

# Information Transfer Purpose

## SYNOPSIS

This chapter will cover the following topics

1. Molecular basis of coding and decoding genetic information
2. Molecular basis of information transfer
3. DNA as a genetic material
4. Hierarchy of DNA structure—from single stranded to double helix to nucleosomes
5. Concept of genetic code
6. Universality and degeneracy of genetic code
7. Gene in terms of recombination

## DNA: A DOUBLE STRANDED HELIX

After most biologists became convinced that DNA was the genetic material, a race was on to determine how the structure of this molecule could account for its role in heredity. By the beginning of the 1950, arrangement of covalent bonds in a nucleic acid polymer was well established and researchers focused on discovering the three-dimensional structure of DNA. Among the many scientists working on the problem were two scientists American James D. Watson and Englishman Francis Crick.

The brief but celebrated partnership soon solved the puzzle of DNA structure.

## WATSON-CRICK MODEL OF DNA

1. The model says that DNA exists as a double helix. A DNA molecule has two unbranched polynucleotide strands. Each

polynucleotide strand or chain consists of a sequence of nucleotides linked together by phosphodiester bonds. The polynucleotide strands are anti-parallel and seen in opposite direction.

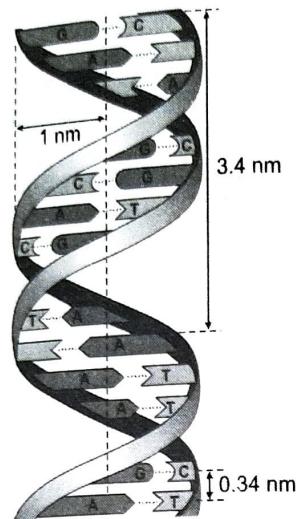
2. The two strands are not coiled upon each other but the double strand is coiled upon itself around a common axis like spiral staircase with base pair forming steps while the backbone of the two strands form railing. Backbone is formed of sugar and phosphate.
3. Base pairing is specific Adenine always pair with thymine and guanine pairs with cytosine, thus all base pair consists of one purine and one pyrimidine. Once the sequence of base in one strand of DNA double helix is known, the sequence of base in another strand can also be known because of specific base pairing. The two strands of DNA are said to be complementary. This is known as complementary base pairing.
4. The two polynucleotide strands are held together by hydrogen bonding between bases in opposite strands. Adenine and thymine are connected with two hydrogen bonds, guanine and cytosine are connected with three hydrogen bonds.
5. One end of strand is called 5' end where fifth carbon of pentose sugar is free and the other end is called 3' end where the third carbon of pentose is free.
6. At each base pair the strand turns 36°, one full turn of the helical strand (360°) would involve ten base pairs, i.e. one turn of 360 m° of helical strand has about ten nucleotides on each strand of DNA. The base pairs in DNA are stacked 3–4 Å apart. Thus pitch of DNA is 34 Å apart as ten base pairs occupy a distance of about 34 Å (Fig. 6.1).

## Chargaff's Rule

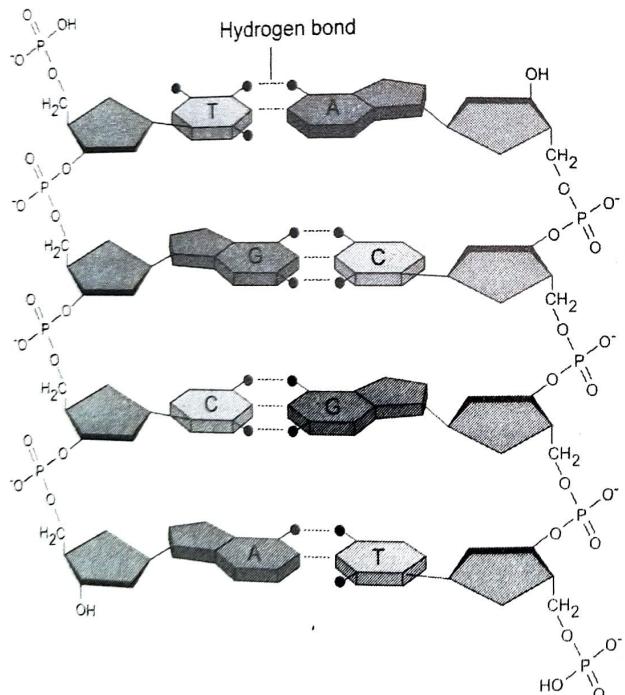
In 1950, Chargaff found that in a DNA molecule:

1. Amount of purines and pyrimidines are equal.

$$A + G = T + C$$



(a) Key features of DNA structure



(b) Partial chemical structure

Fig. 6.1: Structure of DNA double Helix

2. Amount of adenine is equal to thymine and amount of guanine is always equal to cytosine  
 $A = T$  and  $G = C$

3. Deoxyribose sugar and phosphate components occur in equal proportions.

### Nucleosomes

Walter Flemming observed banded objects in the nuclei of stained eukaryote cells. He called the material chromosome which contains DNA plus the various proteins that package the DNA in a more compact form.

In humans, nucleus contains 46 chromosomes. The major protein of chromosomes is called histones. Histones are simple, basic proteins containing numerous lysine and Arginine amino acids whose positive charge allow the proteins to bind to negatively charged sugar phosphate backbone of DNA. Chromosomes unfold when it is treated with a solution of low ionic strength. The extended chromosome looks like beads on a string in an electron micrograph. The beads are DNA histone complex called nucleosome and the string is double stranded DNA. Each nucleosome consists of histone around which DNA is wrapped (Fig. 6.2).

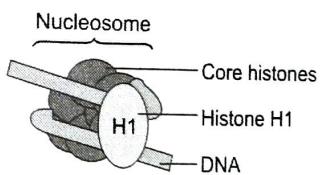


Fig. 6.2: Structure of nucleosome

### GENETIC CODE

As DNA is a genetic material, it carries genetic information from cell to cell and from generation to generation. At this stage, an attempt will be made to determine in what manner the genetic information existed in DNA molecule.

As we know, DNA molecule is composed of three kind of components: (1) Phosphoric acid, (2) deoxyribose sugar, (3) nitrogenous base. Since the sugar phosphate forms the

backbone of DNA and is always same, therefore, it is unlikely that this component of DNA carry the genetic information. Nitrogen bases, however, vary from one segment of DNA molecule to another, so the genetic information will depend on their sequence. These four DNA bases can be considered as four alphabets of DNA molecules.

George Gamow first proposed the basic structural unit of genetic code. The basic problem of genetic code is to indicate how information written in four letter language of nitrogenous base of DNA can be translated into twenty letter language of amino acids of proteins.

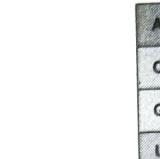
The group of nucleotides that specifies the amino acid is a code word or codon. The simplest possible code is a single code in which one nucleotide code for one amino acid, such a code is inadequate as only four amino acids can be specified. A doublet code, a code of two letters is also inadequate because it could specify only sixteen ( $4 \times 4$ ) amino acids (Fig. 6.3).

A triplet code, a code of three letters, could specify sixty four ( $4 \times 4 \times 4$ ) amino acids. Therefore, it is likely that there are 64 triplet codes for 20 amino acids.

The first experimental evidence on support to the concept of triplet code was provided by Crick and coworkers in 1961. During the experiment when they added or deleted single or double base pairs in a particular region of DNA of T4 bacteriophage of *E. coli*, they found out that such bacteriophages ceased to perform their normal functions. However, bacteriophages with addition or deletion of three base pairs in DNA molecules had performed normal functions, from this experiment they concluded that a genetic code is in triplet form because the addition of one or four nucleotides has put the reading of the code out of order. While the addition of third nucleotides resulted in a return to the proper reading of the message.

### Process for Making a Protein

Genes that provide instructions for proteins are expressed a two-step process.



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

Singlet code:  $4^1 = 4 \times 1 = 4$  codons      Doublet code:  $4^2 = 4 \times 1 = 4$  codons

	U	C	A	G	
U	UUU UUC UUA UUG	UCU UCC UCA UCG	UAU UAC	UGU UGC	UC CA AG
C	CUU CUC CUA CUG	CCC CCA CCG	CAU CAC	CGU CGC	UC CA AG
A	AUU AUC AUA AUG	ACU ACC ACA ACG	AAU AAC	AGU AGC	UC CA AG
G	GUU GUC GUA GUG	GCU GCC GCA GCG	AAA AAG	AGA AGG	UC CA AG

Triplet code:  $4^3 = 4 \times 4 \times 4 = 64$  codons

Fig. 6.3: Genetic code showing singlet, doublet and triplet code

- In transcription the DNA sequences of a gene are rewritten in the RNA eukaryotes, the RNA must go through additional steps to become a messenger RNA (mRNA).
- In translation the sequence of nucleotides in the mRNA is translated into a sequence of amino acids in a polypeptides (protein chain).

### Pattern to the Genetic Code

A notable pattern emerged when the genetic code was studied. First amino acids with similar structural properties tend to have related codons, thus aspartic acid codons (GAU and GAC) are similar to glutamic acid codons (GAA, GAG), similarly codons for aromatic acid phenylalanine (UAU, UUC), tyrosine (UUU,

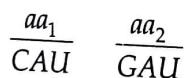
UAC) and tryptophan (UGG) all begin with uracil. This feature of the code is thought to have evolved to minimize the consequences of mistakes made during translation. If an amino acid in a protein is by mistake replaced by one with similar properties, the protein may still be functional.

Second pattern to the code is that for many of the synonym codon specifying the same amino acid, the first two bases of the triplet are constant, whereas third can vary. For example, all codons starting with CC specifying Protein (CCU, CCC, CCA and CCG), this flexibility in the third nucleotide of a codon helps to minimize the consequence of errors.

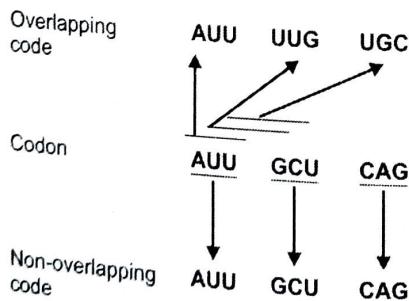
### Characters of Triplet Code

The genetic dictionary of mRNA codons reveal following important features of triplet codons.

1. **The code is non-overlapping:** Since the DNA molecules is a long chain of nucleotides it could be read in an overlapping or non-overlapping manner. In the non overlapping code six nucleotides would code for two amino acids. While in overlapping code four amino acids can be read (Fig. 6.4).



Non-overlapping code C, A, U and G are bases  $aa_1$  and  $aa_2$  are amino acids.

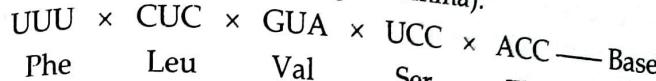


**Fig. 6.4:** Overlapping and non-overlapping code

**Overlapping code:** Studies on gene mutations show the code is non-overlapping type in the tobacco mosaic trees

(TMV) mutation of one base of mosaic acid results in alteration of only one single amino acid.

2. **The code is commaless:** The genetic code reads in an uninterrupted manner from one end of nucleic acid chain to another or are the bases (commas) between successive codons to code with commas can be represented as follows ( $x$  represent a base acting as comma):



A commaless code would have no comma bases and can be represented as:



All the available evidence indicates that the code is commaless. The work of Khorana and his associate gave clear evidence of commaless code.

3. **Code has polarity:** If a gene is to specify the same protein repeatedly it is essential the code must have fixed start and end points. These points are the indication and termination codons respectively. The code must be read in a fixed direction. In other words, code must have polarity. The available evidence indicates code is read in  $5' \rightarrow 3'$  direction in mRNA. The polypeptides chain is synthesized in N  $\rightarrow$  C direction, from amino ( $NH_2$ ) terminal to carboxyl (COOH) terminal.

4. **Codons and anticodons:** During translation the codons of mRNA pair with complementary anticodons of tRNA. Since mRNA is read in polar manner in the  $5' \rightarrow 3'$  directions, the codons are written on  $5' \rightarrow 3'$  direction. The codon AUG is written as  $5' AUG 3'$ . Often to make things simpler, anticodons are written  $3' \rightarrow 5'$  directions so as to bring easier correlation between codon and anticodon bases.

**Codon (mRNA) –  $5' AUG 3'$**

**Anticodon (tRNA) –  $3' UAC 5'$**

5. **Initiation codon:** The starting amino acid in syntheses of most proteins chain is methionine (eukaryotes) or N. formyl

methionine (Prokaryotes). It is said to initiation sites containing AUG codon. This codon is, therefore, called the initiation codon. Less often GUG also function as initiation codon in bacterial protein.

**6. Termination codon:** Three of 64 codons do not specify any tRNA and are hence called nonsense codons. These codons are (UAG) amber, UAA (ochre) and UGA (opal or umber), since they bring about termination of polypeptide chain synthesis, they are also called termination codons.

**7. The code is degenerate:** There are 64 possible codons in a triplet code of which 61 take part to code amino acids. Since only 20 amino acids take part in protein synthesis, it is obvious that there are many more codons types, as a result there are multiple codons for most amino acids. For example: There are six codons for serine, four for glycine and two for lysine. Because of the existence of several codons for most amino acids, the genetic code is said to be degenerate.

Different codons that specify the same amino acid are known as synonymous codons. The degeneracy of genetic code minimises the effect of mutations since changing a single nucleotide often results in a codon that still specifies the same amino acid. In a standard genetic code, the only amino acids with single codon are methionine and tryptophan.

**Wobble hypothesis:** The triplet code is a degenerate one with many more codons than the number of amino acid types coded. An explanation for this degeneracy is provided by the Wobble hypothesis proposed by Crick (1966). Since there are 61 codons specifying amino acids the cell should contain 61 different tRNA molecules each with a different anticodons but the number of tRNA molecules types discovered is much less than 61. This implies that the anticodon of some tRNA read more than one codon in mRNA.

According to the Wobble hypothesis, only the first two positions of the triplet codon on mRNA have a precise pairing with the basis of the tRNA anticodon. The pairing

of the third position bases of the codon may be ambiguous and varies according to the nucleotide present in this position. Thus a single tRNA type is able to recognize two or more codons differing only in the third base. For example, anticodon UCG of serine tRNA recognizes two codons AGC and AGU. The bonding between UCG and AGU follows the usual Watson-Crick pairing. In UCG and AGU pairing hydrogen bonding takes place between G and U which is different from the usual Watson-Crick pairing where G pairs with C and A with U. Such interaction between the third base is referred to as 'Wobble pairing'.

mRNA codon (Serine)	5' AGC 3'	5' AGU 3'
tRNA anticodon	3' UCG 5'	3' UCG 5'

**8. The code is universal:** The genetic code is valid for all organisms ranging from bacteria to man. It is essentially same for all organisms and so it is said to be universal. The universality of the code was demonstrated by Marshall and Nirenberg who found *E. coli* (bacteria), *Xenopus laevis* (amphibian) and guinea pig (mammal) use almost the same code, this showed that the code is essentially universal.

## Techniques of Genetic Engineering

New techniques have been developed to manipulate the genetic material. Genetic engineering started in 1993 by combining a gene from a bacterium with the plasmid of *E.coli*. Genetic engineering techniques have been used for many purposes useful for mankind. This new techniques of genetic engineering are known as recombinant DNA technology which deals with manipulation of genetic material by man *in vitro*.

## Recombinant DNA Technology

It involves two basic processes:

- Formation of recombinant DNA.
- Introduction of recombinant DNA into an appropriate host.

Recombinant DNA is formed by combining DNA from different organisms. For example, insulin gene cut off from rat's DNA and linked to bacterial plasmid gives recombinant DNA.

120

Denaturation and renaturation are the two properties which help in the formation of recombinant DNA. Denaturation is the separation of strands of DNA by breaking of hydrogen bonds in heating. Renaturation is the reunion of complementary strands to form double helix on cooling.

### **Tools of Recombinant DNA Technology**

Three types of biological tools are used in synthesis of recombinant DNA. These include:

1. Enzymes
2. Vectors DNA
3. Passenger DNA

#### **Enzymes**

Many kind of enzymes are used in genetic engineering. These include:

1. Lysing enzyme: These are used to open up the cells to get DNA for genetic experiments. Lysozyme is commonly used to dissolve the bacterial cell wall.
2. Clearing enzymes: These are used to break DNA molecule. These are of three kinds:
  - a. Exonucleases: Which cut off nucleotide in 5' or 3' end of DNA molecule.
  - b. Endonucleases: Which cut off DNA molecule at any part except ends.
  - c. Restriction endonucleases: Which cut off DNA duplex at specific point in such a way that single stranded free ends project from each fragment of DNA duplex. The single stranded free ends are called 'sticky ends' because they can join similar complementary ends of DNA fragment from other source and can be said as a molecular scissors EcoR1 is commonly used restriction enzyme.
3. Synthesizing enzymes: These enzymes play a role in the synthesis of DNA strands on suitable templates. These are of two types:
  - a. Reverse transcriptase: Which helps in the synthesis of complementary DNA strand on RNA template.

- b. DNA polymerase: Which helps in the syntheses of complementary DNA strand on DNA template.
- 4. Joining enzymes: They help in scaling gaps in DNA fragment which are otherwise formed by complementary base pairing. They are the molecular glues. They join DNA fragments by forming phosphodiester bonds.
- 5. Alkaline phosphatases: These cut off phosphate group from 5' end of linearised circular DNA to check its recircularization.

### **Vehicle of Vector DNA**

DNA used as a carrier for transferring fragment of foreign DNA into a suitable host is called vehicle DNA. It is also called a gene carrier. The desired gene is introduced in a vector where recombinant DNA (rDNA) is formed. The vector carrying rDNA divides, thereby forming the several copies of rDNA. Five types of DNA are used as vehicles:

1. Plasmid DNA
2. Bacteriophage DNA
3. DNA of plant and animal virus
4. Transposons (jumping genes)
5. Artificial DNA

Plasmids and bacteriophage DNA are commonly used vehicles.

1. **Plasmids:** These are small double stranded, closed circular symbiotic DNA molecules that naturally occurs in bacteria, outside bacterial chromosome and are regarded as extra chromosomal DNA. A bacteria cell may have one or many copies of plasmids. They are inherited from parent bacterial cell to daughter cells and have the capability for self-replication the cytoplasm of bacteria.

Another feature of plasmids is the presence of specific restriction site where the enzyme restriction endonucleases make a cut so that foreign DNA segment may be joined to the plasmid. These are resistant to antibodies.

2. **Bacteriophage vectors:** They are the virus that infect bacteria by introducing their DNA into bacterial cell. The