information from one strand of the DNA into RNA.
  - Here, adenine pairs with uracil instead of thymine.
  - Both strands are not copied during transcription, because
    - **(a)** The code for protein is different in both strands. This complicates the translation.
    - **(b)** If two RNA molecules are produced simultaneously they would be complementary to each other, hence form a double-stranded RNA. This prevents translation.
- **Transcription Unit**
  - It is the segment of DNA between the sites of initiation and termination of transcription.
  - It consists of 3 regions :
    - (a) **A promoter (Transcription start site) :** Binding site for RNA polymerase.
    - (b) **Structural gene :** The region between promoter and terminator where transcription takes place.
    - (c) **A terminator :** The site where transcription stops.
  - The DNA- dependent RNA polymerase catalyses the polymerization only in 5'→3'direction.
  - 3'→5' acts as the template strand. 5'→3' acts as the coding strand.
  - 3'–ATGCATGCATGCATGCATGCATGC–5' template strand.
    5'–TACGTACGTACGTACGTACGTACG–3' coding strand.
- **Transcription Unit and the Gene**
  - **Gene :** Functional unit of inheritance. It is the DNA sequence coding for RNA molecule.
  - **Cistron :** A segment of DNA coding for a polypeptide.
  - Structural gene in a transcription unit is of two types :
    - **(a) Monocistronic structural genes (split genes) :** It is seen in eukaryotes. Here, the coding sequences (expressed sequences or exons) are interrupted by introns (intervening sequences).
    - **(b) Polycistronic structural genes :** It is seen in prokaryotes. Here, there are no split genes.
  - **Exons and Introns :** In eukaryotes, the monocistronic structural genes have interrupted coding sequences i.e., the genes in eukaryotes are split. The coding sequences or expressed sequences are called as exons. Exons are said to be those sequences that appear in mature or processed RNA. The exons are interrupted by introns. Introns or intervening sequences do not appear in mature or processed RNA.
- **Steps of transcription in prokaryotes**
  - **Initiation :** Here, the enzyme RNA polymerase binds at the promoter site of DNA. This causes the local unwinding of the DNA double helix. An initiation factor (σ factor) present in RNA polymerase initiates the RNA synthesis.
  - **Elongation :** The RNA chain is synthesized in the 5'-3' direction. In this process, activated ribonucleoside triphosphates (ATP, GTP, UTP & CTP) are added. This is complementary to the base sequence in the DNA template.
  - **Termination :** A termination factor (ρ factor) binds to the RNA polymerase and terminates the transcription.
  - In bacteria (Prokaryotes), transcription and translation can be coupled (Translation can begin before mRNA is fully transcribed) because mRNA requires no processing to become active.
  - Transcription and translation take place in the same compartment (no separation of cytosol and nucleus).

Key Words

Introns : The regions of a gene which are removed during the processing of mRNA.

Exons : The regions of a gene which become part of mRNA and code for different regions of proteins.

- **In eukaryotes, there are 2 additional complexities :**
  - **(a) There are three RNA polymerases :**
    - **RNA polymerase I :** Transcribes rRNAs (28S, 18S & 5.8S).
    - **RNA polymerase II :** Transcribes the heterogeneous nuclear RNA (hnRNA). It is the precursor of mRNA.
    - **RNA polymerase III :** Transcribes tRNA, 5S rRNA and snRNAs (small nuclear RNAs).
  - **(b) The primary transcripts (hnRNA) :** They contain both the exons and introns and are non-functional. Hence introns have to be removed. For this, it undergoes the following processes :
    - **Splicing :** From hnRNA, introns are removed (by the spliceosome) and exons are spliced (joined) together.
    - **Capping :** Here, a nucleotide methyl guanosine triphosphate (cap) is added to the 5' end of hnRNA.
    - **Tailing (Polyadenylation) :** Here, adenylate residues (200-300) are added at 3'-end. It is the fully processed hnRNA, now called mRNA.

Mnemonics

1. Concept: Erwin Chargaff's Rule
Mnemonics: Aaya**T**ha; **C**hala**G**ya
Interpretations: A- Adenine **= T-** Thymine
G-Guanine **= C-** Cytosine

2. Concept: Central Dogma of Molecular Biology
Mnemonics: Doctors **R**ecovered **P**atients
Interpretations: DNA **;** RNA → Protein

IMPORTANT DIAGRAMS :

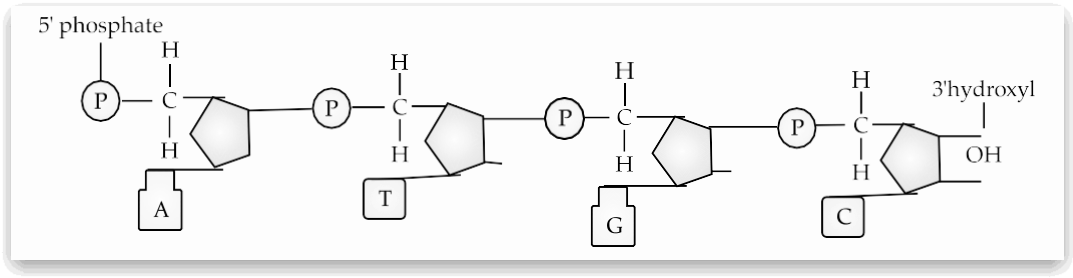

Fig 5.1 : A polynucleotide Chain of RNA

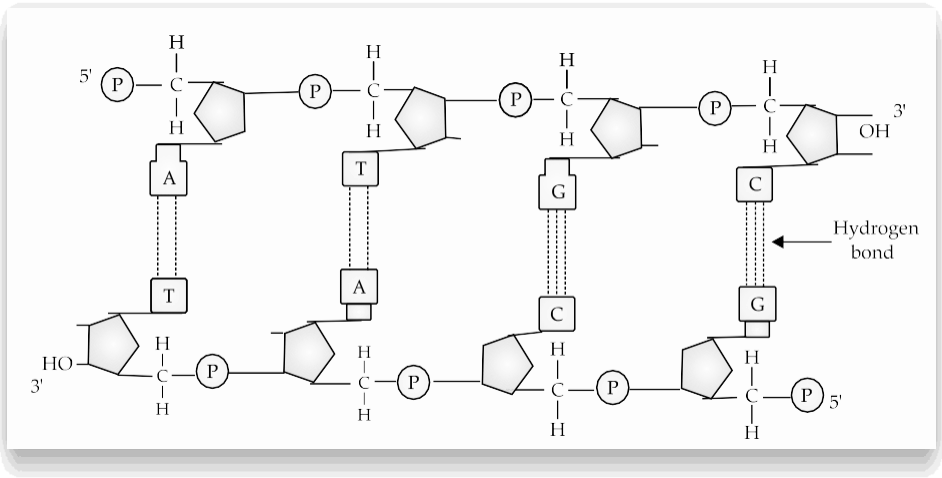

Fig 5.2 : Double Stranded polynucleotide chain

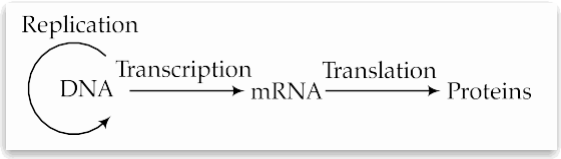

Fig 5.3 : Central Dogma

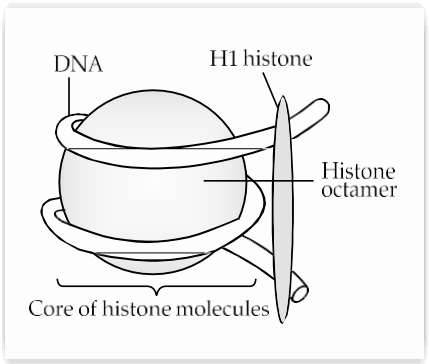

Fig 5.4 : Nucleosome

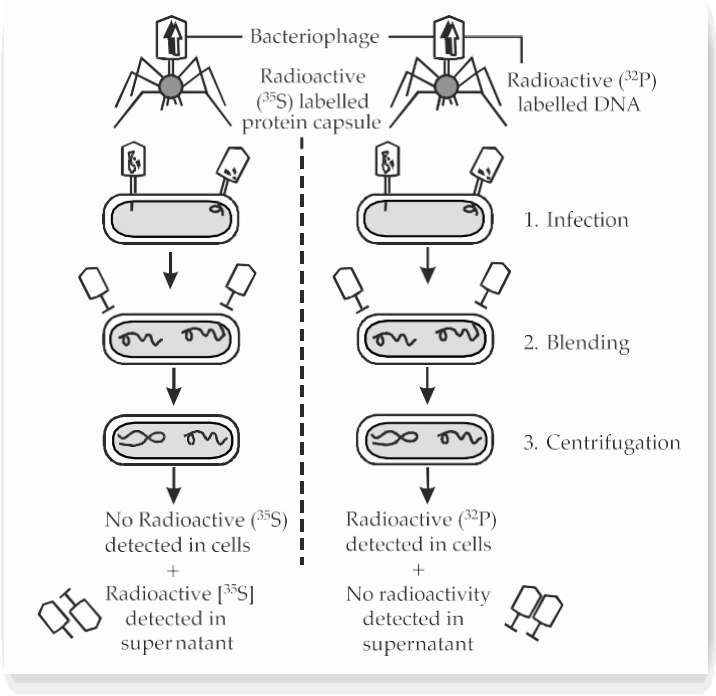

Fig 5.5 : The Hershey and Chase Experiment

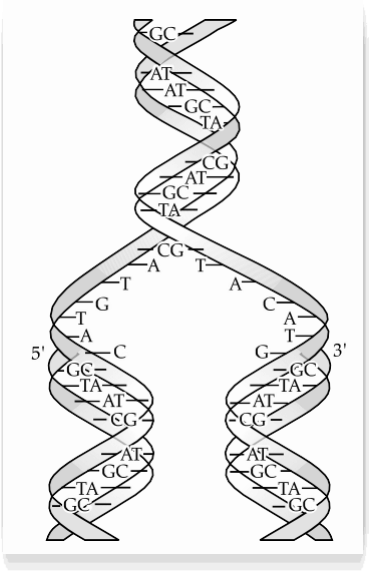

Fig 5.6 : Watson Crick model of Semi-conservative DNA replication

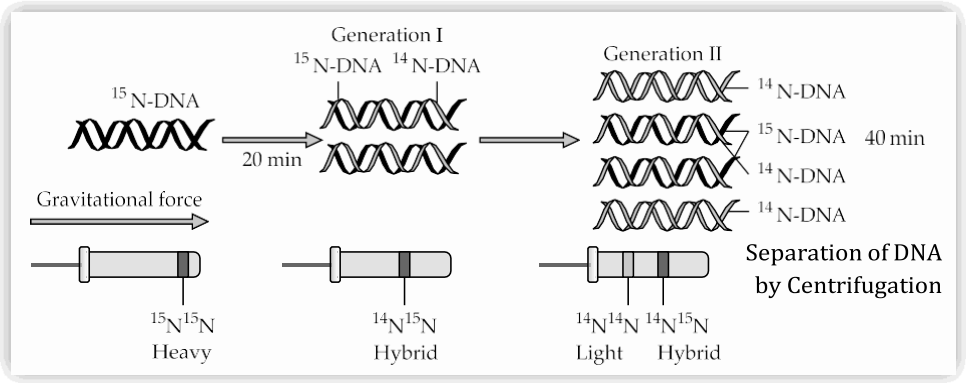

Fig. 5.7 : Meselson and Stahl's Experiment

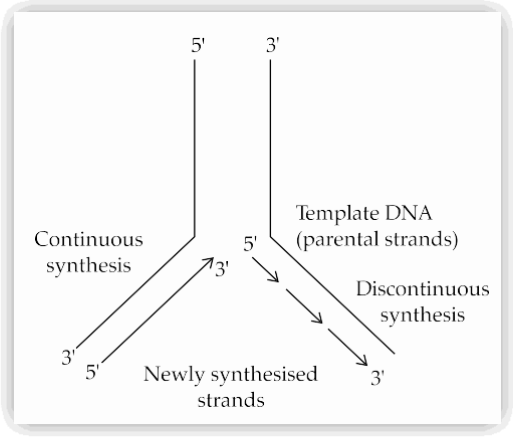

Fig 5.8 : Replicating Fork

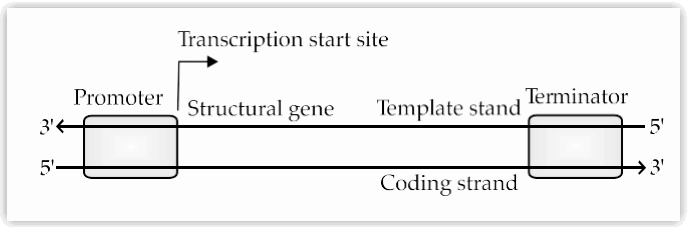

Fig 5.9 : Schematic structure of a transcription unit

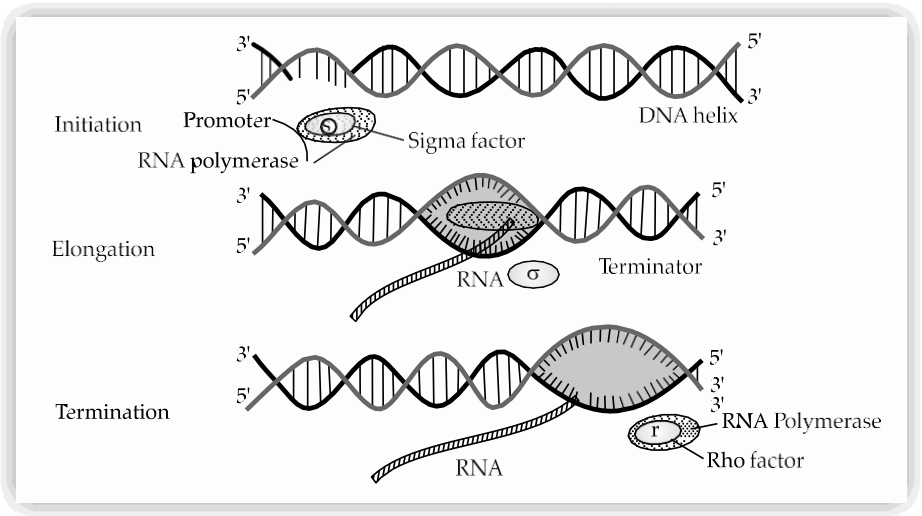

Fig 5.10 : Process of Transcription in Bacteria

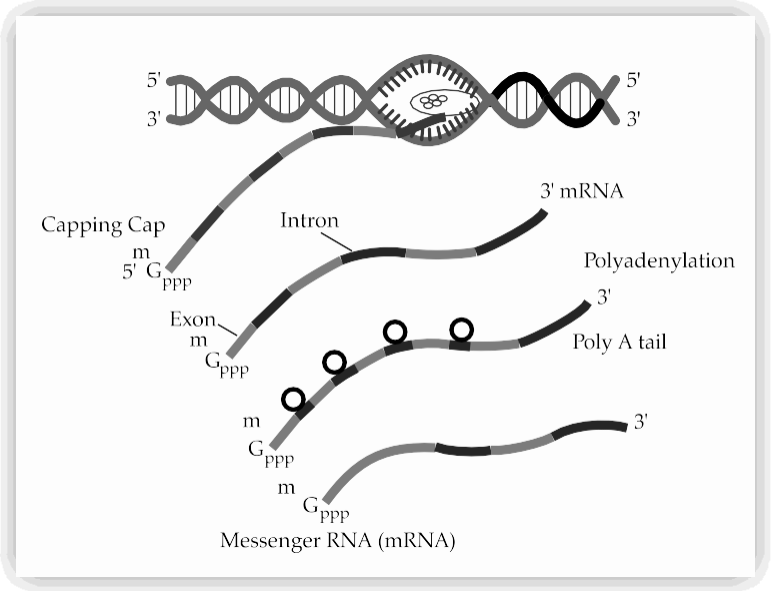

Fig 5.11 : Process of Transcription in Eukaryotes

Topic-2

Genetic Code, Translation, Lac Operon, HGP and DNA Fingerprinting

Concepts Covered • *Genetic code, Translation, Gene expression, Lac Operon, HGP, Rice Genome Project, steps and application of DNA Fingerprinting*

Revision Notes

- **Genetic Code** : It is the sequence of nucleotides in mRNA that contains information for protein synthesis (translation).
- 20 amino acids are involved in translation.
  - **George Gamow :** Suggested that for coding 20 amino acids, the code should be made up of 3 consecutive nucleotides.
  - **Har Gobind Khorana :** Developed the chemical method in synthesizing RNA molecules with defined combinations of bases (homopolymers and copolymers).
  - **Marshall Nirenberg :** Developed a cell-free system for protein synthesis.

- *Severo Ochoa* (polynucleotide phosphorylase) enzyme is used to polymerize RNA with defined sequences in a template-independent manner.

Salient Features of Genetic Code

- The genetic code is a triplet code (three-letter code) where three adjacent nitrogen bases code for a single amino acid.
- 61 **codons** code for amino acids. 3 codons (UAA, UAG and UGA) do not code for any amino acids. They function as stop codons (Termination codons or non-sense codons).
- Genetic code is universal e.g., From bacteria to human UUU codes for Phenylalanine. Some exceptions are found in mitochondrial codons and in some protozoans.
- No punctuations between adjacent codons (comma less code). The codon is read in mRNA in a continuous fashion.
- Genetic code is non-overlapping.
- A single amino acid is represented by many codons (except AUG for methionine and UGG for tryptophan). Such codons are called degenerate codons.
- Genetic code is unambiguous and specific. i.e., one codon specifies for only one amino acid.
- The codon is read in the 5' $\rightarrow$ 3' direction.
- AUG has dual functions. It codes for Methionine (met) and also acts as an initiator codon. In eukaryotes, methionine is the first amino acid and formyl methionine is the first amino acid in prokaryotes.

Key Facts

- RNA was the first genetic material to evolve and DNA was derived from it.
- DNA polymerase can not start DNA synthesis i.e., Denovo synthesis.
- RNA polymerase can start Denovo symthesis

Mutations and Genetic Code

- The relationships between genes and DNA are best understood by mutation studies.
- Effects of large deletions and rearrangements in a segment of DNA may result in loss or gain of a gene and so a function.
- A classical example of point mutation is a change of single base pair in the gene for beta globin chain of haemoglobin that results in the change of amino acid residue glutamate to valine. It results into a diseased condition called sickle cell anaemia.
- Insertion or deletion of one or two bases changes the reading frame from the point of insertion or deletion.
- When there is shifting of the reading frame due to insertion or deletion of the nucleotide, such mutation is known as **frameshift mutation**.
- This forms the genetic basis of proof that the codon is a triplet and is read in a continuous manner.

The Adaptor Molecule – tRNA

- The tRNA is a molecule that has about 60% of its part double-stranded and the rest remains single stranded which has unpaired bases.
- The tRNA has
  - **(a)** An **anticodon** (NODOC) loop that has bases complementary to the CODON with which it gets attached in mRNA.
  - **(b)** An amino acid acceptor end to which amino acid binds. This end or site lies at the 3' end & CCA–OH group. The 5' end bears G.
  - **(c)** **T Ψ C loop :** This is the site for attaching with the ribosome. This has some unusual bases like Ψ (pseudouridine) and ribothymidine.
  - **(d)** **DHU-Loop :** It is the binding site for the enzyme aminoacyl synthetase. It is the largest loop and has Dihydrouridine.
  - **(e)** **Extra arm :** It is a variable side arm lying between T Ψ C and anticodon loop.
- tRNA is called an adaptor molecule because it picks up amino acids from the cytoplasm and transfers them to ribosomes during protein synthesis.
- For initiation, there is another tRNA called initiator tRNA.
- There are no tRNAs for stop codons.
- 2-D structure of tRNA looks like a **clover-leaf** according to Robert Holly (1965). The 3-D structure looks like inverted 'L' according to Klug (1974).

Key Words

Anticodon : A sequence of three nitrogenous bases on tRNA which is complementary to the codon on mRNA.

Codon: A Sequence of three nucleotides present on mRNA which encodes for specific amino acid during translation.

- **Translation – Protein Synthesis**

It takes place in ribosomes. It includes 4 steps :

1. **Charging of tRNA (aminoacylation of tRNA)**
   - Formation of a peptide bond requires energy obtained from ATP.
   - For this, amino acids are activated (amino acid + ATP) and linked to their cognate tRNA in the presence of aminoacyl tRNA synthetase. So, the tRNA becomes charged.
2. **Initiation**
   - It begins at the 5'-end of mRNA in the presence of an initiation factor.
   - The mRNA binds to the small subunit of the ribosome. Now the large subunit binds to the small subunit to complete the initiation complex.
   - Large subunit has 2 binding sites for tRNA- aminoacyl tRNA binding site (A site) and peptidyl site (P site).
   - Initiation codon for methionine is AUG. So, methionyl tRNA complex would have UAC at the anticodon site.
3. **Elongation**
   - At the P-site the first codon of mRNA binds with anticodon of methionyl tRNA complex.
   - Another aminoacyl tRNA complex with an appropriate amino acid enters the ribosome and attaches to A site.
   - Its anticodon binds to the second codon on the mRNA and a peptide bond is formed between first and second amino acids in presence of an enzyme, peptidyl transferase.
   - The uncharged tRNA moves from the P site to the E site and the peptidyl-tRNA moves to the P site. This is called a translocation.
   - Then 3rd codon comes into A site and a suitable tRNA with 3rd amino acid binds at the A site. This process is repeated.
   - A group of ribosomes associated with a single mRNA for translation is called a polyribosome (polysomes).
   - A ribozyme is a ribonucleic acid (RNA) enzyme that catalyses a chemical reaction. The ribozyme catalyses specific reactions in a similar way to that of protein synthesis. Also called catalytic RNA, ribozyme are found in ribosome where they join amino acids together to form protein chains.
4. **Termination**
   - When aminoacyl tRNA reaches the termination codon like UAA, UAG & UGA, the termination of translation occurs. The polypeptide and tRNA are released from the ribosomes.
   - The ribosome dissociates into large and small subunits at the end of protein synthesis.
     An mRNA has additional sequences that are not translated (untranslated regions or UTR). UTRs are present at both 5'-end (before start codon) and 3'-end (after stop codon). They are required for an efficient translation process.

Mnemonics

Concept : Translation process (It include 4 steps)

Mnemonics : Come **I**n **E**vening **T**ime

Interpretations : Charging of tRNA, **I**nitiation, **E**longation, **T**ermination

- **Regulation of Gene Expression**

Gene expression results in the formation of a polypeptide. In eukaryotes, the regulation includes the following levels :

- Transcriptional level (formation of primary transcript).
- Processing level (regulation of splicing).
- Transport of mRNA from the nucleus to the cytoplasm.
- Translational level.

- **Importance of regulation of gene expression:**
  - Gene regulation is the process to switch off or switch on the genes as per the requirement of the organism.
  - Gene regulation is required so that there is no waste of energy in expressing the genes not required at the time.
  - However, there are housekeeping genes that are always expressed in the cell.

The metabolic, physiological and environmental conditions regulate the expression of genes. e.g.,

- In *E. coli*, the enzyme beta-galactosidase hydrolyses lactose into galactose and glucose. In the absence of lactose, the synthesis of beta-galactosidase stops.
- The development and differentiation of an embryo into an adult the result of the regulation of several set of genes.

- **Operon Concept :** This is a regulatory system that is observed in bacteria.
  - *"Each metabolic reaction is controlled by a set of genes".*

- All the genes regulating a metabolic reaction constitute an *Operon* e.g., *lac* operon, *trp* operon, *ara* operon, *his* operon, *val* operon etc.
- When a substrate is added to growth medium of bacteria, a set of genes is switched on to metabolize it. This is called induction.
- When a metabolite (product) is added, the genes to produce it are turned off. This is called repression.

The Lac Operon

- **Lac operon in *E. coli* :** The operon controlling lactose metabolism. It consists of a regulator gene, 3-structural genes, an operator gene, promoter gene, a repressor and an inducer.
  - **(a) A regulatory or inhibitor gene :** Codes for the repressor.
  - **(b) 3 structural genes :**
    - **(i) z gene :** Codes for β-galactosidase (hydrolyze lactose to galactose and glucose).
    - **(ii) y gene :** Codes for permease (increase permeability of the cell to lactose).
    - **(iii) a gene :** Codes for a transacetylase.

Key Word

Operon : A group of genes which control a metabolic pathway.

- The genes present in the operon function together in the same or related metabolic pathway. There is an operator region for each operon.
- If there is no lactose (inducer), lac operon remains switched off. In the absence of inducer, repressor gene is active. The regulator gene synthesizes mRNA to produce the repressor protein, this protein binds to the operator genes and blocks RNA polymerase movement. So, the structural genes are not expressed.
- In the absence of glucose, If lactose is provided in the growth medium, the lactose is transported into the *E. coli* cells by the action of permease. Lactose (inducer) binds with repressor protein.
- So, repressor protein cannot bind to operator gene. The operator gene becomes free and induces the RNA polymerase to bind with promoter gene then transcription starts. Regulation of lac operon by repressor is called negative regulation.

Human Genome Project (HGP)

- The entire DNA in the haploid set of chromosome of an organism is called a Genome.
- In human genome, DNA is packed in 23 chromosomes.
- Human Genome Project (1990-2003) is the first effort in identifying the sequence of nucleotides and mapping of all the genes in the human genome.
- Human genome contains about 3×10^9 bp.

Goals of HGP

(a) To identify all the estimated genes in human DNA.
(b) To determine the sequences of the 3 billion chemical base pairs that make up human DNA.
(c) To store this information in databases.
(d) To improve tools for data analysis.
(e) To transfer related technologies developed during the project of society to other sectors of society.
(f) To address the ethical, legal and social issues **(ELSI)** that may arise from the project.

HGP was Closely Associated with Bioinformatics

The application of computer science and information technology to the field of biology and medicine helps in analysing DNA sequence data.

Methodologies of HGP

There are two major approaches namely, ESTs and sequence annotation.

- **Expressed Sequence Tags (ESTs) :** Focused on identifying all the genes that are expressed as RNA and sequencing the same.
- **Sequence annotation :** Sequencing whole set of the genome containing all the coding & non-coding regions and later assigning functions to different regions.

Procedure :

Isolate total DNA from a cell → Convert into random fragments of smaller size → Clone in suitable host (e.g., BAC – bacterial artificial chromosomes & YAC – yeast artificial chromosomes) for amplification through PCR (polymerase chain reaction) → Fragments are sequenced using Automated DNA sequencers (using Frederick Sanger method) → Sequences are arranged based of the overlapping regions → Alignment of sequences using computer-based programs → Genetic and physical maps on the genome were generated using the information on polymorphism of restriction endonuclease recognition sites and some repetitive DNA sequences (micro-satellites).

Salient Features of Human Genome

(a) Human genome contains 3164.7 million nucleotide bases pairs.

(b) Total number of genes = about 25,000.
(c) Average gene consists of 3000 bases, but sizes vary. The largest known human gene (dystrophin on X-chromosome) contains 2.4 million bases.
(d) 99.9% of nucleotide bases are identical in all people. It is 0.1% which makes each of us unique.
(e) Functions of over 50% of discovered genes are unknown.
(f) Chromosome I has the most genes (2968) and Y has the fewest (231).
(g) Less than 2% of the genome codes for proteins.
(h) Repeated sequences make up a very large portion of the human genome. Repetitive sequences are stretches of DNA sequences that are repeated many times. They have no direct coding functions but they shed light on chromosome structure, dynamics and evolution.
(i) About 1.4 million locations where single-base DNA differences (SNPs- Single nucleotide polymorphism or 'snips') occur in humans.

- **Rice Genome Project**

Rice is one of the most largely consumed foods in India. Also, the population is increasing with a rapid pace, so, to meet this requirement, Rice genome project has been launched to increase the production of rice. Rice has the smallest genome of 430Mb nucleotides located on chromosome 12.

- **Rice Genome :** It is a joint project of National Institute of Aerobiological Sciences (NIAS), forestry and fisheries (STAFF), Ministry of Agriculture, Forestry and Fisheries (NAFF), Society for Techno-innovation of Agriculture genome research program.

Arabidopsis is an experiment plant of rice genome because it has fast life cycle and can be easily grown. It has smaller genome and high diversity and helps in enhancing the molecular products.

Key Word

ESLI : Ethical, legal and social implications of genetic and genomic research.

Need for sequencing rice genome:
- To know the functioning of genes by accurate gene sequencing.
- It is important for agronomic traits which requires mapping of genomic sequences.
- Improvement of other cereals will become easier.

- **DNA Fingerprinting (DNA profiling)**
  - It is the technique to compare the DNA fragments of two individuals.
  - Developed by **Alec Jeffreys (1985)**. He is considered as the father of DNA fingerprinting. **Lalji Singh** is the Father of Indian DNA fingerprinting.
- **Basis of DNA Fingerprinting**
  - DNA carries some non-coding sequences called repetitive sequence [Variable Number of Tandem Repeats **(VNTR)**].
  - Number of repeats is specific. It varies from person to person and is specific to a person.
  - The size of VNTR varies from 0.1 to 20 kb.
  - Repetitive DNA is separated from bulk genomic DNA as different peaks during density gradient centrifugation.
  - The bulk DNA forms a major peak and the other small peaks are called satellite DNA.
  - Satellite DNA is classified into many categories (micro-satellites, mini-satellites, etc.) based on the base composition (A-T rich or G-C rich), length of segment and number of repetitive units.
  - An inheritable mutation observed in a population at high frequency is called DNA polymorphism (variation at genetic level).
  - Polymorphism is higher in non-coding DNA sequence. This is because mutations in these sequences may not have any immediate effect on an individual's reproductive ability.
  - These mutations accumulate generation after generation and cause polymorphism. For evolution & speciation, polymorphisms play an important role.
- **Steps of DNA Fingerprinting (Southern Blotting Technique)**

(a) Isolate DNA (from any cells like blood stains, semen stains or hair roots).
(b) Make copies (amplification) of DNA by Polymerase Chain Reaction (PCR) if the amount of isolated DNA is small.
(c) Digest DNA by restriction endonucleases.
(d) Separate DNA fragments by gel electrophoresis over agarose polymer gel.
(e) Treat with alkali solution (NaOH) to denature DNA bonds so as to split them into single-stranded DNAs in the gel.
(f) Transfer (blotting) single-stranded DNA fragments to synthetic membranes such as nitrocellulose or nylon, and then baked in a vacuum oven at 80°C for 3-5 hours (to fix the DNA fragment on the membrane).
(g) Nitrocellulose filter membrane is placed in a solution containing a radioactive labelled single-stranded DNA probe. The DNA probes are small radioactive synthetic DNA segments of known sequences of nitrogen bases. These DNA probe binds with the complementary sequences of the DNA fragment on the membrane to form a hybridized DNA.
(h) The filter paper is washed to remove unbound probe.

(i) The hybridised DNA is photographed on to an X-ray film by autoradiography. The image (in the form of dark & light bands) obtained is called a DNA fingerprint. This gives the characteristic pattern of an individual's DNA.

- **Applications of DNA Fingerprinting are :**
  - Forensic tool to solve paternity, rape, murder, etc.
  - For the diagnosis of genetic diseases.
  - To determine the phylogenetic status of animals.

Key Word

VNTR: Variable Number of Tandem Repeats

Key Fact

- DNA finger printing is based upon principle of polymorphism in DNA sequence.

IMPORTANT DIAGRAMS :

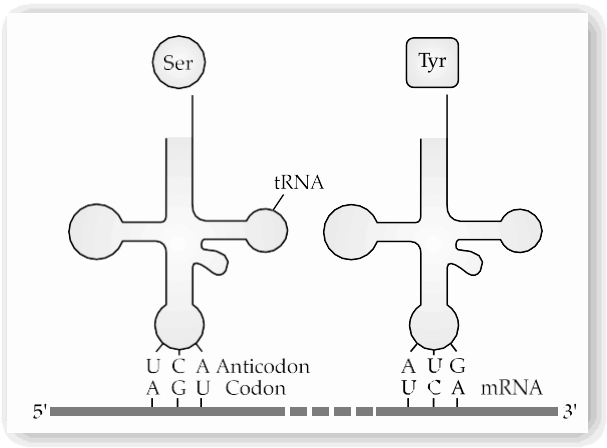

Fig 5.12: tRNA- the adapter molecule

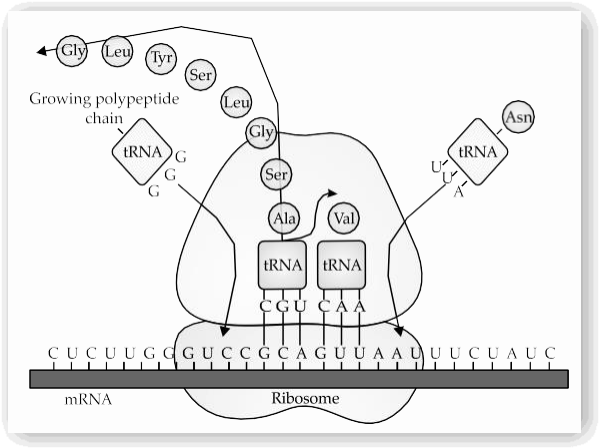

Fig 5.13: Translation

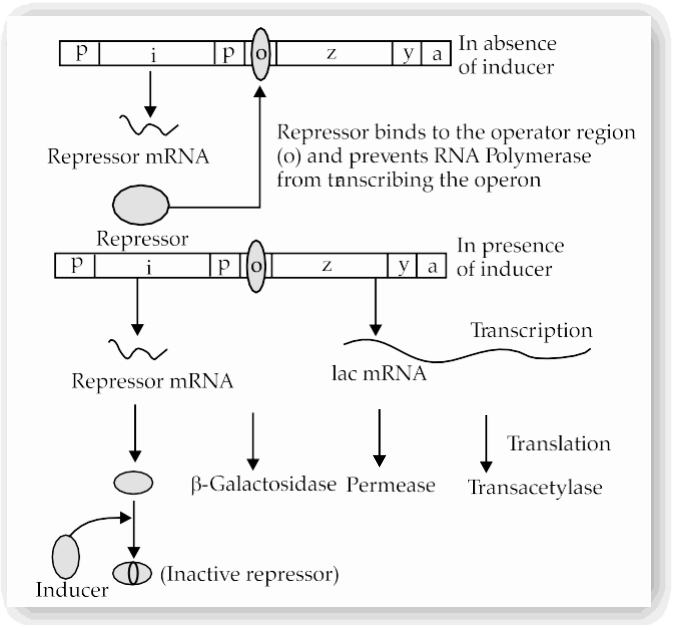

Fig 5.14: The *lac* Operon

Example 2

Q. When does lac operon get switched off ?

Sol. The lac operon comprises of one regulatory gene or inhibitor gene (i), are promoter gene, one operator gene and three structural genes. Regulator gene codes for a protein known as repressor protein, it is synthesised all the time from the i-gene.

The operon gets switched off when repressor protein produced by regulatory or inhibitor gene binds to operation gene. RNA polymerase gets blocked, so there is no transcription.

Repressor protein + Operator gene $\rightarrow$ Switched off