COURSE: B Sc (Hons) ,Part -3

PAPER – 1- VIII

TOPIC- Genetics - 1

PREAPARED BY: Ajai Kishore Sharan

TOPICS COVERED

1 Structure of DNA and RNA

2. RNA

3. Genetic code

4. Chargaff rule

5. Polymerase Chain Reaction

6. Tools used in Genetic Engineering

TOPIC-1

STRUCTURE OF DNA AND RNA

Chemical Nature of DNA and RNA

The concept of Nucleic acid started during 1869 when Friedrick Miescher isolated a new molecule from the pus cells and called it as Nuclein. Miescher determined that chemically is made up of hydrogen, oxygen, nitrogen and phosporus. Miescher could also report an unique ratio of phosphorus of nitrogen in the Nuclein. In his further attempt Nuclein was also found in Salmon sperm. In 1889 Richard Altman discovered that Nuclein have acidic properties and hence the name was changed to Nucleic acid. In 1891 Albert Kossel discovered that Nuclein is made up of four bases and sugar molecule for which he got noble prize in medicine. In 1897 Eduard Zachans reported that Nuclein in an integral part of chromosome and his (1897) proposed the concept of chromatin with the chemical substance as Nuclein. Until 1940-1950, Nuclein as considered to be protein, until Avery et al (1944) and Hershey and chase (1952) proved that DNA is the genetic material in Escherichia coli (Bacteria) and T2 virus (A Bacteriophage of E.Coli).

Basic Structure of DNA and RNA

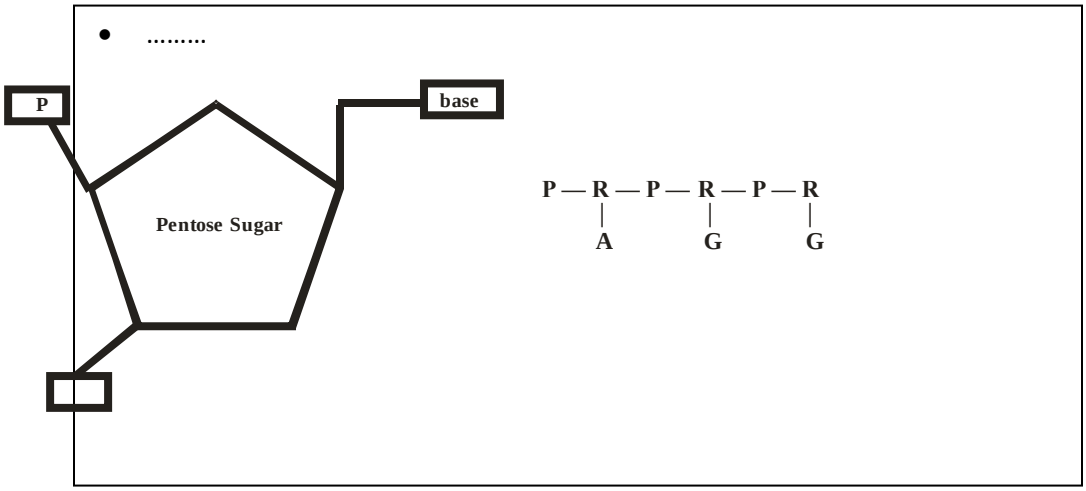
The basic structure of DNA and the RNA is the same and has the following structural configuration.

- Thread like in structure
- Made up of long chain of Polynucleotide.
- Each nucleotide consists of a nitrogen containing aromatic base attached to a pentose (five carbon) sugar, which in turn attached to a phosphate group.
- Each Nucleic acid contains four of five nitrogen base such as:

1. Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) – in DNA.
2. Adenine (A), Guanine (G), Cytosine (C) and Uracil (U) – in RNA.

- A and G are categorized as purines and C, T and U are collectively categorized as Pyrimidines.
- A, T, C, G are common for DNA A, U, C, G makes the RNA.
- The Pentose sugar of DNA differs from pentose sugar (Ribose) by the absence of a hydroxyl group (–OH) from carbon position two and hence the name deoxyribose.

• The Phosphate group connects to the sugar group by 5` – hydroxyl group (known as 5` prime end) and the 3` – hydroxyl group (known as 3` prime end), by two ester bonds called as phospho-di-ester bond. This phosho-di-ester bond is common for both DNA and RNA. The arrangement plan has been shown below



. **Structural plan of Nucleoside of DNA/RNA**

Chemical Components of Nucleotides in DNA and RNA

The Nucleotides of DNA and RNA are made up of Nitrogenous bases, Pentose sugar and Phospate group.

Biochemical Properties of DNA

Besides being the genetic material DNA has certain other roles to place, such as

Denaturation: During adoption of various biotechnological principles, denaturation or melting of DNA is required. The principle of DNA denaturation is adopted during hybridization of DNA with complementary strands of DNA, a process called as renaturation. This practice is often used during manipulation of DNA. In denaturation, the DNA is heated in a solution to break the hydrogen bond, and cooled again for renaturation. The thermal denaturation profile of DNA is often used for establishing a genetic homology between two genetically unrelated species. It also helps in determining A, T rich and G, C rich biological systems, because G, C rich system is often considered more advanced.

Methylation

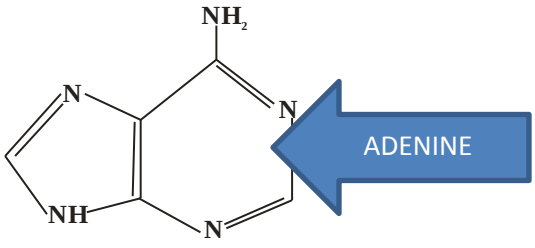
DN A molecules are prone to be affected during oxidation, ionizing radiation and also by carcinogens. Methylation can occur at two of the nucleotides, cytosine and adenine, when carbon molecule is replaced with CH3 molecule. This change leads to epigenetic change (not of heritable nature but acquired due to influence of environment). This change leads to perform various biological activity was hitherto, has not been carried by heitable genes. The methylation leads to modify the function of DNA. When present it acts to repress native gene of heritable change. DNA methyl transferase is the enzyme which performs this function. Methylation is the process which is concerned with reprogramming of gene regulation. Two of the DNA's four nucleotides, cytosine and adenine can be methylated. This rate of change differs in different plant and animal species. DNA methylation also forms the basis of chromatin structure which enables to single cell to grow into multiple organs or perform multiple functions. DNA methylation also plays important role in cancer development, cardio-vascular disease (antherosclerosis) and in aging. Methylation and demethylation occurs frequently in living system, which leads to abnormal activity of the cells.

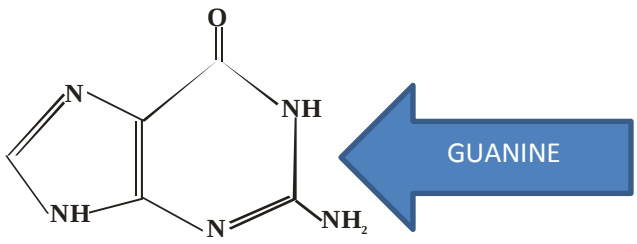
Mutation

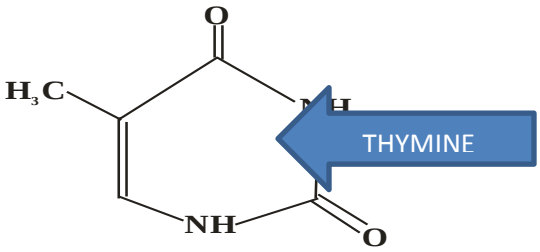
Chemical modification of DNA can lead to mutations in the genetic material. Cytosine can change to uracil which changes the genetic message (confined to Genetic Code). Many chemical and physical mutagens are known which can change the chemistry of genes and thus bringing about heritable changes.

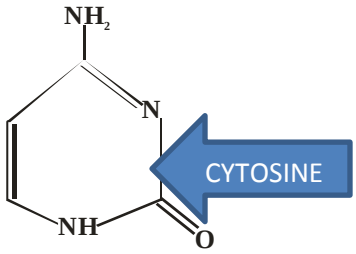
Chemistry of Coiling and Super Coiling of DNA

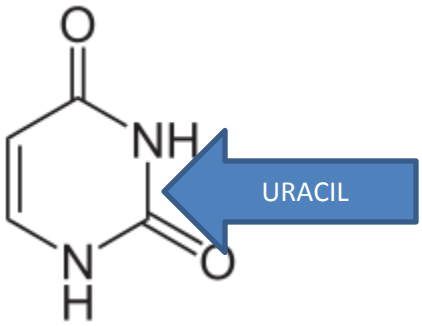
The human DNA can stretch to three metres if it is unfolded from the 23 pairs of chromosome. It has to be packed into a nucleus which is 10 micro-meters in diameter. The enzyme gyrases and topo isomerases catalyze the winding and relaxation of DNA. Holding the long stretch of DNA, both in Prokaryotes and in Eukaryotes, is the function of these two enzymes present closely associated to DNA molecules.

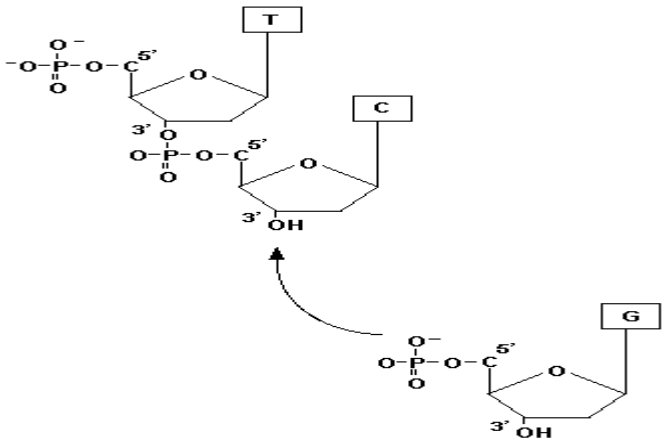



O
N
NH
NH
N
NH2
GUANINE


O
H3C
NH
O
THYMINE


NH2
N
NH
O
CYTOSINE


O
NH
N
H
O
URACIL



Phosphodiester bond connection between 5'prime and 3'prime of Pentose sugar

Three-dimensional Structure of DNA

In 1953, Watson and Crick, on the basis of data collected from X-ray crystallography and chargaff's observation proposed a structure of DNA. Watson and Crick proposed that in DNA.

- In DNA two stands of polynucleotide coil around each other, forming a double helix.
- The two strands run in opposite directions due to their orientation of the 5` and 3` posho diester bonds.
- The phosphate sugar association runs along the outside of the strands and the bases remain on the inside.
- The Nitrogenous bases remain bonded by hydrogen bonds, which forms a stable association.
- The normal DNA is called the B–DNA. This rotates along the axis in a right handed manner. The helix which is twined around each other takes a turn approximately at 10 base pairs.
- B-DNA has two principal grooves, a wide groove called major groove and a narrow groove called as minor groove.

A configurational change occurs in the structure of DNA, the data which could be collected from X-ray diffraction of the crystalline forms of DNA. The isolated DNA was converted into crystalline form for X-ray diffraction. Before conducting crystallography, the sample was hydrated and diffraction pattern was observed. The yielded two forms of pattern

named B-form and elongated microcrystals (in more drier samples) as A form. Continued analysis of fibre pattern of B and A form could reveal the following facts:

- That is B-form, the second carbon molecule (C2) of ribose sugar was out of plane. This could in influence the rotation of the helix around the main axis of DNA. This also creates a distance of 7A0 between two phosphate group located on 3` and 5` end.

- That in A DNA the third carbon molecule (C3) was found to be out of plane. This resulted into shortening of the sugar-phosphate back-bone. This also resulted into displacement of base pair and hence a wider helix was found. This distance between the two adjacent phosphate group is also reduced to 5.8 – 6A0.

X-ray diffraction technique for analyzing the structure of DNA remained as the only means to determine the structure of form of B-DNA and A-DNA. The chemical methods for synthesizing oligonucteolides became successful. Ris and Boom undertook a collaborative study to synthesize oligonucleodite using G – C base pairs (being more stable). A CGCGCG sequence of oligonucleotide was created a duplex DNA was created using this sequence which an unusual form was found which revealed.

- Left hander form of the double-helix.

- Two ant parallel chains held together A–T and G–C base pairs; but different from B–DNA.

- It is not stable and difficult to study.

- It was elongated and thinner molecule that had only one groove.

- This DNA was name as Z–DNA or Zigzag DNA.

- In B–DNA and A–DNA the associated substituent (atoms or group of atoms) are located in opposite directions. Whereas, in Z–DNA the substituents are located in the same direction resulting into thinning of the helix.

- It has sugar in C3 conformation (like A DNA) and Guanine base in same conformation close to each other) and hence different from A and B form.

- Duplex in Z–DNA has to accommodate the distortion of nucleotide G in the same conformation.

- The cytosine in the adjacent nucleotide of Z DNA is in the C-2 endo anti conformation.

Topic-2

RNA

Chemical Nature and Types of Ribonucleic Acid (RNA)

The nucleic acid in the cell has two important functions to perform. DNA is responsible for inheritance and its transfer to next generation; RNA is concerned with carrying out metabolic function. DNA can remain present in the cell, since birth. RNA has to be synthesized by DNA to carry out metabolic function. A sustainable life, appears to be interaction of DNA and RNA.

It is often postulated that life began as RNA molecule, concept becomes evident as some of the viruses (Influenza virus, foot and mouth virus, Rous sarcoma virus, Reovirus and Bacteriophage, Tobacco mosaic virus) contain RNA as genetic material.

Thus the RNA can be

1. Genetic
2. Non-Genetic

1. The Genetic RNA

The basic structure of RNA

Every RNA has two aspects of structure and function. The structural aspect is concerned with its unique assemblage of molecules to construct it, while the functional aspect deals with the modification of structure to perform an assigned job.

The RNA is a single stranded nucleic acid made up of four nucleotides; A, C, G and U joined together with a back alternating sequence of Phosphate and ribose sugar. It shows resemblance to DNA molecule so far as union of the three molecules is concerned. The Pentose sugar of RNA is Ribose sugar and the nitrogenous base Thymine of DNA is replaced by Uracil. It appears that functional DNA acquires the structure of RNA when functional aspect has to deal in the cell. By doing so the DNA (heritable material of cell) conserves itself. RNA is concerned with performing various metabolic functions of the cells, and hence acquires various forms and shapes to encounter various types of enzymes (acting as catalytic agent).

Types of RNA

The Non-genetic role of RNA are many, and the role with RNA has to play originates form the DNA. So, RNA is made by DNA by a process called as Transcription. The three RNA formed are

a. Messenger RNA (m RNA)

b. Ribosomal RNA (r RNA)

c. Transfer RNA (t RNA)

In order to remain functional, the cell, with the interaction of DNA and RNA has to perform many metabolic functions for growth and development. Various kinds of proteins have to be made to perform this function. The DNA transfers knowledge for synthesis of proteins and the three RNA molecules helps in effective performance of this activity. The three RNA molecules are assigned with three separate jobs, such as:

- The mRNA carries the information from the DNA.
- The tRNA is concerned with supply of amino acids and
- The rRNA serves as a factory for alignment of different amino acids for the synthesis of a specific protein.

Messenger RNA (mRNA)

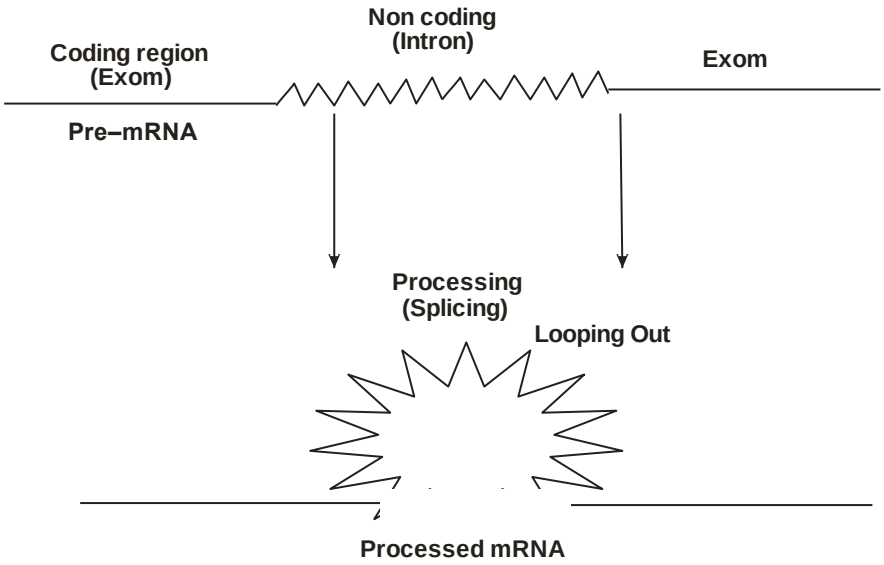
Messenger RNA is synthesized by DNA when metabolic activity of the cell begins. This RNA acquires information form DNA, in a coded language, regarding biochemical activity of the cell. This information can later on decode into a required protein. In Prokaryotes, m RNA's contain an exact transcribed copy of the DNA. This contains a terminal 5` end and a 3` end. In eukaryotes the terminal 5` end is further esterifies to form a cap. An additional segment of long adenosine residues (Poly A) are added enzymatic ally at the 3` end. This Poly A sequence is not encoded in the DNA. Due to cap and Poly A sequences, the eukaryotic RNA becomes more stable, the eukaryotic RNA becomes more stable, whereas, the prokaryotic mRNA is very short lived.

Processing of mRNA

Eukaryotic DNA is represented by functional DNA (required for the cell) called exon or coding sequence and intron (carried over DNA of the ancestors) called non-coding sequence.

The mRNA synthesized by this DNA, first forms a Pre-mRNA. This Pre-mRNA is processed during mRNA maturation, by a process called as splicing. The splicing is performed

by group of enzymes called splicesome. This splicing adopts the principle of cleavage (cutting of mRNA) and rejoining. This way the Pre-mRNA is processed into original mRNA, corresponding to the functional genes of the system.



Removal of Non coding sequence (intron) during processing of mRNA

rRNA

Mature rRNAs make up to 50-60 to of each ribosome. Some of the rRNA are purely structural, whereas others have catalytic activities to play. The eukaryotic ribosomes is composed of two sub units: a large sub-unit (60 S) and a small sub-unit (40S). The 60 S submit is composed of the 28 S rRNA, 5.8S rRNA, 5S rRNA and 50 proteins. The 40 S sub-units is composed of the 18S rRNA and 30 proteins. The bacterial (Prokaryotic) ribosome is composed of two similar sub units, with slightly different components. The bacterial large sub unit is called the 50 S sub-unit and is composed of the 23 S rRNA, 5 S rRNA and 31 proteins, while the bacterial small sub-unit is called 30S sub-unit and is composed of the 16S rRNA and 21 proteins.

The sub-units join to constitute the functional ribosome.

Processing of rRNA

In prokaryotes, which lack a nucleus, few rRNA genes helps synthesis of 50S, 30S sub-units as well as the protein sub-units. After synthesis of these two components of the ribosome,

they unite to form the small and the large sub-units of functional nature. However, in Eukaryotes, the 28S, 5.8S, and 18S are synthesized in the Nucleolus (within the nucleus) while the protein (50) is synthesized in the cytoplasm. This protein synthesized in cytoplasm is then transported to the nucleus for sub-assembly (large and small sub-units) in the nucleolus. After processing these sub-units are returned to the cytoplasm for final assembly.

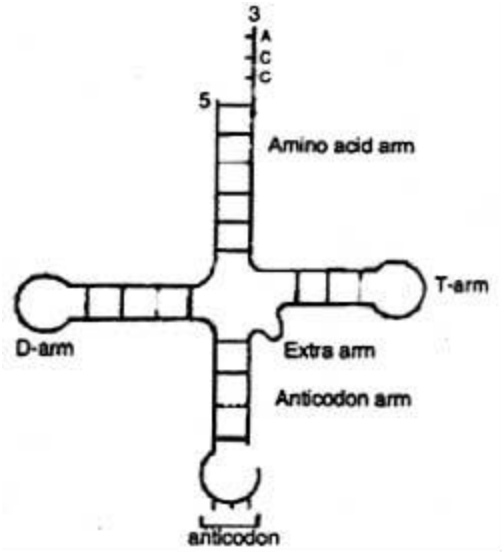
Role of 16S rRNA

A bacterial DNA sequence which is most amenable to change is that which encodes for 16S rRNA. This 16S RNA gene is present is all bacteria, and also in eukaryotes. Analysis of the 16S rRNA sequences from many organisms has revealed that some portions of the molecule undergo rapid genetic changes, this creates a background for distinguishing different species within the genus, other positions of gene change very slowly, and this facilitates various levels of taxonomic significance, and hence used in modern classification.

Structure of tRNA

tRNA is a small type of tRNA with a size around 4S, consisting of less than 70-80 nucleotides. There are 40 to 50 known types, which represents about 5% of overall molecules. Transfer RNA carries individual amino acids into the ribosome for assembly of amino acids to make protein. Specialized tRNA exists for each of the 20 amino acids needed for synthesis of protein. Sometimes more than one tRNA for each amino acid is present for example a total of 40 different t RNAS are used to translate 61codons provided by the t RNA. Hence, t RNA has unique property to change its nucleotide sequence to accommodate all the 61 codons. The union of amino acid with tRNA is mediated by specialized enzymes called aminoacyl tRNA synthetases, usually one synthetase for each amino acid.

Each tRNA binds to a specific amino acid and transfers it to the ribosome. Mature to RNA made up of 70-80 nucleotides acquires a three dimensional form in a such a way that the position of the amino acid binding site is at one end and the anticodon an unpaired loop of nucleotides at the other. The anticodon is a three nucleotide sequence, unique to each different tRNA, which interacts with a mRNA by forming a complementary base pair.



Sructure of tRNA

Processing of tRNA

tRNA molecules are synthesized in the cell (synthesized in nucleus and transported to cytoplasm in Eukaryotes but in the cytoplasm in Prokaryotes) as Pre-tRNA molecule. This Pre-tRNA requires multiple processing steps before the mature t-RNA is formed for effective translation. This processing is less frequent in Prokaryote as Pre-tRNA is transcribed as a single RNA moiety. The first step involves digestion of the RNA to release individual Pre tRNA. The process by which Pre-tRNA get converted to mature tRNA involves following steps.

1. The 5` end of Pre-tRNA, called the 5` leader sequence, is cleaved off.

2. The 3` end of the Pre-tRNA is cleaved off.

3. In Eukaryotes and many of the bacterial system, a sequence of CCA nucleotide is added to the 3` end. This CCA sequence at the mature tRNA is the site at which the amino acids are added.

4. Many nucleotides in the Pre-tRNA are chemically modified by altering their nitrogen bases. About 12 nucleotides are modified, as a result of which adenine (A) is modified to

Pseudouridine , adenine is also modified to inosine (1). In the same way Uridine may modified to dihydrouridine.

5. In eukaryotes the pre-tRNA have introns (non-coding regions) which get spliced out during processing.

The mature tRNA with the help of enzyme, aminoacyl tRNA gets attached to the specific amino acid, henceforth called as charging of tRNA. All tRNA acquire the same structural configuration because they have to interact with ribosome on the same site.

Some Other Forms of RNA

Ribozyme

In 1989, Thomas cech and Sidney Altman shared the Nobel Prize, for discovery of certain RNA which solely acts as enzyme, called as Ribozymes. The enzymes activity of Ribozyme was noticed in Protozoans, when to make the rRNA functional, some part of the rRNA (non-coding sequences or intron) could be removed itself from the rRNA molecule. This class of enzyme was named as Ribozyme due to its catalytic function which other-wise is the property of enzyme (proteins).

Antisense RNAS

In eukaryotes, the DNA synthesizes RNA having codons for protein synthesis. These codons are ultimately translated into respective proteins and hence named as sense codons. Sometimes the modified derivatives of RNA or DNA forms a sequence which is complementary to mRNA which blocks the synthesis of translated protein, and hence called as Antisense RNA (Opposite to RNA). The antisense RNA can block the synthesis of translation of protein by forming a double-stranded structure. Antisense RNA can be synthesized artificially and can be introduced in the cell to block the functioning of disease prone mRNA sequence. This RNA has thus great therapeutic value.

Topic-3

GENETIC CODE

Cell is the structural and functional unit. Whereas, the structure is made by unique assembly of different and complex compounds, the function is carrying out of the metabolic processes in order to provide inputs for life. Both the processes are the adoptions of guided principle provided by nature. When a cell is structurally formed its next job is to carry out metabolic process. Making of a cell and its functional attribute is carrying out of complex chemical reaction in which several new chemical molecules are synthesized and broken to maintain the life. It has often been said that a cell is a chemical factory which requires various inputs to function. DNA is a genetic material which is the key molecule guiding the process of inheritance. One more function which is attributed to this DNA is that it is the carrier of genetic information. This information is being carried from cell to cell and from generation to generation. The manner in which the genetic information exists in DNA molecule is chemical. So, it is necessary to understand the chemical nature of DNA before understanding Genetic Code.

A DNA molecule is composed of three kinds of moieties:

(i) Phosphoric acid

(ii) Deoxyribose sugar and

(iii) Nitrogen bases

The genetic information may be written in any one of the three moieties of DNA. On close examination of DNA molecule and its union during making it appears that sugar (Deoxyribose) and phosphate (Phosphoric acid) the two structural component of DNA is similar in all the nucleotides and hence it appears unlikely that these moieties of Sugar-phosphate carry the genetic information. Option falls upon nitrogen bases, the nature of which varies from one nucleotide to another. It appeared that genetic information might well depend on their sequences. The sequences of nitrogen bases (A, T, C, G) of a given segment of DNA molecule, actually has been found to be identical to linear sequence of amino acids in a protein molecule. The first proof of this collinearity between DNA nitrogen bases and amino acid sequence in protein molecule was first obtained during assembly of head protein of bacteriophage T4 and in *Escherichia coli*. The collinearity of amino acid molecule and DNA nucleotides has given the clue that the specific arrangement of four nitrogen based (e.g. A, T, C and G) in DNA polynucleotide chains determines the sequence of amino acids in protein molecules. Hence, it

was proposed that these four DNA bases can be considered as four alphabets of DNA molecule concerned with coded language of DNA. It was further suggested by molecular biologists that using alphabets as the genetic information may invite complex system of adoption of principles

of english grammar and hence it appears that genetic information's existed in DNA molecule in the form of certain language of code words. These code words utilize the services of four nitrogen bases of DNA for its symbols.

Assignment of code to amino acid

The basic problem of a four lettered genetic code is to indicate now information written in a four letter language (A, T, C, G) can be translated into twenty letter language (twenty amino acids of proteins). The simplest code is a singlet code (a code of single letter) in which one nucleotide codes for one amino acid. Such a code is inadequate for twenty amino acids as only four amino acid will be utilized. A doubled code is also inadequate because it could specify only sixteen (4 X 4) amino acids.

	A	G	C	U
A	AA	AG	AC	AU
G	GA	GG	GC	GU
C	CA	CG	CC	CU
U	UA	UG	UC	UU

A triplet code could specify sixty four (4X4X4) amino acids. There it is likely that there may be 64 triplet codes for 20 amino acids.

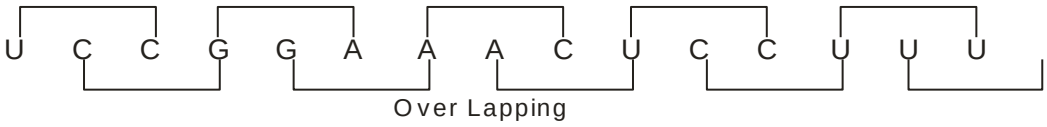
FIRST LETTER ↓	Second Letter				
	U	C	A	G	3rd Letter ↓
U	PHE	SER	TYR	CYS	U
	PHE	SER	TYR	CYS	C
	LEU	SER	**STOP**	**STOP**	A
	LEU	SER	**STOP**	TRY	G
C	LEU	PRO	HIS	ARG	U
	LEU	PRO	HIS	ARG	C
	LEU	PRO	GLN	ARG	A
	LEU	PRO	GLN	ARG	G
A	ISO	THR	ASN	SER	U
	ISO	THR	ASN	SER	C
	ISO	THR	LYS	ARG	A
	MET	THR	LYS	ARG	G
G	VAL	ALA	ASP	GLY	U
	VAL	ALA	ASP	GLY	C
	VAL	ALA	GLU	GLY	A
	VAL	ALA	GLU	GLY	G

The first evidence to support the concept of triplet code is provided by crick and his coworkers in 1961. When they added or deleted single or double base pairs in a bacteriophage T4 infecting E. coli, bacteriophages ceased to perform normal functions. However, bacteriophages with addition or deletion of three base pairs in DNA molecule had performed normal functions. The experiment carried out by crick proposed that genetic code is in triplet form because addition of one or two nucleotides has put the reading of the code out of order, while addition of third nucleotide resulted in a return to the proper reading message. The group of nucleotides that specifies one amino acid is code word or codon.

Experimental proof that genetic code is triplet

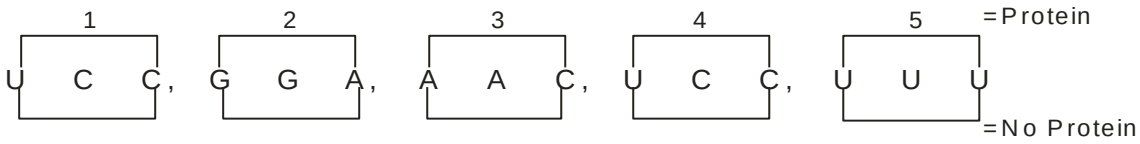
Three hypotheses were put forward as a perspective idea to prove that genetic code is triplet in nature.

(i) The genetic code is overlapping



This hypothesis was negated as this can cause statistical regularities between adjacent acid residues. A single point mutation may lead to change in two adjacent residues. This possibility was eliminated.

(ii) Punctuated code



(iii) Unpunctuated code

Deletions of three nucleotides (or multiple of there) will restore the reading frame. This is the actual form of the code.

Most of our current knowledge of the general nature of the genetic code and the nucleotide composition has been obtained from the four main types of experimental approaches.

A. Polyuridytic acid method

Initial breakthrough in deciphering of the genetic code was achieved by Marshall Nierenberg and Heinrich matthaei in 1961. They used enzyme polynucleotide phosphorylase to construct a synthetic polyribonucleotide containing only one kind of base (Uracil – found in mRNA) and added it to a cell free amino-acid-incorporating system from *E. coli*. A polyuracil (UUU.....UUU) was prepared and to which they added different amino acids labelled with C^{14}. Hence, this poly U system specifically stimulated the formation of a polypeptide which contained only the amino acid phenylalanine. This was concluded that code word for the amino acid phyenylalanine was a sequence of three uracil nucleotides (UUU). Nirenberg influenced by the success of this work got deeply involved in this problem with Severo Ochoa to assign codon for lysine as AAA (poly A) and CCC (poly C) for proline. Attempt to assign poly G as a codon for specific amino acid failed, because GGG soon acquires a secondary structure and cannot attach to ribosomes.

B. Copolymer method

In this method, Nirenberg used mixtures of two or more ribonucleotides diphosphates and with the help of enzyme phosphorylase, prepared polyribonucleotides. Using UDP (uracil diphosphate) and cytosine diphosphate (CDP) in the ratio 3:1, he obtained a polynucleotide which contained the triplets UUU, UUC, UCU and CUU. The triplets containing 2Cs and 1U

were the least frequent. In such a poly-UC system, Nirenberg obtained polypeptide containing the amino acids phenylalanine and serine in the ratio of 3:1. This method, however, did not give the exact sequence of three bases.

C. Binding technique

Marshall W. Nirenberg and Philip Leeder in 1964 found that if a synthetic trinucleotide for a known sequence is used with ribosome and a particular aminocy–tRNA (t–RNA having its own specific amino acid attached), a complex will be formed provided the used codon codes for the amino acid attached to the given aminoacyl tRNA.

Codon + Ribosome + Amino acid (AA) + t RNA =- Ribosome-codon-AA-t RNA (complex)

Whereas, AA-tRNA can pass through nitrocellulose membrane, the ribosome - Codon - AA - t RNA complex adsorbs to the membrane. If a mixture only one of the amino acids is made radioactive, then the presence or absence of the radioactivity on the nitrocellulose membrane will show whether there is a relationship between the Codon and the amino acid which was made radioactive. Such a treatment was given by Nirenberg and his coworkers to all the 64 synthetic codons and their respective amino acids were identified. The binding of AA-tRNA was not equally efficient in all cases; hence, sequences of bases in only 45 cordons could be worked out by this method.

4. Repetitive sequencing technique

This method of confirming the genetic code was proposed by Har Gobind Khorana and is considered to be the most authentic one. The method involves *in vitro* (chemical synthesis) of short segments of DNA of known base sequence with the help of enzyme DNA polymerase. From this DNA, RNA of defined base sequence is transcribed with the help of enzyme RNA polymerase. The RNA sequentially helps in the synthesis of a Polypeptide (A chain of amino acid). The principle of central Dogma is thus followed as described below: The chain of amino acid thus formed is then determined and correlated with the base sequence of DNA and RNA. Similar type of sequence of DNA thus synthesized same sequence of amino acid. Khorana, developed a method for synthesizing polyribonucleotide's with different but repeating sequences to obtain serine, arginine and glutamic acid. Hence determined the importance of repeating units of short segments of DNA and suggested that codon AAG codes for lysine, AGA for arginine and GAA for glutamic acid. The study made by Khorana thus proved that:

1. The base sequence in DNA specifies the amino acids in proteins.
2. The information contained in DNA is conveyed through RNA.
3. The genetic code is triplet is literature, and is also non-overlapping in nature.

Properties and Characteristics of the genetic code

By various experimental evidences the characteristic properties of genetic code have been established as described below:

1. The genetic code in literature is triplet

The coded language of DNA for transcription is a three lettered word. There are altogether 64 codons, for 20 amino acids. The codon assignment for different amino and has been shown below:

No. of Codons	Amino acid	Total
One Codon each	Methione, Tryptophan	02
Two Codons each	Aspargine, Asparatic acid, cysleine, Glutamine, Glutamic acid, Histidine, Lysine, Phenylalanine, Tyrosine	18
Three Codons each	Isoleucine	03
Four Codons each	Alanine, Glycine, Proline, Threonine, Valine	20
Six Codons each	Arginine, Leucine, Serine	18
Terminator Codons		03

The code dictionary reveals certain trend of patterns

- Amino acids with similar structural properties have related codons. The difference lies in third codon. e.g. asparatic acid has a codon GAU whereas, glutamic acid has a codon GAA.

- All codons with U in the second position specify hydrophobic amino acids (Isoleucine, Leucine, Methionine, Phenylalanine and Valine).

- All the acidic (Asparagine and Glutamic acid) and basic (Arginine and Lysine) amino acids have A or G as the second base.

2. Degenerate

A code in which several code words have the same meaning is known as degenerate. The genetic code is degenerate because there are many instances in which different codons specify the same amino acid. A genetic code in which some amino acids may each be encoded by more than one codon is therefore known as degenerate. For example two codons for 18 aminoacids,four codons for 20 amino acids.

3. Non overlapping

The genetic code is non overlapping, this means that adjacent codons do not overlap. In actual practice, six based code for not more than two amino acids for example UUUCCC on mRNA will code only for 2 amino acids i.e. Phenylalanine (UUU) and Proline (CCC).

4. Comma less

There is no comma (,) at the end of one codon and the begining of the next. This means that after one amino acid is coded, the second amino acid will automatically coded by the next three letters.

5. Non-ambiguity

Non-ambiguity of the codon means, a particular codon will always code for the same amino acid.

6. Universality

The genetic code is same for the entire living organism. This means that same sequence of 3 bases encode for same amino acids in all life forms. Thus, UUU codes for phenylalanine and GUC for valine in all living things. The genetic code which was developed in the bacteria about 3 billion (300 crore) years ago has not undergone any change and has been preserved in its almost original form in the course of evolution.

7. Polarity

The genetic code has polarity (direction), that is, the code is always read in a fixed direction i.e. in the 5` 3` direction.

5`UUGAUC GUC UCG3`

Leu Ile Val Serine

8. Chain initiation codons

The triplet codons AUG and GUG play vital role in *Escherichia coli* when these codons occur in between the two ends of a functional gene (cistron) they code for the amino acid methionine and valine. When they occur immediately after a terminaton codon (stop codon) they act as chain initiation or starter codons for the synthesis of a polypeptide chain (chain of amino acids). The codon AUG which codes for amino acid have two functions to perform:

1. When initiates protein synthesis it acts as formylated methionine but when lies in the intermediate position.

2. Hence there is a separate species of tRNA for initiating methionine called as tRNA met and intermediate methionine.

9. Chain termination codon

The triplets UAA, UAG and UGA do not recognize any amino acid. So, they are called as non-sense codons, as against the remaining 61 codons, which are termed as sense codons. These three codons perform the function of punctuating genetic message like a full stop at the end of the sentence. The codons UAA and UAG were discovered in bacteria and were respectively associated with the ochre and amber mutations. The codon UGA is also called as opal. The termination codon causes the release (break) of the polypeptide chain from the ribosome hence the name terminator.

Multiple recognition of codons and Wobble hypothesis

Crick (1966) proposed the 'wobble hypothesis' to explain the degeneracy of the genetic code. He was of the opinions that except for tryptophan and methionine, more than one codon direct the synthesis of one amino acid. There are 61 codons that synthesize amino acids, therefore, there must be 61 tRNAs each having different anticodons. But the total number of tRNAs is less than 61.This may be explained if it is considered that the anticodons of some tRNA read more than one codon. In addition, identity of the third codon seems to be unimportant. For example CGU,

CGC, CGA and CGG all code for arginine. It appears that CG specifies arginine and the third letter is not important. Conventionally, the codons are written from 5′ end to 3′ end. Therefore, the first and second bases specify amino acids in some cases. According to the Wobble hypothesis, only the first and second bases of the triple codon on 5′ → '3 mRNA pair with the bases of the anticodon of tRNA i.e A with U, or G with C.The pairing of the third base varies according to the base at this position, for example G may pair with U. The conventional pairing (A = U, G = C) is known as Watson-Crick pairing and the second abnormal pairing is called wobble pairing.

This was observed from the discovery that the anticodon of yeast alanine-tRNA contains the nucleoside inosine (a deamination product of adenosine) in the first position (5′ → 3′) that paired with the third base of the codon (5′ → 3′). Inosine was also found at the first position in other tRNAs e.g. isoleucine and serine.

The purine, inosine, is a wobble nucleotide and is similar to guanine which normally pairs with A, U and C. For example a glycine-tRNA with anticodon 5′-ICC-3′ will pair with glycine codons GGU, GGC, GGA and GGG (Fig 7.2). Similarly, a seryl-tRNA with anticodon 5′-IGA-3′ pairs with serine codons UCC, UCU and UCA (5-3′). The U at the wobble position will be able to pair with an adenine or a guanine.

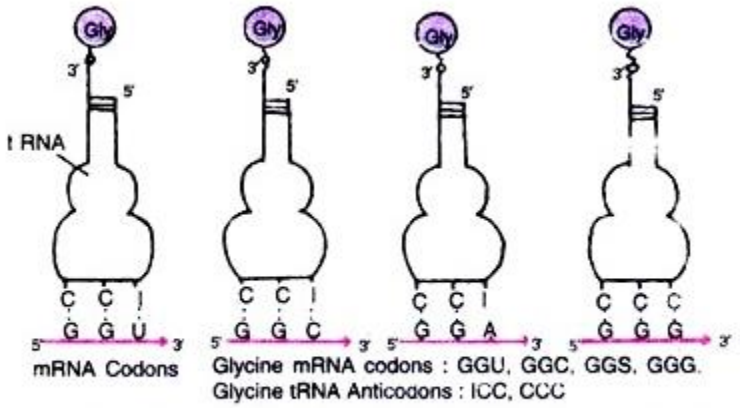
Due to the Wobble base pairing one tRNA becomes able to recognise more than one codons for an individual amino acid. By direct sequence of several tRNA molecules, the wobble hypothesis is confirmed which explains the pattern of redundancy in genetic code in some anticodons (e.g. the anticodons containing U, I and G in the first position in 5'→ 3′ direction)

Three types of wobble pairings have been proposed:

(i) U in the wobble position of the tRNA anticodon pairs with A or G of codon,

(ii) G pairs with U or C, and

(iii) 1 pairs with A, U or C.



Wobble pairing of one glycine tRNA with three codons of mRNA due to Wobble in 5'→3' direction.

Topic-4

DNA and Chargaff's rule

Avery and his associates proved that DNA is the genetic material. This opened up new scope of research in DNA biology. Several DNA samples were collected and char gaff and his associates proceeded to separate DNA. It was primarily hydrolyzed with acid, which would break the phosphodiester bonds to separate the components of DNA. The technique of paper chromatography and UV- spectrophotometer was used on Nitrogenous bases. During this it was observed that the purine bases (Adenine and Guanine) and pyrimidine bases (cytosine and Thymine) are not in equal amounts.

Chargaff further observed that GC equals to the concentration of AT. However, in RNA, Thymine is replaced with Uracil. Chargaff observed that the amount of certain bases were equal to each other. The base Guanine was equivalent to the amount of cytosine and the same held true for Adenine and Thymine. This means:

Total Amount of Purine = Total Amount of Pyrimidine

A + G = T + G

The ratio of A : T = 1

The ratio of G : C = 1

A is parried to thymine by 2 hydroxyl bonds and C to G by 3 hydrogen bonds.

Chargaff's conclusion of the significance of A, T, C and G, revolutionized the biological sciences. This helped James D. Watson and Francis crick in determination of structure of DNA. The idea that Adenine and Thymine always exist in the same proportion, they must remain bonded, and the same is true for Cytosine and Guanine. So, a proposal that DNA is a double helical structure was established.

Topic-5

POLYMERASE CHAIN REACTION(PCR)

Kary Banks Mullis (December 28, 1944 – August 7, 2019) was an American biochemist. His invention became a central technique in biochemistry and molecular biology, described by The New York Times as "highly original and significant, virtually dividing biology into the two epochs of before PCR and after PCR."

A stepping stone in the field of molecular genetics was laid by James D. Watson and Francis Crick in the year 1953 by proposing a double helix structural model of DNA (Watson and Crick, 1953). In the early 1960's, significant advances were made in elucidation of the genetic code and synthetic oligonucleotides were used as primer templates for DNA polymerase by Dr. H.

Gobind Khorana, for which he was also awarded the Nobel prize in the year 1968 . In1971, Kjell Kleppe, a researcher in Khorana's laboratory, described the replication of a segment of DNA by a two-primer system (Kleppe et al., 1971).Polymerase chain reaction is an *in vitro* technique that enables replication and amplification of a DNA sequence to billions fold amplitude. The technique of polymerase chain reaction (PCR) was invented by Kary Banks Mullis in the year 1983, while he was working as a biotechnologist in Cetus Corporation, Emeryville, California, USA. In 1985, a joint venture was established between Cetus Corporation and Perkin-Elmer, another US based Biotech Company to design thermal cycler instruments and reagents for PCR and in 1987, a press release announced the availability of the "PCR-1000 Thermal Cycler" and "AmpliTaq DNA Polymerase" commercially. The invention won him laurels of the Nobel Prize in Chemistry as well as the Japan Prize in the year 1993.

The following is a typical PCR thermocycler profile:

Initialization.

1 Denaturation (repeated 15-40 times)

2 Annealing (repeated 15-40 times)

3 Elongation or Extension (repeated 15-40 times)

4 Step 2-4 are then repeated 15-40 times.

5 Final elongation.

Polymerase chain reaction (PCR) is a common laboratory technique used to make many copies (millions or billions!) of a particular region of DNA. The goal of PCR is to make several copies

of the target DNA region that it can be analyzed or used in some other way. For instance, DNA used for sequencing, for making a DNA profile using gel electrophoresis, or for cloned into a vector like a plasmid for further experiments.

PCR is used in many areas of biology and medicine, including molecular biology research, medical diagnostics, and even some branches of ecology.

Taq polymerase

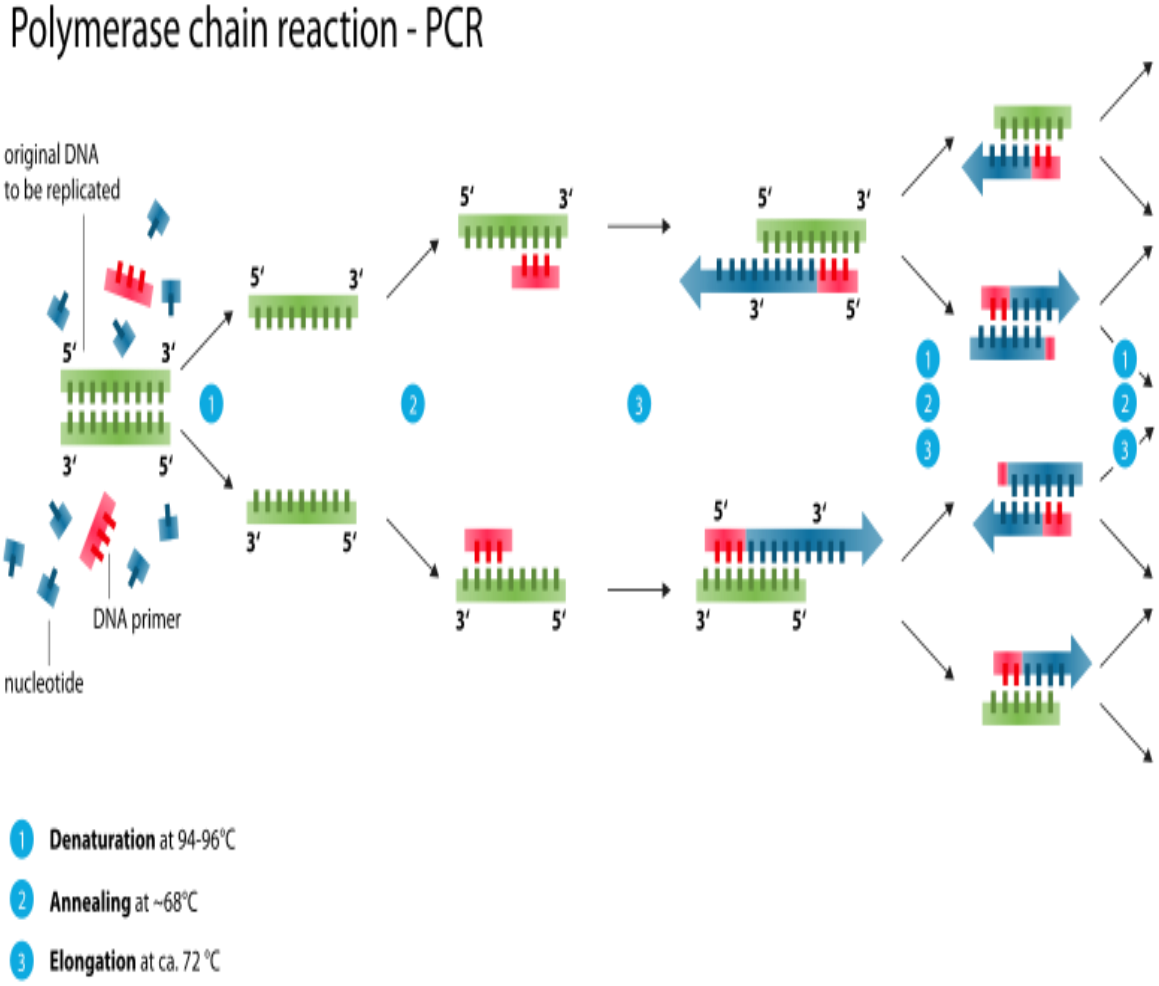
PCR technique for DNA amplification requires occasional heating of the DNA and also cooling. During heating in the heating mixture the DNA strands get separated to become single stranded. In this single stranded DNA Primer is added to make it a double stranded structure. This requires involvement of DNA polymerase enzyme. Most of the DNA polymerase are temperature sensitive to overcome this problem a DNA polymerase enzyme which can stand high temperature is used. This enzyme is obtained from a heat-tolerant bacterium Thermus aquaticus from which it was isolated. Heat tolerant enzyme DNA polymerase can be obtained from this bacterium.

T. aquaticus lives in hot springs and hydrothermal vents. Its DNA polymerase is very heat-stable and is most active around 70 ° (a temperature at which a human or *E. coli* DNA polymerase would be nonfunctional). This heat-stability makes Taq polymerase ideal for PCR. As we'll see, high temperature is used repeatedly in PCR to denature the template DNA, or separate its strands.

PCR primers

Like other DNA polymerases, Taq polymerase can only make DNA if it's given a primer, a short sequence of nucleotides that provides a starting point for DNA synthesis. In a PCR reaction, the experimenter determines the region of DNA that will be copied, or amplified, by the primers she or he chooses.

PCR primers are short pieces of single-stranded DNA, usually around 202020 nucleotides in length. Two primers are used in each PCR reaction, and they are designed so that they flank the target region (region that should be copied). That is, they are given sequences that will make them bind to opposite strands of the template DNA, just at the edges of the region to be copied. The primers bind to the template by complementary base pairing.


Polymerase chain reaction - PCR
original DNA
to be replicated
DNA primer
nucleotide
5′
3′
1
2
3
Denaturation at 94-96°C
Annealing at ~68°C
Elongation at ca. 72 °C

TOPIC- 6

TOOLS USED IN GENETIC ENGINEERING

The term genetic engineering initially referred to various techniques used for the modification or manipulation of organisms through the processes of heredity and reproduction. As such, the term embraced both artificial selection and all the interventions of biomedical techniques, among them artificial insemination, in vitro fertilization (e.g., "test-tube" babies), cloning, and gene manipulation. In the latter part of the 20th century, however, the term came to refer more specifically to methods of recombinant DNA technology (or gene cloning), in which DNA molecules from two or more sources are combined either within cells or *in vitro* and are then inserted into host organisms in which they are able to propagate.

The possibility for recombinant DNA technology emerged with the discovery of restriction enzymes in 1968 by Swiss microbiologist Werner Arber. The following year American microbiologist Hamilton O. Smith purified so-called type II restriction enzymes, which were found to be essential to genetic engineering for their ability to cleave a specific site within the DNA (as opposed to type I restriction enzymes, which cleave DNA at random sites). Drawing on Smith's work, American molecular biologist Daniel Nathans helped advance the technique of DNA recombination in 1970–71 and demonstrated that type II enzymes could be useful in genetic studies. Genetic engineering based on recombination was pioneered in 1973 by American biochemists Stanley N. Cohen and Herbert W. Boyer, who were among the first to cut DNA into fragments, rejoin different fragments, and insert the new genes into *E. coli* bacteria, which then reproduced.

DNA technology utilizes a number of biological tools to achieve its objectives, most important of them being the enzymes.

Important biological tools for rec DNA technology are:

(A) Enzymes:

a. Restriction Endonucleases

b. Exonucleases

c. DNA ligases

d. DNA polymerase

(B) Cloning Vector

(C) Host organism

(D) DNA insert or foreign DNA

(E) Linker and adaptor sequences.

An account of all these biological tools of genetic engineering is given below:

(A) ENZYMES:

A number of specific enzymes are utilized to achieve the objectives of rec DNA technology.

The enzymology of genetic engineering includes the following types of enzymes:

(a) Restriction Endonuclease:

These enzymes serve as important tools to cut DNA molecules at specific sites, which is the basic need for rec DNA technology.

These are the enzymes that produce internal cuts (cleavage) in the strands of DNA, only within or near some specific sites called recognition sites/recognition sequences/ restriction sites 01 target sites. Such recognition sequences are specific for each restriction enzyme. Restriction endonuclease enzymes are the first necessity for rec DNA technology.

The presence of restriction enzymes was first of all reported by W. Arber in the year 1962. He found that when the DNA of a phage was introduced into a host bacterium, it was fragmented into small pieces. This led him to postulate the presence of restriction enzymes. The first true restriction endonuclease was isolated in 1970s from the bacterium E. coli by Meselson and Yuan.

Another important breakthrough was the discovery of restriction enzyme Hind-II in 1970s by Kelly, Smith and Nathans. They isolated it from -the bacterium Haemophilus influenza. In the year 1978, the Nobel Prize for Physiology and Medicine was given to Smith, Arber and Nathans for the discovery of endonucleases.

Types of Restriction Endonucleases:

There are 3 main categories of restriction endonuclease enzymes:

Type-I Restriction Endonucleases

Type-II Restriction Endonucleases

Type-III Restriction Endonucleases

Type-I Restriction Endonucleases:

These are the complex type of endonucleases which cleave only one strand of DNA. These enzymes have the recognition sequences of about 15 bp length .

Type-II Restriction Endonucleases:

These are the most important endonucleases for gene cloning and hence for rec DNA technology. These enzymes are most stable. They show cleavage only at specific sites and therefore they produce the DNA fragments of a defined length. These enzymes show cleavage in both the strands of DNA.Such enzymes are advantageous because they don't require ATP for cleavage and they cause cleavage in both strands of DNA. Only Type II Restriction Endonucleases are used for gene cloning due to their suitability. The recognition sequences for Type-II Restriction Endonuclease enzymes are in the form of palindromic sequences with rotational symmetry, i.e., the base sequence .n the first half of one strand of DNA is the mirror image of the second half of other strand of that DNA double helix . Important examples of Type-II Restriction endonucleases include HindIII, EcoRI, PvuII, Alul, Haelll etc.

Type-III Restriction Endonucleases:

These are not used for gene cloning. They are the intermediate enzymes between Type-I and Type-II restriction endonuclease. They require Mg++ ions and ATP for cleavage and they cleave the DNA at well-defined sites in the immediate vicinity of recognition sequences, e.g. Hind III, etc.

Nature of cleavage by Restriction Endonucleases:

The nature of cleavage produced by a restriction endonuclease is of considerable importance.

(b) Exonucleases:

Exonuclease is an enzyme that removes nucleotides from the ends of a nucleic acid molecule. An exonuclease removes nucleotide from the 5′ or 3′ end of a DNA molecule. An exonuclease never produces internal cuts in DNA.

(c) DNA ligase:

The function of these enzymes is to join two fragments of DNA by synthesizing the phosphodiester bond. They function to repair the single stranded nicks in DNA double helix and in rec DNA technology they are employed for sealing the nicks between adjacent nucleotides. This enzyme is also termed as molecular glue.

(d) DNA polymerases:

These are the enzymes which synthesize a new complementary DNA strand of an existing DNA or RNA template. A few important types of DNA polymerases are used routinely in genetic engineering. One such enzyme is DNA polymerase which, prepared from *E coli*. The Klenow fragment of DNA polymerase-I .is employed to make the protruding ends double-stranded by extension of the shorter strand. Another type of DNA polymerase used in genetic engineering is Taq DNA polymerase which is used in PCR (Polymerase Chain Reaction).

Reverse transcriptase is also an important type of DNA polymerase enzyme for genetic engineering. It uses RNA as a template for synthesizing a new DNA strand called as cDNA

(Complementary DNA). Its main use is in the formation of cDNA libraries. Apart from all these above mentioned enzymes, a few other enzymes also mark their importance in genetic engineering.

Genetic engineering has advanced the understanding of many theoretical and practical aspects of gene function and organization. Through recombinant DNA techniques, bacteria have been created that are capable of synthesizing human insulin, human growth hormone, alpha interferon, a hepatitis B vaccine, and other medically useful substances. Plants may be genetically adjusted to enable them to fix nitrogen, and genetic diseases can possibly be corrected by replacing dysfunctional genes with normally functioning genes. Nevertheless, special concern has been focused on such achievements for fear that they might result in the introduction of unfavourable and possibly dangerous traits into microorganisms that were previously free of them—e.g., resistance to antibiotics, production of toxins, or a tendency to cause disease. Likewise, the application of gene editing in humans has raised ethical concerns, particularly regarding its potential use to alter traits such as intelligence and beauty.