Molecular Basis of Inheritance

- The operon controlling lactose metabolism
  - • Structural gene
  - • Operator gene
  - • Promoter gene
  - • Repressor
  - • Inducer
  - Lac operon
    - In absence of inducer: Repressor mRNA → Repressor; Repressor binds to the operator region (o) and prevents RNA polymerase from transcribing the operon
    - In presence of inducer: Repressor mRNA + lac mRNA (Transcription); Translation → β-galactosidase, permease, transacetylase; Inducer → (Inactive repressor)
- Human Genome Project
  - • To identify all the estimated genes in human DNA
  - • To determine the sequences of the 3 billion base pairs
  - • To provide tools for data analysis
- Polynucleotide Chain
  - Made of Nucleotide
    - Phosphate group
    - Sugar
      - Ribose in RNA
      - Deoxyribose in DNA
    - Nitrogenous Base
      - C, T (only in DNA), U (Only in RNA)
      - A, G
- DNA packaged Nucleosome
  - DNA, H1 histone, Histone octamer, Core of histone molecules
- Central dogma
  - DNA —transcription→ mRNA —translation→ Protien
- Gene expression and regulation
  - Genetic Code
    - Definition: Sequence of nucleotides (nitrogen bases) in mRNA that contains information for protein synthesis (translation).
    - Features
      - •Triplet
      - •Non-ambiguous
      - •Degenerate
      - •Comma less
      - •Universal
  - Translation
    - Growing polypeptide chain: Gly, Leu, Tyr, Ser, Leu, Gly, Ser, Ala, Val
    - tRNA, tRNA, tRNA, tRNA
    - mRNA: C U C U U G G G U C C G C A G U U A A U U C U A U C
    - Ribosome
  - Transcription
    - In Prokaryotes
    - In Eukaryotes
    - Post transcriptional modification
      - Capping: Cap, Intron, Exon, RNA Splicing, Polydenylation, Poly A Tail, Mssenger RNA
      - Promoter, RNA polymerase, Sigma factor, DNA helix, Inittation, RNA, Terminator, Elongation, Termination, RNA, RNA polymerase, Rho factor
- Replication
  - Process: Template DNA (parental strands); Continuous synthesis; Discontinuous synthesis; Newly synthesised strands
  - Search for genetic material
    - Griffith exp
      - • S-strain → Inject into mice → Mice die
      - • R-strain → Inject into mice → Mice live
      - • S-strain (Heat killed) → Inject into mice → Mice live
      - • S-strain (Heat killed) + R-strain (live) → Inject into mice → Mice die
      - • Conclusion: Some 'transforming principle', transferred from heat-killed S-strain to R-strain.
    - Semi-Conservative DNA Replication
      - Meselson and Stahl experiment
        - Gravitational force
        - Generation I: ^{15}N-DNA; ^{15}N-DNA ^{14}N-DNA; 20 min
        - Generation II: ^{14}N-DNA, ^{14}N-DNA, ^{15}N-DNA, ^{14}N-DNA; 40 min
        - ^{15}N^{15}N Heavy; ^{14}N^{15}N Hybrid; ^{14}N^{14}N Light; ^{14}N^{15}N Hybrid
      - Biochemical Nature
    - Hershey and Chase exp.
      - Bacteriophage; Radioactive (*S) labelled protein capsule; Radioactive (*P) labelled DNA
      - 1. Infection; 2. Blending; 3. Centrifugation
      - No Radioactive (*S) detected in cells; Radioactive (*P) detected in cells
      - Radioactive [*S] detected in supernatant; No radioactivity detected in supernatant
      - •Discovered by Oswald Avery, Colin MacLeod and Maclyn McCarty.
      - •Conclusion: DNA is the hereditary material.

CHAPTER

28 MOLECULAR BASIS OF INHERITANCE

Chapter Objectives

Search for genetic material and DNA as genetic material, Structure of DNA and RNA, DNA packaging,DNA replication, Central dogma, Transcription, Genetic code, Translation, Gene expression and regulation – Lac operon, Genome and Human and Rice genome projects, DNA fingerprinting.

STUDY MATERIAL

I. Concept Clarified

1. INTRODUCTION

- Hereditary molecules of inheritance are Deoxyribonucleic acid (DNA) and **Ribonucleic acid (RNA)** are the two types of nucleic acids. In most of the organisms, DNA is the genetic material. In some viruses, RNA acts as the genetic material.

Search for genetic material and DNA as genetic material

- The research work of **Gregor Mendel**, **Sutton** and **Boveri**, **Thomas Morgan** etc. had restricted the search to the chromosomes located in the nucleus of most cells.

Griffith Transforming Principle

- **Frederick Griffith**; working on *Streptococcus pneumoniae* made an accidental discovery of **transformation** in the DNA of bacteria.
- He observed that the bacteria produced two types of colonies on culture plates – **the smooth strain** (**S**) which secreted a mucopolysaccharide coat and **rough strain** (**R**) which lacked such a coat.

  When he inoculated smooth strains into the experimental mice, it produced the disease while that of the rough strains did not. He concluded from his observations that:

  S strain → when Injected into mice → Mice died

  R strain → when injected into mice → Mice lived.
- He further observed that, **heat killed S strains** did not kill the mice. But, when the mice were injected with a mixture of **live R** and **heat killed S strains**, the mice died. He represented this observations as :

  (Heat killed) S strain→ injected into mice→ Mice live

  S strain (heat killed) + R strain → injected into mice → Mice die (live)
- From these observations, **Griffith** concluded that there was some **transformation in the DNA of** the S strain which was transferred to the R strain, that caused it to secrete the polysaccharide coat and get converted into the S strain.

  This is known as **transformation.** This was the first evidence regarding the genetic material (DNA).

2. BIOCHEMICAL ANALYSIS OF TRANSFORMING PRINCIPLE

- **Oswald Avery**, **Colin Mcleod** and **McCarty** (1933 – 44) purified biochemicals (proteins, DNA, RNA etc) from the heat – killed S cells to see which ones could transform live R cells into S cells. They observed that digestion with proteases and RNAases did not affect transformation so the transforming principle was not a protein, or RNA. But the digestion with **DNAase** did inhibit transformation, suggesting that **DNA caused** the **transformation**.

  DNAase is the enzyme responsible for the digestion and separation of DNA from the nucleosomes.

3. DNA AS THE GENETIC MATERIAL OF TRANSFORMATION

- **Alfred Hershey** and **Martha Chase** in 1952 established that **DNA** was the **genetic material**, on the basis of their work conducted on **bacteriophages.**

Mechanism of Transformation in Bacteriophage:

- The bacteriophage attaches itself to the host bacteria and its genetic material enters the bacterial cell. The bacterial cell treats the viral genetic material as if it was its own and subsequently manufactures more virus particles.
- They grew viruses in a culture medium containing **radioactive phosphorus** and **sulphur**. The viruses which were grown in presence of radioactive phosphorus had radioactive DNA but not radioactive protein. Similarly those grown in the presence of radioactive sulphur had radioactive protein but not radioactive DNA. These viruses were then allowed to infect *E.coli* cells. The viruses were then separated from bacterial cells by centrifugation.
- The bacteria infected with viruses having radioactive DNA were found to be radioactive indicating that the DNA had been injected into the bacterial cells while those infected with radioactive protein were not radioactive.

This observation made it clearly evident that the **DNA was the genetic material.**

4. CHARACTERISTICS OF THE GENETIC MATERIAL:

Any substance which form the heritable material must fulfil some essential requirements and DNA was found to fulfil them all.

1. It is stable
2. It is able to carry and transcribe information which are required to control the processes which give the organism its specificity (transcription).
3. It is capable of replicating exactly, so that the genetic determinants are transmitted down from cell to cell and from generation to generation unchanged.
4. It is able to mutate to give more variations.

5. RIBONUCLEIC ACID

- RNA was the first genetic material. There are several evidences which suggest that major life processes have evolved around RNA. Apart from functioning as a genetic material, RNA also serves as a catalyst.
- But, RNA being a catalyst was reactive and unstable. Therefore, DNA has evolved from RNA with chemical modifications, that make it more stable. DNA being double stranded and having complementary strand further resists changes by evolving a process of repair.

Some Organisms have RNA as Genetic Material:

- Some viruses have RNA as genetic material Coat. Notable human diseases caused by RNA viruses include SARS, Influenza. Hepatitis C, Measels, Mumps and Rabies, Rous Sarcoma Virus (causing cancer) and HIV (causing AIDS) both are Retroviruses using reverse Transcription for DNA synthesis in host cell.

6. STRUCTURE OF DNA AND RNA

The Basics of the DNA – Deoxyribonucleic acid:

- DNA is a complex molecule of heredity. It is a long polymer of deoxyribonucleotides. The length of DNA is usually defined as the number of nucleotides present in it, which subsequently is characteristic of an organism. The number of base pairs differ in different organisms from simple to complex level. For e.g : ***E. coli*** has **4.6×10^6 bp**, **bacteriophage** has **48502 bp**, bacteriophage ϕ **174 (ϕ x 174)** has **5386** nucleotides and human DNA's haploid content is **3.3×10^9bp**.

7. Composition of Nucleotide:

- The nucleotide is the basic unit or the **building block of polynucleotide chain.** A molecule of nucleotide has 3 components – **a nitrogenous base**, **a pentose sugar** (deoxyribose or ribose) and a **phosphate group**. Nitrogenous bases are of two types– **Purines** (Adenine and Guanine) and **Pyrimidines** (Cytosine, Uracil, Thymine). Nitrogenous base is linked to the pentose sugar through **N – glycosidic linkage** to form a **nucleoside**.

- When a phosphate group is linked to 5′ – OH of a nucleoside through a **phosphodiester** bond, a **nucleotide is formed**. Two nucleotides are linked through 3′ – 5′ phosphodiester linkage to form a **dinucleotide**. Many nucleotides are joined together to form a polynucleotide chain. The resulting chain will have a 5′ – end of pentose sugar, which is referred to as 5′ – end (5′ prime end).
- The 3′ end of the polynucleotide chain has a free 3′ – OH group and is termed as 3′ – end (3 – prime end). The sugar and phosphate forms the back bone of the polynucleotide chain.

In RNA, the following differences are found:

(a) An additional – OH group at the 2′ – position in ribose sugar.

(b) Uracil base is present instead of thymine base.

(c) It is single stranded.

Outline Composition of Nucleic Acid

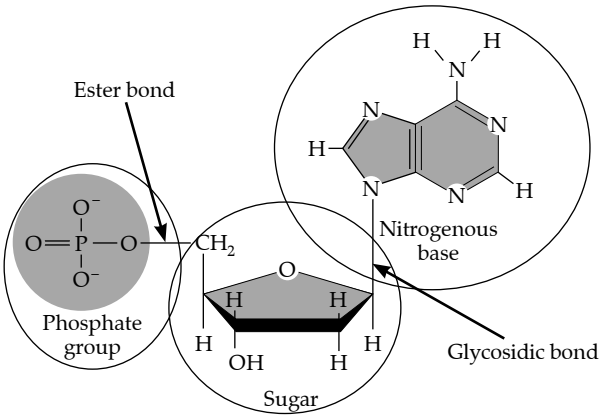

Fig.: Structure of a Nucleotide

8. BASICS OF DOUBLE HELICAL STRUCTURE OF DNA

In 1983 James Watson and Francis crick based on the X-ray diffraction data produced by Wilkins and Rosalind Franklin proposed a very simple but famous Double Helix Model for the structure of DNA.

The salient features of double – helix structure of DNA are:

(i) It is made of **two polynucleotide chains**, with **sugar – phosphate** linkage forming the **backbone** and the bases (A, G, T & C) are projecting inside.

(ii) **Anti - Parallel polarity** of the two chains due to characteristic phosphate linkage and N_2 – base pairing to the sugar molecule.

(iii) Bases of the two strands are paired through **Hydrogen bonding**. It shows A = T and G ≡ C pairing.

(iv) The two chains are coiled in a **right handed** fashion.

(v) The pitch of the helix is **3.4** nm and there are approximately **10 base pairs** in each turn (gyre). Thus, one gyre shows the base pair at a distance of 0.34 nm.

(vi) The plane of one base pair stacks over the other in the double helix.

9. Configurational Forms of DNA's Double Helices:

DNA undergoes different configurational changes, which are the alternative forms of double helices. These forms differ in features like :

(a) Number of residues per turn or

(b) The spacing of residues along the helical axis. The conformation of DNA depends on the nature of the molecules with which it is interacting. The three main types of DNA seen are B-DNA, A-DNA and C-DNA. D forms and E forms are also found as extreme variants which have only 8 to 7 ½ base pairs per turn and seen in DNA molecules lacking guanine. All these types of DNA have right handed helix. Another form of DNA, **Z-DNA exists which is left handed helix.**

B – DNA:

1. The helical coiling in this DNA is right handed and this type of DNA is called B-form.
2. The structure of DNA described above is the one commonly found in living systems.

10. DNA PACKAGING

Coiling of DNA Double Helix and DNA Proteins:

- Packaging or coiling of DNA helix is different in prokaryotic and eukaryotic organisms. In prokaryotic cells, like *E.coli*, the DNA is confined in a particular area within the cell, called the '**nucleoid'.**
- In eukaryotic cells, the organisation of DNA is more complex with a well-defined nucleus. The nucleus contains certain specialised protein units called **histones**.
- **Packaging of DNA** is dependent on these positively charged histones.DNA which is **negatively charged,** is wrapped around these positively charged histones to form a strucure called the **nucleosome**.
- The nucleosome consists of nearly 200 bp of DNA. These nucleosome form the repeating unit of a structure in the nucleus called the **chromatin**.
- The DNA strand between the two nucleosome is called the linked DNA, the length of which varies from 8 to 114 base pair. H_1 is associated with the linker DNA.

11. HETEROCHROMATIN AND EUCHROMATIN

Chromatin are thin thread-like, **darkly staining** structures within the nucleus, which **condense** during the **metaphase** stage of the cell division to form **chromosome**.

- The additional set of chromatin packaging are collectively referred to as **Non-Histone Chromosomal Proteins** (**NHC**).

Regions of the **chromatin** which is **loosely packed** and **stains lightly,** are called **euchromatin** which is active in transcription and those regions which are **densely packed** and **stain darkly** are called **heterochromatin**. Heterochromatin is transcriptionally inactive.

12. CENTRAL DOGMA

Flow of Genetic Information:

- The flow of genetic information was observed by **Francis Crick's Central Dogma of Molecular Biology**. The **genetic information flows** from **DNA** to **RNA** and then to **protein**. It is called as Central Dogma. It is represented as :

$$\text{DNA} \underset{\text{Reverse}}{\overset{\text{Transcription}}{\rightleftharpoons}} \text{RNA} \xrightarrow{\text{Translation}} \text{Protein}$$

Central dogma

13. DNA REPLICATION

Mechanism of The Flow of Genetic Information (Central Dogma)

The flow of genetic information takes place in three major steps such as:

(i) Replication, (ii) Transcription and (iii) Translation

Replication:

- It is the process of the synthesis of new DNA strands from the segment of gene on the parent DNA (Template DNA) with the help of DNA Polymerase enzyme.
- Apart from proposing the helical structure of DNA, Watson and Crick also proposed a scheme for the replication and suggested that the two strands would separate and act as a template for the synthesis of new complementary strands. Thus, each new molecule of DNA would have one **parental strand** and one **newly synthesized** strand. This scheme is referred to as the **semi-conservative DNA replication**.

Messelson and Stahl's Semi conservative replication of DNA:

- They grew *E.coli* in a medium, in which the only source of nitrogen was $^{15}NH_4Cl$ for several generations.
- As a result, the newly synthesized cells contained DNA incorporated with ^{15}N, which could easily be distinguished from normal DNA by **CsCl density gradient centrifugation**.
- Then, the cells were transferred to medium with normal $^{14}NH_4Cl$ and samples were collected at regular intervals as the cells multiplied.
- It was seen that DNA extracted from cells, after transfer from ^{15}N to N^{14} medium after one generation, had a **hybrid** or **intermediate** density (N^{14} and ^{15}N together in the DNA).
- DNA extracted from the culture after another generation had equal amount of hybrid DNA and light DNA.

14. REPLICATION FORK AND ENZYMES OF REPLICATION:

Replication starts at a specific point called **origin of replication**.

- At this site the two strands separate and form a replicating fork in the form of Y.

  There are two types of replication forks on the basis of direction of replication.,

  1. Unidirectional replication fork
  2. Bidirectional replication fork

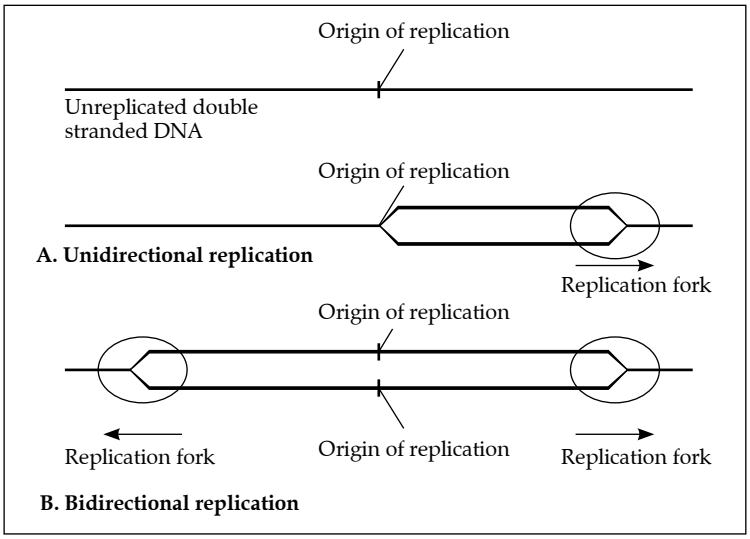

Fig.: Unidirectional and bidirectional replication fork

Any mistake during replication can lead to mutations. Replication of DNA is a duplication process by which DNA molecule forms exact copies of its own structure.

- Replication occurs inside the chromosomes.
- It occurs during the **S phase** of interphase.
- The two strands separate and act as template for the synthesis of new DNA strand.

Enzymes Involved in Replication

There are several enzymes involved in the process of replication of which the most important one is the **DNA polymerase** enzyme. The enzyme is attributed with properties like :

(a) Catalysis of polymerisation of nucleotides

(b) Speed of catalysis

(c) Accuracy

The entire strand does not separate as it requires a large amount of energy. This separation is brought about by enzyme **helicase.**

- DNA synthesis requires a **primer** which is a **short RNA polynucleotide chain.**
- The synthesis of new DNA strand takes place by the **addition of DNA nucleotide to the 3′ OH group of the RNA primer.**
- The DNA-dependent DNA **polymerase** catalyses the polymerisation only in one direction i.e. 5′→3′.
- Thus, **one strand is synthesised in a continuous manner** (the template of which has polarity 5′→3′) while on the other is synthesised in small fragments (which has template with polarity 5′→3′).
- The discontinuous synthesis of DNA strand produces fragments known as **Okazaki fragments** which are joined by an enzyme called **DNA ligase**.
- The RNA primer is degraded at the end of replication.
- The newly formed DNA is **complementary** to the template chain.

The **DNA polymerase III** is the most important enzyme involved in the DNA replication. DNA polymerase I and II also take part in the replication.

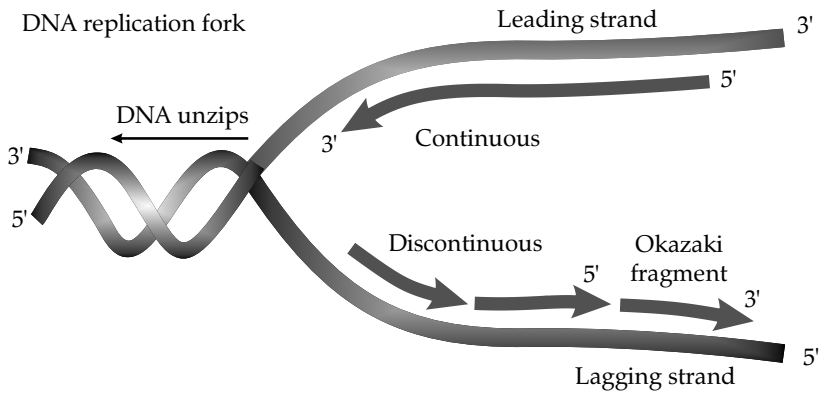

Fig.: DNA Replication

15. TRANSCRIPTION

The process of flow of genetic information from DNA to the RNA with the help of **DNA dependent RNA Polymerase** enzyme is called as Transcription.

Complementary principle governs the process of transcription, except that adenosine pairs with uracil instead of thymine. Unlike in the process of replication, only a segment of DNA and **only one of the strands is copied into RNA in transcription.**

In transcription, **both the strands are not copied** because,

(i) If both strands act as a template, they would code for RNA molecule with different sequences. (with respect to Adenine and Thymine).

(ii) Two RNA molecules if produced simultaneously would be complementary to each other, and hence would result in **a double stranded RNA**; which would prevent it from being translated into a protein.

Transcriptional Unit in DNA

A transcription unit in DNA consist of the following 3 regions :

(i) Promoter; (ii) The structural gene; (iii) Terminator

DNA dependent RNA polymerase catalyses the polymerisation reaction only in one direction (5′ → 3′). So the strand with polarity 3′ → 5′ acts as a **template strand**. The other strand which has the polarity and sequence same as RNA, is displaced during transcription. This is called the **coding strand**, though it does not code for anything.

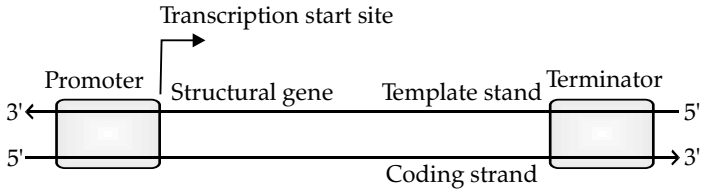

Fig. : Schematic structure for a transcription unit

The template strand has a special region called **promoter.** This region is located in the 3′ end. The RNA synthesis starts from the promoter. In DNA a particular site called **terminator** site is present. The promoter and terminator are present on the either sides of the structural gene, in the transcription unit.

A DNA sequence provides binding site for RNA polymerase, and the presence of a promoter defines the template and coding strands.

Transcriptional Unit in Prokaryotes and Eukaryotes:

- **Gene** is often described as the functional unit of inheritance.
- A **cistron** is defined as the segment of DNA coding for a polypeptide.
- The structural gene in a **transcription unit** found mostly **in the eukaryotes** is said to be **monocistronic**, i.e., code for a polypeptide chain.
- However, the condition is **polycistronic** in the case of **bacteria** and other **prokaryotes**.

In eukaryotes, the monocistronic structural genes are split with interrupted coding sequences These coding sequences are called **exons** (these appear in mature or processed RNA). The exons are interrupted by sequences called **introns**, which do not appear in mature or processed RNA.

- The **promoter** and the **regulatory sequences** of a structural gene also affect the inheritance of a character. This is the reason why some regulatory sequences are referred to as regulatory genes, though they do not code for any RNA or protein.

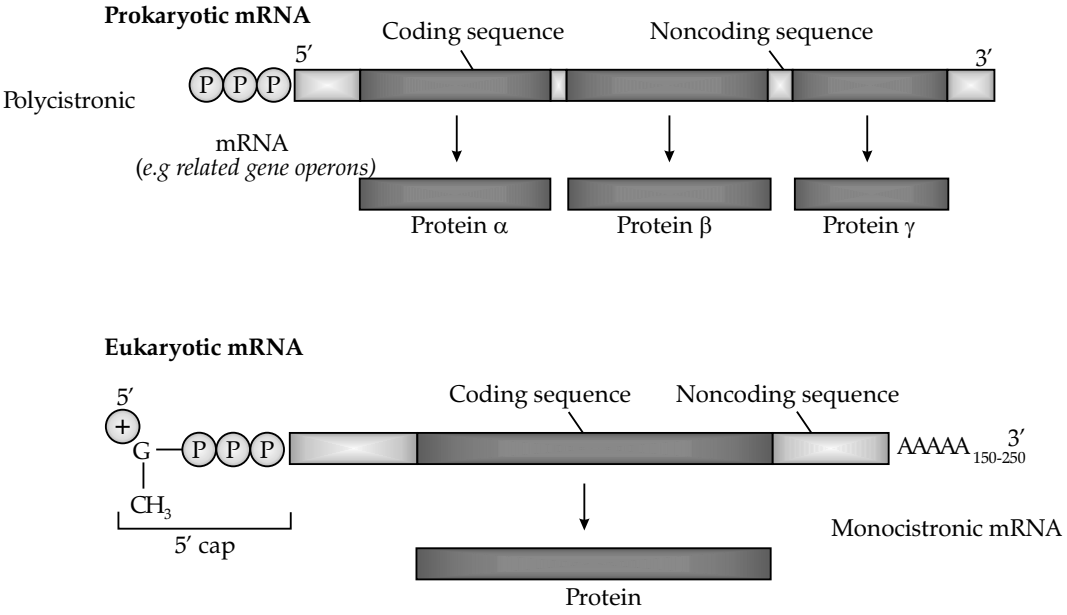

Fig. : Polycistronic and monocistronic mRNA

16. TYPES OF RNA

There are three types of RNA on the basis of their configurational changes in their structure.

Bacterial RNA can be basically of 3 different types;

(a) mRNA (messenger RNA) →provides template

(b) tRNA (transfer RNA) → brings amino acids and reads genetic code

(c) rRNA (ribosomal RNA) → structural and catalytic role during translation

Types of RNA Polymerase Enzymes:

There are 3 RNA polymerases in the nucleus:

- **RNA polymerase I** transcribes **rRNAs -**
- **RNA polymerase II** transcribes the precusor of **mRNA** (heterogenous nuclear RNA – hnRNA).
- **RNA polymerase III** transcribes **tRNA**, **5s rRNA**, (**snRNAs**, **small nuclear RNA**)

RNA Polymerases in eukaryotes

Form	Product	Location
I	rRNA	Nucleolus
II	mRNA, SRNA	Nucleoplasm
III	5S rRNA, TRNA	Nucleoplasm

Fig.: Prokaryotic RNA Enzyme - A Holoenzyme

Mechanism of Transcription

The process of transcription is very complex, and it takes place in a sequential steps headed under, (i) Initiation; (ii) Elongation and (iii) Termination.

(i) RNA polymerase binds to the promoter and initiates transcription. This is termed as **Initiation**. The recognition is done by the **sigma factor (σ)** present in RNA polymerase.

(ii) Using nucleoside triphosphate as a substrate, the polymerisation is carried out in a template dependent fashion.

(iii) The nucleotide complementary to the first base of initiation site of template DNA gets attached to RNA polymerase.

(iv) The helix is opened gradually and the **elongation** is continued.

(v) **Termination:** When the polymerase reaches the terminator region (in termination site the sequence is a poly A i.e., AAAAA), the nascent RNA and the RNA polymerase falls off and the process is terminated.

- When the elongated RNA chain is about 10 base long, the sigma factor dissociates from the core enzyme and is available to bind to free core enzymes to initiate new rounds of transcription.
- The **rho** factor (ρ) associated with RNA polymerase recognises the stop sequence on the DNA strand.

  In fact, the RNA polymerase enzyme can only catalyse elongation reaction when, it associates itself, with two factors, namely initiation factor (σ) and termination factor (ρ), to catalyse the initiation and termination process respectively.

  In bacteria, since the mRNA does not require any processing to become active, and also since transcription and translation takes place in the same compartment, the two can be coupled.

Rho-dependent Transcription Termination

Post - Transcriptional Modifications in Eukaryotes:

1. The primary transcripts are non – functional as they contain both **introns** and **exons**. Therefore, they have to undergo a treatment called **splicing** by means of which the introns are removed and the exons are joined together to form the functional transcript, mRNA.

2. The hnRNA undergoes two additional processings called **capping** and **tailing**.

In capping, an **unusual nucleotide** (**methyl guanosine triphosphate**) is added to the 5′ end of the hnRNA.

In tailing, about **200 – 300 adenylate** residues are added to the 3′ end in a **template independent** manner.

It is the fully processed hnRNA, that is transported out of the nucleus for translation.

17. GENETIC CODE

- **George Gamow**, a physicist proposed that a genetic code should be constituted by a combination of bases. He also added that in order to code for 20 amino acids, a code should **contains three nucleotides** (**Triplet codon**). Further the chemical method devised by **Hargobind Khorana** helped to **synthesize RNA molecules** with defined combination of bases.
- Finally the **genetic code** was **deciphered by the cell free system** for protein synthesis put forward by **Marshall Nirenberg**.

Salient Features of Genetic Code:

(i) The genetic **codon** is **triplet** in nature.

(ii) The **genetic code** is **unambiguous** and specific.

(iii) It **degenerates** and hence more than **one codon codes for the same amino acid** (also known as Wobble hypothesis).

(iv) The codon is **read** in mRNA in a **continuous fashion**. (from 5′ – end to 3′ – end).

(v) The code is **universal**.

(vi) AUG has **dual function** i.e., it codes for **methionine** and also acts as the **initiator codon**.

Amino Acids Coding : There are 64 codons in the genetic code. Out of these **61** are **coding** (**sense codons**) and **3** are **non – coding** or **stop** or **terminator codons** (UAA, UAG, UGA). The **AUG codon** codes for the **formyl methionine** in prokaryotes eukaryotes it codes for **methionine.**

Second base

First base	U	C	A	G	Third base
U	UUU, UUC — Phenylalanine **F** UUA, UUG — Leucine **L**	UCU, UCC, UCA, UCG — Serine **S**	UAU, UAC — Tyrosine **Y** UAA — Stop codon UAG — Stop codon	UGU, UGC — Cysteine **C** UGA — Stop codon UGG — Tryptophan **W**	U C A G
C	CUU, CUC, CUA, CUG — Leucine **L**	CCU, CCC, CCA, CCG — Proline **P**	CAU, CAC — Histidine **H** CAA, CAG — Glutamine **Q**	CGU, CGC, CGA, CGG — Arginine **R**	U C A G
A	AUU, AUC, AUA — Isoleucine **I** AUG — Methionine **M** start codon	ACU, ACC, ACA, ACG — Threonine **T**	AAU, AAC — Asparagine **N** AAA, AAG — Lysine **K**	AGU, AGC — Serine **S** AGA, AGG — Arginine **R**	U C A G
G	GUU, GUC, GUA, GUG — Valine **V**	GCU, GCC, GCA, GCG — Alanine **A**	GAU, GAC — Aspartic acid **D** GAA, GAG — Glutamic acid **E**	GGU, GGC, GGA, GGG — Glycine **G**	U C A G

Fig.: Genetic code

Mutations : These are the sudden changes (heritable) that occur in the genetic constitution of an organism.

- They can occur due to **deletion, substitution** or **insertion** (**duplication**) of DNA sequences. A classical example is that of **sickle cell anaemia,** where **glutamic** acid is replaced by **valine** in the sixth position of the β - chain of the **haemoglobin** molecule. Haemoglobin consists four polypeptide chains as **two** α and **two** β.
- **Insertion** or **deletion** of one or two **bases** can change the **reading frame** from the point of insertion or deletion. Such mutations are called **frame shift mutations**. However insertion or deletion of bases in 3 or multiples of 3 will not alter the reading frame.

	Substitution	Insertion	Deletion
Original sequence	T G G C A G	T G G C A G	T G G C A G
Mutated sequence	T G G T A G	T G G T A T C A G	T G G G

Fig.: Types of mutations

18. TRANSLATION

The process of formation of proteins from m-RNA is called as translation.

The type of protein depends on the sequence of amino acids involved, which in turn depends on the number and sequences of bases on the **mRNA**. The amino acids are joined together by peptide bond, the joining of which requires energy.

Clover- Leaf Structure of tRNA : Francis Crick postulated the presence of an adaptor molecule that would, on the, one hand read the code and on the other hand, would bind to specific amino acids. The **tRNA** was known to have existed, even before the genetic code was postulated.

The tRNA has an **anticodon** loop which contains **bases** complementary to the genetic code. It also has an **acceptor end** to which it specifically binds the amino acids. Initiation is specifically done by the **initiator tRNA**. However there are no tRNAs for the stop codon.

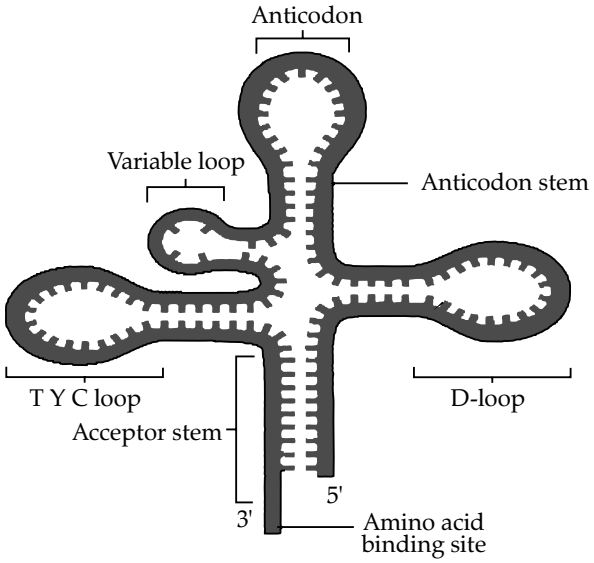

Fig.: Clover-leaf structure of tRNA

Mechanism of Translation

The process of translation takes place in three steps:

1. **Initiation**
2. **Elongation**
3. **Termination**

Activation of amino acids in the presence of ATP and their linking to the tRNA – this process is called the **charging** or **aminoacylation** of tRNA.

- The formation of a peptide bond between two such charged tRNAs is energetically favoured.
- The process of translation begins with the encounter of the **smaller ribosomal subunit** with mRNA.
- The larger ribosomal subunit has **two sites** (A and P sites), for amino acids to bind. **The first aminoacyl tRNA always binds to P site.** The next aminoacyl tRNA binds to the A site followed by peptide bond formation between the two amino acids.
- The **translator unit** in mRNA is a sequence of RNA flanked by a **start codon** (AUG) and a **stop codon**. It may also have some additional sequences, which are not translated (**untranslated regions - UTR**). .
- The ribosome moves from codon to codon along the mRNA. Amino acids are added one by one and translated into polypeptide.
- Finally the **release factor** binds to the stop codon, terminating the translation and releasing the polypeptide chain from the ribosome.

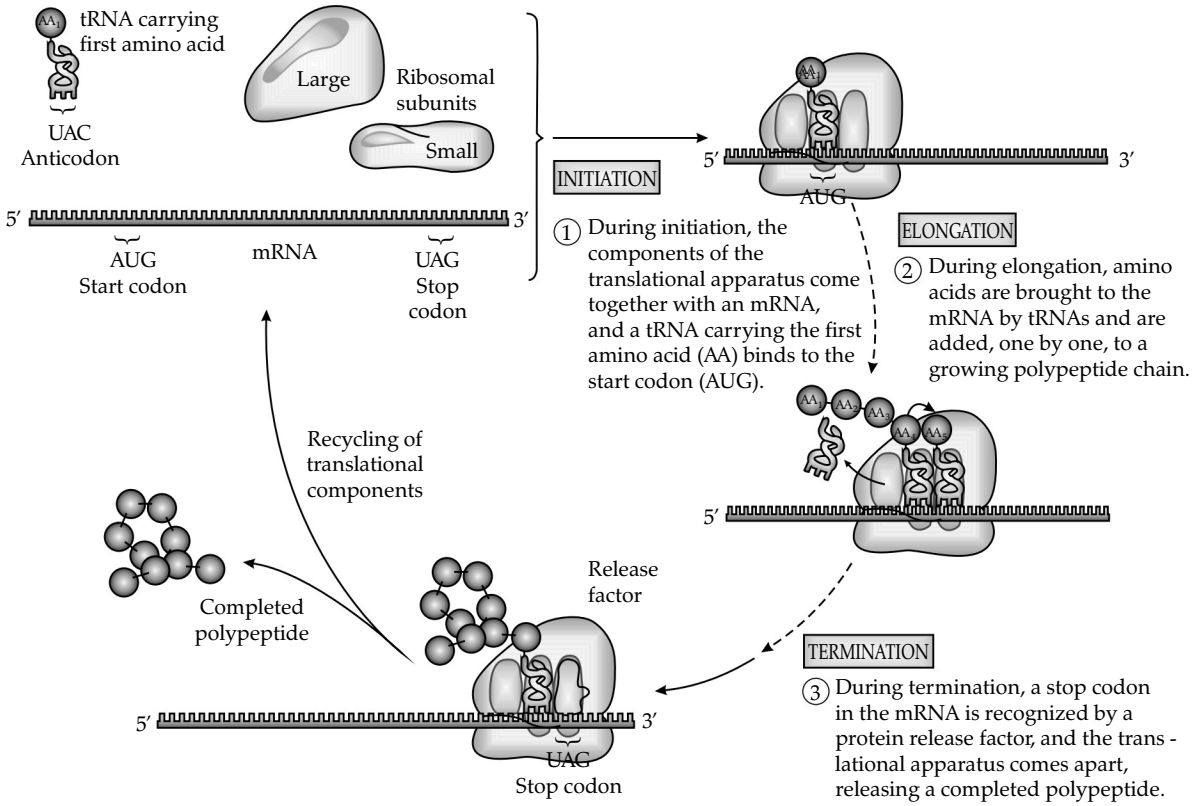

Fig.: Process of translation

Enzymes involved in translation:

1. Activation of amino acid-**Aminoacyl synthetase.**
2. Attachment of activated amino acid to tRNA-**Aminoacyl tRNA synthetase .**
3. Enzyme involved in peptide bond formation- **Peptide synthetase (it is an integral part of 50 S subunits).**

Regulation of Translation:

Protein synthesis is the ultimate result of gene expression and this process of synthesis can be regulated at various levels. In **eukaryotes** the regulation could be exerted :

(i) During the formation of primary transcript.

(ii) During splicing

(iii) During transportation of mRNA from nucleus to the cytoplasm.

(iv) During translation

It is the **metabolic**, **physiological** and **environmental** factors that control or regulate the gene expression.

In **prokaryotes**, transcriptional level is the primary area of control of gene expression.

19. GENE EXPRESSION AND REGULATION

Regulation of Gene Expression- The Lac Operon:

- **Jacob** and **Monod** studied the regulation of gene expression for the first time in the intestinal bacteria, ***E.coli***. The regulation could be done either by **induction** or by **repression**. The scheme proposed by these workers to explain induction and repression of enzyme synthesis is called **operon concept**.
- The first well documented gene regulation in bacteria was the genetic determination of three enzymes that cause the **breakdown of lactose**. These were the **structural genes** (segments of DNA that codes for the synthesis of enzymes). The number of structural genes corresponds to the number of enzymes required for the breaking down of **lactose**. The three genes involved are **z, y** and **a**.
- The sequence of DNA which controls the transcription is called **operator gene**. This gene functions as the control unit in switching on and off the activity of the structural genes.

- When switched on, mRNA is transcribed and protein synthesis begins. Protein synthesis is inhibited when the gene is switched off.
- A small region next to the operator gene is called the **promoter gene,** which reacts with enzyme that causes transcription. The operator gene is under the control of another gene called regulator gene.
- **Regulator gene** produces a specific protein called repressor that binds with operator gene and renders it inactive. This prevents the enzyme bound to the promoter gene from progressing to the structural genes, thus blocking transcription.
- Occasionally, substances in the cytoplasm called **inducers** may bind with the repressor. When this occurs, the operator gene becomes active and transcription will continue until the repressor binds with the operator.
- In lac operon, it was found that **lactose** (**inducer**) interacts with the **repressor** and the operator gene becomes active and simulates the structural genes to produce mRNA that determine the synthesis of three different enzymes required for the breakdown of lactose. The three enzymes are, β **- galactosidase**, **galactoside permease, galactoside transacetylase**. As lactose is metabolised, repressor molecules are free to bind with the operator gene, and transcription by the structural genes stops.

  Since the system involves metabolism of lactose, this operon is called **Lactose Operon** or **Lac Operon**.

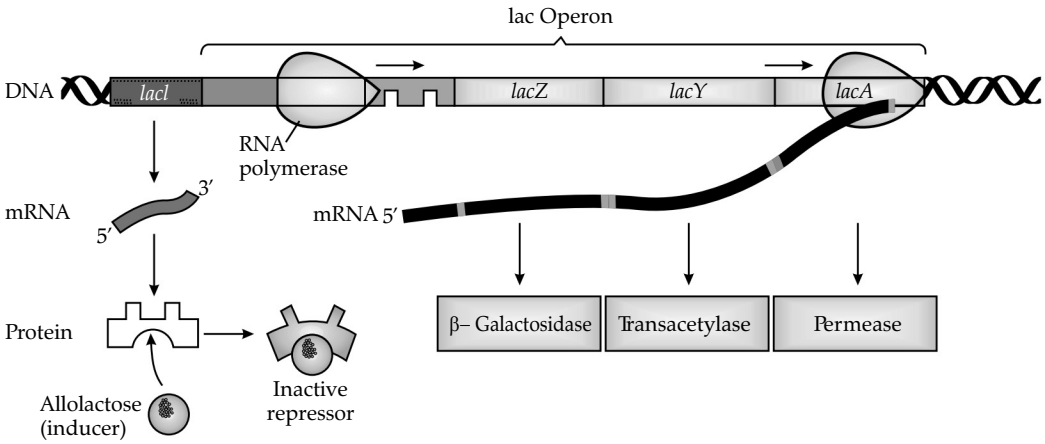

Fig. : Regulation of gene expression in *lac*-operon

20. HUMAN GENOME PROJECT

The Human Genome Project (HGP) was the international, collaborative research program Its goal was the complete mapping and understanding of all the genes of human beings. All our genes together are known as our "genome." It was launched in the year 1990. It was a **13 year** project coordinated by the **U.S Department of Energy** and **National Institute of Health**. The project was completed in **2003**.

Highlights of HGP:

(i) Identification of nearly 20,000 – 25,000 genes in human DNA.

(ii) Determination of sequences of 3 billion chemical base pairs.

(iii) Storing this information in databases.

(iv) Improve tools for data analysis.

(v) Transfer the information to other sectors, like industries.

(vi) Address the ethical, legal and social issues that may arise from the project.

The salient features of human genome are:

(1) The human genome contains 3164.7 million nucleotide bases.

(2) The average gene consists of 3000 bases.

(3) The total number of genes is estimated to be 30,000.

(4) Functions of over 50% of the discovered genes are unknown.

(5) Less than 2% of the genome codes for proteins.

(6) A large portion of the genome is made of repeated sequences (commonly termed as VNTR).

(7) Repetitive sequences have no direct coding functions, but it gives an idea on chromosome structure, dynamics and evolution.

(8) Chromosome 1 has the most genes (2968) and Y has the fewest (231).

(9) Scientists have identified about 1.4 million locations where single base DNA differences (SNPs →Single Nucleotide Polymorphism) occur in humans.

Methodologies

It involves two major approaches:

(i) Identifying all the genes that are expressed as RNA-referred to as Expressed Sequence Tags (ESTs)

(ii) Sequencing the whole set of genome that contained all the coding and non-coding sequences and latter assigning different regions with functions - Sequence Annotation.

(iii) For sequencing, the total DNA from a cell is isolated and converted into random fragments of smaller sizes and cloned in suitable host using vectors. The commonly used hosts were bacteria and yeast and the vectors were called as BAC (Bacterial Artificial Chromosomes) and YAC (Yeast Artificial Chromosomes).

(iv) The fragments were sequenced using automated DNA sequencers that worked on the principle of a method developed by Frederick Sanger.

(v) These sequences were then arranged based on some overlapping regions present in them. This required generation of overlapping fragments for sequencing with the help of computer based programs.

(vi) These sequences were subsequently annotated and were assigned to each chromosome.

(vii) Assigning the genetic and physical maps on the genome was generated using information on polymorphism of restriction endonuclease recognition sites.

Advantages of HGP:

(i) It helps in **diagnosis**, **treatment** and **prevention** of thousands of human disorders.

(ii) It provides clues to **Human Biology**.

(iii) It provides knowledge on capabilities of non – human organisms and on how to apply them in solving the problems of **health care**, **agriculture**, **energy production** and **environmental remediation**.

(iv) Non – human model organisms like **bacteria**, **yeast**, ***Drosophila*** and *Arabidopsis* have also been sequenced.

21. RICE GENOME PROJECT

- Rice, *Oryza sativa*, is one of the most important crop. Rice, wheat, and maize together account for about half of the world's food production. Rice itself is the principal food of the half of the world's population.
- For the first whole genome sequencing of a cereal crop, the rice genome is well mapped and well characterized, and it is the smallest of the major cereal crop genomes at an estimated 400 to 430 Mb.
- Rice an excellent model cereal, as it is an easiest of the cereal plants to transform genetically due to the genome size of 430 Mb. The rice genome is 3.5 times the size of the **Arabidopsis** genome and the third largest public genome project undertaken to date, behind the human and mouse genomes.

22. DNA FINGERPRINTING

- This techniques was invented in 1984 by Professor Sir Alec Jeffreys.
- DNA finger printing involves identifying difference in some specific regions in DNA sequences called as repetitive DNA. Repetitive sequences are small stretches of repeating strands of DNA. These repeating units are separated from the bulk genomic DNA by **density gradient centrifugation**. In the process, the bulk DNA forms a **major peak** and the **smaller peaks** are called the **satellite DNA**.
- The probability of having two people with the same DNA fingerprint that are not identical twins is very small.
- Depending on the **length of the segment**, and number **of repetitive units**, satellite DNA is classified into many categories. e.g., **micro satellites**, **mini satellites** etc. which do not code for proteins, though they constitute a large portion of the human genome. The sequences show an increased degree of polymorphism which inturn forms the basis of DNA **fingerprinting**. Every part of the body in a given individual, whether it be **blood**, **skin**, **bone** or **saliva,** has the same degree of polymorphism and hence is of great significance in **forensic science**.
- The gene polymorphisms are also inherited from parents to children, so it forms the basis of paternity testing in cases of dispute.
- **The human genome mapping** also makes use of DNA polymorphism. Polymorphism also arises due to mutations. Mutations that happen in germ cell do not seriously impair the **reproductive capability** of the individual. The mutation can be transmitted to other members of the population through sexual reproduction. **Inheritable mutations** of high frequency seen in a population are called **DNA polymorphism**. These variations are more commonly observed in **non – coding** DNA sequences. These mutations accumulate throughout

generations and cause **polymorphism**. Hence, they hence play a key role, by serving as raw materials for **Evolution** and **Speciation**.

- Alec Jeffreys used DNA sequences with high degree of polymorphism called the **Variable Number of Tandem Repeats** (**VNTR**) as probes in his technique of DNA fingerprinting. The various steps involved can be summarized as follows :
  (i) DNA isolation.
  (ii) Fragmentation of the DNA sequences using **restriction enzymes.**
  (iii) Transfer of these fragments blotting to synthetic **membrane** such as **nitrocellulose** or **nylon**.
  (iv) Hybridisation using the VNTR **probe**.
  (v) Detection of hybridised DNA fragments by autoradiography.

II. Know the Facts

- **Friedrich Meisher** was the first to identify DNA as an acidic substance present in the nucleus in 1869 and he called it nuclein. But, it was only in 1953, that **James Watson** and **Francis Crick** proposed the **double helix structure of DNA** based on the **X ray – diffraction** studies.
- **Erwin Chargaff**, put forward the rule : In a molecule of DNA, **the amount of purines will always be equal to that of the pyrimidines**. This concept is termed as CHARGAFF'S rule given as., [A]+[G] = [C]+[T].
- **Reverse transcription** is proposed by Temin and Baltimore (1970).
- It is common in Tumour viruses, and HIV etc, generally termed as **Retroviruses**.
- Histones are **positively charged** protein units, the strength of which depends on the abundance of basic amino acid residues with positive charges on their side chains.
- These histones in turn are organised to form a unit of eight molecules called as **Histone Octamer.**
- Histones Octamer are of four pairs as H_2A, H_2B, H_3 and H_4. An additional linker protein called H_1 – histone is also present in the periphery of nucleosome.
- Both, DNA and RNA can function as genetic material but DNA being more stable stores genetic information more effectively and RNA is better as far as transfer of genetic material is concerned.
- Genetic material of different organisms:

Double Stranded DNA. (dsDNA)	Higher animals and plants Bacteria. Polyoma virus and small-pox virus.
Single Stranded DNA. (ssDNA)	The bacteriophage ϕ X174 and several bacterial viruses.
Double Stranded RNA. (dsRNA)	Reo group of viruses. Wound tumour virus.
Single Stranded RNA. (SS/Lc)	Tobacco mosaic virus. Influenza virus, Poliomyelitis virus.

- **Types of DNA Polymerase in Prokaryotes and Eukaryotes**

Name	Function
Prokaryotic Polymerases	
DNA polymerase I	Erases primer and fills in gaps on lagging strand
DNA polymerase II (error-prone polymerase)	DNA repair
DNA polymerase III	Primary enzyme of DNA synthesis
Eukaryotic Polymerases	
DNA polymerase α	Initiator polymerase
Primase sub-unit	Synthesizes the RNA primer
DNA polymerase unit	Adds stretch of about 20 nucleotides to the primer
DNA polymerase β (error-prone polymerase)	DNA repair
DNA polymerase δ	Primary enzyme of DNA synthesis

- If DNA repair is required for the correct base pairing, then the information flows in a reverse direction i.e., from RNA to DNA in the presence of Reverse Transcriptase enzyme.
- There are three known **stop codons** named as, UAG (amber), UGA (opal), and UAA (ochre). These are also called as **termination codons** as they do not have tRNA anti-codons, instead they bind release factors and release polypeptide chains from the ribosome due to binding of release factors.
- Post- translational modifications involves,

  (i) Phosphorylation, (ii) Glycosylation, (iii) Lipidation, (iv) Acetylation, (v) Ubiquitination and (vi) Disulfide bond
- Catabolic breakdown product of glucose prevents activation of the *lac* operon by lactose, so this effect was originally called catabolite repression. The effect of the glucose catabolite is exerted on an important cellular constituent called *cyclic adenosine monophosphate (cAMP).*
- Human DNA, as it turns out, is largely junk—that is, 98.6 percent does not code for proteins. Half of the junk DNA consists of repeated sequences of various types, most of which are parasitic elements inherited from our distant evolutionary past. Only 1.1 percent to 1.4 percent constitute sequences that code for proteins that function as genes.
- Pieces of DNA which are cut by restriction enzymes are separated according to size by a process called as **'gel electrophoresis'**, which uses Agarose gel helping in the movement of shorter pieces of DNA easily.

❑❑❑