# CHAPTER – 6 : MOLECULAR BASIS OF INHERITANCE

## Structure of DNA:

Watson and Crick proposed a double helical model for DNA, based on X-ray crystallography of the molecule. Each strand (helix) is a polymer of nucleotides, each nucleotide consisting of a deoxyribose sugar, a nitrogen base and a phosphate. The sugar – phosphate chain is on the outside and act as back bone and the bases are on the inside (like in ladder). The two strands are held together by weak hydrogen bonds between the nitrogen bases. A purine base, always pairs with a pyrimidine base, i.e., adenine (A) pairs with thymine (T) and guanine (G) pairs with cytosine (C). So the two strands are complementary to each other and run in antiparallel direction with one chain having 5' – 3' orientation and the other having a 3' – 5' orientation. The purine and pyrimidine bases are stacked 0.34 nm apart in the chain and the helix makes a turn after ten base pairs, i.e., 3.4 nm.

## Central dogma of molecular biology:

Crick proposed the Central dogma in molecular biology, which states that the genetic information flows from DNA --> RNA --> Protein. In some viruses like retroviruses, the flow of information is in reverse direction that is from RNA --> DNA --> mRNA --> Protein.

## Packaging of DNA helix:

In prokaryotes, negatively charged DNA is held with some positively charged proteins and form as nucleoid.

In eukaryotes, negatively charged DNA is held with positively charged proteins called Histones (octomer) and form a structure called Nucleosome.

## The search for Genetic Material:

### 1. Bacterial Transformation (Transforming Principle) :

Fredrick Griffith conducted his experiment on *Streptococcus pneumoniae*, the pneumonia causing bacterium. He observed that there are two strains of this bacterium, one forming smooth colonies (S-type) with capsule (virulent) and the other forming rough colonies (R – type) without capsule (avirulent).

**Experiment:**

a) Smooth type bacteria were injected into mice. These mice died as a result of pneumonia caused by bacteria.
b) Rough type bacteria were injected into mice. These mice lived and pneumonia was not produced.
c) Smooth type bacteria which normally cause disease were heat killed and then injected into the mice. The mice lived and pneumonia was not caused.
d) Rough type bacteria (living) and heat killed S-type were injected together into mice. The mice died due to pneumonia and virulent smooth type living bacteria could also be recovered from their bodies.

This indicates that some factor from the dead S-cells converted the live R-cells into S-cells (transformation).

Later Avery, MacLeod and McCarty (1944) found out that when DNA isolated from the heat killed S-cells was added to R-cells in a culture, the R-cells changed into S-cells and pathogenic.

**Evidence from experiments with bacteriophage:**

This experiment was devised by Hershey and Chase with two different preparations of $T_2$ phage. In one preparation, the protein part was made radioactive and in the other, nucleic acid (DNA) was made radioactive. These two phage preparations were allowed to infect the culture of E.coli. Soon after infection, before lysis of cells, the *E.coli* cells were gently agitated in a blender, to loosen the adhering phage particles and the culture was centrifuged. The heavier infected bacterial cells pelleted to the bottom and the lighter viral particles were present in the supernatant. It was found that when $T_2$ phage containing radioactive DNA was used to infect *E.coli*, the pellet contained radioactivity. If T2 phage containing radioactive protein coat was used to infect *E.coli*, the supernatant contained most of the radioactivity. This suggests that during infection by the virus, the viral DNA enters the bacterial cell and that has the information for the production of more viral particles. It proves that DNA and not proteins, is the genetic material in bacteriophage.

**Properties of Genetic Material**:

a) It should be able to generate its replica (replication)
b) It should chemically and structurally be stable.
c) It should provide the scope for slow changes (mutation) that are required for evolution.
d) It should be able to express itself in the form of 'Mendelian Characters'.

**Replication:**

The Watson – Crick model of DNA immediately suggested that the two strands of DNA should separate. Each separated or parent strand now serves as a template (model) for the formation of a new but complementary strand. Thus, the new or daughter DNA molecules formed would be made of one old or parental strand and another newly formed complementary strand. This method of formation of new daughter DNA molecules is called semi-conservative method of replication.

**The Experimental Proof:**

Meselson and Stahl conducted an experiment to prove that DNA replication is semi conservative. They grew bacterium E. coli in a medium containing nitrogen salts ($^{15}NH_4Cl$) labeled with radioactive $^{15}N$. $^{15}N$ was incorporated into both the strands of DNA and such a DNA was heavier than the DNA obtained from *E.coli* grown on a medium containing $^{14}N$. Then they transferred the *E.coli* cells on to a medium containing $^{14}N$. After one generation, when one bacterial cell has multiplied into two, they isolated the DNA and evaluated its density. Its density was intermediate between that of the heavier $^{15}N$-DNA and the lighter $^{14}N$-DNA. This is because during replication, new DNA molecule with one $^{15}N$-old strand and a complementary $^{14}N$-new strand was formed (semi-conservative replication) and so its density is intermediate between the two.

**Mechanism of DNA replication**:

The intertwined DNA strands start separating from a particular point called origin of replication (single in prokaryotes and many in eukaryotes). This unwinding is catalysed by enzymes called Helicases. Enzymes called Topoisomerases break and reseal one of the strands of DNA, so that the unwound strands will not wind back. When the double stranded DNA is unwound upto a point, it shows a Y-shaped structure called Replication Fork. Enzyme DNA dependent DNA polymerase catalyses the joining of Deoxyribonucleotides (A, G, C and T) in the 5' – 3' direction. The enzyme forms one new strand in a continuous stretch (leading strand) in the 5' – 3' direction, on one of the template strands. On the other template strand, the enzyme forms short stretches (discontinuous) strand of DNA also in the 5' – 3'. The discontinuous fragments are later joined by DNA-ligase to form a leading strand. The two strands are held together by hydrogen bonds between nucleotides.

**Transcription:**

Transcription is the process by which DNA gives rise to RNA. It can also be defined as, the process of copying genetic information from one strand of the DNA into RNA is termed as Transcription.

***Transcription Unit:***

A transcription unit in DNA is defined primarily by the three regions in the DNA;

- A Promoter
- The Structural gene
- A Terminator

**Mechanism of Transcription:**

Transcription involves the binding of RNA-polymerase at the promoter site on DNA. As it moves along (through structural gene), the DNA unwinds and one of the two strands acts as template to synthesize a meaningful RNA and other strand act as non-coding. A complementary RNA strand is synthesized with A, U, C and G as bases. RNA synthesis is terminated when the RNA-polymerase falls off a Terminator sequence on the DNA.

**Transcription Unit and the Gene:**

A gene is defined as the functional unit of inheritance. In eukaryotes, DNA consists of both coding and non-coding sequences of nucleotides. The coding sequences / expressed sequences are defined as Exons. Exons are said to be those sequence that appear in mature / processed RNA. These exons are interrupted by non-coding sequences called Introns. These introns do not appear in mature RNA.

**Types of RNA:**

In prokaryotes, a single RNA polymerase enzyme (composed of different subunits) catalyses the synthesis of all types of RNA(mRNA, tRNA and rRNA) in bacteria.

Where as in eukaryotes, there are three different RNA polymerase enzymes I, II and III, they catalyse the synthesis of all types of RNA.

RNA polymerase I – rRNAs
RNA polymerase II - mRNA
RNA polymerase III – tRNA

**Process of transcription in Prokaryotes:**

RNA polymerase binds to promoter and initiates transcription. RNA polymerase associates with initiation factor and termination factor to initiate and terminate the transcription respectively. In prokaryotes, since the mRNA does not require any processing, the transcription and translation take place in the same compartment and can be coupled.

**Process of transcription in Eukaryotes:**

In eukaryotes, the primary RNA contains both the exons and introns and is non-functional. Hence, these non-coding introns will be removed by the process called Splicing. Then this mature RNA undergoes ***Capping*** (addition of unusual nucleotide methyl guanosine triphosphate at 5' –end) and ***Tailing*** (addition of adenylate residues at 3' –end). Now, this fully matured RNA will be transported out of the nucleus for translation.

**Genetic Code:**

Genetic code refers to the relationship between the sequence of nucleotides (nitrogen bases) on mRNA and the sequence of amino acids in proteins. Each code is known as Codon with three nucleotides (triplet). It has been deciphered by Nirenberg, Khorana, Severo Ochoa and Crick.

Salient features of Genetic code:

- The codon is triplet. 61 codons code for 20 different amino acids and 3 codons do not code for any amino acids, hence they function as Stop codons (UAG, UGA and UAA).
- One codon codes for only one amino acid, hence, it is unambiguous and specific.
- Some amino acids are coded by more than one codon, hence the code is degenerate.
- The codon is read in mRNA in a contiguous fashion. There are no punctuations.
- The code is nearly universal. For example, from bacteria to human, UUU would code for Phenylalanine (phe) amino acid.
- AUG has dual function. It codes for Methionine (met), and it also act as Initiator codon.

**Mutations and Genetic Code:**

Mutation caused due to insertion / deletion of single base pair is known as Point mutation. Effect of point mutations that inserts or deletes a base in structural gene can be better understood by following simple example;

Consider a statement that is made up of the following words each having three letters like genetic code;

**RAM HAS RED CAP**

If we insert a letter B in between HAS and RED and rearrange the statement, it would read as follows;

**RAM HAS BRE DCA P**

Similarly, if we now insert two letters at the same place, say BI'. Now it would read,

**RAM HAS BIR EDC AP**

Now we insert three letters together, say BIG, the statement would read,

**RAM HAS BIG RED CAP**

The conclusion is, insertion or deletion of one or two bases changes the reading frame from the point of insertion or deletion. Insertion or deletion of three or its multiple bases insert or delete one or multiple codon hence one or multiple amino acids, and reading frame remains unaltered from that point onwards. Such mutations are referred to as ***Frame-shift insertion or deletion mutations.***

**Structure of t-RNA : The Adapter Molecule:**

tRNA molecule appears like a clover leaf , but in actual structure, the tRNA is a compact molecule which looks like inverted L.

tRNA has three loops,

a) an anticodon loop that has bases complementary to the codon.
b) An amino acid accepter end to which it binds to amino acids.
c) Ribosomal binding loop.

tRNAs are specific for each amino acid. There are no tRNAs for stop codons.

**Translation:**

It refers to the process of polymerization of amino acids to form a polypeptide. The order and sequence of amino acids are defined by the sequence of bases in the mRNA. The amino acids are joined by a bond which is known as a peptide bond.

It involves four steps namely

- Activation of amino acids (charging of tRNA / aminoacylation of tRNA)
- Initiation of polypeptide synthesis
- Elongation of polypeptide synthesis
- Termination of polypeptide synthesis

a) *Activation of amino acids*: In this process, a particular amino acid becomes to a specific tRNA molecule.

b) *Initiation of polypeptide chain*: The initiator methionyl-tRNA charged with amino acid methionine and anticodon UAC interacts with the initiation codon by codon-anticodon interaction. With the initiator methionyl-tRNA at P site, the larger subunit binds to the smaller subunit, thus forming an initiation complex.

c) *Elongation of polypeptide chain*: A second tRNA charged with an appropriate amino acid enters the ribosome at the A site, close to the P site. A peptide bond is formed between the first amino acid and the second amino acid. Then the first tRNA is removed from the P-site and the second tRNA at the A site, now carrying a dipeptide, is pulled along with mRNA to the P-site (translocation). Now the A-site is occupied by a third codon and an appropriate aminoacyl tRNA will bind to it. This process of peptide bond formation and translocation will be repeated and the polypeptide chain grows in length.

d) *Termination of polypeptide chain*: When untranslated regions / termination codons come at the A-site, no amino acid would be added, as it is not recognized by any tRNA. So protein synthesis will stop. At the end, a release factor binds to the stop codon, terminating translation and releasing the complete polypeptide from the ribosome.

**Regulation of Gene Expression:**

All the genes are not needed constantly. The genes needed only sometimes are called regulatory genes and are made to function only when required and remain non-functional at other times. Such regulated genes, therefore required to be switched 'on' or 'off' when a particular function is to begin or stop.

**The Lac operon:**

Jacob and Monod (1961) proposed a model of gene regulation, known as operon model. Operon is a co-ordinated group of genes such as structural genes, operator genes, promoter genes, regulater genes and repressor which function or transcribed together and regulate a metabolic pathway as a unit.

There are three structural genes, lac Z, lac Y and lac A, coding for galactosidase, permease and transacetylase respectively. These three genes are controlled by a single switch called operator. The operator switch is controlled by the repressor protein which coded by the regulator gene.

When the repressor binds to the operator, the genes are not expressed (switched off). When the operator switch is on, the three structural genes transcribe a long polycistronic mRNA catalysed by RNA – polymerase.

A few molecules of lactose (inducer) enter the cell by the action o enzyme permease. They are converted into an active form of lactose which binds to the repressor and changes its configuration and prevents it from binding to the operator. Beta-galactosidase breaks lactose into glucose and galactose. (Fig. Text book p.117).

**Human Genome Project:**

**Goals of HGP:**

- Identify all the approximately 20,000-25,000 genes in human DNA;
- Determine the sequences of the 3 billion chemical base pairs that make up human DNA
- Store this information in databases;
- Improve tools for data analysis;
- Transfer related technologies to other sectors, such as industries.

**Methodologies:**

The methods involved two major approaches. One approach focused on identifying all the genes that expressed as RNA referred as ***Expressed Sequence Tags*** (ESTs). The other approach is blind approach of simply sequencing the whole set of genome that contained all the coding and non-coding sequence, and later assigning different regions in the regions in the sequence with functions, referred as ***Sequence Annotation.***

**Steps involved in sequencing:**

a) Isolation of total DNA from a cell and converted into random fragments.
b) Cloning of DNA fragments can be performed by using cloning vectors like BAC (Bacterial Artificial chromosomes) and YAC (yeast artificial chromosomes).
c) The fragments were sequenced using automated DNA sequencers that worked on the principle of a method developed by Frederick Sanger.
d) These sequences were then arranged based o some overlapping regions present in them.

**Salient features of Human Genome:**

a) The human genome contains 3164.7 million nucleotide bases.
b) The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million bases.
c) Less than 2 per cent of the genome codes for proteins.
d) Repeated sequences make up very large portion of the human genome.
e) Repetitive sequences are stretches of DNA sequences that are repeated many times, sometimes hundred to thousand times.
f) Chromosome 1 has most genes (2968), and the Y has the fewest (231).
g) Scientists have identified about 1.4 million locations where single base DNA differences (SNPs – single nucleotide polymorphism) occur in humans.

**DNA Fingerprinting:**

DNA fingerprinting involves identifying differences in some specific regions in DNA sequence called as repetitive DNA, because in these sequences, a small stretch of DNA is repeated many times. These repetitive DNA are 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 referred to as satellite DNA. These sequence

show high degree of polymorphism (variation at genetic level) and form the basis of DNA fingerprinting.

*Polymorphism* can be defined as, an inheritable mutation is observed in a population at high frequency, it is referred to as DNA polymorphism.

The technique of DNA fingerprinting was initially developed by Alec Jeffreys. He used a satellite DNA as probe that shows very high degree of polymorphism. It was called Variable Number of Tandem Repeats (VNTRs).

**Mechanism of DNA fingerprinting :**

***Extraction:*** DNA is extracted from the small amounts of blood, semen or hair bulbs available.
***Amplification:*** Many copies of this DNA are made by a technique called Polymerase Chain Reaction (PCR).
***Restriction Digestion:*** DNA is cut into desired reproducible segments using restriction enzymes.
***Separation***: These DNA sequences (restriction fragments) are separated by Gel Electrophoresis.
***Southern Blotting***: The separated DNA sequences are transferred from Gel onto a nitrocellulose membrane.
***Hybridisation*** with probe, the DNA sequence complementary to VNTR sequences.
Exposure of the membrane to X-ray film, whose specific bands are developed.

**Applications:**
It is used effectively in forensic science for identifying;

a) the biological father (in case of paternity disparity)
b) the criminals such as murderers and rapists.