Chapter 16

Molecular Basis of Inheritance

Chapter Contents

Introduction

"Factors/Genes" were first detected and analyzed by Mendel and subsequently many other scientists, by following their patterns of transmission from generation to generation. These studies, while greatly elucidating the nature of inheritance in living organisms, provided no insight into the structure or molecular composition of "factors". In 1926, the quest to determine the mechanism for genetic inheritance reached the molecular level and the nature of the putative genetic material was investigated culminating in the realisation that DNA-deoxyribonucleic acid is the genetic material at least for the majority of organisms. This is the substance which controls the inheritance of traits from one generation to the next and it is also able to express its effect through the formation and functioning of traits.

Nucleic acid is of two types in all living systems *i.e.*, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA is genetic material in all organisms except some viruses. RNA is genetic material in riboviruses. In others, RNA functions as a messenger carrying genetic information, an adapter for picking up amino acids, structural and catalytic molecule in some cases.

In this chapter, we are going to discuss the structure of DNA, its replication, the process of making RNA from DNA (transcription), genetic code that determines the sequence of amino acids in proteins, the process of protein synthesis (translation) and elementary basis of their regulation. The essentials of human genome sequencing and its consequences will also be discussed in the last section.

THE DNA

DNA is a long polymer of deoxyribonucleotides. It is an acidic substance present in nucleus, which was **first identified by Friedrich Meischer in 1869**. He named it as **"Nuclein". Altmann** found these substances to be acidic in nature, hence he named it **nucleic acid**. The **length of DNA** is usually defined as number of nucleotides or a pair of nucleotide referred to as base pairs (bp) present in it. **This also is the characteristic of an organism.**

Few examples are given below :

Organism	Genetic material	No. of nucleotides or bp
ϕ174 bacteriophage	ssDNA, Circular	5386 bases
Lambda (λ) phage	dsDNA, Linear	48502 bp
Escherichia coli	dsDNA, Circular	4.6×10^6 bp
Human genome	dsDNA, Linear	3.3×10^9 bp

Structure of Polynucleotide Chain

The basic unit of DNA is a nucleotide which has three components – **a nitrogenous base**, a **pentose sugar** (deoxyribose) and a **phosphate** group. There are two types of nitrogenous bases :

(i) Purines : Heterocyclic, 9-membered double-ring structure with N at position 1, 3, 7 and 9, *e.g.*, Adenine (A) and Guanine (G).

(ii) Pyrimidines : Heterocyclic, 6-membered single-ring structure with N at 1 and 3 position, *e.g.*, Cytosine (C), Thymine and Uracil. Cytosine is common in both DNA and RNA; thymine is present in DNA and uracil is present in RNA at the place of thymine.

A polynucleotide chain shows following types of linkage or bond in its components :

(i) N-glycosidic linkage : A nitrogenous base is linked to the pentose sugar through a N-glycosidic linkage to form a nucleoside. Purine nucleosides have 1′ — 9 glycosidic linkage (carbon 1′ of sugar and 9 position of A/G). Pyrimidine nucleosides have 1′ — 1 linkage *i.e.*, sugar carbon 1′ and 1 position of T/C).

(ii) Phosphoester linkage : When a phosphate group is linked to 5′ – OH of a nucleoside through phosphoester linkage a corresponding nucleotide is formed. Two nucleotides are linked through **3′— 5′ phosphodiester linkage** to form a **dinucleotide**.

A polymer thus formed has a free phosphate moiety at 5′-end of sugar, which is referred as 5′-end of polynucleotide chain. Similarly, at the other end of the polymer the sugar has a free 3′-OH group which is referred to as 3′-end of polynucleotide chain. The backbone in a polynucleotide chain is formed due to sugar and phosphates. The nitrogenous base linked to sugar moiety projects from the backbone.

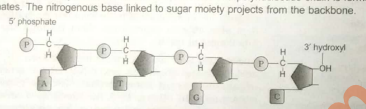


Fig. : A Polynucleotide chain

Types of Nucleosides in DNA	Types of Nucleotides in DNA
(i) Deoxyadenosine (A + S)	+ P = dAMP (deoxyadenosine monophosphate)
(ii) Deoxyguanosine (G + S)	+ P = dGMP (deoxyguanosine monophosphate)
(iii) Deoxycytidine (C + S)	+ P = dCMP (deoxycytidine monophosphate)
(iv) Deoxythymidine (T + S)	+ P = dTMP (deoxythymidine monophosphate)
Types of Nucleosides in RNA	**Types of Nucleotides in RNA**
(i) Adenosine (A + S)	+ P = AMP (adenosine monophosphate)
(ii) Guanosine (G + S)	+ P = GMP (Guanosine monophosphate)
(iii) Cytidine (C + S)	+ P = CMP (Cytidine monophosphate)
(iv) Uridine (U + S)	+ P = UMP (Uridine monophosphate)

Note : *S and P represents sugar and phosphate respectively. Sugar is ribose ($C_5H_{10}O_4$) in RNA and deoxyribose ($C_5H_{10}O_4$) in DNA*

Derivation of DNA Structure

Two lines of investigations helped in derivation of DNA structure *i.e.*,

(a) X-ray Crystallography and

(b) Chargaff's rule

(a) X-ray Crystallography : Maurice Wilkins and Rosalind Franklin obtained very fine X-ray diffraction pictures of DNA. It was suggested that structure of DNA was sort of helix with 3.4 Å periodicity. But they had not proposed a definitive model for DNA.

(b) Erwin Chargaff's Rules : Chargaff's along with his colleagues, performed base composition studies and put forward certain generalisations for double-stranded DNA, called as Chargaff's rule (Not applicable for single-stranded DNA).

(i) Purines and pyrimidines occur in equal amounts.

(ii) Purines found in DNA are adenine and guanine. Pyrimidines of DNA are thymine and cytosine.

$$A + G = T + C$$

(iii) $\frac{A+G}{T+C} = 1$, this value is constant for all species.

(iv) Base ratio $\frac{A+T}{C+G}$ is specific for a species. It is used to identify the species. It is less than one in prokaryotes, *e.g., E. coli* = 0.92 and more than one in eukaryotes, *e.g.*, Humans = 1.52.

(v) Sugar deoxyribose and phosphate residues occur in equal number.

(vi) Purine adenine is equimolar with pyrimidine thymine.

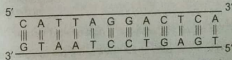
(vii) Purine guanine is equimolar with pyrimidine cytosine.

James Watson and Francis Crick on the basis of previous informations proposed a very simple but famous double helix model for the structure of DNA. One of the **hallmarks of their proposition was base pairing** between the two strands of polynucleotide chains. However this proposition was based on the observations of Erwin Chargaff. The base pairing confers a very unique property to the polynucleotide chains. They are said to be complementary to each other and therefore if the sequence of bases in one strand is known then the sequence in other strand can be predicted. Thus if one DNA strand has A, the other would have T and if one has G, the other would have C. Therefore, if the base sequence of one strand is CATTAGGAC, the base sequence of other strand would be GTAATCCTG. Also, if each strand from a DNA acts as template for synthesis of a new strand, the two double-stranded DNA or daughter DNA produced would be identical to the parental DNA molecule.

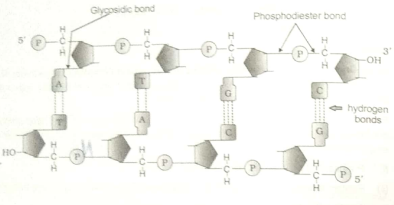
Salient features of the double helix structure of DNA are :

(i) DNA consists of two polynucleotide chains. The backbone is constituted by sugar-phosphate and the bases project inside.

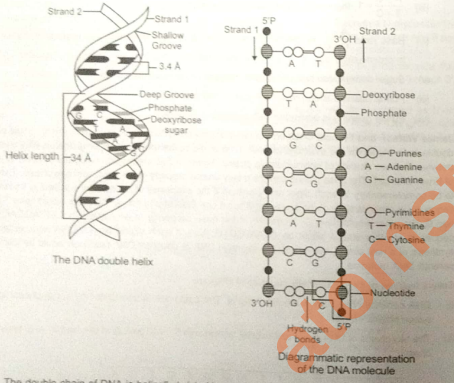
(ii) The two chains of DNA run in anti-parallel fashion with 5′ → 3′ polarity in one and 3′ → 5′ polarity in other chain.



(iii) The bases in two strands are paired through hydrogen bonds forming base pairs (bp). Adenine forms two H-bonds with thymine from opposite strand and viceversa. Similarly, guanine is bonded with cytosine with three H-bonds. As a result, always a purine comes opposite to a pyrimidine. This generates approximately uniform distance between the two strands of helix.



Double-stranded polynucleotide chain



(iv) The double chain of DNA is helically twisted in a right-handed fashion.

(v) Each turn of double helix or the pitch of the helix is 3.4 nm (a nanometer is one billionth of a meter, that is 10^{-9} m). It has roughly 10 base pairs in each turn. The distance between two adjacent base pairs is approximately equal to 0.34 nm.

(vi) The plane of one base pair stacks over the other in double helix. This, in addition to H-bonds, confers **stability to the helical structure**.

Knowledge Cloud

(i) **Types of DNA and their comparison.**

DNA types	Base pairs per turn (*n*)	Rotation	Vertical rise per bp	Helical diameter bp (*h*)
A	11	Right handed	2.56 Å	23 Å
B	10	Right handed	3.4 Å	20 Å
C	9.33	Right handed	3.3 Å	19 Å
Z	12	Left handed	3.8 Å	18.4 Å

(ii) **Linear double-stranded DNA** in eukaryotes and PPLO (Monerans)

(iii) **Repetitive DNA :** It is the part of DNA which contains the same sequence of nitrogen bases repeated more than once in genome. The area with long sequence of short repetitive DNA is called **satellite DNA** because it separates out during density gradient ultracentrifugation as small dark bands.

5′ A A A A A A A 3′
3′ T T T T T T T 5′

(iv) **Palindromic DNA :** It has base sequence which reads the same on both strands either in 5′ → 3′ or 3′ → 5′ direction. Different types of palindromic sequences are recognized by restriction endonucleases, e.g.,

5′–G A A T T C–3′
3′–C T T A A G–5′

(v) **Denaturation and Renaturation :** Separation of two strands of DNA from each other due to breakage of H-bonds when it is exposed to high temperature, acid or alkali is called denaturation or melting. Reassociation of separated DNA by H-bonds formation is called renaturation or annealing. DNA with more A = T has low melting areas and denatured more easily. DNA with more G ≡ C than A = T has high melting areas.

(vi) **C-value :** Total amount of DNA per genome.

(vii) The amount of DNA is expressed in **picogram**.

1 pg = 10^{-12} gm.

(viii) DNA functions :

(a) Hereditary information

(b) Variations : It occurs due to crossing over at the time of meiosis.

(c) Mutations : Sudden inheritable variations due to change in genetic material.

(d) Autocatalytic function or DNA replication *i.e.*, DNA → DNA synthesis

(e) Heterocatalytic function : DNA → RNA, proteins, hormones synthesis

(f) Control of metabolism, Growth and differentiation

(g) DNA fingerprinting

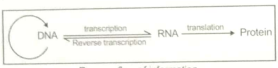
Central Dogma of Molecular Biology

It explains one way or unidirectional flow of information from master copy DNA to working copy RNA and from RNA to building molecule or trait expressing molecule polypeptide. Central dogma of molecular biology was proposed by Francis Crick.

Central dogma

Reverse Central Dogma or Teminism

An exception to this one way flow of information was reported in 1970 by **H. Temin and D. Baltimore**. They independently discovered reverse transcription in some viruses. These viruses produce an enzyme reverse transcriptase which can synthesize DNA over RNA template. This discovery was important in understanding cancer and, hence, these two scientists were awarded Nobel prize. The modified flow of information now can be shown as follows :



Reverse flow of information

Packaging of DNA Helix

The distance between two consecutive base pairs is 0.34 nm (0.34×10^{-9} m) then length of DNA for a human diploid cell is 6.6×10^9 bp $\times$ 0.34×10^{-9} m = 2.2 metres. This length is far greater than the dimension of a typical nucleus which is approximately 10^{-6} m.

Similarly, the number of base pairs in *E. coli* is 4.6×10^6 so the total length comes out to be 1.36 mm which is placed in a cell having size 1 µm. So, the long sized DNA can be accommodated in small area only through packaging or compaction.

DNA Packaging in Prokaryotes

In prokaryotes, DNA is not scattered throughout the cell although they do not have a defined nucleus. DNA is found in cytoplasm in super coiled stage. The coils are maintained by non-histone basic protein polyamines which have positive charge. The packaged structure of DNA is called **nucleoid** or **genophore**.

DNA Packaging in Eukaryotes

In eukaryotes, this organisation is much more complex and is carried out by a set of positively charged basic proteins called **histones**. Histones are rich in the basic amino acids residues **lysines and arginines** with charged side chains. There are five types of histone proteins *i.e.*, H_1, H_2A, H_2B, H_3 and H_4. Four of them occur in pairs to produce histone octamer or *nu-body* (two copies of each H_2A, H_2B, H_3 & H_4). The negatively charged DNA is wrapped around the positively charged histone octamer to form a structure called **nucleosome**.

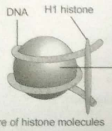


Fig. : Nucleosome

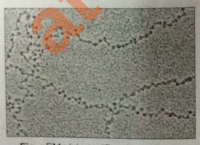


Fig. : EM picture - "Beads-on-String"

About 200 bp of DNA is wrapped over *nu-body* to complete about 1¾ turns. This forms nucleosome of size 110 × 60Å. DNA present between two adjacent nucleosome is called linker DNA with about 80 bp. The nucleosome constitute repeating unit of a structure in nucleus called **chromatin**. The nucleosomes in chromatin gives a **'beads on string'** appearance under electron microscope. The nucleosomes further coils to form **solenoid/chromatin fibre**. It has diameter of 30 nm. Chromatin fibres are further coiled and condensed at metaphase stage of cell division to form chromosomes. The packaging of chromatin at higher level requires additional set of proteins that collectively are referred as **Non-histone chromosomal (NHC) proteins**.

Chromatin is differentiated into two regions, on the basis of staining behaviour in a typical nucleus :

1. **Heterochromatin**
   - (i) It is darkly stained region
   - (ii) Chromatin is densely packed
   - (iii) Transcriptionally it is inactive
2. **Euchromatin**
   - (i) Lightly stained region
   - (ii) Loosely packed chromatin
   - (iii) Transcriptionally it is active

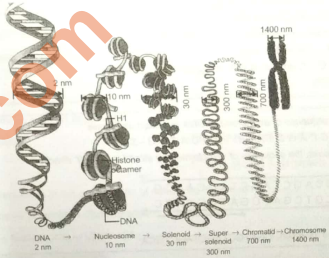


Fig.: Various steps in the folding and superfolding of basic chromatin components to generate an eukaryotic chromosome

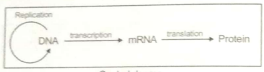
Knowledge Cloud

1. **Non-histone chromosomal proteins** are of three types :
   - (i) Scaffold or structural NHC
   - (ii) Functional NHC protein *e.g.*, DNA polymerase, RNA polymerase
   - (iii) Regulatory NHC protein *e.g.*, HMG (High mobility group proteins that controls gene expression).
2. **Chemical composition of chromosome**

DNA	–	40%
RNA	–	1.2%
Histone proteins	–	50%
Acidic proteins	–	8.5%
Lipid	–	Traces
Ca^{+2}, Mg^{+2}, Fe^{+2}	–	Traces

Central Dogma of Molecular Biology

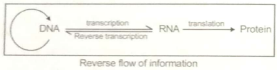
It explains one way or unidirectional flow of information from master copy DNA to working copy RNA and from RNA to building molecule or trait expressing molecule polypeptide. Central dogma of molecular biology was proposed by Francis Crick.



Central dogma

Reverse Central Dogma or Teminism

An exception to this one way flow of information was reported in 1970 by **H. Temin and D. Baltimore**. They independently discovered reverse transcription in some viruses. These viruses produce an enzyme reverse transcriptase which can synthesize DNA over RNA template. This discovery was important in understanding cancer and, hence, these two scientists were awarded Nobel prize. The modified flow of information now can be shown as follows :



Reverse flow of information

Packaging of DNA Helix

The distance between two consecutive base pairs is 0.34 nm (0.34×10^{-9} m) then length of DNA for a human diploid cell is 6.6×10^9 bp $\times$ 0.34×10^{-9} m = 2.2 metres. This length is far greater than the dimension of a typical nucleus which is approximately 10^{-6} m.

Similarly, the number of base pairs in *E. coli* is 4.6×10^6 so the total length comes out to be 1.36 mm which is placed in a cell having size 1 µm. So, the long sized DNA can be accommodated in small area only through packaging or compaction.

DNA Packaging in Prokaryotes

In prokaryotes, DNA is not scattered throughout the cell although they do not have a defined nucleus. DNA is found in cytoplasm in super coiled stage. The coils are maintained by non-histone basic protein polyamines which have positive charge. The packaged structure of DNA is called **nucleoid** or **genophore**.

DNA Packaging in Eukaryotes

In eukaryotes, this organisation is much more complex and is carried out by a set of positively charged basic proteins called **histones**. Histones are rich in the basic amino acids residues **lysines and arginines** with charged side chains. There are five types of histone proteins *i.e.*, H_1, H_2A, H_2B, H_3 and H_4. Four of them occur in pairs to produce histone octamer or *nu-body* (two copies of each H_2A, H_2B, H_3 & H_4). The negatively charged DNA is wrapped around the positively charged histone octamer to form a structure called **nucleosome**.

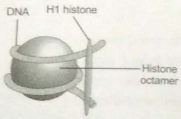


Fig. : Nucleosome

Fig. : EM picture - "Beads-on-String"

About 200 bp of DNA is wrapped over *nu-body* to complete about 1¾ turns. This forms nucleosome of size 110 × 60Å. DNA present between two adjacent nucleosome is called linker DNA with about 80 bp. The chromatin gives a '**beads on string**' appearance under electron microscope. The nucleosomes in to form **solenoid/chromatin fibre**. It has diameter of 30 nm. Chromatin fibres are further coiled and condensed at metaphase stage of cell division to form chromosomes. The packaging of chromatin at higher level requires additional set of proteins that collectively are referred as **Non-histone chromosomal (NHC) proteins**.

Chromatin is differentiated into two regions, on the basis of staining behaviour in a typical nucleus :

1. **Heterochromatin**
   - (i) It is darkly stained region
   - (ii) Chromatin is densely packed
   - (iii) Transcriptionally it is inactive
2. **Euchromatin**
   - (i) Lightly stained region
   - (ii) Loosely packed chromatin
   - (iii) Transcriptionally it is active

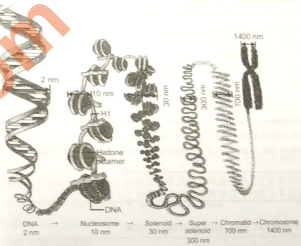


Fig.: Various steps in the folding and superfolding of basic chromatin components to generate an eukaryotic chromosome

Knowledge Cloud

1. **Non-histone chromosomal proteins** are of three types :
   - (i) Scaffold or structural NHC
   - (ii) Functional NHC protein e.g., DNA polymerase, RNA polymerase
   - (iii) Regulatory NHC protein e.g., HMG (High mobility group proteins that controls gene expression).
2. **Chemical composition of chromosome**

DNA	–	40%
RNA	–	1.2%
Histone proteins	–	50%
Acidic proteins	–	8.5%
Lipid	–	Traces
Ca^{+2}, Mg^{+2}, Fe^{+2}	–	Traces

Example 1 : *If a DNA molecule has 2000 bp then calculate the,*
(a) Number of sugar and phosphate molecules
(b) Number of N-glycosidic linkage

Solution : (a) Total number of nucleotides = 4000
Nucleotide → Nitrogen base + Sugar + Phosphate
Thus, total no. of sugar = 4000, and phosphate = 4000.
(b) N-glycosidic linkage occurs in between nitrogen base and sugar.
Total no. = 4000

Example 2 : *DNA was extracted from Streptococcus bacterium. The proportion of Adenine was found to be 28%, then calculate the amount of cytosine.*

Solution : According to Chargaff rule :
Equimolar concentration of A = T and G = C
A = T → 28 + 28 = 56%
Thus G = C amount is 44% with 22% G and 22%C.
C = 22%

Example 3 : *If the sequence of one strand of DNA is written as follows :*
5'– T G C A G C T A G C T A G C A T C G –3'
Write down the sequence of complementary strand in 5' → 3' direction.

Solution :
5' T G C A G C T A G C T A G C A T C G 3'
3' A C G T C G A T C G A T C G T A G C 5'

In 5'→3' direction; CGATGCTAGCTAGCTGCA

Example 4 : *Enumerate the number of beaded structures (nucleosomes) present in the nucleus of diploid eukaryotic cell which possess 2.4 × 10^6 bp.*

Solution : One nucleosome has 200 bp.

$$\frac{2.4\times10^6}{200} = 1.2\times10^4 \text{ or } 12\times10^3 \text{ nucleosomes}$$

Try Yourself

1. Select true or false statement :
   (a) Two nucleotides in a strand are linked through H-bond to form a dinucleotide.
   (b) The pitch of the DNA helix is 3.4 nm.
   (c) Deoxythymidine is monomer nucleotide of DNA.
   (d) DNA is packaged with non-histone basic protein to form nucleoid.
2. Can you suggest simple name to the process of RNA → DNA synthesis?

EXERCISE

1. Which of the following bond is not associated with a deoxyribonucleotide?
   (1) Phosphoester bond (2) Glycosidic bond
   (3) Phosphodiester bond (4) More than one option is correct
2. RNA possess additional _____ group at _____ position in the sugar than the DNA.
   (1) OH, 5′ (2) H, 2′
   (3) OH, 2′ (4) H, 5′
3. Hallmark of the Watson and Crick three dimensional DNA model was based upon the findings of
   (1) Wilkins and Franklin (2) Erwin Chargaff
   (3) Hershey and Chase (4) Meselson and Stahl
4. Which of the following DNA form has maximum number of base pairs per turn?
   (1) A-DNA (2) B-DNA
   (3) C-DNA (4) Z-DNA
5. Which of the following is a part of *nu-body*?
   (1) Histone octamer (2) DNA + Core of nucleosome
   (3) H1 protein (4) $1\frac{3}{4}$ turn of DNA + H1 protein
6. Choose the correct steps in the organisation of eukaryotic chromosome
   (1) Nucleosome → Solenoid → Supersolenoid
   (2) Solenoid → Nucleosome → Chromatid
   (3) DNA → Solenoid → Nucleosome
   (4) Chromatin → Solenoid → Nucleosome
7. Heterochromatin
   (1) Is transcriptionally active (2) Is densely packed
   (3) Replicated during early S-phase (4) Stains lightly
8. Non-histone proteins
   (1) Are of five types (2) Are involved in nucleosome formation
   (3) Control gene expression (4) Are basic proteins
9. The number of glycosidic bonds associated with DNA of diploid human cell are
   (1) 6.6×10^9 (2) $2 \times 6.6 \times 10^9$
   (3) 3.3×10^9 (4) $3.3 \times 10^9 - 2$
10. Which of the following does not confer stability to the helical structure of DNA?
    (1) Phosphodiester bond (2) H-bond
    (3) N-glycosidic linkage (4) More than one option is correct

THE SEARCH FOR GENETIC MATERIAL

Even though the discovery of nuclein by Meischer and the proposition for principle of inheritance by Mendel were almost at the same time, but that the DNA acts as a genetic material took long to be discovered and proven.

The experiments given below prove that DNA is the genetic material :

1. **Transforming Principle**: The transformation experiments, conducted by Frederick Griffith in 1928, are of great evidence in establishing the nature of genetic material. He performed series of experiments by selecting two strains of bacterium *Streptococcus pneumoniae* (also called *Pneumococcus*) namely, S-III and R-II.

   (i) S-III strain/smooth or capsulated type have a mucous (Polysaccharide) coat and produce smooth shiny colonies in culture plate. These are virulent and cause pneumonia.

   (ii) R-II strain/rough or non-capsulated type have no mucous coat and produce rough colonies. These are non-virulent and do not cause pneumonia.

   The experiment can be described in following four steps :

   (a) S strain → Injected into mice → Mice die
   (b) R strain → Injected into mice → Mice live
   (c) S strain (heat-killed) → Injected into mice → Mice live
   (d) S-strain (heat-killed) + R-strain (live) → Injected into mice → Mice die

   Griffith was able to kill bacteria by heating them. He observed that heat-killed S-strain bacteria injected into mice did not kill them. When he injected a mixture of heat-killed S and live R-bacteria, the mice died. Moreover, he recovered living S-bacteria from the dead mice.

   He concluded that the R-strain bacteria had somehow been transformed by the heat-killed S-strain bacteria. This occurred perhaps due to absorption of some transforming principle or substance by rough type bacteria from heat-killed smooth bacteria. It had enabled the R-strain to synthesize a smooth polysaccharide coat and become virulent. This must be due to the transfer of the genetic material. However, the biochemical nature of genetic material was not defined from his experiments.

   Biochemical characterisation of Transforming Principle: Oswald Avery, Colin Macleod and Maclyn McCarty (1944) repeated the experiment *in-vitro* to identify the biochemical nature of transforming substance. They proved that this substance is DNA. Prior to their work the genetic material was thought to be protein.

   They purified biochemical *i.e.*, proteins, DNA and RNA from the heat-killed S-cells to see which ones could transform live R-cells into S-cells. They discovered that DNA alone from S-bacteria caused R-bacteria to become transformed.

   They also discovered that protein-digesting enzymes *i.e.*, protease and RNA-digesting enzymes *i.e.*, RNase did not affect transformation, so the transforming substance was not a protein or RNA. Digestion with DNase did inhibit transformation, suggesting that the DNA caused the transformation. They concluded that DNA is the hereditary material, but not all biologists were convinced.

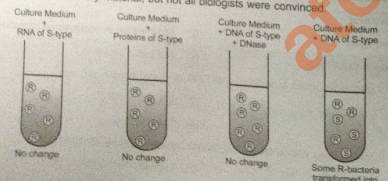


Fig. : Experiment of Avery *et. al.* to prove that DNA is transforming principle

2. **Evidence from Experiments with Bacteriophage:** The unequivocal proof that DNA is the genetic material came from the experiments of Alfred Hershey and Martha Chase (1952). They worked with virus (T_2 bacteriophage) that infects bacterium *Escherichia coli* and multiples inside it. T_2 phage is made up of DNA and protein coat. Thus, it is the most suitable material to determine whether DNA or protein contains information for the production of new virus particles.

   The functions of DNA and proteins could be found out by labelling them with radioactive tracers. DNA contains phosphorus but not sulphur. Therefore, phage DNA was labelled with P^{32} by growing bacteria infected with phages in culture medium containing $^{32}PO_4$. Similarly, protein of phage contains sulphur but no phosphorus. Thus, the phage protein coat was labelled with S^{35} by growing bacteria infected with phages in another culture medium containing $^{35}SO_4$. After labelling, three steps were followed *i.e.*, **infection, blending** and **centrifugation**.

   (i) **Infection :** Both types of labelled phages were allowed to infect normally cultured bacteria in separate experiments.

   (ii) **Blending :** These bacterial cells were agitated in a blender to break the contact between virus and bacteria.

   (iii) **Centrifugation :** The virus particles were separated from the bacteria by spinning them in a centrifuge.

   After the centrifugation the bacterial cells showed the presence of radioactive DNA labelled with P^{32} while radioactive protein labelled with S^{35} appeared outside the bacterial cells *i.e.*, in the medium. Labelled DNA was also found in the next generation of phage. Bacteria that were infected with viruses that had radioactive proteins were not radioactive. This indicated that proteins did not enter the bacteria from the viruses. DNA is therefore the genetic material that is passed from virus to bacteria.

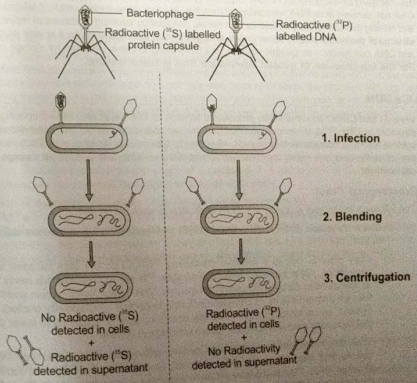


Fig. : The Hershey-Chase Experiment

Properties of Genetic Material (DNA versus RNA)

Now it is clear that the debate between proteins versus DNA as the genetic material was unequivocally resolved from Hershey-Chase experiment. However, it subsequently become clear that in some viruses RNA is the genetic material *e.g.,* Tobacco Mosaic viruses, QB bacteriophage, etc.

A molecule that can act as genetic material must fulfill the following criteria :

(i) It should chemically and structurally be stable.

(ii) It should be able to generate its replica (replication).

(iii) It should provide the scope for slow mutation that are required for evolution.

(iv) It should be able to express itself in the form of Mendelian characters.

The genetic material should be stable enough not to change with different stages of life cycle, age or with change in physiology of the organism. DNA being more stable is preferred as genetic material, as

(a) Free 2'OH of RNA makes it more labile and easily degradable. Therefore DNA in comparison is more stable.

(b) Presence of thymine (5-Methyl uracil) at the place of uracil also confers additional stability to DNA.

(c) RNA being unstable, mutates at a faster rate. Consequently, viruses having RNA genome can directly code for the synthesis of proteins, hence can easily express the characters.

RNA WORLD

RNA was the first genetic material. There are evidences to suggest that essential life processes, such as metabolism, translation, splicing etc. evolved around RNA. RNA used to act as a genetic material as well as a catalyst. There are some important biochemical reactions in living systems that are catalysed by RNA catalyst (ribozyme) and not by protein enzymes *e.g.,* Ribonuclease P (Cleavage), Snurps (Splicing), Peptidyl transferase (peptide bond formation). But, RNA being a catalyst was reactive and hence 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. RNA is **adapter, structural molecule** and in some cases **catalytic**. From above discussion it is very much clear both RNA and DNA can function as genetic material, but DNA being stable is preferred for storage of genetic material. **For the transmission of genetic information RNA is better material.**

REPLICATION

Watson and Crick had immediately proposed a scheme for DNA replication while proposing the double helical structure of DNA. The scheme suggested that the two strands would separate and act as template for the synthesis of new complementary strands. After the completion of replication, each DNA molecule would have one parental and one newly synthesised strand. This scheme was termed as **semiconservative DNA replication**.

The Experimental Proof

The following experiment suggests that DNA replication is semiconservative :

(A) Matthew Meselson and Franklin Stahl (1958) performed following experiment using heavy nitrogen (^{15}N) in *E. coli*.

(i) They grew *E. coli* in a medium containing $^{15}NH_4Cl$ as the only nitrogen source for many generations. ^{15}N is the heavy isotope of nitrogen. ^{15}N was incorporated into newly synthesised DNA as well as other nitrogen-containing compounds. This heavy DNA molecule could be distinguished from the normal DNA by centrifugation in a cesium chloride (CsCl) density gradient. A dense solution of CsCl, on centrifugation, forms density gradient bands of a solution of lower density at the top that increases gradually towards bottom with highest density.

(ii) Then they transferred the cells into a medium with normal $^{14}NH_4Cl$ and took samples at various definite time intervals as the cells multiplied, and extracted the DNA that remained as double-stranded helix. The various samples were separated independently on CsCl gradients to measure the densities of DNA.

(iii) Thus, the DNA that was extracted from the culture after 1st generation *i.e.,* just after 20 minutes had a hybrid or intermediate density. DNA extracted from the culture after another generation *i.e.,* 2nd generation or 40 minutes was composed of equal amounts of this hybrid DNA ($N^{14}N^{15}$) and of light DNA ($N^{14}N^{14}$). Increase in the amount of light DNA and decrease in hybrid DNA amount can be possible due to semiconservative mode of replication.

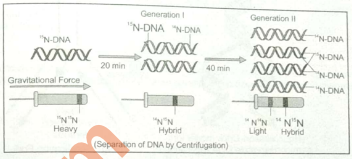


Fig. : Meselson and Stahl's Experiment

(B) Taylor *et. al.* have proved semiconservative mode of chromosome replication in eukaryotes using tritiated thymidine (^{3}H-thymidine) in root of *Vicia faba* (Faba beans).

Knowledge Cloud

Cairns proved semiconservative mode of replication in *E. coli* by using tritiated thymidine (H^3-tdR) in autoradiography experiment. He proposed θ-model for replication in circular DNA.

The Machinery and the Enzymes/DNA Replication Mechanism

The process of replication in living cells requires a set of enzymes. The main enzyme is referred to as **DNA-dependent DNA polymerase**. It is highly efficient with the ability to polymerise some 2000 bp per second. Not only do these polymerases have to be fast, but they also have to catalyse the reaction with high degree of accuracy. Any mistake during replication would result into mutations. The whole genome of *Escherischia coli* having 4.6×10^6 bp is replicated within **38** minutes. DNA replication completes in following steps :

(i) Origin of Replication : Replication begins at a particular region of DNA which is called origin of replication. It is because of the requirement of the origin of replication that a piece of DNA if needed to be propagated during recombinant DNA procedures, requires a vector. The vectors provide the origin of replication. **Prokaryotes** have **single origin of replication**. It is called *ori-c* in *E. coli*. On the other hand, eukaryotes have **several thousands origins of replication**.

(ii) Activation of deoxyribonucleotides : Four types of deoxyribonucletides, namely, dAMP, dGMP, dTMP and dCMP are activated by phosphate, energy and enzyme **phosphorylase** into **triphosphate** state. **Deoxyribonucleoside triphosphates serve dual purposes. In addition to acting as substrates, they provide energy for polymerisation reaction, because the two terminal phosphates in a deoxynucleoside triphosphates are high energy phosphates, same as in case of ATP.**

$$\begin{matrix} dAMP \\ dGMP \\ dCMP \\ dTMP \end{matrix} + 2H_2PO_4 \xrightarrow[\text{Energy}]{\text{Phosphorylase}} \begin{matrix} dATP \\ dGTP \\ dCTP \\ dTTP \end{matrix}$$

(iii) Unwinding of helix : Unwinding of double helical parental molecule is brought about by enzyme **helicase**, which is ATP dependent.

Unwinding of DNA molecule into two strands results in the formation of Y-shaped structure, called **replication fork**. These exposed single strands are stabilised with the help of single strand binding

proteins (SSBP). Due to unwinding, a supercoiling gets developed on the end of DNA opposite to replicating fork. This tension is released by enzyme **topoisomerase**. In prokaryotes, **DNA gyrase** has **topoisomerase** activity.

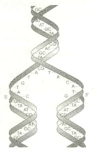


Fig. : Watson-Crick model for semiconservative DNA replication

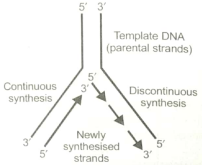


Fig. : Replicating Fork

(iv) **Formation of primer strand :** A new strand is now to be synthesised opposite to the parental strands. DNA polymerase III is the true replicase in *E. coli*, which is incapable of initiating DNA synthesis, *i.e.*, it is unable to deposit the first nucleotide in a daughter strand. Another enzyme, known as primase, synthesizes a short primer strand of RNA. The primer strand then serves as a stepping stone to start errorless replication. Once the initiation of DNA synthesis is completed, this primer RNA strand is then removed enzymatically.

(v) **Elongation of new strand :** The DNA dependent DNA polymerases catalyse polymerisation only in one direction, that is 5′ → 3′. This creates some additional complications at the replicating fork. Consequently, the replication is **continuous** on one template strand with polarity 3′ → 5′. It is now known as leading daughter strand. The replication is **discontinuous** in the form of short **Okazaki fragments** on other template strand with polarity 5′ → 3′, this is called lagging daughter strand. The discontinuously synthesised fragments are latter joined by the enzyme **DNA ligase.**

In eukaryotes, the replication of DNA takes place at S-phase of cell-cycle. The replication of DNA and cell division cycle should be highly coordinated. A failure in cell division after DNA replication results into polyploidy.

Knowledge Cloud

1. DNA polymerases are of 3 main types *i.e.*, DNA polymerase I, II and III in prokaryotes. All have exonuclease as well as polymerase activity.

	Exonuclease activity	Polymerase activity
DNA polymerase I (Kornberg enzyme) :	5′ → 3′ [removal of RNA primer & T = T (thymine dimer) formed by UV rays.] and 3′ → 5′	5′ → 3′ Adds 1000 nucleotides per minute
DNA polymerase II :	3′ → 5′	5′ → 3′ Rate → 50 nucleotides/minute
DNA polymerase III :	3′ → 5′	5′ → 3′ with polymerisation rate of 2000 bp per second

2. In eukaryotes, DNA polymerases are of 5 types, these are DNA polymerase α, β, γ, δ and ε.
3. Synthesis of leading or continuous strand is fast with the help of single primer; while the synthesis of lagging or discontinuous strand is slow and requires many primers.
4. **Main polymerizing enzyme is DNA polymerase III.**

Example 5 : *Fill in the blanks :*

(a) Viruses grown in the presence of radioactive phosphorus contained radioactive _____ but not radioactive _____.

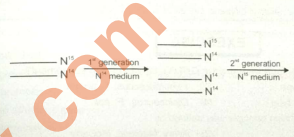
(b) RNA is labile and easily degradable due to the presence of _____ group in sugar.

Solution : (a) DNA, protein

(b) 2′OH

Example 6 : *If hybrid DNA is allowed to replicate for one generation in medium containing N^{14} and for second generation in medium containing N^{15} then what is the proportion of light, heavy and hybrid DNA obtained respectively?*

Solution :

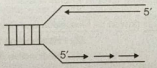


Heavy DNA = 25%

Hybrid DNA = 75%

Light DNA = 0%

Example 7 : *Consider the given diagram and answers the questions*



(a) What is the polarity of template strand which forms continuous complementary strand?

(b) Mention the polarity of template strand which forms Okazaki's fragments.

Solution : (a) 3′ → 5′ polarity in template strand

(b) 5′ → 3′ polarity in template strand

Example 8 : *During DNA synthesis in bacteria which of the following enzyme is not required?*

(1) DNA dependent DNA polymerase

(2) DNA dependent RNA polymerase

(3) RNA dependent DNA polymerase

(4) DNA gyrase

Solution : Answer (3)

RNA-dependent DNA polymerase is required for DNA synthesis in Retroviruses.

Try Yourself

3. Fill in the blanks :
   (a) DNA in chromosomes replicate semiconservatively was experimentally proved in *Vicia faba* by _____ using _____.
   (b) Deoxyribonucleoside triphosphates acts as _____ and they provide _____ for polymerisation.
   (c) The discontinuously synthesised fragments are joined by the enzyme _____ during DNA replication.
4. (i) What is the site of DNA replication in cell cycle?
   (ii) What will be the result of failure in cell division after DNA replication?

EXERCISE

11. Which of the following types of bacteria were used in Griffith's transformation experiment?
   (1) *Diplococcus*, R-III and S-II type (2) *Pneumococcus*, T_2 phage
   (3) *Streptococcus*, R-II and S-III type (4) *Diplococcus*, *E. coli*
12. The biochemical nature of transforming principle was defined by
   (1) Griffith (2) Avery, Macleod, McCarty
   (3) Watson and Crick (4) Taylor
13. In Hershey and Chase experiment, the protein of T_2 phage was made radioactive by using
   (1) S^{32} (2) P^{31}
   (3) S^{35} (4) P^{32}
14. Choose the correct option w.r.t. RNA.
   (1) Presence of thymine in place of uracil (2) Absence of free 2'OH in sugar
   (3) Mutates at faster rate (4) Is non-catalytic
15. Semiconservative DNA replication was proved by Messelson and Stahl, in which DNA was made
   (1) Radioactive using N^{15} (2) Heavy using N^{14}
   (3) Heavy using $^{15}NH_4Cl$ (4) Radioactive using $^{14}NH_4Cl$
16. During DNA replication, strand separation by breaking the H-bonds is performed by
   (1) Topoisomerase (2) Gyrase
   (3) Helicases (4) More than one option is correct
17. RNA primer is removed by
   (1) DNAP-I (2) DNAP-II
   (3) DNAP-III (4) Primase
18. How many types of DNA polymerases are associated with eukaryotic cell?
   (1) Three (2) Four
   (3) Five (4) Two
19. Which of the following acts as substrate as well as provide energy for DNA polymerisation?
   (1) Ribonucleoside (2) Deoxyribonucleoside
   (3) Ribonucleotide (4) Deoxyribonucleoside triphosphate
20. DNA replication is
   (1) Semi-conservative, continuous, unidirectional (2) Conservative, continuous
   (3) Semi-continuous, semi-discontinuous (4) Semi-continuous, conservative

TRANSCRIPTION

The process of copying genetic information from one strand of the DNA into RNA is known as **transcription**.

Like DNA replication, the **principle of complementarity governs the process of transcription**, except the adenosine which forms base pair with uracil instead of thymine.

But, unlike DNA replication where total DNA of an organism gets duplicated, in transcription only a segment of DNA and only one of the strands is copied into RNA. Here only one strand is template strand while in replication both strands are template.

There are two explanations for both the strands of DNA not being copied during transcription.

(1) If both strands act as template, they would code for RNA molecule with different sequences. And in turn, if they code for proteins, the sequence of amino acids in the proteins would be different. Hence, one segment of the DNA would be coding for two different proteins. This would complicate the genetic information transfer machinery.

(2) The two RNA molecules if produced simultaneously would be complementary to each other, hence would form a double-stranded RNA. This would prevent the translation of RNA into protein.

Transcription Unit

The segment of DNA that takes part in transcription is called **transcription unit**. It has three components :

(i) A promoter
(ii) The structural gene
(iii) A terminator

Template Strand and Coding Strand

There is a convention in defining the two strands of the DNA in the structural gene of a transcription unit. Since the two strands have opposite polarity and the DNA-dependent RNA polymerase also catalyse the polymerisation in only one direction *i.e.*, 5′ → 3′ polarity. The strand that has the polarity 3′ → 5′ acts as template, and is called **template strand** or **non-coding strand**. The other strand with polarity 5′ → 3′ and the sequence same as RNA, except thymine at the place of uracil, is displaced during transcription. And this strand is called **coding strand** or **sense strand** or **non-template strand**.

Structural genes are flanked on both sides by a promoter and a terminator in transcription unit.

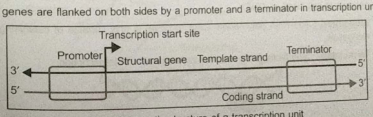


Fig. : Schematic structure of a transcription unit

Promoter sequences are present upstream towards 5′ end of the structural gene of transcription unit (the reference is made with respect to the polarity of coding strand). It is a DNA sequence that provides binding site for RNA polymerase. It is the presence of a promoter in a transcription unit that also defines the template and coding strands. By switching its position with terminator, the definition of template and coding strand could be reversed. The binding sites for RNA polymerase lie within the promoter sequence. Certain short sequences within the promotor sites are conserved, known as **recognition sequence**. (See knowledge cloud).

The terminator is present at 3′ end (downstream) of coding strand and it usually defines the end of the process of transcription.

Transcription Unit and the Gene

A gene is defined as the functional unit of inheritance. Genes are located on the DNA and it is difficult to literally define a gene in terms of DNA sequence. The DNA sequence coding for tRNA or rRNA molecule also define a gene. Cistron is defined as a functional unit of gene. It is a segment of DNA coding for a polypeptide.

The structural gene in a transcription unit is monocistronic (mostly in eukaryotes) and polycistronic (mostly in prokaryotes or bacteria). Monocistronic gene synthesises one type of polypeptide or protein. Polycistronic gene synthesises different proteins or polypeptides.

The monocistronic structural genes have interrupted coding sequences *i.e.* the genes in eukaryotes are split. The coding sequences or expressed sequences are defined as **exons** which appear in mature or processed RNA. The exons are interrupted by **introns**. Introns are intervening sequences that do not appear in mature or processed RNA. The split-gene arrangement further complicates the definition of a gene in terms of a DNA segment.

Types of RNA and Process of Transcription

There are three major types of RNA : mRNA (messenger RNA), tRNA (transfer RNA) and rRNA (ribosomal RNA).

S. No.	mRNA	rRNA	tRNA
1.	5% of total RNA in cell	80%	15%
2.	Longest	Smaller	Smallest
3.	It is called template/nuclear/ messenger or informational RNA as it carries genetic information provided by DNA	Has structural (forms ribosome) and catalytic role during translation	Soluble or adapter RNA and carries amino acids

Thus all three RNA's are needed to synthesise protein in a cell.

(A) Transcription in Prokaryotes : It occurs in cytoplasm with the help of trancripting enzyme.

The transcripting enzyme *i.e.*, DNA-dependent RNA polymerase is only of one type and transcribe all types of RNAs *i.e.*, mRNA, tRNA and rRNA. All three RNAs are needed to synthesize a protein in a cell.

RNA polymerase is a holoenzyme that is made of polypeptides ($\alpha_2\beta\beta'\omega$)$\sigma$. The enzyme without σ subunit is referred to as core enzyme. The process of transcription completes in 3-steps :

(i) **Initiation :** It is catalysed by sigma (σ) factor or initiation factor. It binds to the promoter site of DNA and confers specificity. In the absence of σ-factor, transcription starts non-specifically by core enzyme at any base on DNA.

(ii) **Elongation :** The RNA polymerase (core enzyme) is only capable of catalysing the process of elongation.

(iii) **Termination :** Rho factor (ρ) is required for termination of transcription.

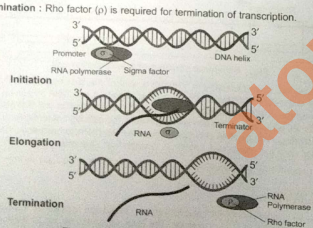


Fig. : Process of Transcription in Bacteria

RNA polymerase binds to promoter region of the DNA and the process of transcription begins. It uses nucleoside triphosphates as substrate and polymerises in a template-depended fashion following the rule of complementarity. It also helps in the opening of helix and continues elongation, only a short stretch of RNA is attached to the enzyme. Once the polymerase reaches the terminator region, the nascent RNA and RNA polymerase falls off and it results in termination of transcription.

Following points can be summarised for bacterial transcription :

(i) mRNA does not require any processing to become active.

(ii) Transcription and translation take place in the same compartment as there is no separation of cytosol and nucleus.

(iii) Many times the translation can begin much before the mRNA is fully transcribed. Thus, the transcription and translation can be coupled in bacteria.

(B) Transcription in Eukaryotes : There are three types of transcripting enzymes *i.e.* RNA polymerases in the nucleus in addition to RNA polymerase found in the organelles. There is a clear-cut division of labour. Functions of different RNA polymerases in eukaryotes are given below :

(i) **RNA polymerase I :** 5.8S, 18S, 28S rRNA synthesis

(ii) **RNA polymerase II :** hnRNA (heterogeneous nuclear RNA)

(iii) **RNA polymerase III :** tRNA, ScRNA, 5S rRNA and SnRNA (small nuclear RNA) synthesis

The nascent RNA synthesised by RNA polymerase II is called hnRNA or primary transcript. It contains both unwanted base sequences (introns) alternated with useful base sequences (exons).

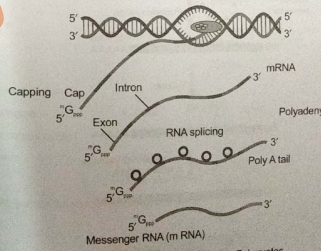


Fig. : Process of Transcription in Eukaryotes

This primary transcript is converted into functional mRNA after post-transcriptional processing which involves 3 steps :

(i) **Modification of 5′ end by capping :** Capping at 5′ end occurs rapidly after the start of transcription. An unusual nucleotide *i.e.* methyl guanosine triphosphate is added to the 5′-end of hnRNA. It is catalysed by guanyl transferase. Cap is essential for formation of mRNA-ribosome complex. Translation is not possible if cap is lacking because cap is identified by 18S rRNA of ribosome unit.

(ii) **Tailing and Splicing:** Tailing is the addition of adenylate residues about 200-300 at 3′ end in a template-independent manner on newly formed hnRNA with the help of Poly A polymerase. Splicing is the process of removal of introns and joining of exons in a defined order. Introns are removed by small nuclear RNA (SnRNA) and protein complex called small nuclear ribonucleoproteins or SnRNPs (Snurps).

The fully processed hnRNA is now called mRNA and it is transported out of the nucleus for translation.

The split-gene arrangements represent probably an ancient feature of genome. The presence of introns is reminiscent of antiquity, and the process of splicing represents the dominance of **RNA-world**.

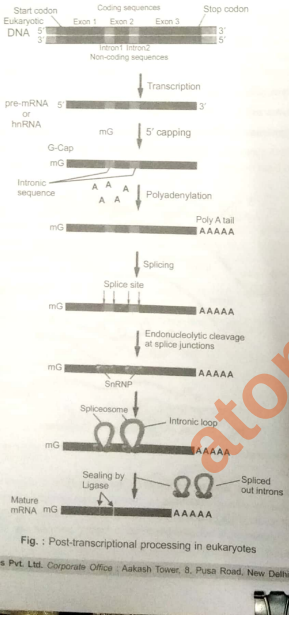


Fig. : Post-transcriptional processing in eukaryotes

EXERCISE

21. Which of the following is a genetic RNA?
(1) mRNA (2) rRNA
(3) hn-RNA (4) RNA present in plant viruses

22. The mRNA of prokaryotes is
(1) Polycistronic (2) Monocistronic
(3) Formed by splicing of hnRNA (4) Carries genetic message to DNA

23. Capping in hnRNA is catalysed by
(1) Poly A polymerase (2) SnRNA
(3) Guanyl transferase (4) Catalytic RNA

24. Which of the following type of ribosomal RNA is not present in eukaryotic cytoplasm?
(1) 18S (2) 28S
(3) 5.8S (4) 16S

25. Mark the correct option (w.r.t. function of RNAP-I)
(1) 5.8S rRNA (2) 5S rRNA
(3) SnRNA (4) ScRNA

26. Soluble RNA is
(1) tRNA (2) mRNA
(3) rRNA (4) hnRNA

27. Find the incorrect match
(1) Central dogma : F. Crick
(2) Reverse central dogma : Temin and Baltimore
(3) Split genes : Kornberg
(4) mRNA : Jacob and Monad

28. Recognition sequence for transcription in prokaryotes is
(1) TATATAT (2) TATAAT
(3) TATAAAT (4) CAAT

29. Transcription starts non-specifically in the absence of
(1) Sigma factor (2) Rho factor
(3) Core enzyme (4) DNA polymerase

30. Tailoring of hnRNA is done by
(1) Snurps (2) Introns
(3) Exons (4) 18 SrRNA

GENETIC CODE

DNA (or RNA) carries all genetic information. It is expressed in the form of proteins which are made up of 20 different types of amino acids. The information about the number and sequence of these amino acids forming protein is present in DNA and is passed on to mRNA during transcription. Thus, genetic code is inter-relationship between nucleotides sequence of DNA or mRNA and amino acids sequence in a polypeptide. It is a mRNA sequence containing coded information for one amino acid and consists of 3 nucleotides.

The proposition and deciphering of genetic code were most challenging. In a very true sense, it required involvement of scientists from several disciplines – physicists, organic chemists, biochemists and geneticists. It was **George Gamow**, a physicist, who coined the term genetic code and argued that since there are only 4 bases and if they have to code for 20 amino acids, the code should constitute a combination of bases. He suggested that in order to code for all the 20 amino acids, the code should be made up of 3 nucleotides. This was a very bold proposition, because a permutation combination of $4 \times 4 \times 4$ (4^3) would generate 64 codons, generating many more codons than required.

The important discovery was the result of experiments by **Marshall W. Nirenberg** and **J. Heinrich Matthaei** and later by **H.G. Khorana**. Nirenberg and Matthaei used a synthetic poly U RNA and deciphered the code by translating this as polyphenylalanine. The chemical method developed by Har Gobind Khorana was instrumental in synthesising RNA molecules with defined combinations of bases (homopolymers and copolymers). Using synthetic DNA, he prepared polynucleotide with known repeating sequence *e.g.*, CUCUCUCUCUCU, which produced only two amino acids, leucine (CUC) and serine (UCU).

Severo ochoa enzyme is **polynucleotide phosphorylase**, it was also helpful in **polymerising RNA** with defined sequences in a **template-independent manner** *i.e.*, enzymatic synthesis of RNA. Finally a checker-board for genetic code was prepared which is given below :

Table : The Codons for the Various Amino Acids

First position	Second Position U	Second Position C	Second Position A	Second Position G	Third position
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

Salient Features of Genetic Code

(i) **Triplet code :** Each codon is made of three adjacent nitrogen bases. 61 codons code for amino acids and 3 codons do not code for any amino acids, hence they function as stop codons.

(ii) **Non-ambiguous and specific codons :** One codon codes for only one amino acid, hence it is unambiguous and specific.

(**Note :** GUG is ambiguous codon, it normally codes for valine but at initiating position, codes for methionine)

(iii) **Commaless nature :** The codon is read in mRNA in a contiguous fashion without any punctuations.

(iv) **Degeneracy of code :** Some amino acids are coded by more than one codon, hence the code is degenerate *e.g.* serine, leucine, arginine by 6 codons, proline, valine, glycine, alanine and threonine by 4 codons etc.

Exception - AUG (Met.) and UGG (Trp.) are non-degenerate codons.

(v) **Universal code :** The code is nearly universal, *e.g.* UUU would code for phenylalanine in all organisms. Some exceptions to this rule have been found in mitochondria and protozoa. (See knowledge cloud).

(vi) **Initiation codon/start signal :** AUG has dual functions, it codes for methionine, and it also acts as initiator codon.

(vii) **Stop signals :** Polypeptide chain termination is signalled by three termination codons - UAA (ochre), UAG (amber) and UGA (opal). They do not specify any amino acids, hence called as nonsense codons or stop codons.

(viii) **Non-overlapping codon :** Each codon is independent and one codon does not overlap the next codon.

Mutations and Genetic Code

The relationship between genes and DNA are best understood by mutation studies. Effects of large deletions and rearrangements in a segment of DNA are easy to comprehend. It may result in loss or gain of a gene and so a function. A classical example of gene mutation or point mutation is a change of single base pair in the gene for beta-globin chain that results in the change of amino acid residue glutamate to valine. It results into a diseased condition called as **sickel cell anemia**. 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 mutations. This forms the genetic basis of proof that codon is a triplet and it is read in a contiguous manner.

tRNA – The Adapter Molecule

The existence of tRNA was postulated by Francis Crick. It was also known as soluble RNA (sRNA) before the genetic code was postulated. These constitute about 15% of the total cellular RNA.

Crick postulated the presence of an adapter molecule that would on one hand read the code and on other hand would bind to specific amino acid. It acts as intermediate molecule between triplet code of mRNA and amino acid sequence of polypeptide chain. All tRNA's have almost same basic structure. There are over 60 types of tRNA. The three-dimensional structure of the tRNA was proposed to be **inverted L-shaped** (by Kim and Klug). This is the actual structure of tRNA. The secondary structure of tRNA has been depicted that it looks like a **clover-leaf**. All tRNA molecules have a guanine residue at its 5′ terminal end. At its 3′ end, unpaired -CCA **sequence** is present. Amino acid gets attached at this end only. tRNAs are specific for each amino acid. For **initiation, there** is another specific tRNA that is known as initiator tRNA. There are no tRNAs for stop codons.

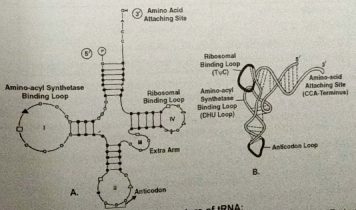


Fig. : Structure of tRNA:
A. Clover leaf model to show basic plan of tRNA secondary structure or 2D structure
B. Three-Dimensional Structure showing inverted L-shaped configuration

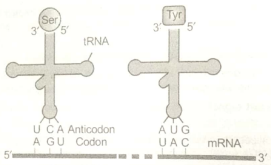


Fig. : tRNA – the adapter molecule

There are three loops in tRNA :

(i) Aminoacyl synthetase binding loop or DHU loop (dihydroxyuridine loop) - 1st loop from 5′ end.

(ii) Ribosomal binding loop with 7 unpaired bases - It is 1st loop from 3′ end also called as TΨC loop.

(iii) Anticodon loop with 7 unpaired bases. Out of the 7 bases in anticodon loop, 3 bases act as anticodon for a particular triplet codon present on mRNA.

Knowledge Cloud

1. In prokaryotes, recognition sequence is present in promotor region at upstream for RNA polymerase binding e.g., (i) **TATAAT** or **Pribnow box** (–10 sequence or 10 bp upstream from the start point) **(ii) TTGACA** : –35 sequence as recognition sequence.
2. In eukaryotes, **TATA box** or **Hogness box** (7 bp long - TATATAT or TATAAAT) located 20 bp upstream to the start point; another sequence is **CAAT box** present between –70 and –80 bp.
3. Split gene was discovered by **Richard J. Roberts** and **Philip Sharp**. Eukaryotic gene with exons and introns as intervening sequence is split gene.
4. Difference between universal genetic code and mitochondrial genetic code.

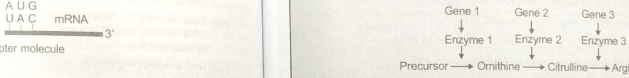
S. No.	Universal	Mitochondrial code (mammals, yeast)
1.	55 anticodons (tRNA)	22 anticodons (tRNA)
2.	3 termination codons	4 termination codons
	UAA, UAG, UGA	UAA, UAG, AGA, AGG
	(a) UGA = Termination codon	(a) UGA code for tryptophan
	(b) AGA, AGG code for arginine	(b) AGA and AGG are termination codons

5. **Wobble hypothesis :** A change in nitrogen base at the 3rd position of a codon does not normally cause any change in the expression of the codon because the codon is mostly read by the first two nitrogen bases. The position of the third nitrogen base in a codon which does not influence the reading of the codon is termed as Wobble position. It helps one tRNA to read more than one codon and thus, provides economy in number of tRNA molecules at the time of translation.
6. **Shine Dalgarno (SD) sequence :** It is 5′-AGGAGGU-3′ sequence at 5′-end near initiation codon in prokaryotes. It helps in binding of 30S subunit of ribosome on it.
7. Ribosomal RNA (rRNA) is most stable type of RNA and is constituent of ribosome.

   In eukaryotes, rRNAs are of 4 types *i.e.*, 5S, 5.8S, 28S and 18S.

   In prokaryotes - 5S, 23S, 16S types of rRNA.
8. **Genetic RNA** : RNA is genetic material in most plant viruses.

9. **Beadle and Tatum** put forward a theory–**one gene-one enzyme** in support of the earlier hypothesis that enzymes are proteinaceous in nature and each is produced by a single gene. They conducted experiments on the nutritional strains of pink mould, *Neurospora crassa*. This fungus grows on simple **prototroph**. An organism that is unable to synthesize a particular cellular metabolite, such as an amino acid or coenzyme is called **auxotroph**. Beadle and Tatum produced arginine (an amino acid) auxotrophs (mutants of *Neurospora* unable to synthesize arginine) by giving X-rays treatment to the cells. **Arginine synthesis** passes through the following path :



They found that any step of this metabolic chain could be blocked by a mutation in a specific enzyme catalyzing the reaction, each enzyme representing a different gene product. Thus, Beadle and Tatum reached a **conclusion that each gene functions to produce a single enzyme.**

Some proteins, *e.g.*, haemoglobin and other quaternary proteins are made up of two or more than two polypeptide chains. **Ingram** suggested the "**one gene-one polypeptide hypothesis** to explain the genetic determination of synthesis of the peptide chains of the haemoglobin. Later **Jacobson and Baltimore** proposed **one mRNA – one polypeptide** hypothesis.

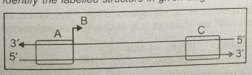
Example 9 : *If the sequence of coding strand in a transcription unit is written as follows:*

5′ - C G T A T C G A T C G G T T A C G A - 3′

Write down the sequence of complementary strand in 3′ → 5′ direction.

Solution : Complementary strand : 3′ – G C A T A G C T A G C C A A T G C T – 5′

Example 10 : *Identify the labelled structure in given diagram :*



Solution : A – Promoter

B – Transcription start site

C – Terminator

Example 11 : *When the polymerase reaches the terminator region, the nascent RNA and enzyme are released by which polypeptide molecule?*

Solution : Rho-factor (ρ-factor)

Example 12 : *Select incorrectly matched pair :*

(1) Catalytic RNA – Soluble RNA

(2) Snurps – SnRNA

(3) Capping – Guanosyl transferase

(4) Ambiguous codon – GUG

Solution : Answer (1)

Catalytic RNA – 23S/28S rRNA as the structural part of peptidyl transferase (ribozyme)

Try Yourself

5. What should be the nature of genetic code if there would have been 65 amino acids?
6. Fill in the blanks :
   (a) The transcription and translation can be coupled in _____.
   (b) _____ do not appear in mature or processed RNA.
   (c) The terminator is located at _____ end of non-template strand.
   (d) The codon is read in mRNA in a contiguous fashion, there are no _____.
   (e) Gene mutation involving insertion or deletion of one nitrogen base is _____ mutation.

TRANSLATION

It refers to the polymerisation 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 cellular factory responsible for synthesising proteins is the ribosome. The ribosome consists of structural RNAs and about 80 different proteins. In its inactive state, it exists as two sub units, a large subunit and a small sub unit. Ribosomes have two sites for binding amino acyl tRNA, P-site (peptidyl site) and A-site (aminoacyl). When the small subunit encounters an mRNA, the process of translation of the mRNA to protein begins.

The steps of translation mechanism are :

(a) Activation of amino acids : In the presence of enzyme aminoacyl - tRNA synthetase (E), specific amino acid (AA) bind with ATP.

$$AA_1 + ATP \xrightarrow{E,\ Mg^{2+}} AA_1\text{-AMP-}E_1 \text{ complex} + PPi$$

(b) Charging of tRNA : The AA_1-AMP-E_1 complex reacts with specific tRNA. Thus, amino acid is transferred to tRNA. As a result, the enzyme and AMP are liberated. It is also called as **aminoacylation** of tRNA.

$$AA_1 - AMP\text{-}E_1 \text{ complex} + tRNA_1 \longrightarrow AA_1 - \underset{(\text{Charged tRNA})}{tRNA_1} + AMP + E_1$$

(c) Formation of polypeptide chain : It completes in three steps :

(1) Chain initiation : It requires 3 initiation factors in prokaryotes and 9 initiation factors in eukaryotes.

(i) Binding of mRNA with smaller subunit of ribosomes (30S/40S)

$$30S + mRNA \longrightarrow 30S\text{-mRNA complex}$$

In eukaryotes, there is formation of 40S-mRNA complex.

(ii) Binding of 30S-mRNA with $tRNA_{Meth.}$, non-formylated methionine is attached with tRNA in eukaryotes and formylated in prokaryotes.

$$30S\text{-mRNA} + tRNA_{t\text{-}Met} \xrightarrow{GTP} 30S\text{-mRNA-}tRNA_{t\text{-}Met}$$

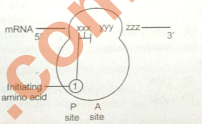
(iii) Attachment of larger subunit of ribosome. It is 50S in prokaryotes & 60S in eukaryotes.

(2) Chain elongation : After the formation of complete ribosome - mRNA-tRNA complex, an aminoacyl acceptor site (A-site) is established next to the P-site. It exposes mRNA codon next to the initiation codon. A new aminoacyl tRNA complex reaches the A-site and forms codon – anticodon bonding. This requires elongation factor and energy *i.e.*, GTP. A peptide bond is formed between COOH group of first amino acid (methionine) and NH_2 group of second amino acid. If two charged tRNAs are close enough the formation of peptide bond between them would be favoured energetically. The

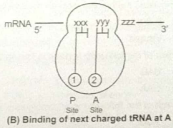
presence of a catalyst would enhance the rate of peptide bond formation. It is catalysed by enzyme peptidyl transferase (a type of ribozyme – catalytic RNA *i.e.*, 23S rRNA in bacteria and 28 SrRNA in eukaryotes). The elongation factors are required in this process. Translocation is movement of ribosome on mRNA. The ribosome move from codon to codon along the mRNA. Amino acids are added one by one, translated into polypeptide sequences **dictated by DNA** and **represented by mRNA**.

A translational unit in mRNA is the sequence of RNA that is flanked by the start codon (AUG) and the stop codon and codes for a polypeptide. An mRNA also has some additional sequences that are not translated and are referred as **untranslated regions (UTRs)**. The UTRs are present at both 5′-end before start codon and 3′-end after stop codon. They are required for **efficient translation process.**

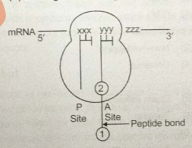
(3) Chain termination : The termination of polypeptide is signalled by one of the three termination codons (UAA, UAG, UGA). A GTP-dependent factor known as release factor is associated with termination codon. It is eRF_1 in eukaryotes and RF_1 and RF_2 in prokaryotes, that help in terminating translation and releasing the complete polypeptide from the ribosome.



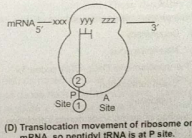
(A) Binding of first charged tRNA at P site



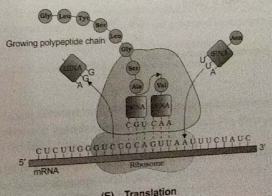
(B) Binding of next charged tRNA at A site



(C) Peptide bond formation by peptidyl transferase



(D) Translocation movement of ribosome on mRNA, so peptidyl tRNA is at P site. A site is open for the next charged t-RNA



(E) Translation

Knowledge Cloud

There are some metabolic inhibitors which inhibit the synthesis of proteins in bacteria. These antibiotics are used in checking the growth of certain bacteria which are pathogenic in nature e.g.,

Tetracycline	-	Inhibits binding of aminoacyl-tRNA to ribosomes
Streptomycin	-	Inhibits initiation of translation and causes misreading
Chloramphenicol	-	Inhibits peptidyl transferase activity
Erythromycin	-	Inhibits translocation of mRNA along ribosomes
Neomycin	-	Inhibits interaction between tRNA and mRNA.

EXERCISE

31. Formylated methionine acts as translation initiation in
(1) Eubacteria (2) Eukaryotes
(3) Viruses (4) Archaebacteria
32. Which of the following codons is known as ochre?
(1) UAG (2) UGA
(3) UAA (4) UUU
33. Which of the following is an ambiguous codon?
(1) AUG (2) GUG
(3) UAG (4) GAG
34. Which property of genetic code is utilised in wobble hypothesis?
(1) Degeneracy (2) Non-overlapping
(3) Non-ambiguous (4) Universal
35. In the mitochondrial DNA, UGA codes for
(1) Chain termination (2) Chain initiation
(3) Tryptophan (4) Tyrosine
36. Activation of amino acids during translation is done by
(1) Peptidyl transferase (2) Aminoacyl-tRNA synthetase
(3) Methionine (4) Initiation factors
37. Movement of ribosome on mRNA is called
(1) Transcription (2) Translation
(3) Translocation (4) Protein synthesis
38. The elongation factors required for prokaryotes are
(1) EF-Tu and EF-Ts (2) eEF_1
(3) eIF_2 (4) eEF_2
39. Which of the following inhibits binding of amino-acyl tRNA to ribosomes?
(1) Neomycin (2) Erythromycin
(3) Streptomycin (4) Tetracycline
40. The mechanism by which a gene is able to express itself in the phenotype of an organism is called
(1) Gene expression (2) RNA synthesis
(3) Translocation (4) Formylation

REGULATION OF GENE EXPRESSION

Regulation of gene expression refers to a very broad term that may occur at various levels. Considering that gene expression results in the formation of a polypeptide, it can be regulated at several levels.

In eukaryotes, the regulation of gene expression could be exerted at four levels.

(i) Transcriptional level : Formation of primary transcript.

(ii) Processing level : Regulation of splicing.

(iii) Transport of mRNA from nucleus to the cytoplasm.

(iv) Translational level.

The genes in a cell are expressed to perform a particular function or a set of functions. In eukaryotes, functionally related genes do not represent an operon but are present on different sites, chromosomes. Here structural gene is called split gene which is a mosaic of exons and introns, *i.e.*, the base triplet - amino acid matching is not continuous. The entire split gene is transcribed to form a continuous strip of mRNA. The removal of non-coding intronic part and fusion of exonic coding parts of RNA is called RNA splicing. About 50-90% of primary transcribed RNA is discarded during processing. The development and differentiation of embryo into adult organisms are also a result of the coordinated regulation of expression of several sets of genes.

It is metabolic, physiological or environmental condition that regulates the expression of gene.

Britten-Davidson gene battery model: It is most popular for eukaryotic genes expression. It proposes the occurrence of 5 types of genes - producer, receptor, integrator, sensor and enhancer silencer.

In prokaryotes, control of the rate of transcriptional initiation is the predominant site for control of gene expression. In a transcription unit, the activity of RNA polymerase at a given promoter is regulated by interaction with accessory proteins, which affects its ability to recognise start sites. These regulatory proteins can act both activators (positively) and repressor (negatively). The functioning of operator depends upon the protein products.

Operon Concept

Francois Jacob (a geneticist) and Jacques Monod (a biochemist) proposed a model of gene regulation, known as operon model in bacteria. Operon is a co-ordinated group of genes such as structural gene, operator gene, promoter gene, regulator gene which function together and regulate a metabolic pathway as a unit, *e.g.*, *lac* operon, *trp* operon, *ara* operon, *his* operon, *val* operon etc.

(i) Regulator gene : It synthesises a biochemical or regulator protein which can act positively as activator and negatively as repressor. It control, the activity of operator gene.

(ii) Operator gene : It is a gene which receives the product of regulator gene. It allows the functioning of the operon when it is not covered by the biochemical produced by regulator gene.

(iii) Promoter gene : Provides attachment site for RNA polymerase.

(iv) Structural gene : Transcribes mRNA for polypeptide synthesis.

The Lac Operon

The *lac* operon (*lac* refers to *lactose*) consists of one **regulatory gene** or **inhibitor gene** (*i*), one **promoter gene**, one **operator gene** and three **structural genes**. A polycistronic structural gene is regulated by a common promoter and regulatory gene.

In *E. coli*, breakdown of lactose requires three enzymes. These enzymes are synthesised together in a co-ordinated manner by functional unit of DNA i.e., *lac* operon. Since the addition of lactose itself stimulates the production of required enzymes, thus it is called **inducible system**.

ac Operon Genes

1. **Structural genes** : Three structural genes are :
   (i) ***lac z*** : The *z* gene codes for β-galactosidase which is primarily responsible for the hydrolysis of the disaccharide, lactose into its monomeric units galactose and glucose.
   (ii) ***lac y*** : The *y* gene codes for permease, which increases permeability of the cell to β-galactosides.
   (iii) ***lac a*** : The *a* gene codes for transacetylase which can transfer acetyl group to β-galactoside.
2. **Operator gene** : It interacts with a protein molecule or regulator molecule, which prevents the transcription of structural genes.
3. **Promoter gene** : The gene possess site for RNA polymerase attachment.
4. **Regulator gene (*i*)** : The gene codes for a protein known as **repressor protein**, it is synthesised all the time from the *i*-gene, thats why it is constitutive gene which is functional always.

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

Repressor protein + Operator gene → Switched off Negative

Regulation of *lac* operon by repressor is referred to as **negative control or regulation.**

If lactose is provided in the growth medium of the bacteria, the lactose is transported into the cells through the action of permease. A very low level of expression of *lac* operon has to be present in the cell all the time, otherwise lactose cannot enter the cells. In the presence of an inducer, such as lactose or allolactose, the repressor is inactivated by interaction with inducer. This allows RNA polymerase access to the promoter and transcription proceeds.

Inducer (Lactose) + Repressor → Switched on Positive

Lac operon is under control of positive regulation as well.

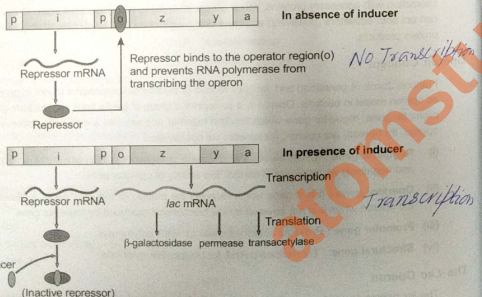


Fig. : The *lac* Operon

Tryptophan Operon - Repressible Operon System

Operon model can also be explained using feed-back repression. In tryptophan (*trp*) operon, three enzymes are necessary for the synthesis of amino acid tryptophan. These enzymes are synthesized by the activity of five different genes in a co-ordinated manner. The addition of tryptophan, however, stops the production of these enzymes. Thus, the system is known as **repressible system**.

In this system, there are five structural genes, trp A, trp B, trp C, trp D and trp E, coding for three enzyme needed for the synthesis of tryptophan, an amino acid. Regulatory gene (R) produces repressor protein which is known as **apo-repressor** because it does not get bound to the operator directly. Hence, the operator gene remains in '**switched on**' position.

Tryptophan when added, binds to the apo-repressor and is called **co-repressor**. This apo-repressor and co-repressor complex (**activated repressor**) now binds to operator gene and blocks the function of RNA polymerase. Thus, transcription would not occur and tryptophan operon would be in '**switched off**' position. Feed-back repression is functional when there is no further need of the end product and, hence, there is no requirement of continuation of this anabolic pathway. This operon stops the process.

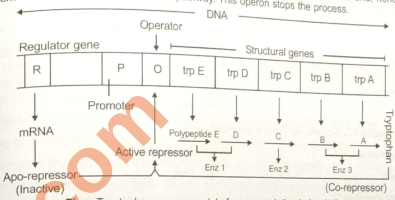


Fig. : Tryptophan operon model of gene regulation in bacteria

In the prokaryotes, control of the rate of transcriptional initiation is the predominant site for control of gene expression.

Knowledge Cloud

1. **Constitutive genes** are those genes which are constantly expressing themselves in a cell because their products are required for the normal cellular activities, *e.g.*, genes for glycolysis, ATPase., i gene as *lac* operon.
2. **Nonconstitutive genes** or **Luxury genes** are not always expressing themselves in a cell. They are switched on or off according to the requirement of cellular activities, *e.g.*, gene for nitrate reductase in plants. Lactose system in *E. coli*. (z, y, a gene)
3. *Trp* operon is repressible operon with five structural genes. The operon normally remains switched on and functional in anabolic pathway for tryptophan synthesis. Feedback repression occurs when there is no need of end products and their concentration is increased.
4. **Homeotic Genes**
   - These genes **control development**.
   - Such genes have been studied extensively in Drosophila.
   - The control is exerted through **homeodomain proteins**.
   - These proteins are formed by a conserved sequence or **homeobox**.
   - Homeobox are present **in the coding region** of homeotic genes
   - Mutation in homeotic genes results in **conversion of one body part into another** e.g., an insect leg may change into antenna.
5. **Oncogenes** : The tumour forming property of cells is due to presence of tumour forming genes known as **oncogenes**. Oncogenes present in virus are known as **v-oncogene** while those present in host cell are known as **c-oncogene**. **c-oncogene** are usually present in the form of **proto-oncogene**. Any mutation may convert these genes into oncogenes.

EXERCISE

41. The genes which are constantly expressing themselves in cell are called as
(1) Luxury genes (2) Constitutive genes
(3) Non-constitutive genes (4) More than one option is correct

42. How many structural genes are present in *lac*-operon of *E. coli*?
(1) 4 (2) 3
(3) 2 (4) 1

43. In *lac*-operon, β-galactosidase enzyme is made by
(1) *lac*-y (2) *lac*-a
(3) *lac*-z (4) *lac*-i

44. Inducer molecule in *lac*-operon of *E coli* is chemically a/an
(1) Disaccharide (2) Amino acid
(3) Protein (4) RNA

45. Tryptophan operon is
(1) Catabolic system (2) Repressible system
(3) Inducible system (4) Having three structural genes

46. Choose the correct option w.r.t. the chemical nature of apo-repressor and co-repressor respectively in *trp*-operon?
(1) Protein, Amino acid (2) Amino acid, Protein
(3) Lipoidal, Sugary (4) Sugary, Lipoidal

47. Gene battery model was proposed by
(1) Jacob and Monad (2) Gamow
(3) H.G. Khorana (4) Britten and Davidson

48. An insect leg may change into antenna due to mutation in
(1) c-oncogene (2) v-oncogene
(3) Homeotic genes (4) Proto-oncogene

49. In repressible operon system, co-repressor molecule is
(1) Lactose (2) Tryptophan
(3) Galactoside (4) Glucose

50. Select incorrectly matched pair
(1) Lac z – Constitutive gene
(2) Operator gene – Smallest gene of lac operon
(3) Lac a – Transacetylase
(4) Promotor gene – RNA polymerase

HUMAN GENOME PROJECT (HGP)

Genetic make-up of an organism or an individual lies in the DNA sequences. The differences in two individuals will naturally be reflected in the differences of their nucleotide sequences. They can be known only if the entire human genome is mapped. With the establishment of genetic-engineering techniques where it was possible to isolate and clone any piece of DNA and availability of simple and fast techniques for determining DNA sequences, a very ambitious project of sequencing human genome was launched in the year 1990.

HGP was the international, collaborative research program whose goal was the complete mapping and understanding of all the genes of human beings. All genes together of haploid set of chromosomes are known as genome.

Human genome project as "Mega project" was a 13-year-project, co-ordinated by the US Department of Energy and the National Institute of Health. Soon Welcome Trust (UK) joined the project as major partner, additional contributions came from Japan, France, Germany, China and others. The project was completed in 2003. HGP has been called a megaproject due to

(i) Huge cost estimated to be 9 billion US dollars, the cost of sequencing 1 bp is US$3.
(ii) Very large number of base pairs (3×10^9 bp) to be identified and sequenced.
(iii) Requires a large number of scientists, technicians and supporting staff.
(iv) Storage of data generated which requires some 3300 books, each with 1000 pages and each page having 1000 typed letters. However, high-speed computational devices for storage, retrieval and analysis of data made it easier to do the same.
(v) The science of **Bioinformatics** also developed during this period and helped HGP.

Goals of HGP

Following are the important goals of HGP :

(i) Identification of all the approximately 20,000-25,000 genes in human DNA.
(ii) To determine the sequences of the 3 billion chemical base pairs that make up human DNA.
(iii) To store this information in databases.
(iv) To improve tools for data analysis.
(v) Transfer-related technologies to other sectors, such as industries.
(vi) **ELSI :** To solve any ethical, legal and social issues.
(vii) **Bioinformatics** *i.e.*, close association of HGP with the rapid development of a new area in biology.
(viii) **Sequencing of model organisms :** Non-human organisms DNA sequences can lead to an understanding of their natural capabilities that can be applied towards solving challenges in health-care, agriculture, energy production, environmental remediation. Many non-human model organisms such as bacteria, yeast, *Caenorhabditis elegans* (a free-living non-pathogenic nematode), *Drosophila*, plants like rice and *Arabidopsis*, etc., have been sequenced. As for examples :

Organisms	Base pairs	No. of genes
E. coli	4.7 million	4,000
Saccharomyces cerevisiae	12 million	6,000
Caenorhabditis elegans	97 million	18,000
Drosophila melanogaster	180 million	13,000
Arabidopsis	130 million	25,000
Oryza sativa	430 million	32000 – 50000

Methodologies

The methods involved two major approaches :

(i) **ESTs/Expressed Sequence Tags :** Identifying all genes that are expressed as RNA.

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

For sequencing, the total DNA from a cell is isolated and converted into random fragments of relatively smaller sizes (recall DNA is a very long polymer, and there are technical limitations in sequencing very long pieces of DNA) and cloned in suitable host using specialised vectors. The cloning resulted into amplification of each piece of DNA fragment so that it subsequently could be sequenced with ease. The commonly used hosts were bacteria and yeast, and the vectors were called as **BAC** (bacterial artificial chromosomes), and **YAC** (yeast artificial chromosomes).

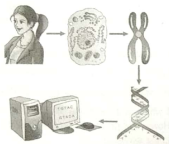


Fig. : A representative diagram of human genome project

The fragments were sequenced using automated DNA sequencers that worked on the principle of a method developed by **Frederick Sanger**. Sanger is also credited for developing method for determination of amino acid sequence in proteins. These sequences were then arranged based on some overlapping regions present in them. This required generation of overlapping fragments for sequencing. Alignment of these sequences was humanly not possible. Therefore, specialised computer-based programs were developed. These sequences were subsequently annotated and were assigned to each chromosome. The sequence of chromosome 1 was completed only in May 2006 (this was the last of the 24 human chromosomes – 22 autosomes and X and Y – to be sequenced).

Salient Features of Human Genome

Some of the salient observations drawn from human genome project are as follows :

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

(ii) The average gene consists of 3000 bases, but size varies greatly, with the largest known human gene being dystrophin as 2.4 million bases and TDF gene as smallest gene with 14 bases.

(iii) The total number of genes is estimated at 30,000 much lower than previous estimates of 80,000 to 1,40,000 genes. Almost all (99.9 percent) nucleotide bases are exactly the same in all people.

(iv) The functions are unknown for over 50 percent of discovered genes.

(v) Less than 2 percent of the genome codes for proteins.

(vi) Repeated sequences make up very large portion of the human genome.

(vii) Repetitive sequences are stretches of DNA sequences that are repeated many times, sometimes hundred to thousand times. They are thought to have no direct coding functions, but they shed light on chromosome structure, dynamics and evolution.

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

(ix) Scientists have identified about 1.4 million locations where single base DNA differences occur in humans. This is known as SNPs - **single nucleotide polymorphisms**, pronounced as '**snips**'. This information promises to revolutionise the process of finding chromosomal locations for disease-associated sequences and tracing human history.

Applications and Future Challenges

1. Completion of first phase of human genome project has been compared to discovery of antibiotics because it has opened a vast data base of knowledge about various aspects of human genome.
2. Soon we shall be mapping all the human genes, all sequences, transposons and junk DNA.
3. There are more than 1200 genes that cause common cardiovascular ailments, endocrine disease like diabetes, Alzheimer's disease, cancers and other neurological ailments. After taking their snapshots, it will be possible to know the method to alter them and remove the possibility of the disorders.
4. Single gene defects produce a number of hereditary diseases, that can be corrected.
5. It will be possible to study interactions between various genes, proteins, as well as mechanism of forming tissues, organs, tumours or switch over to different developmental stages.
6. It holds promise of healthier and longer living, designer drugs and genetically modified diets according to needs of individual human beings.

DNA FINGERPRINTING

It is the technique used for determining nucleotide sequences of certain areas of DNA which are unique to each individual. DNA fingerprinting can distinguish one human being from another with the exception of monozygotic twins. 99.9 percent of base sequences among humans is the same. They have 0.1% of genome or 3×10^6 differences in the base sequence. The differences occur not only in genes but also in repetitive DNA.

DNA fingerprinting involves identifying differences in some specific regions in DNA sequence called as repetitive DNA. 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.

Satellite DNA is classified into following categories on the basis of **base composition** (A : T rich or G : C rich), **length of segment** and **number of repetitive** units. These categories are :

1. **VNTRs** (Variable Number of Tandem Repeats) or **minisatellites** surrounded by conserved restriction sites. A small DNA sequence is arranged tandemly in many copy numbers. The copy numbers varies from chromosome to chromosome in an individual. The number of repeats show very high degree of polymorphism. As a result the size of VNTR varies from 0.1 to 20 kb.
2. **SSRs** (Single Sequence Repeats) or **STRs** (Short Tandem Repeats) or **microsatellites** with 1 – 6 bp.

These sequences normally do not code for any proteins, but they form a large portion of human genome. These sequences show high degree of polymorphism and form the basis of DNA fingerprinting. **Polymorphism** is the variation at genetic level. Since DNA from every tissue (such as blood, hair-follicle, skin, bone, saliva, sperm etc.), from an individual show the same degree of polymorphism, they become very useful identification tool in forensic applications. Further, as the polymorphisms are inheritable from parents to children, DNA fingerprinting is the basis of paternity testing, in case of disputes.

As polymorphism in DNA sequence is the basis of genetic mapping of human genome as well as of DNA fingerprinting, it is essential that we understand that what DNA polymorphism means in simple terms. Polymorphism (variation at genetic level) arises due to mutations. Allelic sequence variation has traditionally been described as a DNA polymorphism if more than one variant (allele) at a locus occurs in human population with frequency greater than 0.01. In simple terms, if an inheritable mutation is observed in a population at high frequency it is referred to as DNA polymorphism. The probability of such variation to be observed in non-coding DNA sequence would be higher as mutations in these sequences may not have any immediate effect in an individual reproductive ability. These mutations keep on accumulating generation after generation and form one of the basis of variability/polymorphism. There is a variety of different types of polymorphisms ranging from single nucleotide change to very large scale changes. For evolution and speciation, such polymorphism play very important role.

Thus, the basis of DNA fingerprinting is **VNTR (a satellite DNA as probe that shows very high degree of polymorphism)**. The technique of DNA fingerprinting was developed by **Alec Jeffreys**. The technique, as used earlier, involved Southern blot hybridisation using radiolabelled VNTR as probe. It included :

(i) Isolation of DNA.

(ii) Digestion of DNA by restriction endonucleases.

(iii) Separation of DNA fragments by electrophoresis, or **(RFLP Restriction Fragment Length Polymorphism)**

(iv) Transferring (blotting) of separated DNA fragments to synthetic membranes, such as nitrocellulose or nylon,

(v) Hybridisation using labelled VNTR probe.

(vi) Detection of hybridised DNA fragments by autoradiography.

After hybridisation with VNTR probe, the autoradiogram gives many bands of different sizes. These bands give a characteristic pattern for an individual DNA.

The sensitivity of the technique has been increased by use of polymerase chain reaction (PCR). Consequently, DNA from a single cell is enough to perform DNA fingerprinting analysis.

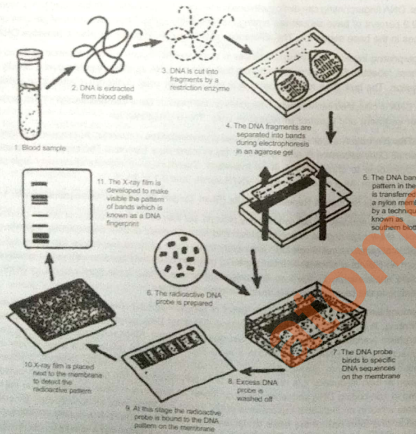


Fig. : The DNA fingerprinting process

Practical Applications :

1. Paternity-maternity disputes
2. Criminal identification and forensics

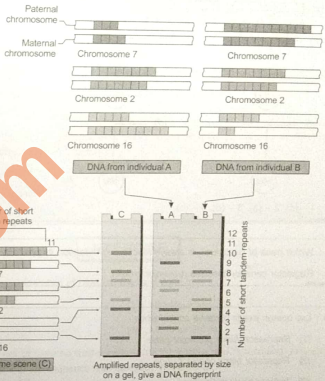


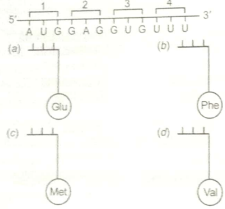
Fig. : Schematic representation of DNA fingerprinting

3. Personal identification
4. Close relations of an intending immigrant

Knowledge Cloud

1. The term genomics was introduced by **Thomas Roderick**.
2. **Structural genomics** involves mapping and sequencing of genes.
   Functional genomics involves the identification of function of a particular gene.
   Application genomics involves use of genomics information for crop improvement etc.
3. **cDNA :** It stands for complementary DNA, a synthetic type of DNA generated from mRNA. By using mRNA as template, scientists use enzymatic reactions to convert its information back into cDNA and then clone it as cDNA library.
4. **Mb :** Mb stands for megabase, a unit of length equal to 1 million base pairs and roughly equal to 1 cM.
5. **Lalji Singh** and **V.K. Kashyap** are Indian experts in the field of DNA fingerprinting.
6. **Microarray :** Micro arrays are devices used in many types of large scale genetic analysis. They can be used to study how large number of genes are expressed as messenger RNA in a particular tissue and how a cell's regulatory network control vast batteries of gases simultaneously.

Example 13 : *Arrange charged tRNA molecules according to given mRNA sequence*



Solution : 1 – (c), 2 – (a), 3 – (d), 4 – (b)

Example 14 : *Which gene of lac operon is always functioning? Mention its product.*

Solution : Regulator gene synthesises repressor protein.

Example 15 : *Lac operon exerts negative control when*

(1) Repressor binds promoter gene

(2) Repressor binds operator gene

(3) Inducer binds repressor

(4) Repressor binds with structural gene

Solution : Answer (2)

When repressor binds operator gene.

Example 16 : *State True or False :*

(a) VNTR is non-radioactive and probe is radioactive.

(b) Less than 2% of genome contain non-coding sequences.

(c) Sequencing of whole genome with both coding and non-coding regions is ESTs.

(d) Transfer of VNTR from gel to nylon paper is blotting technique.

Solution : (a) – True

(b) – False

(c) – False

(d) – True

Try Yourself

7. Select true or false statement :
   (a) Attachment of smaller unit of ribosome on mRNA brings the initiation codon at A site.
   (b) Peptidyl transferase is RNA enzyme formed by 23S rRNA of larger subunit of 70S ribosome.
   (c) Gene regulation is exerted at four levels in eukaryotes.
   (d) VNTR varies in size from 0.01 to 200 kb.
8. Mention the correct sequence of steps followed after separation of DNA fragments by electrophoresis in DNA fingerprinting.
   (a) Hybridization
   (b) Autoradiography
   (c) Blotting

EXERCISE

51. During DNA fingerprinting, separation of DNA fragments is done by
   (1) Autoradiography (2) Hybridisation
   (3) Denaturation (4) Electrophoresis
52. Sequencing the whole set of genome that contained all the coding and non-coding sequences and later assigning different regions in the sequence with functions is known as
   (1) Sequence annotation (2) PCR
   (3) Northern blot (4) Microarray
53. The last step of DNA fingerprinting is
   (1) Blotting (2) Autoradiography
   (3) Hybridisation (4) Isolation of desired DNA
54. DNA fingerprinting can be used
   (1) To solve cases of disputed paternity and maternity
   (2) For criminal identification and forensics
   (3) For personal identification
   (4) More than one option is correct
55. Human genome is said to have approximately
   (1) 3×10^9 bp (2) 3×10^6 bp
   (3) 6.6×10^6 bp (4) 3.3×10^6 bp

56. How many total number of genes are found in human genome?
(1) 18,000 (2) 30,000
(3) 13,000 (4) 4,000

57. _____% of the genome codes for protein in human beings.
(1) 98% (2) 50%
(3) 24% (4) < 2%

58. In humans, the largest gene is present on
(1) Chromosome-1
(2) Y-chromosome
(3) X-chromosome
(4) Chromosome-7

59. TDF gene is the smallest gene in humans with
(1) 231 bp (2) 14 bp
(3) 2968 bp (4) 3000 bp

60. SNPs stands for
(1) Single nucleoside polymorphism
(2) Simple nucleotide polymorphism
(3) Single nucleotide polymorphism
(4) Simple nucleoside polymorphism

Some Important Definitions

- **DNA:** Long polymer of deoxyribonucleotides.
- **Dinucleotide:** Binding of two nucleotides through 3' → 5' phosphodiester linkage.
- **Central dogma:** Flow of genetic information from DNA → RNA → Protein.
- **Histones:** Positively charged, basic proteins in chromosomes of eukaryotes.
- **Nucleosome:** Negatively charged DNA is wrapped around the positively charged histone octamer.
- **Chromatin:** Nucleosomes constitute the repeating unit of a structure in nucleus.
- **Euchromatin:** Chromatin that is loosely packed and stains light.
- **Heterochromatin:** Chromatin that is densely packed and stains dark.
- **Semiconservative DNA replication:** After the completion of replication, each DNA molecule would have one parental and one newly synthesised strand.

- **Replication fork:** Small opening of the DNA where the replication occurs.
- **Origin of replication:** Definite region (s) in DNA where the replication originates.
- **Transcription:** Process of copying genetic information from one strand of the DNA into RNA.
- **Coding strand:** DNA strand of transcription unit which has the polarity 5' → 3' and the sequence same as RNA, except T at the place of U.
- **Cistron:** A segment of DNA coding for a polypeptide.
- **Exons:** Coding sequences or expressed sequences in DNA.
- **Introns:** Non-coding sequences or intervening sequences in DNA.
- **Splicing:** Removal of introns from primary transcript.
- **Frame shift mutation:** Insertion or deletion of one or more bases changes the reading frame from the point of insertion or deletion.
- **Translation:** Process of polymerisation of amino acids to form a polypeptide.
- **Aminoacylation of tRNA:** Amino acids are activated in the presence of ATP and linked to their cognate tRNA.
- **UTR:** Specific sequences in mRNA that are not translated.
- **Operators:** Accessibility of promoter regions of prokaryotic DNA is in many cases regulated by the interaction of proteins with sequences, termed operators.
- **ESTs:** Identifying all the genes that are expressed as RNA.
- **Sequence annotation:** Sequencing of whole set of genome that contained all the coding and non-coding sequence, and later assigning different regions in the sequence with functions.
- **Repetitive DNA:** A small stretch of DNA is repeated many times in genome.

Quick Recap

1. Nucleotide monomers constitute a polymer called **nucleic acid.** It is of two types–RNA and DNA.
2. While DNA is **store house** of information, RNA helps in **transfer** and **expression** of information.
3. As DNA is structurally and chemically more **stable**, it is better genetic material. Although both DNA and RNA serve as genetic material.
4. RNA was **first** to **evolve**, and DNA was derived from it.
5. Bases in two DNA strands show hydrogen bonding (A = T, G = C) and follows **Chargaff's rule,** so that both the strands are complementary and its replication is semiconservative.
6. Segment of DNA that codes for an RNA is known as **gene**. During transcription one DNA strand acts as template which directs the synthesis of complementary RNA.

7. In prokaryotes, transcription and translation is a continuous process. In eukaryotes the genes are split **exons are interrupted by introns.** Introns are removed and exons are joined, to produce functional RNA,

8. The mRNA contains genetic code in combination of three (triplet code) to code for an amino acid. This genetic code is read by tRNA which acts as a **adapter molecule.**

9. There is specific tRNA for each amino acid. Each tRNA binds to amino acid at one end and with codons by H-bonding at another end.

10. Translation occurs at ribosome, here **ribozyme** (rRNA enzyme) acts as catalyst which helps in peptide bond formation. Process of translation has evolved around RNA, which shows that life began around RNA.

11. Since transcription and translation are **energetically** very expensive they are tightly regulated *e.g.*, ***Lac* operon** which is regulated by amount of lactose in medium *i.e.*, regulation of enzyme synthesis by its substrate.

12. Human genome project aimed for sequencing every base in human genome.

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

❑ ❑ ❑