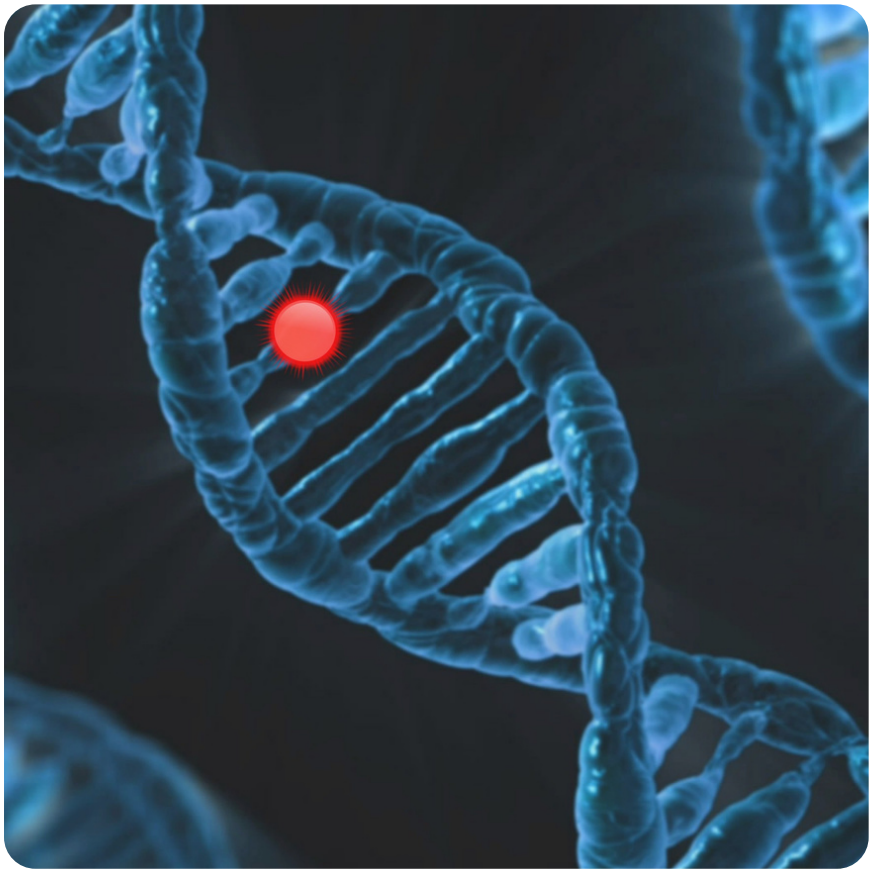
ALLEN CAREER INSTITUTE KOTA (RAJASTHAN)

path to success

PRE-MEDICAL

ZOOLOGY

ENTHUSIAST | LEADER | ACHIEVER



STUDY MATERIAL

Genetics : Molecular basis of inheritance

ENGLISH MEDIUM

Copyright Statement

JAMES DEWEY WATSON was born in Chicago on 6 April 1928. In 1947, he received B.Sc. degree in Zoology. During these years his interest in bird-watching had matured into a serious desire to learn genetics. This became possible when he received a Fellowship for graduate study in Zoology at Indiana University, Bloomington, where he received his Ph.D. degree in 1950 on a study of the effect of hard X-rays on bacteriophage multiplication. He met Crick and discovered their common interest in solving the DNA structure. Their first serious effort, was unsatisfactory. Their second effort based upon more experimental evidence and better appreciation of the nucleic acid literature, resulted, early in March 1953, in the proposal of the complementary double-helical configuration.

FRANCIS HARRY COMPTON CRICK was born on 8 June 1916, at Northampton, England. He studied physics at University College, London and obtained a B.Sc. in 1937. He completed Ph.D. in 1954 on a thesis entitled "X-ray Diffraction: Polypeptides and Proteins". A critical influence in Crick's career was his friendship with J. D. Watson, then a young man of 23, leading in 1953 to the proposal of the double-helical structure for DNA and the replication scheme. Crick was made an F.R.S. in 1959. The honours to Watson with Crick include: the John Collins Warren Prize of the Massachusetts General Hospital, in 1959; the Lasker Award, in 1960; the Research Corporation Prize, in 1962 and above all, the Nobel Prize in 1962.

MOLECULAR BASIS OF INHERITANCE

01. NUCLEIC ACIDS

- **Nucleic acids**
- **The DNA**
- **The Search for Genetic Material**
- **RNA World**
- **Replication**
- **Transcription**
- **Genetic Code**
- **Translation**
- **Regulation of Gene Expression**
- **Human Genome Project**
- **DNA Fingerprinting**

- F. Meischer (1869) discovered nucleic acid in nucleus of pus cell and called it "nuclein". The term nucleic acid was coined by "**Altmann**".

- Nucleic acids are polymer of **nucleotides**.

  Example : DNA and RNA.

  Nucleotide = Nitrogen base + pentose sugar + phosphate.

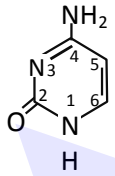
  Nucleoside = Nitrogen base + pentose sugar.

(1) NITROGEN BASE

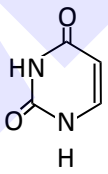
Skeletal of N bases contain heterocyclic ring.

On the basis of structure nitrogen bases are broadly of two types :–

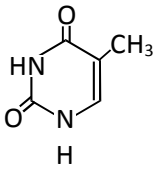
(A) Pyrimidines – Consist of one pyrimidine ring (2N + 4C). Skeleton of ring composed of two nitrogen and four Carbon atoms. e.g. Cytosine, Thymine and Uracil.



CYTOSINE

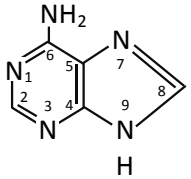


URACIL (in RNA)
(demethylated thymine)

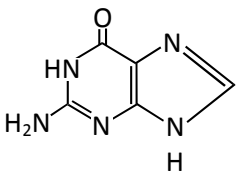


THYMINE (in DNA)
(5-methyl uracil)

(B) Purines – Consist of two rings i.e. one pyrimidine ring (2N + 4C) and one imidazole ring (2N + 3C) e.g. Adenine and Guanine.



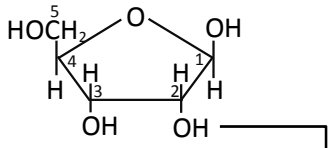
ADENINE



GUANINE

(2) PENTOSE SUGAR (Number of carbon = 5)

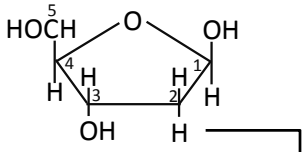
(1)



Ribose (in RNA)

Oxygen present → More reactive → Less stable

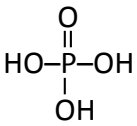
(2)



Deoxy ribose (in DNA)

Oxygen absent → Less reactive → More stable

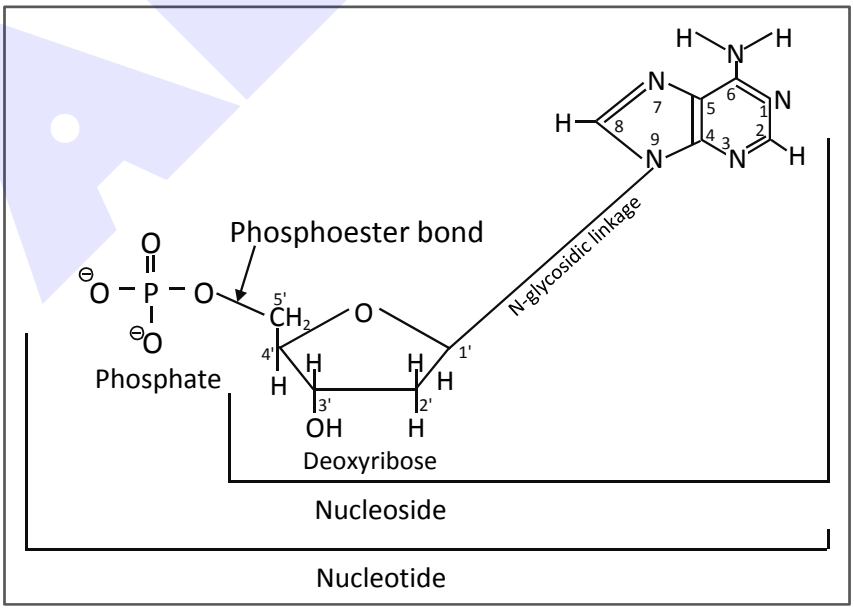
(3) PHOSPHATE

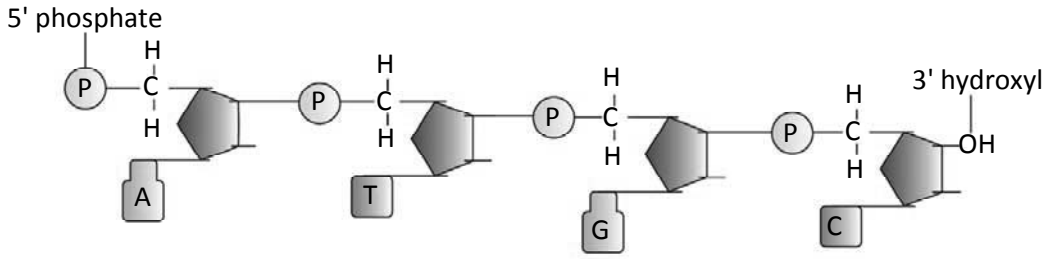


- Acidic
- Negative charged

(4) NUCLEOSIDE AND NUCLEOTIDE

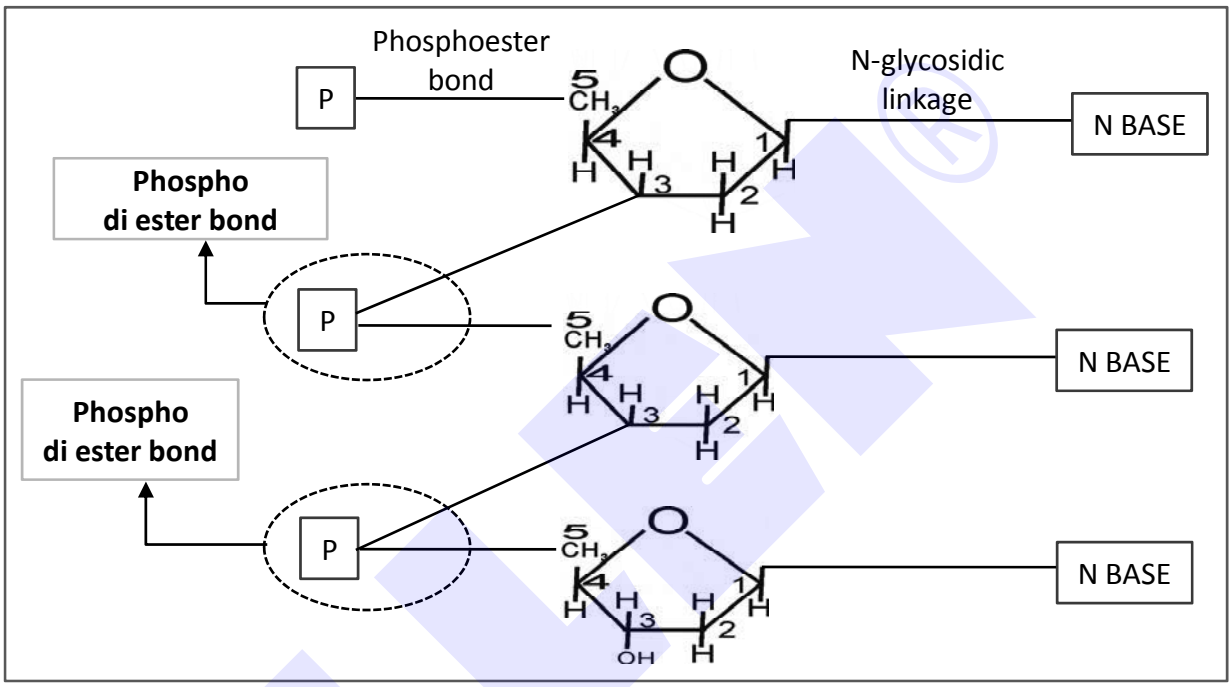
- Nitrogen base forms bond with first carbon of pentose sugar to form a nucleoside.
- Nitrogen of **first place** (N_1) forms bond with sugar in case of pyrimidines while in purines nitrogen of **ninth place** (N_9) forms bond with sugar.
- Phosphate forms ester bond (covalent bond) with fifth Carbon of sugar to form a complete nucleotide.





A Polynucleotide chain

NCERT XII Page No. 96, Figure No. 6.1



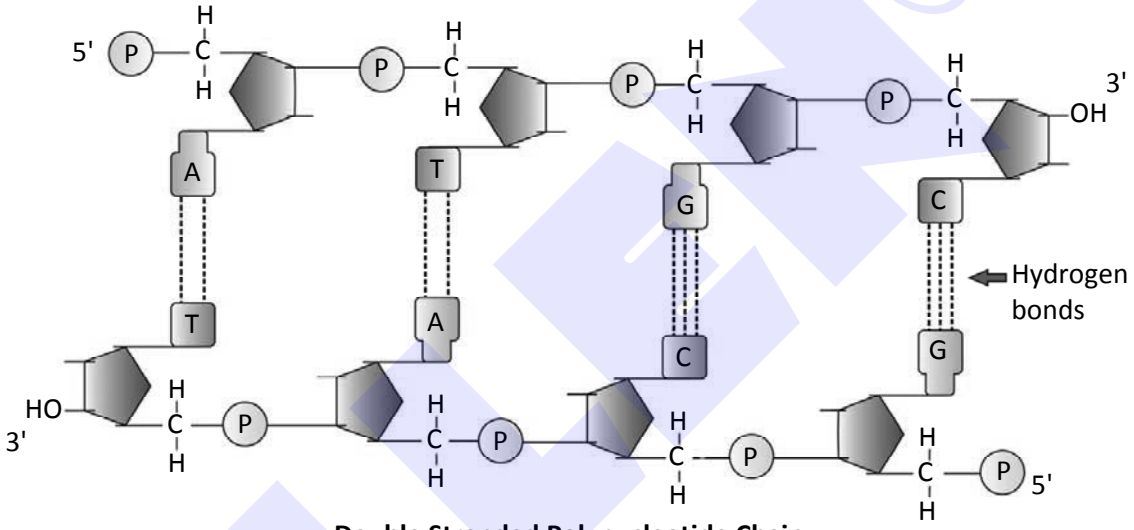
Types of nucleosides and nucleotides

N-base	Nucleosides = N base + Ribose sugar	Nucleotides = Nucleoside + P
Adenine	Adenosine	Adenylic acid
Guanine	Guanosine	Guanylic acid
Cytosine	Cytidine	Cytidylic acid
Uracil	Uridine	Uridylic acid

N-base	Nucleosides = N base + Deox Ribose Sugar	Nucleotides
Adenine	Deoxyadenosine	Deoxyadenylic acid
Guanine	Deoxyguanosine	Deoxyguanylic acid
Cytosine	Deoxycytidine	Deoxycytidylic acid
Thymine	Deoxythymidine	Deoxythymidylic acid

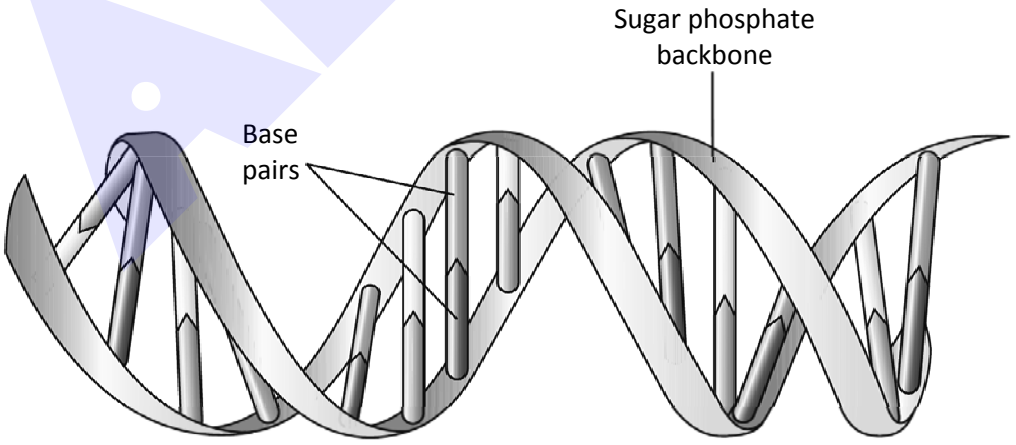
02. DNA

- DNA as an acidic substance present in nucleus was first identified by **Friedrich Meischer** in 1869.
- DNA term was given by - **Zacharias**
- DNA is long polymer of deoxyribonucleotides.
- DNA is negatively charged.
- In DNA, pentose sugar is deoxyribose and four types of nitrogen bases are A,T,G,C
- **Wilkins** and **Franklin** studied DNA molecule with the help of **X-Ray diffraction data/X-Ray crystallography.**



Double Stranded Polynucleotide Chain

NCERT XII Page No. 98, Figure No. 6.2



DNA Double Helix

NCERT XII Page No. 98, Figure No. 6.3

- With the help of this study, **Watson** and **Crick** (1953) proposed a double helix model for DNA. For this model Watson, Crick and Wilkins were awarded by Noble Prize in 1962.
- One of the hallmarks of their proposition was base pairing between the two strands of polynucleotide chains.
- According to this model, DNA is composed of two polynucleotide chains.
- Both polynucleotide chains are complementary and antiparallel to each other.
- In both strand of DNA direction of phosphodiester bond is opposite.
- In one strand of DNA, polarity is 5' $\rightarrow$ 3' and in other strand polarity is 3' $\rightarrow$ 5'.
- Both strand of DNA are held together by hydrogen bonds. These hydrogen bonds are present between nitrogen bases of both the strands.
- Adenine binds to thymine by two hydrogen bonds and cytosine binds to guanine by three hydrogen bonds.
- In a DNA molecule one purine always pairs with a pyrimidine. This generates approximately uniform distance between the two strands of DNA helix.
- In DNA plane of one base pair stacks over the other in double helix. This, in addition to H-bonds, confers stability to the helical structure of DNA.
- **Chargaff's equivalency rule –**
  - In a double stranded DNA amount of purine nucleotides is equal to amount of pyrimidine nucleotides.
  - The ratio between adenine and thymine, guanine and cytosine are constant and equals one.

    Purine = Pyrimidine

    [A] + [G] = [T] + [C]

    $\frac{[A]+[G]}{[T]+[C]}=1$ or $\frac{[T]+[C]}{[A]+[G]}=1$ or $\frac{[A]}{[T]}=1$ or $\frac{[G]}{[C]}=1$ or $\frac{[T]}{[A]}=1$ or $\frac{[C]}{[G]}=1$
- **Base ratio** = $\frac{A+T}{G+C}$ = constant for a given species. i.e. species specific.
  - In a DNA A + T > G + C $\Rightarrow$ A – T type DNA. Base ratio of A – T type of DNA is more than one. eg. Eukaryotic DNA
  - In a DNA G + C > A + T $\Rightarrow$ G – C type DNA. Base ratio of G – C type of DNA is less than one. eg. Prokaryotic DNA
- Melting point of DNA depends on G – C contents.
  - More G – C contents means higher melting point.
  - T_m = Temperature of melting.
  - T_m of prokaryotic DNA > T_m of Eukaryotic DNA
- DNA absorbs U.V. rays of 2600Å wavelength.

- **Denaturation and renaturation of DNA –**
  - If a normal DNA molecule is placed at high temperature (80-90°C) then both strands of DNA will separate from each other due to breaking of hydrogen bonds. It is called DNA-denaturation.
  - When denatured DNA molecule is placed at normal temperature then both strands of DNA get attached and recoiled to each other. It is called renaturation of DNA.

Configuration of DNA Molecule :–

- Two strands of DNA are helically coiled like a revolving ladder. Back bone of this ladder (Reiling) is composed of phosphates and sugars while steps (bars) are composed of pairs of nitrogen bases.

  Two chains have anti-parallel polarity. It means, if one chain has the polarity 5'→3', the other has 3'→5'.
- Distance between two successive steps is 3.4 A^0 . In one complete turn of DNA molecule there are such 10 steps (10 pairs of nitrogen bases). So the length of one complete turn is 34 A^0. This is called helix length (Pitch of the helix).
- Diameter of DNA molecule i.e. distance between phosphates of two strands is 20A^0.
- Each step of ascent is represented by a pair of bases. At each step of ascent, the strands turns 36°.
- ϕ × 174 bacteriophage has a single stranded circular DNA.
-

ϕ × 174 (bacteriophage) [Single stranded DNA]	5386 Nucleotides
λ bacteriophage (ds DNA)	48502 base pairs
E.coli (ds DNA)	4.6×10^6 base pairs
Human (Haploid cell/Genome) (ds DNA)	3.3×10^9 base pairs
Human (Diploid cell) ds DNA	6.6×10^9 base pairs

Types of DNA :–

- On the basis of direction of twisting, there are two types of DNA.
  - **Right Handed DNA –** Clockwise twisting e.g. The DNA for which Watson and Crick proposed model, was 'B' DNA.
  - **Left handed DNA –** Anticlockwise twisting e.g. Z-DNA

DNA	Helix Length	No. of base pairs in one turn	Distance between two base pairs	Diameter
'B'	34 Å	10 pairs	3.4 Å	20 Å
'Z'	45.6 Å	12	3.75 Å	18.4 Å

- **Palindromic DNA** – Sequence of nucleotides same from both ends.

  ⟶
  G G T A C C
  C C A T G G
  ⟵

03. PACKAGING OF DNA HELIX

- The average distance between the two adjacent base pairs of 0.34 nm (0.34×10^{-9} m or 3.4 Å). Length of DNA for a human diploid cell is 6.6×10^9 bp $\times$ 0.34×10^{-9} m/bp = 2.2 m. The length is far greater than the dimension of a typical nucleus.
- The number of base pairs in *Escherichia coli* is 4.6×10^6. The long sized DNA is accommodated in small area (about 1 μm in *E. coli*) only through packaging or compaction.

  If the length of E. coli DNA is 1.36 mm, can you calculate the number of base pairs in E.coli ?
- DNA is acidic due to presence of large number of phosphate groups. Compaction occurs by folding and attachment of DNA with basic proteins, polyamines in prokaryotes and histones in eukaryotes.
- **DNA packaging in Prokaryotes :**
  - In prokaryotes, such as, E. coli, though they do not have a defined nucleus, the DNA is not scattered throughout the cell.
  - DNA (being negatively charged) is held with some proteins (that have positive charges) in a region termed as 'nucleoid'.
  - The DNA in nucleoid is organised in large loops held by proteins.
- **DNA packaging in Eukaryotes :**
  - In eukaryotes, packaging organisation is much more complex.
  - In eukaryotes packaging is done by histone protein.

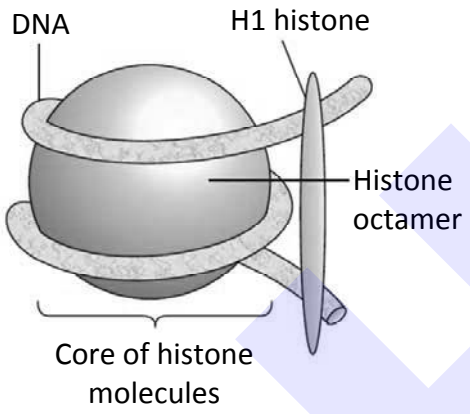
Histone Protein
- → It have abundance of lysine and arginine amino acids.
- → Lysine and arginine are basic amino acids.
- → Lysine and arginine carry positive charges in their side chains.
- → Due to lysine and arginine richness, histone protein have positive charge.
- → Histone is a basic protein.
- → There are five types of histone proteins H_1, H_2A, H_2B, H_3 and H_4

  - Histones are organised to form a unit of eight molecules (2 copies of each H_1, H_2A, H_2B, H_3 and H_4) called histone octamer.
  - Histone core have positively charged ends are directed outside. The negatively charged DNA is wrapped around the positively charged histone octamer to form a structure called nucleosome.

- Nucleosomes constitute the repeating unit of a structure in nucleus called chromatin.
- Chromatin are thread like stained (coloured) bodies seen in nucleus.
- DNA present between two adjacent nucleosomes is called linker DNA. It is attached to H_1 histone protein.
- The nucleosome in chromatin are seen as beads on string appearance under electron microscope (EM).
- A typical nucleosome contains 200 bps of DNA Helix.

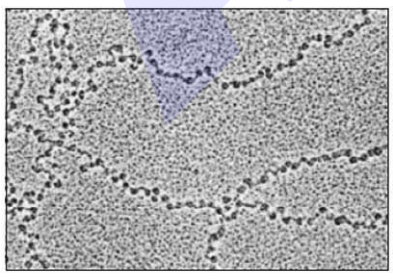
Packaging at higher level :

- The beads on string structure in chromatin is packaged to form chromatin fibres that are further coiled and condensed at metaphase stage of cell division to form chromosomes.
- The packaging at higher level requires additional set of proteins that collectively are referred to as non-histone chromosomal (NHC) proteins.



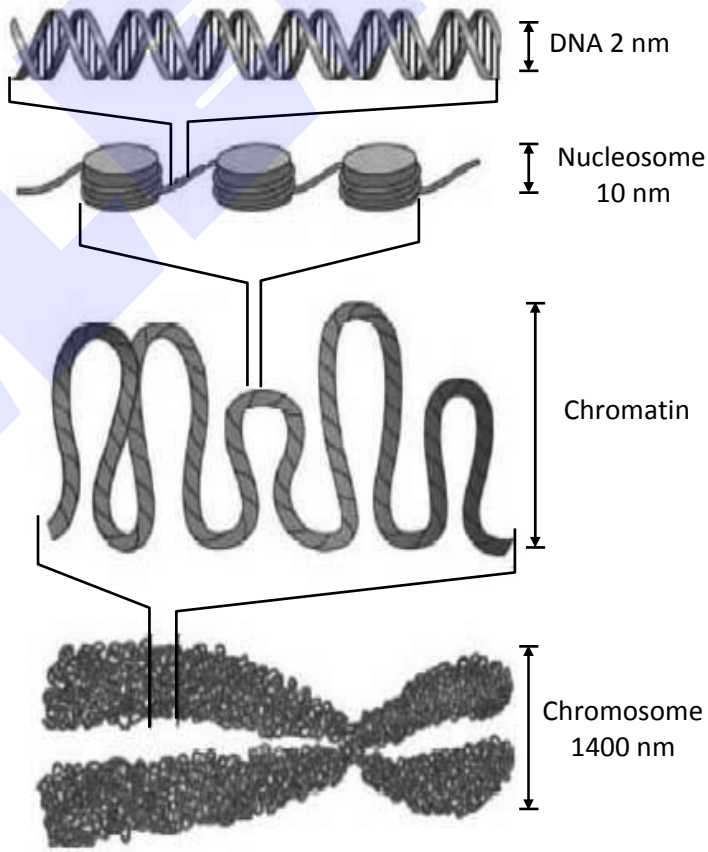
Basic unit of DNA compaction (Nucleosome)

Figure No. 6.4a
NCERT XII Page No. 99



EM picture - "Beads-on-String

Figure No. 6.4b
NCERT XII Page No. 99



Various steps in the folding and super folding of basic chromatin components to generate a eukaryotic chromosome

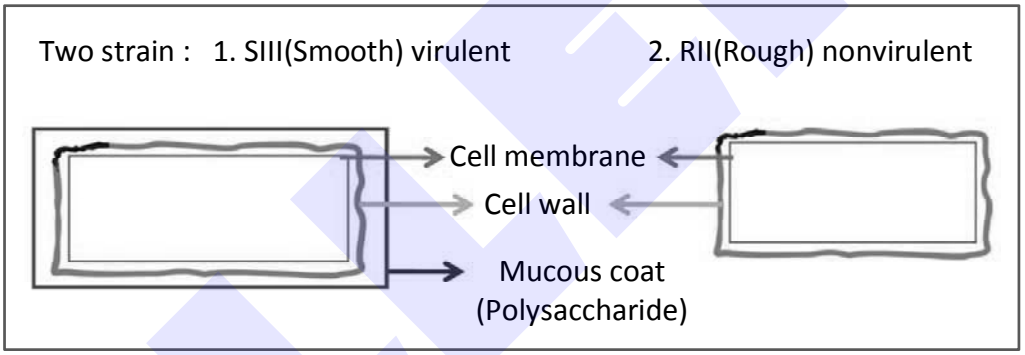
- In a typical nucleus, some region of chromatin are loosely packed (and stains light) and are referred to as euchromatin.
- The chromatin that is more densely packed and stains dark is called as heterochromatin.
- Euchromatin is said to be **transcriptionally active** and heterochromatin is **transcriptionally inactive.**

04. THE SEARCH FOR GENETIC MATERIAL

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

(1) EVIDENCE FROM BACTERIAL TRANSFORMATION

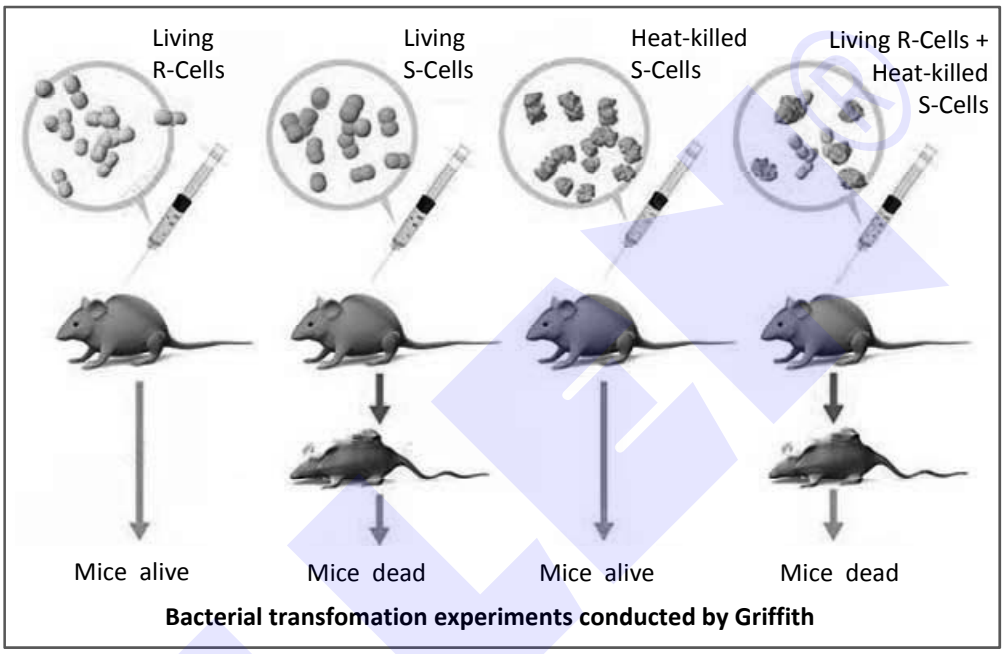
- The transformation experiments conducted by Frederick Griffith in 1928, are of greater importance in establishing the nature of genetic material.
- He used two strains of bacterium *Diplococcus* or *Streptococcus pneumoniae* or *Pneumococcus.* i.e. SIII (Smooth) virulent and RII (Rough) nonvirulent.



- Smooth (S) or capsulated type which have a mucous coat and produce smooth shiny colonies on agar medium culture plate. These bacteria are virulent and cause pneumonia.
- Rough (R) or non-capsulated type in which mucous coat is absent and produce rough colonies. These bacteria are nonvirulent and do not cause pneumonia.

- The experiment can be described in following four steps :
  - Live S strain → injected into mice → Mice die (due to pneumonia)
  - Live R strain → injected into mice → mice live (survived)
  - S strain (heat killed) → Injected into mice → Mice live (survived)
  - S strain (heat killed) + R strain (living) → injected into mice → Mice die.
  - Griffith 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.
- Some 'transforming principle', transferred from the heat-killed S strain, had enabled the R strain to synthesise 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.



Bacterial transfomation experiments conducted by Griffith

(2) BIOCHEMICAL CHARACTERISATION OF TRANSFORMING PRINCIPLE

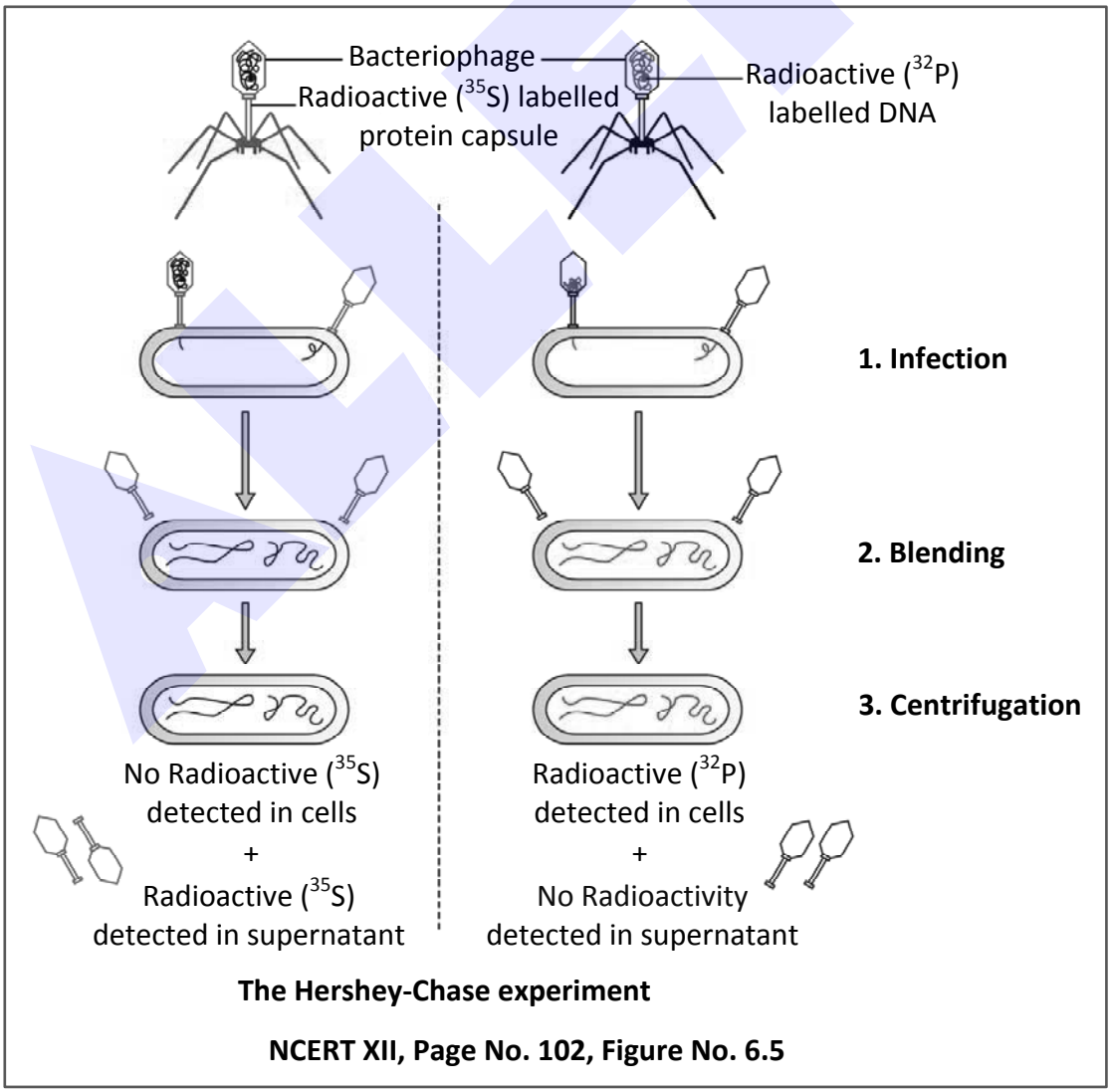
- Prior to the work of Oswald Avery, Colin MacLeod and Maclyn McCarty (1933-44), the genetic material was thought to be a protein. They worked to determine the biochemical nature of 'transforming principle' in Griffith's experiment.
- They purified biochemicals (proteins, DNA, RNA, etc.) from the heat-killed S cells to see which one 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 (proteases) and RNA-digesting enzymes (RNases) 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.
- Can you think of any difference between DNAs and DNase ?

(3) 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 viruses that infect bacteria called bacteriophages.
- T_2 phage is made up of DNA and protein coat.
- T_2 bacteriophage is a virus that infects bacterium *Escherichia coli* and multiplies inside it.
- The bacteriophage attaches to the bacteria and its genetic material then enters the bacterial cell.
- The bacterial cell treats the viral genetic material as if it was its own and subsequently manufactures more virus particles.
- Hershey and Chase worked to discover whether it was protein or DNA from the viruses that entered the bacteria.
- Phosphorous (Component of DNA)
  - $P^{31} \rightarrow$ Nonradioactive isotope
  - $P^{32} \rightarrow$ Radioactive isotope
- Sulfur (Component of protein)
  - $S^{32} \rightarrow$ Nonradioactive isotope
  - $S^{35} \rightarrow$ Radioactive isotope
- They grew some viruses on a medium that contained radioactive phosphorus and some others on medium that contained radioactive sulfur.
- Viruses grown in the presence of radioactive phosphorus contained radioactive DNA but not radioactive protein because DNA contains phosphorus but protein does not.
- Similarly, viruses grown on radioactive sulfur contained radioactive protein but not radioactive DNA because DNA does not contain sulfur.
- After the formation of labelled phages, three steps were followed, i.e., infection, blending, centrifugation.
  - **Infection :** Both type of labelled phages were allowed to infect normally cultured bacteria in separate experiments.
  - **Blending :** These bacterial cells were agitated in a blender to break the contact between virus and bacteria.
  - **Centrifugation :** The virus particles were separated from the bacterium by spinning them in a centrifuge.
  - Radioactive phages were allowed to attach to E. coli bacteria.

- Then, as the infection proceeded, the viral coats were removed from the bacteria by agitating them in a blender.
- The virus particles were separated from the bacteria by spinning them in a centrifuge.
- Bacteria which was infected with viruses that had radioactive DNA were radioactive, indicating that DNA was the material that passed from the virus to the bacteria.
- Bacteria that were infected with viruses that had radioactive proteins were not radioactive.
- This indicates that proteins did not enter the bacteria from the viruses. DNA is therefore the genetic material that is passed from virus to bacteria.
- After the centrifugation the bacterial cells showed the presence of radioactive DNA labelled with P^{32} while radioactive protein labelled with S^{35} appeared on the outside of bacteria cells (i.e., in the medium).
- Labelled DNA was also found in the next generation or phage. This clearly showed that only DNA enters the bacterial host and not the protein.



The Hershey-Chase experiment

NCERT XII, Page No. 102, Figure No. 6.5

Properties of Genetic material :

- It became an established fact that it is DNA that acts as genetic material. However, it subsequently became clear that in some viruses, RNA is the genetic material (for example, Tobacco Mosaic viruses, QB bacteriophage, etc.)
- Following are the properties and functions which should be fulfilled by a substance if it is to qualify as genetic material.

  (a) The genetic material should be able to generate its replica of own kind (replication). Both the nucleic acids (DNA and RNA) have the ability to direct their duplications. The other molecules in the living system, such as proteins fail to fulfill first criteria itself.

  (b) It should chemically and structurally be stable.
  - 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.
  - Stability as one of the properties of genetic material was very evident in Griffith's 'transforming principle' itself that heat, which killed the bacteria, at least did not destroy some of the properties of genetic material. This now can easily be explained in light of the DNA that the two strands being complementary if separated by heating come together, when appropriate conditions are provided.
  - Further, 2'-OH group present at every nucleotide in RNA is a reactive group and makes RNA labile and easily degradable.
  - RNA is also now known to be catalytic, hence reactive. Therefore, DNA chemically is less reactive and structurally more stable when compared to RNA. Therefore, among the two nucleic acids, the DNA is a better genetic material.
  - In fact, the presence of thymine at the place of uracil also confers additional stability to DNA.

  (c) The genetic material should also be capable of undergoing slow changes (mutations) – Both DNA and RNA are able to mutate. In fact, RNA being unstable, mutate at a faster rate. Consequently, viruses having RNA genome and having shorter life span mutate and evolve faster.

  (d) The genetic material should be able to express itself in the form of "Mendelian characters".

  RNA can directly code for the synthesis of proteins, hence can easily express the characters. DNA, however, is dependent on RNA for synthesis of proteins. The protein synthesising machinery has evolved around RNA.
- The above discussion indicate that both RNA and DNA can function as genetic material.
- DNA being more stable is preferred for storage of genetic information.
- For the transmission of genetic information, RNA is better.

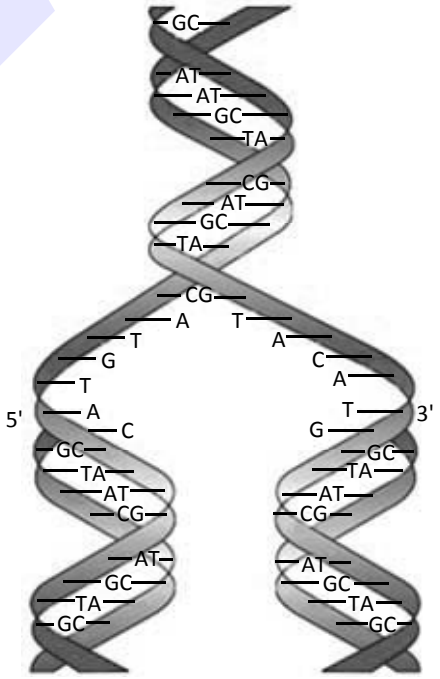
05. RNA WORLD

- 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 catalysts and not by protein enzymes (e.g., splicing) 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. Thus RNA is better material for transmission of information.

06. DNA REPLICATION

- While proposing the double helical structure for DNA,

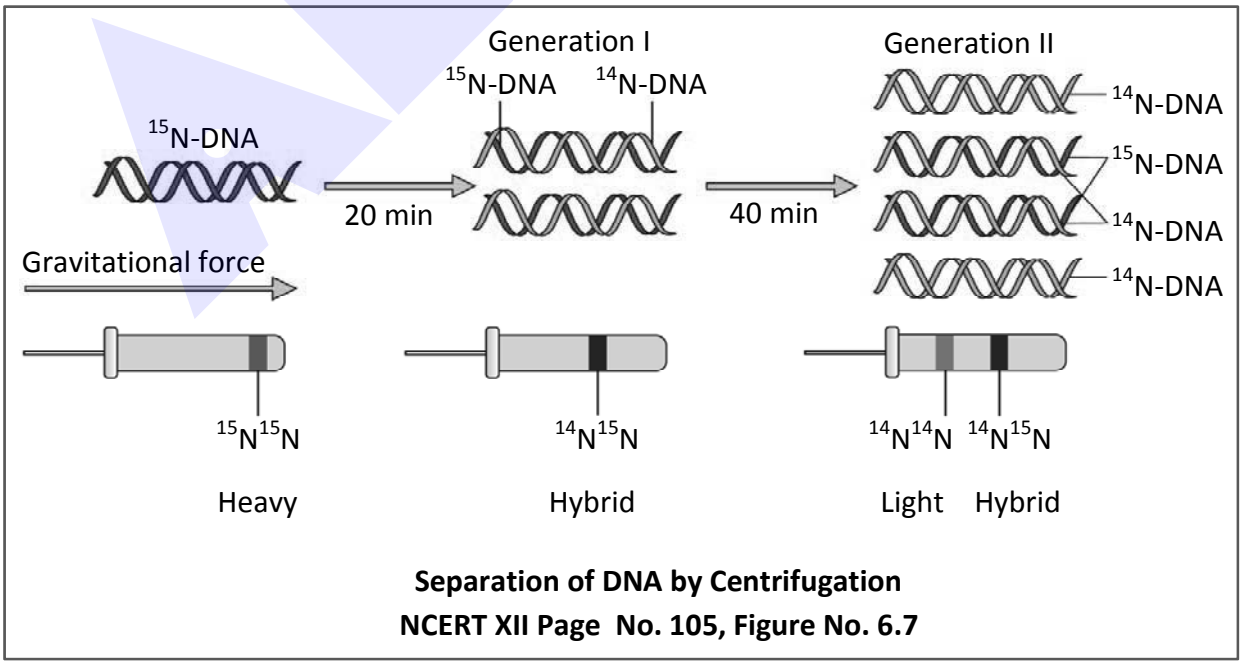
  Watson and Crick had immediately proposed a scheme for replication of DNA. To quote their original statement that is as follows:
- ''It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material'' (Watson and Crick, 1953).
- The scheme suggested that the two strands would separate and act as a 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
- DNA is capable of self duplication.
- DNA replication takes place in "S - Phase" of the cell cycle of eukaryotes.



Watson-Crick model for semiconservative DNA replication
NCERT XII Page No. 104
Figure No. 6.6

07. SEMI CONSERVATIVE MODE OF DNA REPLICATION

- **Method :-** Semi Conservative
- **Idea :-** Watson and Crick
- **Proved by :-** Meselson and Stahl
- **Worked on :-** *E.coli* bacteria
- Use isotope of Nitrogen (N^{14}, N^{15}) both are nonradioactive isotope.
- Use CsCl density gradient centrifugation
- **Proved at chromosome level :-** Taylor
- **Worked on :-** *Vicia faba*
- Use radioactive thymidine
- Matthew Meselson and Franklin Stahl performed the following experiment in 1958 :
  - They grew E. coli in a medium containing $^{15}NH_4Cl$ (^{15}N is the heavy isotope of nitrogen) as the only nitrogen source for many generations. The result was that ^{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 (Please note that ^{15}N is not a radioactive isotope, and it can be separated from ^{14}N only based on density).
  - 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 helices.
  - The various samples were separated independently on CsCl gradients to measure the densities of DNA.
  - Thus, the DNA that was extracted from the culture one generation after the transfer from ^{15}N to ^{14}N medium [that is after 20 minutes; E. coli divides in 20 minutes] had a hybrid or intermediate density. DNA extracted from the culture after another generation [that is after 40 minutes, II generation] was composed of equal amounts of this hybrid DNA and of 'light' DNA.



Separation of DNA by Centrifugation
NCERT XII Page No. 105, Figure No. 6.7

08. MECHANISM OF DNA REPLICATION

The following steps are included in DNA replication :-

(1) UNZIPPING (UNWINDING)

- The separation of 2 chains of DNA is termed as unzipping. And it takes place due to the breaking of H–bonds. The process of unzipping starts at a certain definite region which is termed as **initiation point** or **origin of replication**.
- In prokaryotes there occurs only one origin of replication but in eukaryotes there occurs many origin of replication i.e. unzipping starts at many points simultaneously.
- **Enzyme/protein** of unzipping

  Helicase :- At the place of origin helicase enzyme responsible for unzipping, breaking the hydrogen bonds.
- Cofactor for helicase :- Mg^{+2}
- **SSB** (single stranded DNA binding protein) prevents the reformation of H-bonds.
- **Topoisomerase** (in prokaryotes also called as **DNA gyrase**) release the tension arises due to supercoiling.

  ***Note* :** The process of DNA replication takes a few minutes in prokaryotes and a few hours in Eukaryotes.

(2) FORMATION OF NEW CHAIN

- Direction of new chain formation is always in 5' → 3'.
- Main enzyme :-
- DNA dependent DNA polymerase :-

  This enzyme catalyse polymerisation (chain formation) only in one direction that is 5' → 3'.

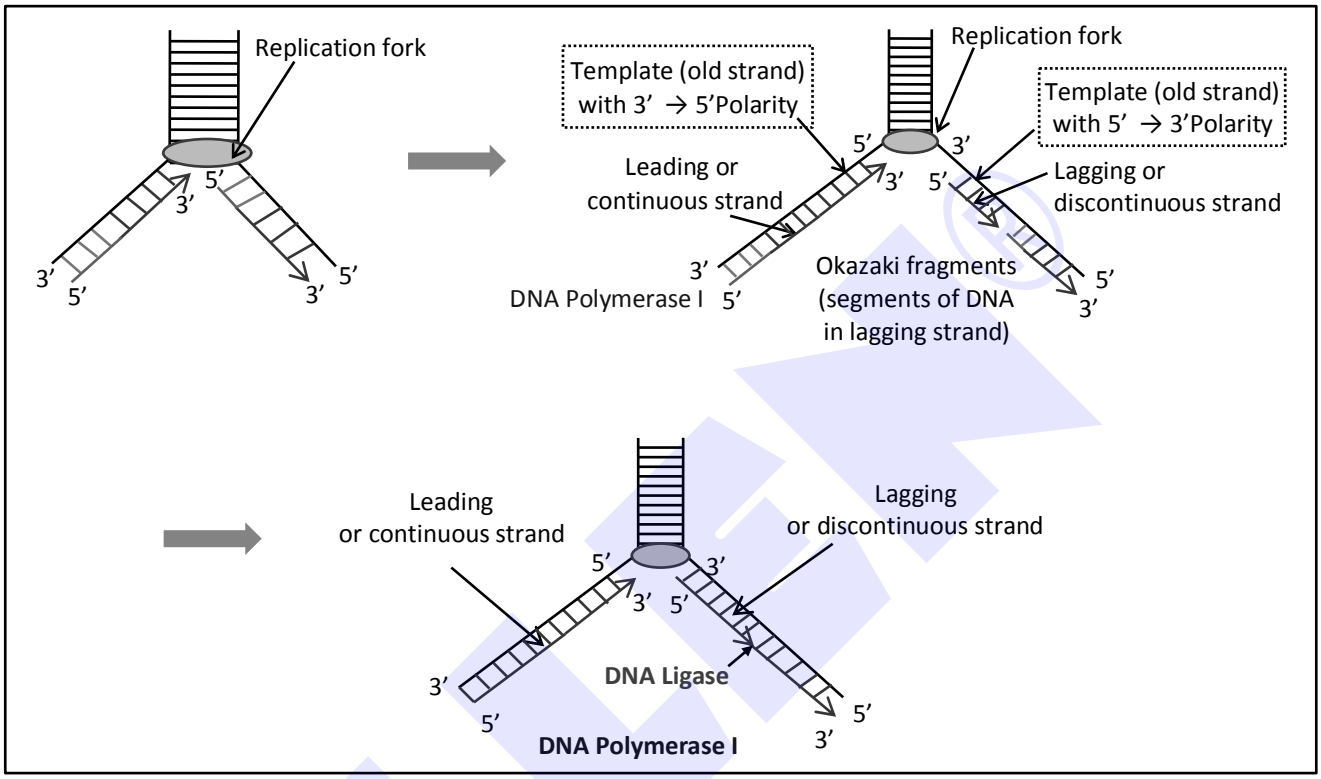
Type of DNA Polymerase	DNA Polymerase I	DNA Polymerase II	DNA Polymerase III
	discovered by Kornberg	least reactive in replication	Main enzyme for replication
5' → 3' Polymerisation (Chain formation)	Yes (Form short chain) (gap filling)	Yes (Form short chain)	Yes (Form long chain)
Remove RNA primer	Yes	No	No

Drawback of DNA polymerase I, II and III :- can't start DNA Replication

- **Primase (RNA Polymerase) :-**

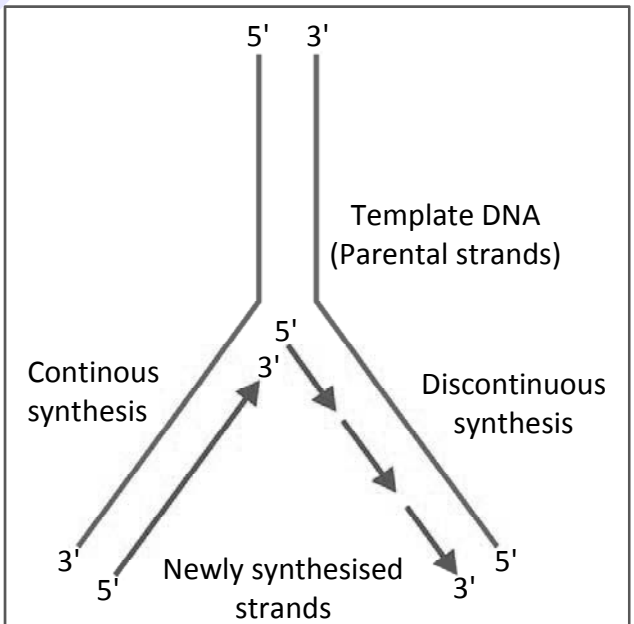
Form RNA Primer → Short chain of RNA
Form RNA Primer → RNA primer Start the process of replication

- **DNA Ligase :-** Join DNA fragments



Mechanism :

- To start the synthesis of new chain, special type of RNA is required which is termed as **RNA Primer**.
- The formation of RNA primer is catalysed by an enzyme - **RNA Polymerase** (primase). Synthesis of RNA–primer takes place in 5' → 3' direction. After the formation of new chain, this RNA is removed. (By DNA polymerase I)
- For the formation of new chain, Nucleotides are obtained from nucleoplasm of eukaryotes. In the nucleoplasm, Nucleotides are present in the form of triphosphates like dATP, dGTP, dCTP, dTTP etc.



Replicating Fork

NCERT XII Page No. 107, Figure No. 6.8

- During replication, the 2 phosphate groups of all nucleotides are separated. In this process energy is released which is consumed (utilized) in DNA replication.
- Energetically replication is a very expensive process.
- During replication deoxyribonucleoside triphosphate serve dual purposes in addition to acting as substrates, they provide energy for polymerisation.
- In addition to DNA-dependent DNA polymerases, many additional enzymes are required to complete the process of replication with high degree of accuracy.
- For long DNA molecules, since the two strands of DNA cannot be separated in its entire length (due to very high energy requirement), the replication occur within a small opening of the DNA helix, referred to as **replication fork**.
- During the process of DNA replication DNA polymerase formed two chains :-
- **Leading strand (Continuous strand) :-**
  - It formed on template (old) strand with 3' $\rightarrow$ 5' direction.
  - It formed in same direction of replication fork.
- **Lagging strand (Discontinuous strand) :-**
  - It formed on template (old) strand with 5' $\rightarrow$ 3' direction.
  - It formed in opposite direction of replication fork.
- The Okazaki fragments are joined together by an enzyme DNA Ligase.
- In the semi conservative mode of replication each daughter DNA molecule receives one chain of polynucleotides from the parental DNA - molecule and the second chain is newly synthesized.

- Replication is semi conservative and bidirectional process
- In E.coli duration of
  1. Cell division:- 20min.
  2. DNA replication :- 18 min.
- In *E.coli* 4.6 × 10^6 bps are present and the whole process of replication complete within 18 minuts.
- The average rate of polymerisation by DNA polymerase enzyme has to be approximately 2000 bp per second
- Means replication is fast and accurate process
- Any mistake during replication would result into mutation.
- Any failure in cell division after DNA replication result into polyploidy

BEGINNER'S BOX
NUCLEIC ACIDS TO MECHANISM OF DNA REPLICATION

1. Core of histone molecule of nucleosome consists of
(1) H_1 and H_2A (2) H_2A and H_2B (3) H_3 and H_4 (4) Both (2) and (3)

2. Radioactive element used to label DNA of bacteriophage in experiment of Hershey and Chase was :-
(1) S^{35} (2) P^{32} (3) N^{15} (4) C^{14}

3. Bonding between deoxyribose and base in pyrimidine nucleoside molecule is :-
(1) 1'-1' glycosidic linkage (2) 1'-6' glycosidic linkage
(3) 1'-9' glycosidic linkage (4) 1'-4' glycosidic linkage

4. T_m (melting temperature) value of DNA is high when it contains
(1) A + T > G + C (2) G + C > A + T (3) A + T = G + C (4) A + G = T + C

5. Select an incorrect statement regarding RNA molecule :-
(1) It has highly reactive 2'-OH group (2) It shows higher rate of mutation than DNA
(3) It is genetic material in some viruses (4) It follows Chargaff rule

6. In Meselson and Stahl's experiment, heavy isotope ^{15}N was used in the form of
(1) $^{15}NaNO_3$ (2) $^{15}NH_4Cl$ (3) $^{15}KNO_3$ (4) $^{15}NH_4NO_3$

7. Assuming that 50 heavy (i.e. containing N^{15}) DNA molecules replicated twice in a medium containing N^{14}, we expect
(1) 100 half heavy and half light and 150 light DNA molecules
(2) 100 half heavy and half light and 100 light DNA molecules
(3) 50 heavy and 150 light DNA molecules
(4) 50 heavy and 100 light DNA molecules

8. The enzyme which shows polymerising activity in 5'→ 3' direction is :
(1) DNA polymerase III (2) DNA polymerase II
(3) DNA polymerase I (4) All of these

9. DNA polymerase I is involved in :-
(1) Removal of RNA primer (2) Filling of gap
(3) Joining of okazaki fragments (4) Both (1) and (2)

10. DNA replication in lagging strand of most of the eukaryotic organisms is :-
(1) Conservative and continous (2) Semi-conservative but discontinous
(3) Conservative and semi-discontinous (4) Semi-conservative but continuous

09. RIBO NUCLEIC ACID (RNA)

- Structure of RNA is fundamentally same as DNA, but there are some differences. The differences are as follows:-
  - In place of De-oxyribose sugar in DNA, Ribose sugar is present in RNA.
  - In place of nitrogen base Thymine in DNA, Uracil is present in RNA.
  - RNA is made up of only one polynucleotide chain i.e. RNA is **single stranded**.

Exception :-

RNA found in Reo - virus is double stranded, i.e. it has two polynucleotide chains.

Types of RNA :

(1) GENETIC RNA OR GENOMIC RNA

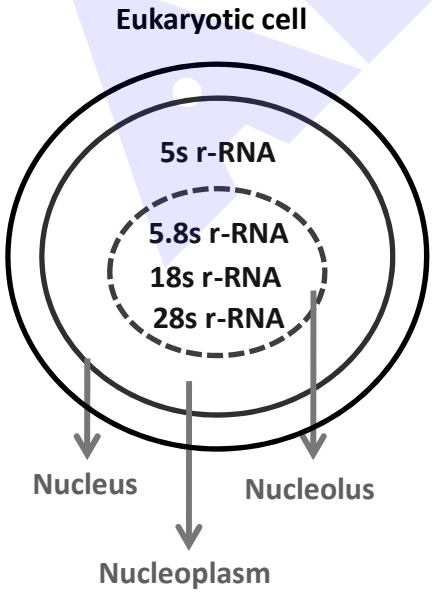
- In some viruses RNA works as genetic material and it transfers informations from one generation to next generation.
  eg. Reo virus, TMV, QB bacteriophage.

(2) NON-GENETIC RNA

- It is mainly of 3 types
  (A) r - RNA (B) t - RNA (C) m - RNA
- RNA functions as adapter, structural and in some cases as a catalyst (Ribozyme)

(A) Ribosomal RNA (r - RNA) :-

- This RNA is 80% of the cell's total RNA (Most abundant RNA)
- It is the most stable form of RNA.

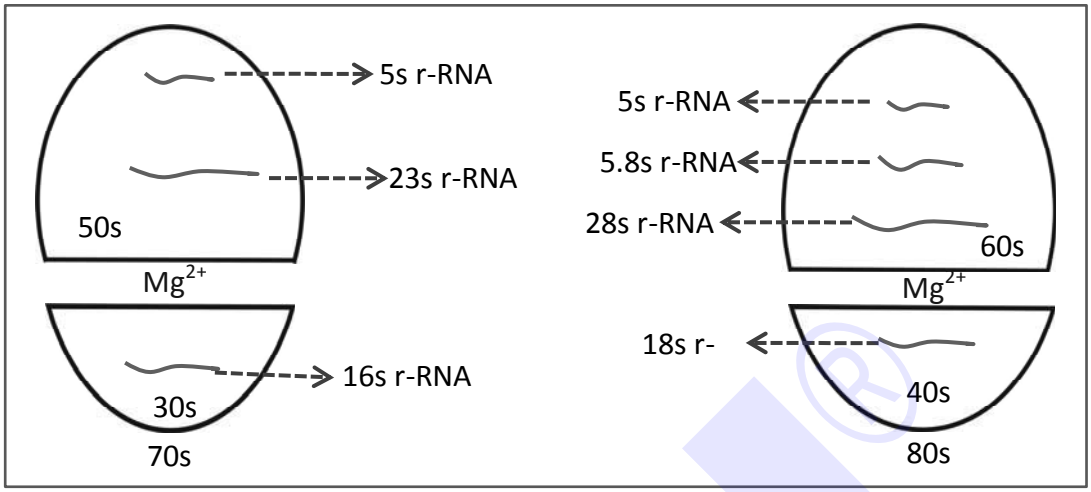


Formation :-

- Eukaryotes :- maximum r-RNA produced in nucleolus except 5s r-RNA.
- 5s r-RNA produced in nucleoplasm
- Prokaryotes :- r-RNA produced in cytoplasm

Function :-

(i) Form structure of Ribosome's (structural RNA)



(ii) Some r-RNA act as a catalyst /enzyme (Ribozyme)/(catalytic RNA)

e.g. 23 s r - RNA(70s)

28 s r - RNA(80s)

(B) Transfer – RNA (t–RNA) :–

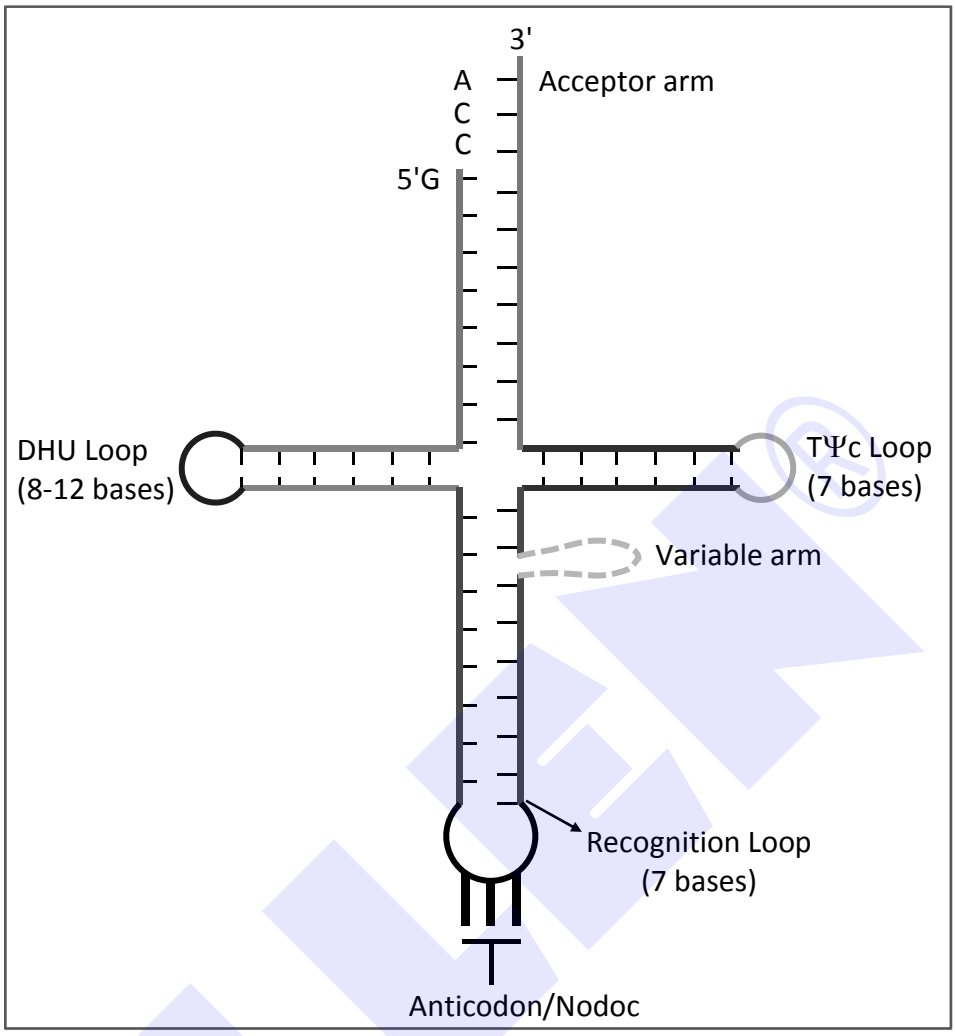
- It is 10-15% of total RNA.
- It is synthesized in the nucleus.
- It is also known as soluble RNA (**s**RNA)
- It is also known as Adapter RNA.
- It is the smallest RNA (4**s**).

Function :– At the time of protein synthesis, it acts as a carrier of amino-acids.

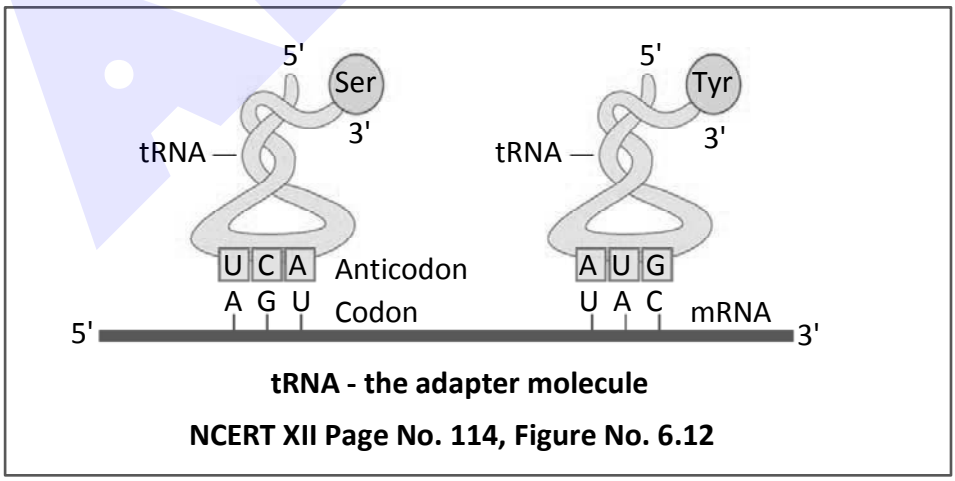
Structure :– The structure of t - RNA is most complicated.

2D :- clover leaf like (by Holley)

3D (Actual structure) :- Inverted L(by Kim)



- There are present three nucleotides in a particular sequence at 3' end of t - RNA and that sequence is CCA.



tRNA - the adapter molecule

NCERT XII Page No. 114, Figure No. 6.12

- At the 5' ends G (guanine) is present.
- 3' end is known as Acceptor end.
- t–RNA accepts amino acids at acceptor points. Amino acid binds to 3' end by its – COOH group.
- The molecule of t - RNA is folded and due to folding some complementary nitrogenous bases come across with each other and form hydrogen bonds.
- There are some places where hydrogen bonds are not formed, these places are known as loop.

Loops :-

There are some abnormal nitrogenous bases in the loops, that is why hydrogen bonds are not formed. ***e.g.*** Inosine (I), Pseudouracil (Ψ), Dihydrouridine (DHU).

- **T Ψ C Loop or Attachment loop :–**

This loop connects t - RNA to the larger subunit of ribosome.

- **Recognition Loop (<u>Anticodon loop</u>) :-**
  - This is the most specific loop of t-RNA and different types of t-RNA are different due to this loop. There is a specific sequence of three nucleotides called **Anticodon,** present at the end of this loop.
  - t–RNA recognizes its place on m - RNA with the help of Anticodon.
  - The anticodon of t-RNA recognises its complementary sequence on m–RNA. This complementary sequence is known as codon.
- **DHU Loop :–**
  - The function of DHU loop is to recognize this specific Aminoacyl synthetase enzyme.
  - There are 20 types of enzymes for 20 types of amino acids.
- **Messenger RNA (m–RNA) :–**
  - The m - RNA is 1 - 5% of the cell's total RNA.
  - The m - RNA is produced by genetic DNA in the nucleus.
  - It is least stable RNA.
  - Template RNA for protein synthesis.

10. TRANSCRIPTION

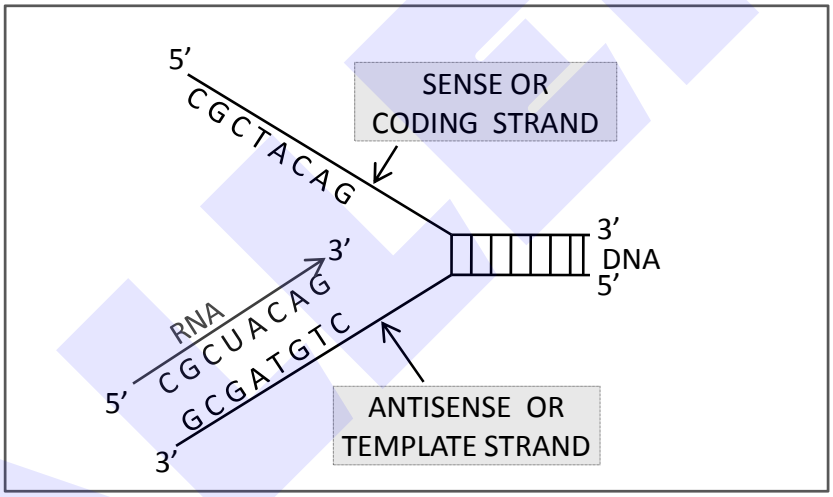
- The process of copying genetic information from one strand of the DNA into RNA is termed as transcription.
- Out of two strand of DNA only one strand participates in transcription.

  Antisense strand or **Template strand** :-

  - This strand code for RNA (means participate in the formation of RNA)
  - It has 3' → 5' polarity

  Sense strand or **coding strand** :-

  - This strand does not code for anything.
  - This strand has 5'→3' polarity.
  - It has sequence same as RNA (except thymine at the place of uracil)



- **Why both the strands are not copied during transcription has the simple answer.**
- First, if both strands act as a template, they would code for RNA molecule with different sequences (Remember complementarity does not mean identical), 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, and this would complicate the genetic information transfer machinery.
- Second, the two RNA molecules if produced simultaneously would be complementary to each other, hence would form a double stranded RNA. This would prevent RNA from being translated into protein and the exercise of transcription would become a futile one.

Enzymes :- DNA dependent RNA polymerase

- **Eukaryotes (Three types) :**

  There are at least three RNA polymerase in the nucleus. (In addition to the RNA polymerase found in the organelles)

  RNA polymerase I
  - 28s rRNA
  - 18s rRNA
  - 5.8s rRNA

  RNA polymerase II
  - hn–RNA (heterogeneous nuclear RNA)
  - m-RNA

  RNA polymerase III
  - t–RNA
  - 5s rRNA
  - SnRNA (small nuclear RNA)

  There is a clear cut division of labour.

- **Prokaryotes (One type) :**

  This RNA polymerase form all types of RNA.

  RNA polymerase
  - r-RNA
  - m-RNA
  - t-RNA

Prokaryotes :- RNA polymerase core enzyme + Sigma factor = RNA polymerase holoenzyme

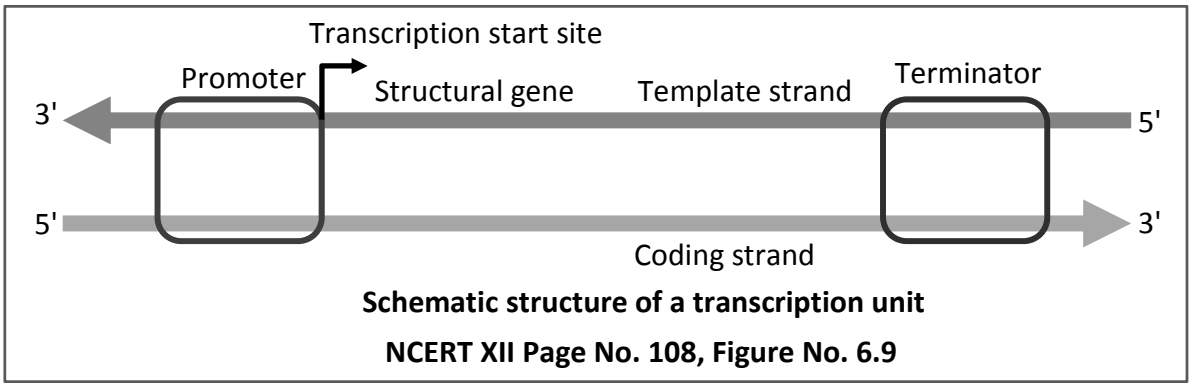
Functions of RNA polymerase :-

- 5'→ 3' Polymerisation – chain formation of RNA
- Break H bonds (Opening of the DNA helix)

Transcription unit :-

- In the process of replication, the total DNA of an organism gets duplicated.
- While in transcription only a segment of DNA and only one of the strands is copied into RNA.
- So it is necessary to define the boundaries of DNA that would be transcribed, this boundary is transcription unit.
- A transcription unit in DNA is defined primarily by three regions in the DNA :-

  (a) A promoter (b) The structural gene (c) A terminator



Schematic structure of a transcription unit

NCERT XII Page No. 108, Figure No. 6.9

Promoter :-

- The promoter and terminator flank the structural gene in a transcription unit.
- It is a DNA sequence that provides binding site for RNA polymerase.
- The promoter is said to be located towards 5'-end (upstream) of the structural gene (the reference is made with respect to the polarity of coding strand).
- 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 coding and template strands could be reversed.

Terminator :-

- It usually defines the end of the process of transcription.
- The terminator is located towards 3'-end (downstream) of the coding strand.

The structural gene :-

- It is a DNA sequence which form any RNA.
- A gene is defined as the functional unit of inheritance.
- The DNA sequence coding for tRNA or rRNA molecule also define a gene.
- Inheritance of a character is also affected by promoter and regulatory sequences of a structural gene.
- Hence, sometime the regulatory sequences are loosely defined as regulatory genes, even though these sequences do not code for any RNA or protein.
- Cistron is a segment of DNA coding for a polypeptide, the structural gene in a transcription unit could be said as monocistronic (mostly in eukaryotes) or polycistronic (mostly in bacteria or prokaryotes).

Mechanism of transcription :-

- The process of copying genetic information from one strand of the DNA into RNA is termed as transcription.
- The principle of complementarity governs the process of transcription, except the adenosine complements now forms base pair with uracil instead of thymine.
- The process of transcription divide in three steps :-
  Initiation, Elongation and Termination.

(1) INITIATION

- RNA polymerase with the help of sigma factor binds to promoter and initiates transcription.
- Sigma factor (σ) recognises the promoter site of DNA.
- In prokaryotes before the 10 N-bases from ***"Starting point"*** a sequence of 6 base pairs (TATAAT) is present on DNA, which is called ***"Pribnow box"***.
- In eukaryotes before the 20 N-bases from ***"Starting point"*** a sequence of 7 base pairs (TATAAAA) or (TATATAT) is present on DNA which is called **"TATA box** or **Hogness box"**
- At start point, RNA polymerase enzyme breaks H–bonds of two strands of DNA and separates them.
- One of them strand takes part in transcription. Transcription (RNA formation) proceeds in $5' \rightarrow 3'$ direction.
- Ribonucleoside triphosphates come to lie opposite of complementary nitrogen bases of anti sense strand.
- These Ribonucleotides are present in the form of triphosphate ATP, GTP, UTP and CTP. When they are used in transcription, pyrophosphatase enzyme hydrolyse two phosphates from each NTP (Triphosphate), this releases energy. This energy is used in the process of transcription.

(2) ELONGATION

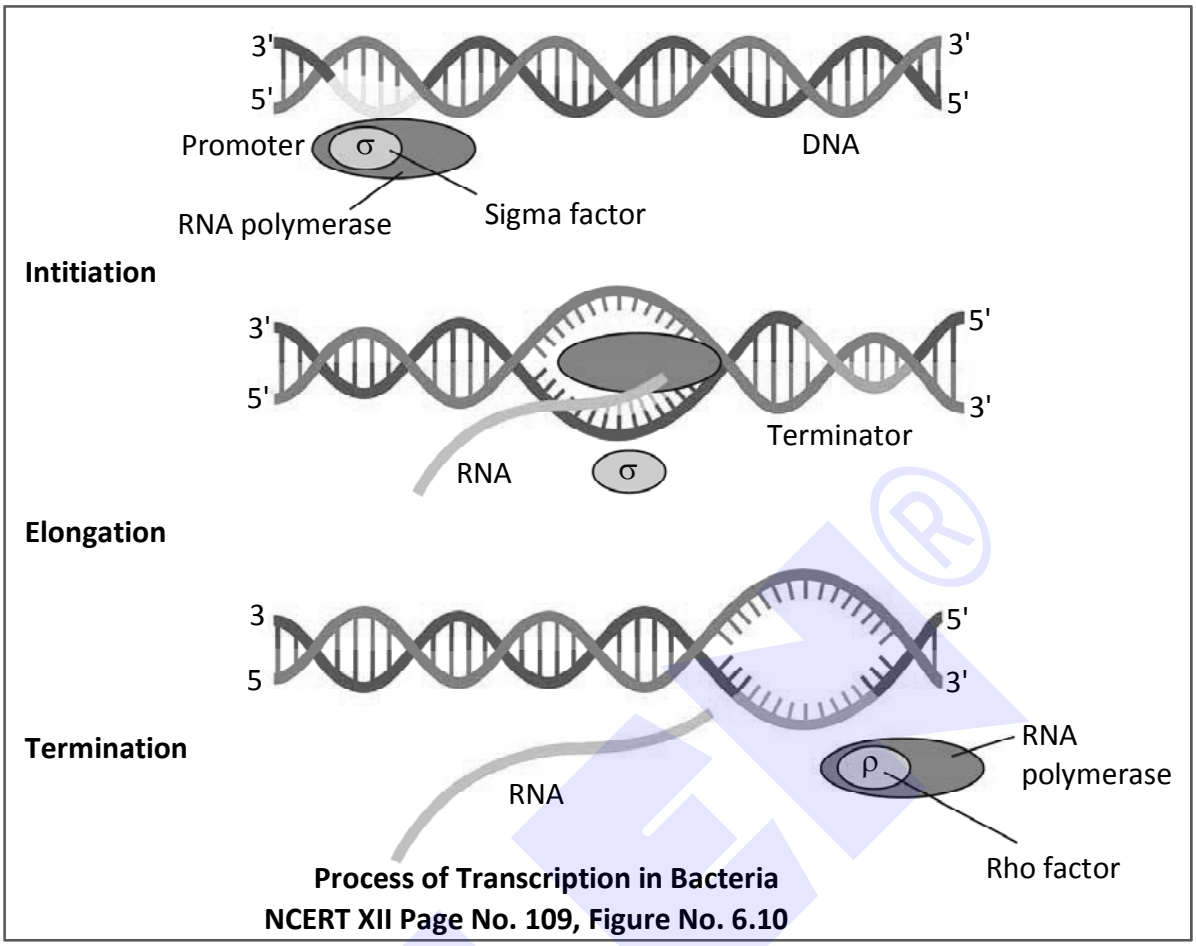
- RNA polymerase uses nucleoside triphosphates as substrate and polymerises in a template depended fashion following the rule of complementarity.
- RNA polymerase somehow also facilitates opening of the helix and continues elongation. Only a short stretch of RNA remains bound to the enzyme.
- RNA polymerase enzyme establishes phosphodiester bond between adjacent ribonucleotides.

(3) TERMINATION

- Once the polymerases reaches the terminator region, the nascent RNA falls off, so also the RNA polymerase. This results in termination of transcription.
- In prokaryotes, the terminator site is recognised with the help of **Rho factor (ρ factor)**.
- Rho (ρ) factor is a specific protein that help in termination.

Special points :-

- An intriguing question is that how is the RNA polymerases able to catalyse all the three steps, which are initiation, elongation and termination.
- The RNA polymerase is only capable of catalysing the process of elongation.
- It associates transiently with initiation-factor (σ) and termination-factor (ρ) to initiate and terminate the transcription, respectively.
- Association with these factors alter the specificity of the RNA polymerase to either initiate or terminate.



Process of Transcription in Bacteria
NCERT XII Page No. 109, Figure No. 6.10

11. SPLIT GENE

- Discovered by Sharp and Roberts. They were awarded Nobel Prize in 1993.
- Gene which contains non coding part along with coding part is known as split gene. Non coding part is called **intron** and coding part is called **exon.**
- In eukaryotes, the monocistronic structural genes have interrupted coding sequences – the genes in eukaryotes are split. The coding sequences or expressed sequences are defined as exons.
- Exons are said to be those sequence that appear in mature or processed RNA. The exons are interrupted by introns.
- Introns or intervening sequences do not appear in mature or processed RNA.
- By transcription split gene produces a RNA which contains coding and non coding sequence and called hn RNA (Heterogenous nuclear RNA). This hn RNA is unstable.

Capping :-

- 7 methyl guanonsine triphosphate (an unusual nucleotide) is added to its 5' end, and a cap like structure is formed. It is called **capping**.

Tailing :-

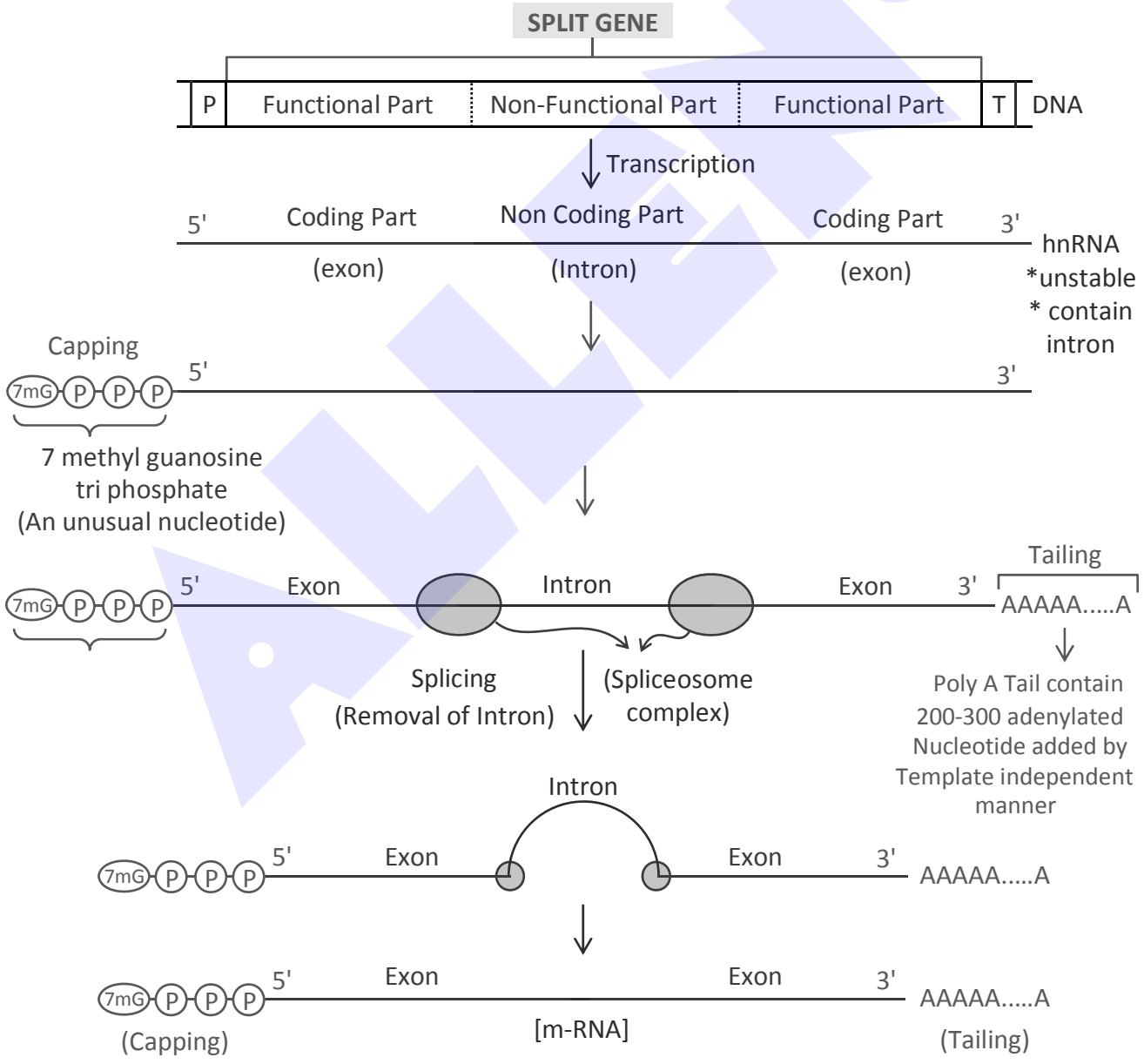
- 200-300 nucleotides of adenylic acid are added in template independent manner to its 3' end, which is called **poly 'A' tail**, now hn-RNA becomes stable.

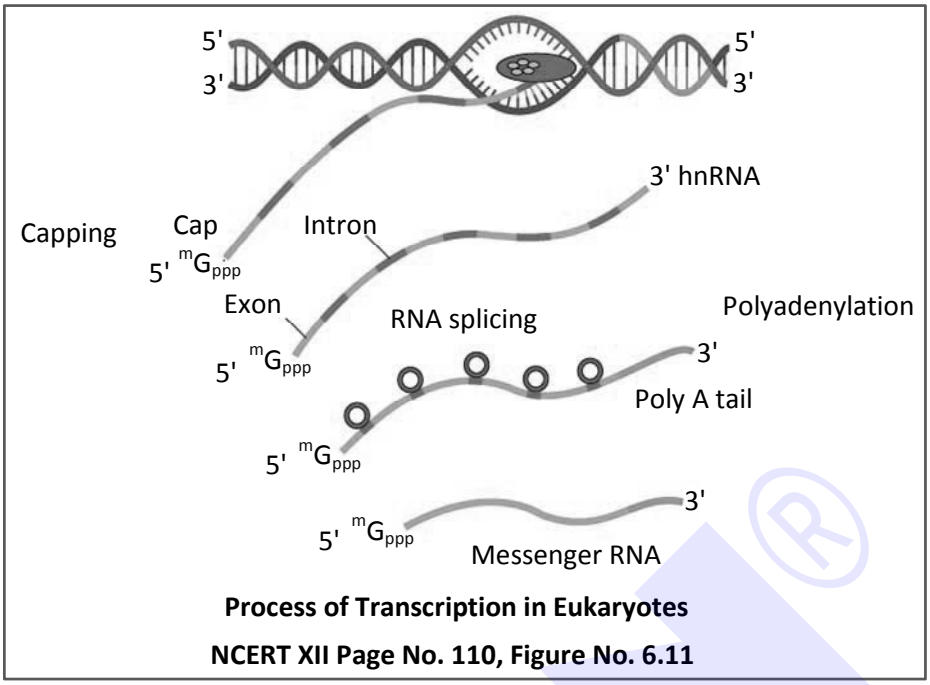
Splicing :-

- In splicing where the introns are removed and exones are joined in a defined order.
- Splicing done by spliceosome complex.

- This complex is made up of Sn RNA and proteins.
- The spliceosome complex uses energy of ATP to cut the RNA, releases the non-coding part and joins the coding-part to produce functional RNA.

- It is the fully processed hnRNA, now called mRNA.
- Now mRNA is transported out of the nucleus for translation.
- Mostly Eukaryotic genes are example of split gene. (Exception :- genes of histone and interferon protein).
- Mostly prokaryotic genes are example of non split gene.
- The split gene represent an ancient (primitive) feature of gene.
- Presence of introns is probably reminiscent of antiquity. [Primitive character]
- The splicing process represents the dominance of RNA world.
- The split-gene arrangement further complicates the definition of a gene in terms of a DNA segment.





Process of Transcription in Eukaryotes

NCERT XII Page No. 110, Figure No. 6.11

12. GENETIC CODE

- Term Given by ***George Gamow***.
- Discovered by Nirenberg, Matthaei and Khorana.
- During replication and transcription a nucleic acid was copied to form another nucleic acid. Hence, these processes are easy to conceptualise on the basis of complementarity.
- The process of translation requires transfer of genetic information from a polymer of nucleotides to synthesise a polymer of amino acids.
- Neither does any complementarity exist between nucleotides and amino acids, nor could any be drawn theoretically.
- There existed ample evidences, though, to support the notion that change in nucleic acids (genetic material) were responsible for change in amino acids in proteins.
- This led to the proposition of a genetic code that could direct the sequence of amino acids during synthesis of proteins.
- The relationship between the sequence of amino acids in a polypeptide chain and nucleotide sequence of DNA or m–RNA is called ***genetic code***.
- m–RNA contains code for each amino acid and it is called codon. A codon is the nucleotide sequence on m–RNA which codes for a particular amino acid ; whereas the genetic code is the sequence of nucleotides on m–RNA molecule, which contains information for the synthesis of polypeptide chain.

(1) TRIPLET CODON

- 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.
- **George Gamow (a physicist) :-** Genetic code is triplet, was first pointed out by Gamow.
- There are four types of N–bases in m–RNA (A, U, G, C) for 20 types of amino acids.
- If genetic code is ***singlet*** i.e. codon is the combination of only one nitrogen base, then only four codons are possible A, C, G and U. These are insufficient to code for 20 types of amino acids.
  - Singlet code = $4^1 = 4 \times 1 = 4$ codons

A
C
G
U

Codons

Singlet Code : 4 × 1 = 4 codons

- If genetic code is ***doublet*** (i.e. codon is the combination of two nitrogen bases) then 16 codons are formed.
  - Doublet code = $4^2 = 4 \times 4 = 16$ codons.
  - 16 codons are insufficient to code for 20 amino acid

AA	AC	AG	AU
CC	CA	CG	CU
GG	GA	GC	GU
UU	UA	UG	UC
Doublet Code : 4 × 4 = 16 codons			

- **Gamow** (1954) pointed out the possibility of three letters codon (Triplet codon).
- Genetic code is triplet i.e. one codon consists of three nitrogen bases
  Triplet code = $4^3 = 4 \times 4 \times 4 = 64$ codons
- In this case there occurs 64 codons in the dictionary of genetic code.
- 64 codons are sufficient to code for 20 types of amino acids.

First position → Third position; Second position

	U	C	A	G	
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

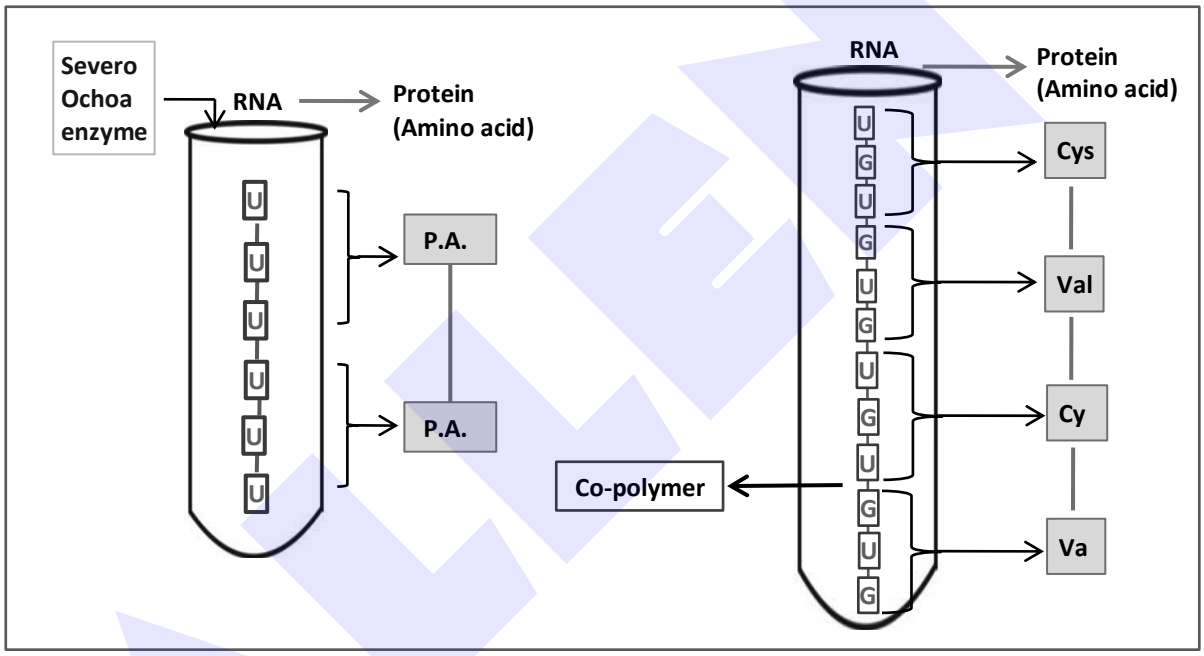
The Codons for the Various Amino Acids
NCERT XII Page No. 112, Table No. 6.1

Marshall Nirenberg's and Matthae :-

- They developed cell-free system for protein synthesis finally helped the code to be deciphered.
- They where synthesized homopolymer RNA like (Poly U, Poly C, Poly A) by the help of Severo Ochoa enzyme (polynucleotide phosphorylase).
- This enzyme helpful in polymerising RNA with defined sequences in a template independent manner (enzymatic synthesis of RNA).
- These RNA were allowed to translate (form protein) in cell free system.

Har Gobind Khorana :-

- The chemical method developed by Har Gobind Khorana was instrumental in synthesising RNA molecules with defined combinations of bases (homopolymers and copolymers).



(2) CHARACTERISTICS OF GENETIC CODE

(A) Triplet in Nature :–

- A codon is composed of three adjacent nitrogen bases which codes or specifies one amino acid in polypeptide chain.

For Ex. :

- In m–RNA if there are total 90 N–bases.
- Then this m–RNA determines 30 amino acids in polypeptide chain.
- In above example, number of nitrogen bases are 90, so codons $\Rightarrow$ 30 and 30 codons decide 30 amino acids in polypeptide chain.

(B) Nearly Universal :–

- The genetic code is nearly universal. The same genetic code is present in all kinds of living organism including viruses, bacteria, unicellular and multicellular organisms.

  Exception :- Mitochondrial codon

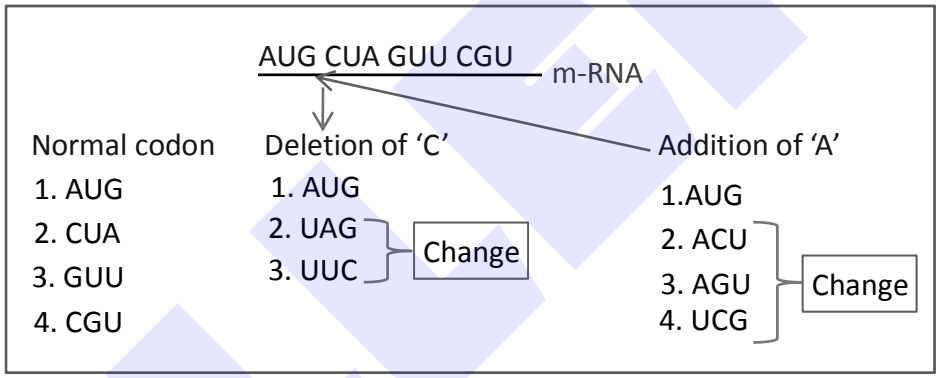
  Some Protozoans

(C) Non – Overlapping :–

- A nitrogen base is a constituent of only one codon.

(D) Commaless :–

- The codon is read in mRNA in a contiguous (continuous) fashion.
- There is no punctuation (comma) between the adjacent codons i.e. each codon is immediately followed by the next codon.
- If a nucleotide is deleted or added, the whole genetic code read differently from the point of change onward.



- Insertion or deletion of one or two bases changes the reading frame from the point of insertion or deletion. However, such mutations are referred to as frameshift insertion or deletion mutations.
- Insertion or deletion of three or its multiple bases insert or delete in one or multiple codon hence one or multiple amino acids, and reading frame remains unaltered from that point onwards.

(E) Unambiguous and Specific :–

- Genetic code is unambiguous i.e. one codon specifies only one amino acid and not any other.

  e.g. UUU $\rightarrow$ code only for phenyl alanine.
- In this case one codon never code for two different amino acids.
- Exception GUG (Ambiguous) codon which may code both valine and methionine amino acids.

(F) Degeneracy (Redundancy) of Genetic code : -

- There are 64 codons for 20 types of amino acids, so most of the amino acids (except two) can be coded by more than one codon.

e.g. Phenyl alanine — UUU, UUC

Glycine — GGG, GGC, GGU, GGA

- Single amino acid coded by more than one codon is called ***"Degeneracy of genetic code"***.
- Only two amino acids ***Tryptophan*** and ***Methionine*** are specified by single codon.
  - UGG for Tryptophan
  - AUG for Methionine.
- All the other amino acids are specified or coded by 2/3/4/6 codons.
- Leucine, serine and arginine are coded by 6–codons (Max.).
- Degeneracy of genetic code is related to third position (3′ – end of triplet codon) of codon. The third base is described as ***"Wobbly base"***.

(G) Chain Initiation Codon :-

- Polypeptide chain synthesis is signalled by initiation codons AUG.
- AUG codon has dual (two) functions :

  (i) Initiation codon (ii) code methionine amino acid.
- AUG codes methionine amino acid in eukaryotes and in prokaryotes AUG codes N–formyl methionine.
- Some times in prokaryotes GUG also functions as start codon. It codes for valine amino acid normally but when it is present at starting position it codes for N-formyl methionine amino acid.

(H) Chain Termination Codon :-

- Out of 64 codons, 3–codons are stop or nonsense or termination codon.

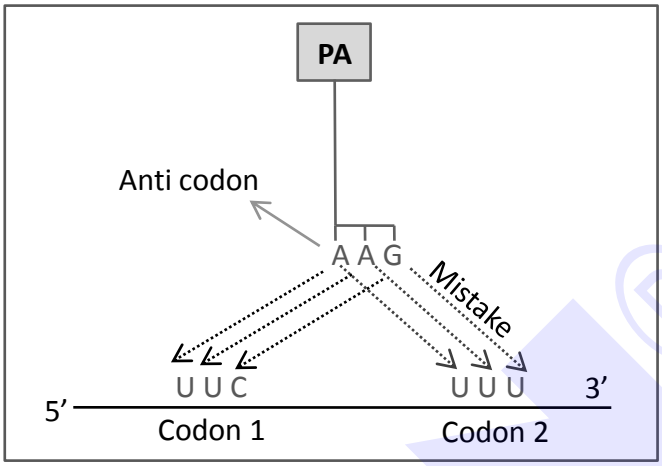
  Nonsense codons do not code any amino acid.

  UAA (Ochre)
  UAG (Amber) } **Non–Sense Codons or Stop codons**
  UGA (Opal)

- So only 61 codons are sense codons which specify 20 amino acids.

(3) WOBBLE HYPOTHESIS

- It was propounded by CRICK.
- Normally an anticodon recognises only one codon, but sometimes an anticodon recognises more than one codon. This is known as Wobbling. Wobbling normally occurs for **third nucleotide** of codon and it is responsible for degeneracy of genetic code.



- For e.g. anticodon AAG can recognise two codons i.e. UUU and UUC, both stands for phenyl alanine.

Types of m-RNA : m-RNA is of 2 types –

(A) Monocistronic :- The m-RNA, in which genetic signal is present for the formation of only one polypeptide chain eg. Eukaryotes.

(B) Polycistronic :- The m-RNA, in which genetic signal is present for the formation of more than one polypeptide chains eg. Prokaryotes.

- Non sense codons are found in middle position in polycistronic m-RNA.

13. CENTRAL DOGMA

- Francis Crick proposed the central dogma.
- The formation (production) of m - RNA from DNA and then synthesis of protein from it, is known as **Central Dogma**.

$$\text{Replication} \circlearrowleft \text{DNA} \xrightarrow{\text{Tanscription}} \text{RNA} \xrightarrow{\text{Tanslation}} \text{Protein}$$

- It means, it includes transcription and translation.

Reverse Transcription :–

- The formation of DNA from RNA is known as **Reverse - transcription**. It was discovered by **Temin and Baltimore** in Rous - sarcoma virus. So it is also called **Teminism.**

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

[Reverse transcriptase (RNA dependent DNA polymerase)]

14. TRANSLATION (PROTEIN SYNTHESIS)

Translation refers to the process of polymerisation of amino acids to form a polypeptide.

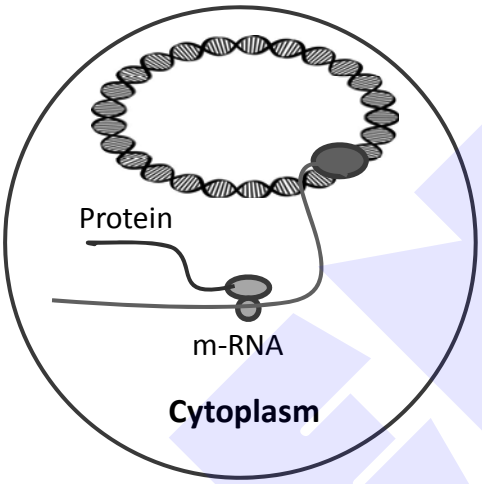
The order and sequence of amino acids are defined by the sequence of bases in the mRNA. Means, in protein the sequence of amino acids are dictated by DNA and represented by mRNA.

In Prokaryotes :- Protein synthesis take place in cytoplasm.

Transcription and translation can be coupled in bacteria because :-

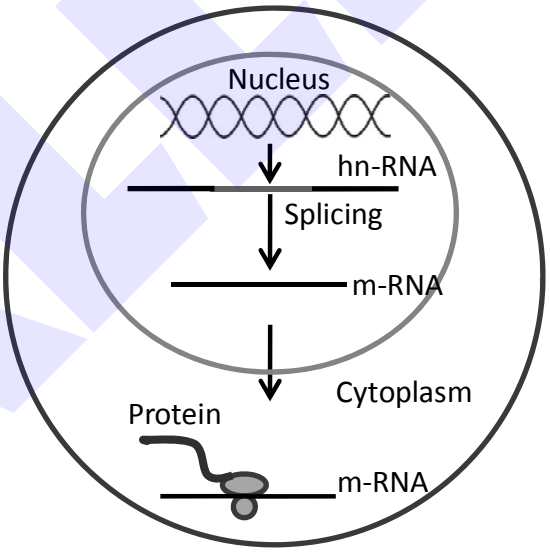
(1) Post transcriptional processing (splicing) is not required.

(2) Transcription and translation take place in the same compartment (cytoplasm).



In Eukaryotes :-

Protein synthesis takes place in cytoplasm, Mitochondria, plastids and surface of RER.



(1) ACTIVATION OF AMINO ACID

- 20 types of amino acids participate in protein synthesis.
- The amino acid are joined by a bond which is known as a peptide bond.
- Formation of a peptide bond requires energy.
- Therefore in first phase itself amino acids are activated in the presence of ATP.

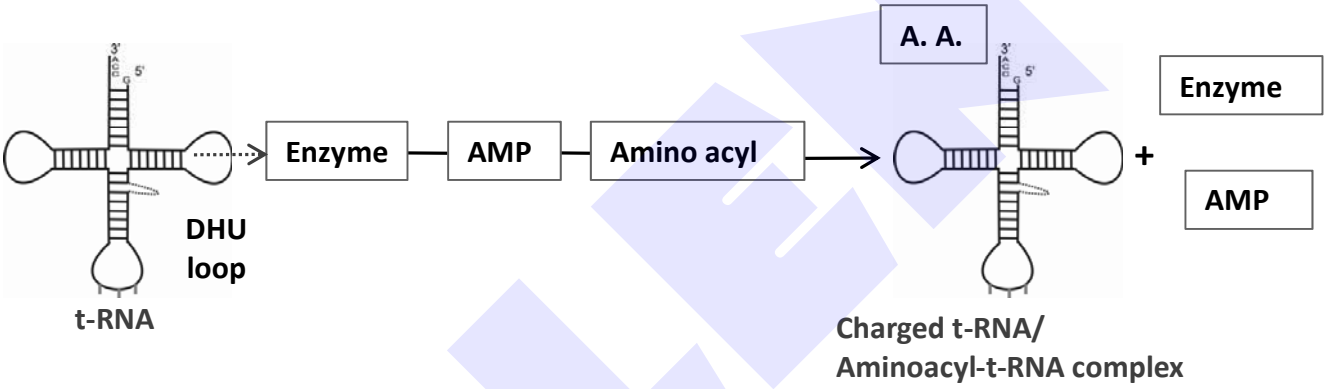
- Amino acid reacts with ATP to form **''Amino acyl AMP enzyme complex''**, which is also known as **'Activated Amino acid'**.

$$\text{Amino acid + ATP} \xrightarrow[\text{t–RNA synthetase enzyme}]{\text{Amino acyl}} \text{Amino acyl AMP–enzyme complex + PPi}$$

- There is a separate **'Amino acyl t–RNA synthetase'** enzyme for each kind of amino acid.

(2) CHARGING OF t–RNA (Loading of t-RNA)/ Aminoacylation of tRNA

- After activation amino acids are linked to their cognate t-RNA a process commonly called as charging / aminoacylation of t RNA.
- Specific activated amino acid is recognised by its specific t–RNA.
- Now amino acid attaches to the **'Amino acid attachment site'** of its specific t–RNA and **AMP** and enzyme are separated from it.



- Now Amino acyl t–RNA moves to the ribosome for protein synthesis.

(3) SYNTHESIS OF POLYPEPTIDE CHAIN

3 steps –

(A) Initiation of Polypeptide Chain :–

Requirements:

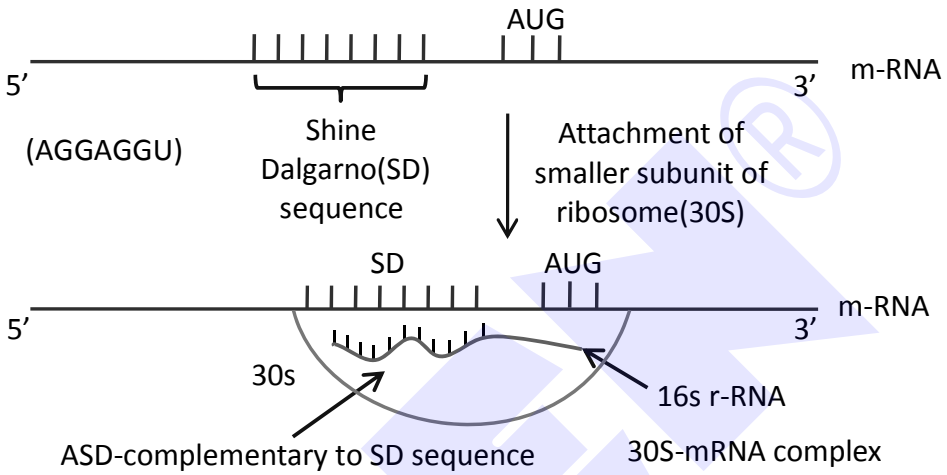
- m-RNA
- charged t-RNA
- 30S and 50S sub units of ribosome
- GTP, Mg^{2+}
- Initiation factors – 3 (in prokaryotes) - IF1, IF2, IF3

Mechanism:

Step (i)

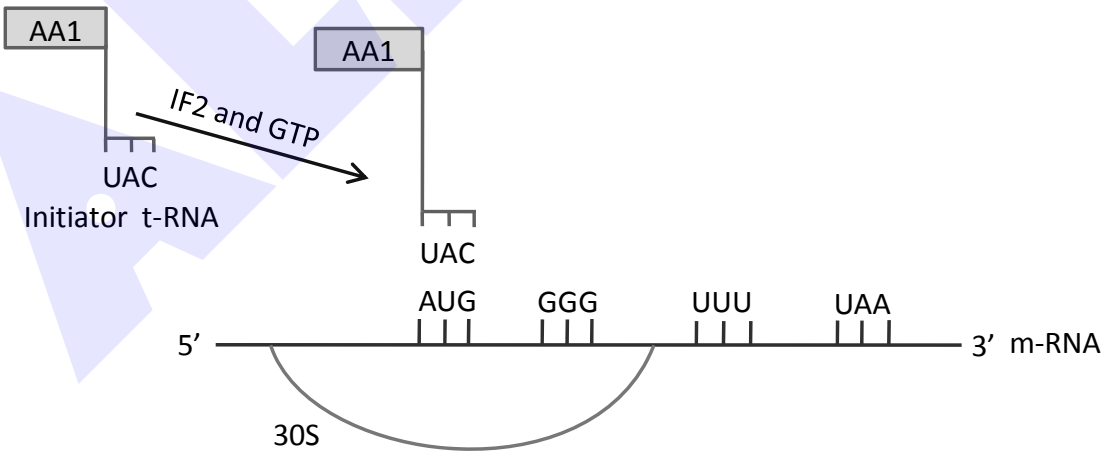
- When the smaller subunit of ribosome encounters an mRNA. The process of translation of the m-RNA to protein begins.

- In prokaryotes with the help of **"S D sequence" (Shine–Dalgarno sequence)** m–RNA recognises the smaller sub unit of ribosome.
- A sequence of few N-bases is present before the 4-12 N-bases of initiation codon on mRNA, called **"SD sequence"**. In Smaller subunit of ribosome, a complementary sequence of **"SD sequence"** is present on 16s rRNA.
- With the help of 'SD' and 'ASD sequence', mRNA and the smaller sub unit of ribosome attached together.



Step (ii)

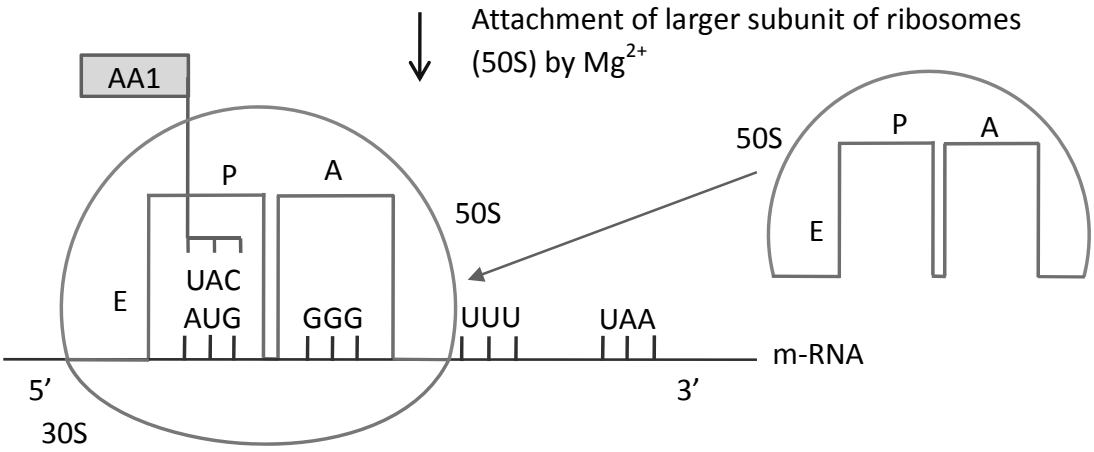
- During initiation, when ribosomal smaller subunit binds to the mRNA. After that the start codon (AUG) that is recognised only by the initiator tRNA.
- First charged tRNA attached by the help of IF_2 and GTP.



30S-mRNA-Nformyl methionyl-tRNA complex

Step (iii)

- In last step of initiation the larger subunit of ribosome (50S) binds with 30S mRNA tRNA complex.



- In larger sub unit of ribosome there are three sites for t–RNA –

  P- site (Peptidyl site) :- It contains t-RNA with growing polypeptide chain.

  A- site (Aminoacyl site) :- Newly arrived t-RNA with amino acid binds here except at initiation codon.

  E- site (Exit site):- t-RNA after donating amino acid move out from here.

(B) Elongation of Polypeptide chain :-

By elongation factors :-

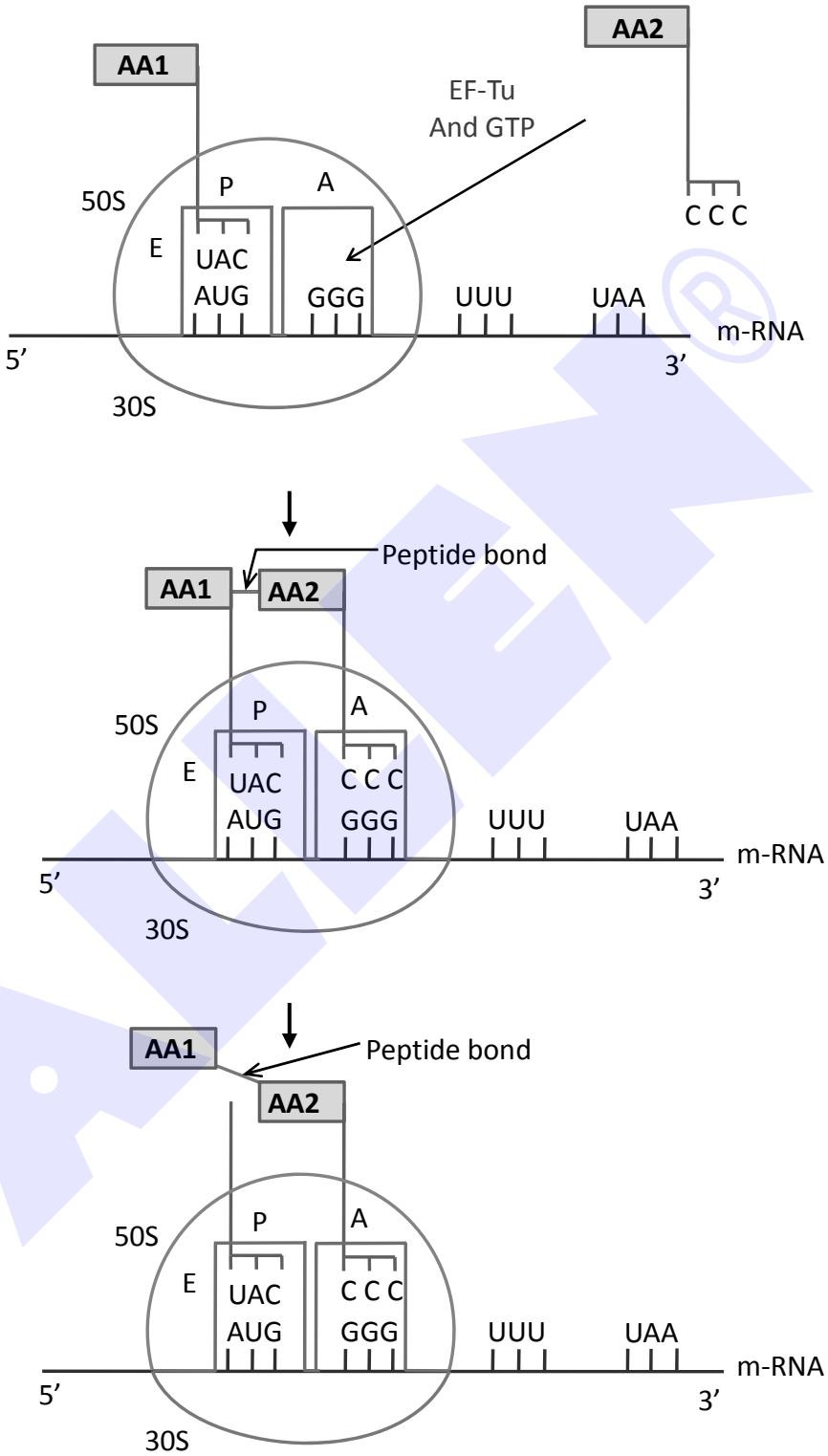
In prokaryotes – 3 factors – (EF – Tu, EF – Ts, EF – G)

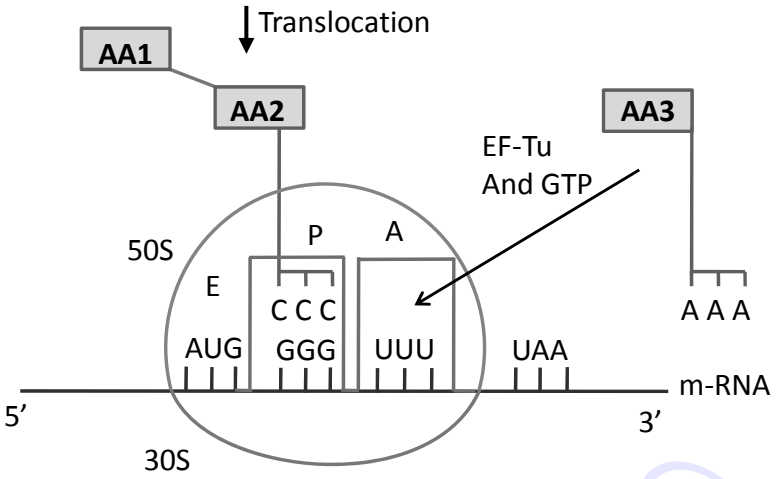
- During elongation new charged tRNA binds at A side of ribosome.
- Peptide bond formation takes place between **– COOH group** of P site amino acid and **– NH_2 group** of A-site amino acid.
- In ribosomal subunit P and A site when two charged tRNA are brought close enough, the formation of peptide bond between them would be favoured energetically.
- The presence of a catalyst would enhance the rate of peptide band formation.
- The ribosome also act as a catalyst (23S rRNA), in bacteria ribozyme enzyme is used for the formation of peptide bond.
- After bond formation t-RNA from P site, exit from the ribosome.
- Now ribosome slides over mRNA strand in 5' $\rightarrow$ 3' direction.

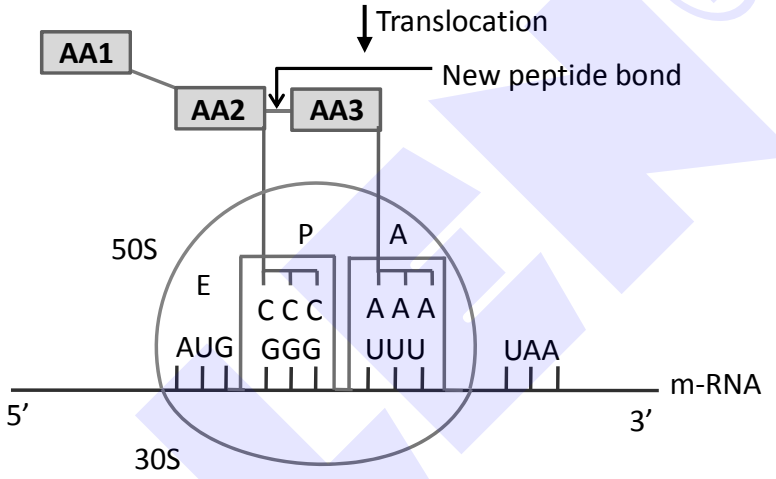
Translocation

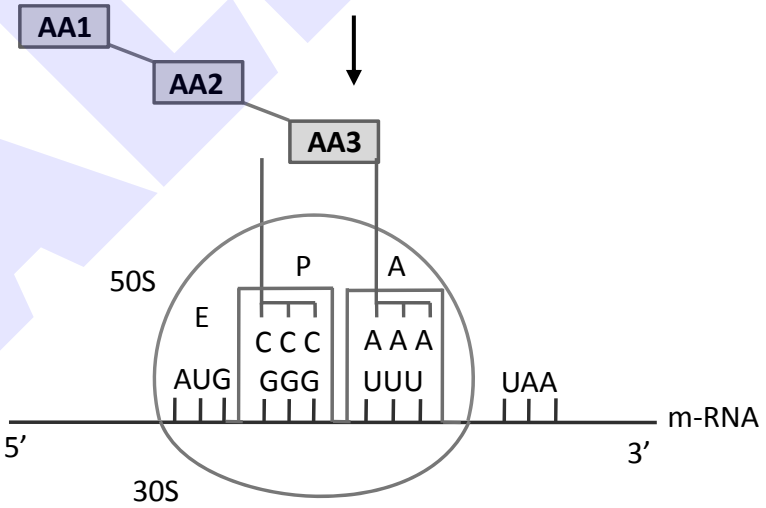
- **Translocase enzyme** is helpful in movement of ribosome (translocation).
- Now t–RNA of A site is transferred to **P site** and **A site** becomes empty.

- Due to continuous sliding of ribosome on m–RNA, new codon of m–RNA continuously available at **A site** of ribosome and according to new codon of m–RNA new amino acid attaches in polypeptide chain.



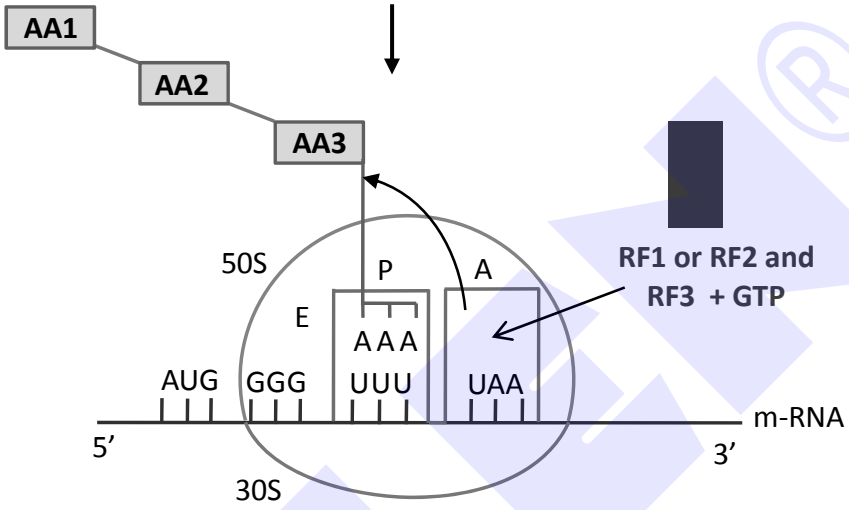

Translocation
AA1
AA2
AA3
EF-Tu
And GTP
50S
P
A
E
C C C
A A A
AUG
GGG
UUU
UAA
m-RNA
5'
3'
30S


Translocation
New peptide bond
AA1
AA2
AA3
50S
P
A
E
C C C
A A A
AUG
GGG
UUU
UAA
m-RNA
5'
3'
30S


AA1
AA2
AA3
50S
P
A
E
C C C
A A A
AUG
GGG
UUU
UAA
m-RNA
5'
3'
30S

(C) Termination of polypeptide chain :–

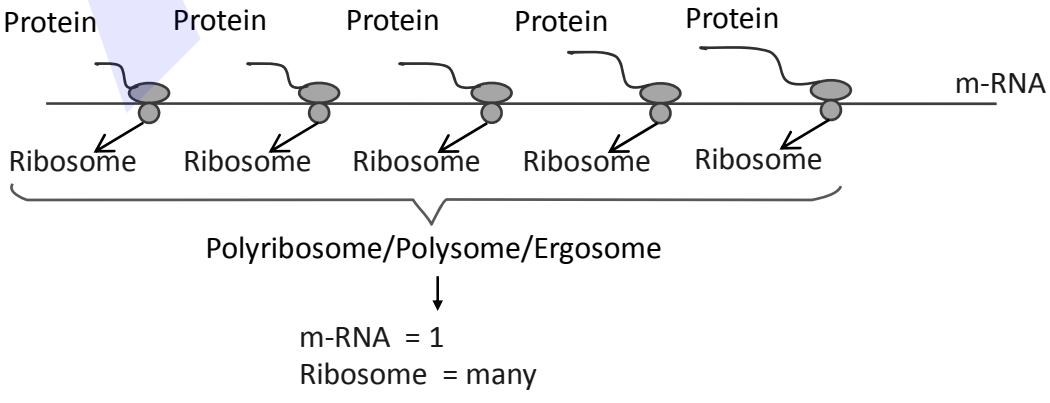
- Due to sliding of ribsome over m–RNA when any **Nonsense codon (UAA, UAG, UGA)** become available at **A site** of ribosome, then polypeptide chain terminates.
- At the end, **releasing factor** binds to the stop codon, terminating translation and releasing the complete polypeptide from the ribosome.
- Release Factors (Prokaryote RF_1, RF_2 and RF_3) :- Break the ester bond between the last t-RNA and amino acid.



- **UTR (untranslated regions) :-** An m-RNA also has some additional sequences that are not translated and are referred to as UTR.
- The UTRS are present at both 5' end (before start codon) and at 3' end (after stop codon).
- UTRs are required for efficient translation process.

Polyribosome / polysome / Ergosome :-

Several ribosomes get attached to single m-RNA.



BEGINNER'S BOX

RIBO NUCLEIC ACID TO TRANSLATION

1. Unidirectional flow of information called central dogma was given by
(1) F.H.C. Crick (2) Temin (3) Baltimore (4) Dulbecco

2. In eukaryotes, RNAPIII catalyses the synthesis of
(1) All rRNA and tRNA
(2) mRNA, HnRNA and SnRNA
(3) 5S rRNA, tRNA and SnRNA
(4) 28S, 18S and 5S rRNA

3. The core enzyme requires a factor for termination of RNA synthesis at some sites. This is known as
(1) Sigma factor
(2) Rho factor
(3) Gamma factor
(4) Alpha particle

4. If one strand of DNA has the base sequence ATCCACGACTAG and the second strand undergoes transcription what would be the base sequence on mRNA ?
(1) TACGTGCTGATC
(2) ATCCACGACTAG
(3) AUCCACGACUAG
(4) AUGCACGACTAG

5. During protein synthesis, amino acid gets attached to tRNA with the help of
(1) mRNA
(2) Aminoacyl synthetase
(3) Ribosome
(4) rRNA

6. The first amino acid in any polypeptide chain of prokaryote is always
(1) Formylated methionine
(2) Formylated arginine
(3) Lysine
(4) Methionine

7. Which site of a tRNA molecule forms hydrogen bonds with mRNA molecule ?
(1) Codon
(2) Anticodon
(3) 5' end of the t-RNA molecule
(4) 3' end of the t-RNA molecule

8. To code the 50 aminoacids in a polypeptide chain, what will be the minimum number of nucleotides in its cistron?
(1) 50 (2) 153 (3) 306 (4) 309

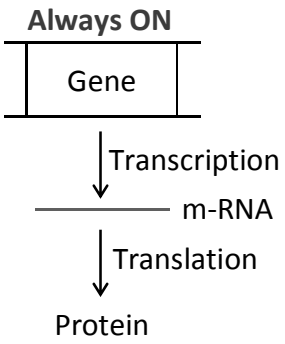
9. A single anticodon can recognize more than one codon of m-RNA. This phenomenon is termed as
(1) Richmond and Lang effect
(2) Gene flow hypothesis
(3) Wobble hypothesis
(4) Transposability

10. The genetic code is called a degenerate code because
(1) One codon has many meanings
(2) More than one codon has the same meaning
(3) One codon has one meaning
(4) There are 64 codons present

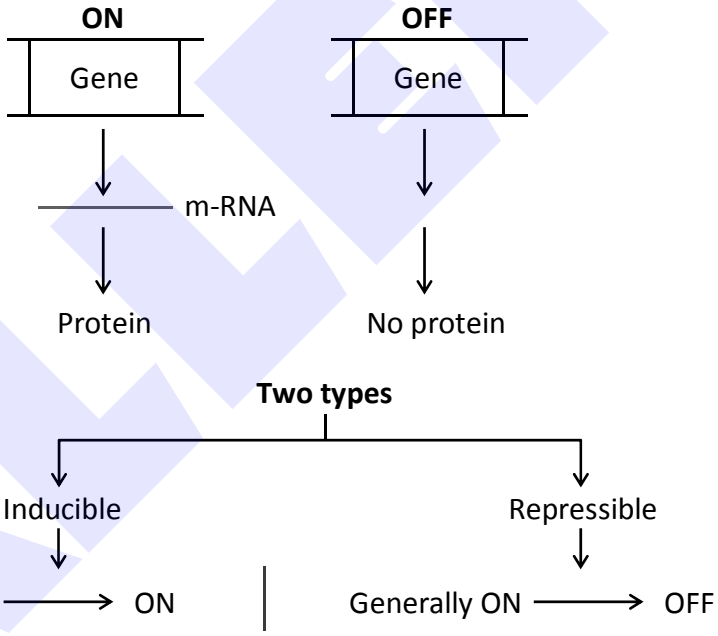
15. REGULATION OF GENE EXPRESSION

Constitutive genes (House-keeping genes) :-

- These genes are always remains active (always ON).
- Their products are constant needed for cellular activity.
- For these genes no needs of regulation e.g. - Genes for enzyme of glycolysis.



- **Non-constitutive genes (Smart gene or Luxury gene) :-**
  - These genes are not always remain active (ON/OFF).
  - Their products are not constant needed for cellular activity.
  - For this genes regulation required.



Gene Regulation :-

1. Gene regulation, includes a wide range of mechanisms that are used by cells to increase or decrease the production of specific gene products (protein or RNA).
2. Since transcription and translation are energetically very expensive processes, these have to be tightly regulated.
3. Gene regulation is the metabolic, physiological or environmental conditions that regulate the expression of genes.
4. The development and differentiation of embryo into adult organisms are also a result of the coordinated regulation of expression of several sets of genes.
5. Gene expression results in the formation of a polypeptide, it can be regulated at several levels.

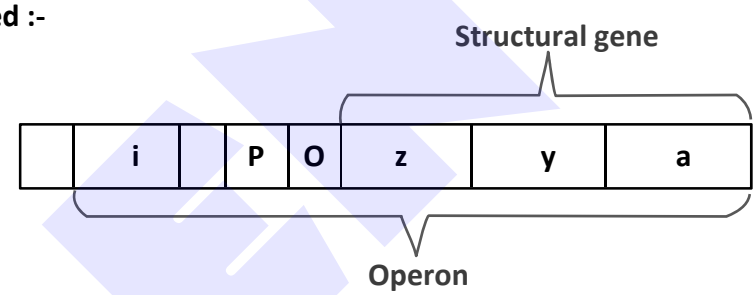
- **In eukaryotes, the regulation could be exerted at :-**
  - (i) Transcriptional level (formation of primary transcript)
  - (ii) RNA processing level (regulation of splicing)
  - (iii) Transport of mRNA from nucleus to the cytoplasm
  - (iv) Translational level.
- **Gene regulation in prokaryotes :-**
  - The predominant site for control of gene expression is at transcriptional initiation level
  - The mechanism of gene regulation in prokaryotes was given by Francis Jacob and Jacques Monad.
  - Jacob and Monad were the first to elucidate a transcriptionally regulated system.
  - They worked on *E.coli* and gave operon model.

OPERON

In bacteria more than one gene is arranged together and regulates in units called as operons.

In operon many genes are involved :-

1. Structural gene
2. Regulator gene (I)
3. Operator gene (O)
4. Promoter gene (P)



Examples of Operon:-

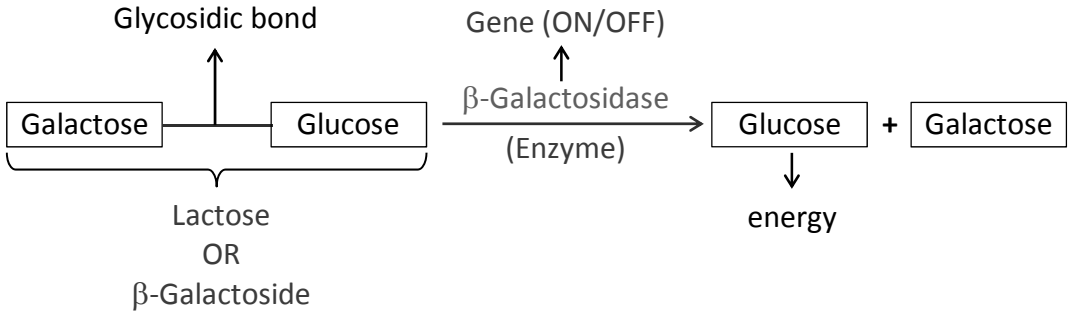
1. lac operon (Lactose)
2. trp operon (Tryptophan)
3. ara operon (Arabinose)
4. his operon (Histidine)
5. val operon (Valine)

Each operon has its specific operator and specific repressor

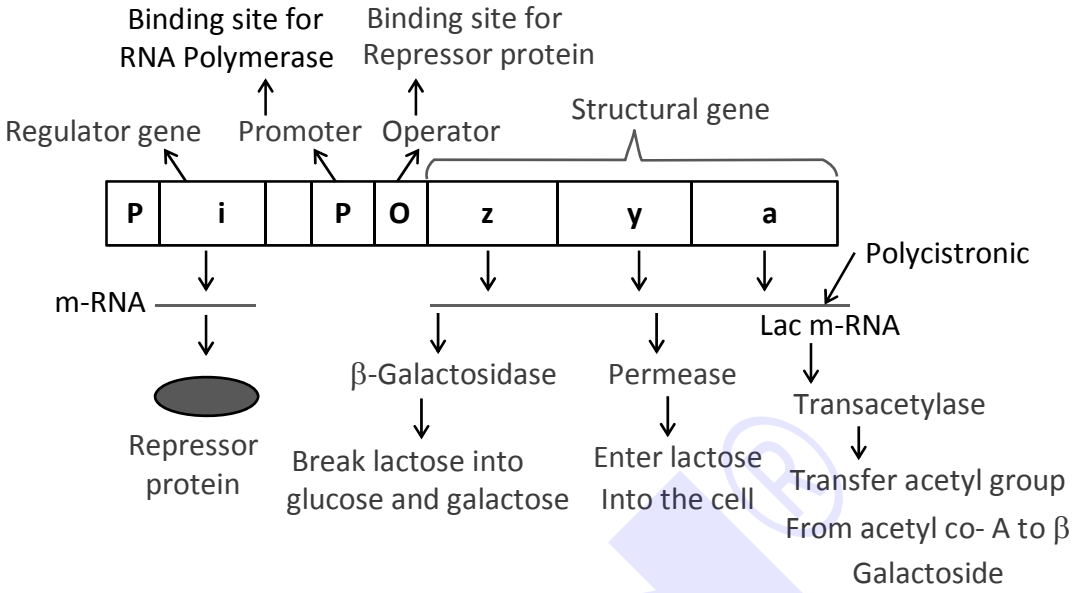
For example, lac operator is present only in the lac operon and it interacts specifically with lac repressor only.

LAC OPERON

- In this operon those genes are regulated which are involved in catabolism (breakdown) of lactose.
- Lactose is a disaccharide compose of glucose and galactose.



Genes of lac operon



The lac operon consists of one regulatory gene (the i gene) and three structural genes (z, y, and a)

Regulator gene (i) -

- The i gene codes for the repressor protein of the lac operon.
- Repressor protein binds to operator gene and stops the working of the structural gene (lac z, lac y, lac a).
- The repressor of the operon is synthesised (all the time-constitutively) from the i gene.
- The i gene - here the term i does not refer to inducer, rather it is derived from the word inhibitor.

Promoter gene :-

- This gene is the site for initial binding of RNA polymerase.
- The enzyme RNA polymerase moves over operator and reaches the structural genes to perform transcription.
- In a transcription unit, the activity of RNA polymerase at a given promoter is in turn regulated by interaction with accessory proteins, which affect its ability to recognise start sites.
- The accessibility of promoter regions of prokaryotic DNA is in many cases regulated by the interaction of proteins (represssor protein) with sequences termed operators.

Operator gene:-

- The operator region is adjacent to the promoter elements.
- In most operons and in most cases the sequences of the operator bind a repressor protein.

Structural gene :-

- The lac operon (lactose operon) of *E. coli* contains three structural genes (lac z, lac y, lac a) (polycistronic structural gene).
- These genes occur adjacent to each other and thus are linked.

- In lac operon a polycistronic structural gene is regulated by a common promoter and regulatory genes.

lac z :-

- The z gene codes for beta-galactosidase (β-gal) enzyme.
- This enzyme is primarily responsible for the hydrolysis of the disaccharide, lactose into its monomeric units, galactose and glucose.

lac y :-

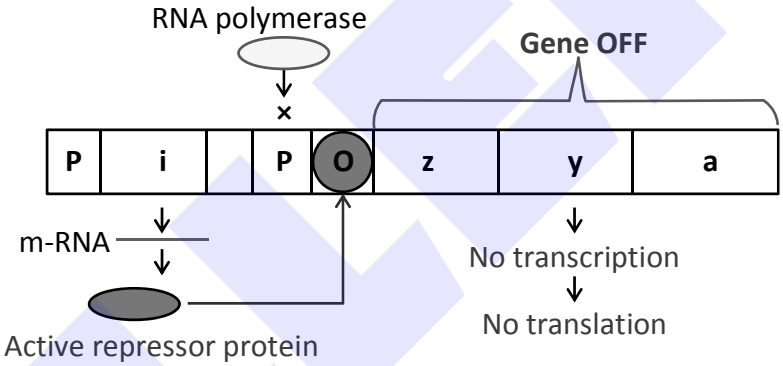
- The y gene codes for permease enzyme .
- This enzyme is increases permeability of the cell to β-galactosides (lactose).

lac a :-

- The a gene encodes an enzyme transacetylase.
- Hence, all the three gene products in lac operon are required for metabolism of lactose.
- In most other operons as well, the genes present in the operon are needed together to function in the same or related metabolic pathway.

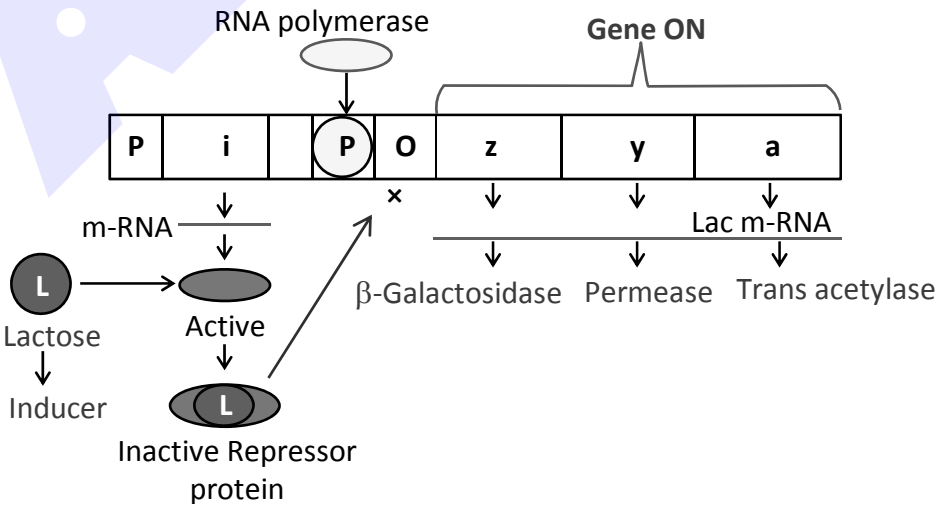
Mechanism of lac operon :-

1. **When lactose is absent (Gene off)**



- In the absence of lactose (inducer), the repressor protein binds to the operator region of the operon and prevents RNA polymerase from transcribing the operon (switching off of the structural gene).

2. **When lactose is present (Gene ON)**

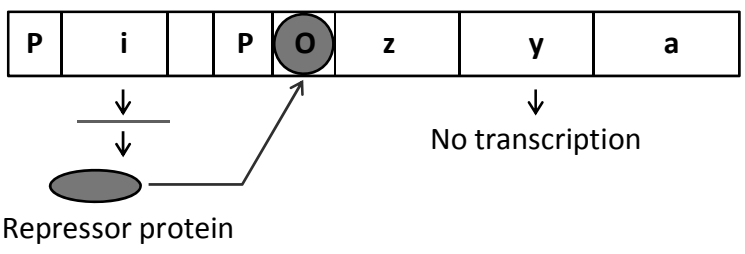


- Lactose is the substrate for the enzyme beta-galactosidase and it regulates switching on and off of the operon. Hence, it is termed as inducer.
- In the absence of a preferred carbon source such as glucose, if lactose is provided in the growth medium of the bacteria, the lactose is transported into the cells through the action of permease.
- In the presence of an inducer, such as lactose or allolactose, the repressor is inactivated by interaction with the inducer.
- This allows RNA polymerase access to the promoter and transcription proceeds.
- Essentially, regulation of lac operon can also be visualised as regulation of enzyme synthesis by its substrate.

Special points of lac operon :-

1. Lac :- lactose
2. i gene :- Constitutive genes
3. z, y and a genes.- non –constitutive genes
4. It is a catabolic pathway
5. Lac operon is an 'Inducible' operon (off $\rightarrow$ on)
6. Inducer:- lactose
7. Real inducer :- allolactose
8. Lac operon is the prototype operon in bacteria.
9. 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.
10. Lac operon show both Negative and Positive regulation

Negative regulation :- When any regulatory molecule binds with DNA and stop transcription called Negative regulation e.g. Regulation by repressor protein.	Positive regulation :- When any regulatory molecule binds with DNA and start transcription called Positive regulation e.g. Regulation by CAP protein. (Catabolic Activator Protein)



Means regulatory proteins can act as both positively (activators) and negatively (repressors).

16. MUTATION

- Mutation is a phenomenon which results in alteration of DNA sequences and consequently results in changes in the genotype and the phenotype of an organism.
- In addition to recombination (crossing over), mutation is another phenomenon that leads to variation in DNA.
- The relationship between genes and DNA are best understood by mutation studies.
- Only those mutation are heritable which occur in **germinal cell** of an organism. While somatic mutations are non heritable.
- Mutation word was given by **Hugo de Vries**.
- **de Vries** studied mutations in the plant ***Oenothera lamarckiana*** (evening primrose).
- **Beadle and Tatum** induced mutations in *Neurospora* by the help of U. V. rays. or X-rays.

  Wild Neurospora $\xrightarrow{\text{U. V. rays}}$ Mutant Neurospora.

- They gave "one gene-one enzyme" concept.
- **M.S. Swaminathan** induced mutations in wheat by the help of γ-rays to obtain good varieties of wheat. eg. Sharbati Sonora, Pusa Lerma.

Types of mutation :

1. CHROMOSOMAL MUTATION
2. GENE MUTATIONS

(1) CHROMOSOMAL MUTATIONS

- Change in number or structure of chromosome.

  Types of chromosomal mutation

  (A) Heteroploidy/Genomatic mutation $\rightarrow$ change in chromosome number.

  (B) Chromosomal aberration $\rightarrow$ change in structure of chromosome.

(A) Heteroploidy / Genomatic Mutation

- Change in number of one or few chromosomes in a set or number of entire set of chromosome.

  It is of two types :

  (i) **Euploidy** $\rightarrow$ Change in number of chromosome sets.

  (ii) **Aneuploidy** $\rightarrow$ Change in number of chromosome in a set.

  (i) Euploidy :

  - Change in number of sets of chromosome i.e. either loss or addition of sets of chromosomes.
  - **Monoploidy (x) -** Presence of one set of chromosomes.
  - **Diploidy (2x) -** Presence of two sets of chromosomes.

- **Polyploidy -** Presence of more than two sets of chromosomes.

  It may be :-

Triploidy (3x)	Tetraploidy (4x)	Pentaploidy (5x)
Hexaploidy (6x)	Heptaploidy (7x)	Octaploidy (8x)

Example

Triticale :- Formed by cross of *triticum aestivum* (wheat) and *Secale cereal* (Rye)

(ii) Aneuploidy :

- Loss or addition of chromosomes in a set of chromosomes.

Types of Aneuploidy :

(a) Hypoaneuploidy (loss)

- **Monosomy (2n – 1) :-** loss of one chromosome.
- **Double monosomy (2n – 1 – 1) :-** loss of two chromosome, but these are non homologous.
- **Nullisomy (2n – 2) :-** loss of two homologous chromosomes.

(b) Hyperaneuploidy (addition)

- **Trisomy (2n + 1) :-** addition of one chromosome.
- **Double Trisomy (2n + 1 + 1) :-** addition of two chromosome but these are non homologous.
- **Tetrasomy (2n + 2) :-** addition of two homologous chromosomes.

- Cause of aneuploidy is chromosomal nondisjunction means chromosomes fail to separate during meiosis.

(B) Chromosomal Aberrations :

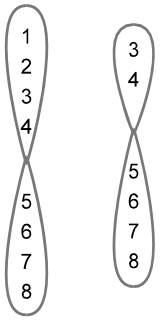
- Loss (deletions) or gain (insertion/duplication) of a segment of DNA, result in alteration in chromosomes.
- Since genes are known to be located on chromosomes, alteration in chromosomes results in abnormalities or aberrations.

(i) Deletion :

- Loss of a part or segment of chromosome which leads to loss of some genes is called as deletion.

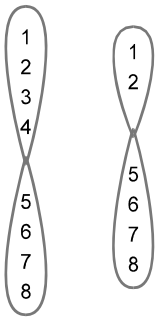
  It is of 2 types :-

  (a) Terminal deletion - Loss of chromosomal segment from one or both ends.



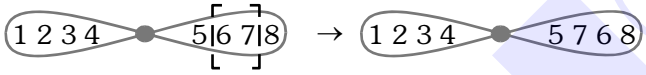
eg. The cri -du-chat syndrome is an example of terminal deletion in short arm of 5th chromosome.

(b) Intercalary deletion - Loss of chromosomal part between the ends.



(ii) Inversion :

Breakage of chromosomal segment but reunion on same chromosome in reverse orders.



(iii) Duplication :

Occurence of a chromosomal segment twice on a chromosome.

A B C D E F → A B C D

A B C D E F → A B C D E F E F

(iv) Translocation :

In this, a part of the chromosome is broken and may be joined with non homologous chromosome.

Types of translocation -

(a) Simple Translocation → When a chromosomal segment breaks and get attached to the terminal end of a non- homologous chromosome.

1 2 3 4 5 6 → 1 2 3 4

A B C D E F → A B C D E F 5 6

(b) Reciprocal Translocation → Exchange of chromosomal segments between two non-homologous chromosome this is also known as **Illegitimate** crossing over (illegal crossing over).

1 2 3 4 5 6 → 1 2 3 4 E F

A B C D E F → A B C D 5 6

eg. **Chronic myeloid leukaemia [C M L] is a type of blood cancer. This disease is a result of reciprocal translocation between 22 and 9 chromosome.**

- Chromosomal aberrations are commonly observed in cancer cells.

(2) GENE MUTATION OR POINT MUTATION

- Mutation arise due to change in a single base pair of DNA. This is known as point mutation.

Two types :-

(A) Substitution

(B) Frame shift mutation.

(A) Substitution :

- Replacement of one nitrogenous base by another nitrogenous base is called as substitution.
- It causes change in one codon in genetic code which leads to change in one amino acid in structure of protein. eg. Sickle cell anaemia
- **Change may not occur some time because for one animo acid more than one type of codons are present.**

Substitution is of two types :-

(i) Transition :

- Replacement of one purine by another purine or replacement of pyrimidine by another pyrimidine.

(ii) Transversion :

- Replacement of **purine** by **pyrimidine** or **pyrimidine** by **purine** is called **transversion**.

(B) Frame Shift Mutation/Gibberish Mutation :

Deletion and insertions of base pair/pairs of DNA, causes frame shift mutation.

Loss or addition of one or rarely more than one nitrogenous bases in structure of DNA.

Frame shift mutation is of two types

(i) Addition / Insertion :

- Addition of one or rarely more than one nitrogenous bases in structure of DNA.

(ii) Deletion :

- Loss of one or rarely more than one nitrogenous bases in structure of DNA.
- Due to frame shift mutation complete reading of genetic code is changed. It leads to change in amino acids after the sequence of addition or deletion of nitrogen base.
- So frame shift mutations are more harmful as compared to substitution.

eg : Thalassemia (lethal genetic disorder)

- Insertion or deletion of one or two bases changes the reading frame from the point of insertion or deletion. However, such mutations are referred to as frameshift insertion or deletion mutations.
- Insertion or deletion of three or its multiple bases insert or delete in one or multiple codon hence one or multiple amino acids, and reading frame remains unaltered from that point onwards.

MUTAGENS :

- Mutagens are those substances which cause mutations.

Ex. **Non ionising radiation :-** U. V. rays.
Gamma rays

★ Golden Key Points ★

- Mostly mutations are harmful.
- Sometimes they are lethal which leads to death of organisms. But sometimes they are beneficial.
- Mostly mutations are recessive and they never eliminate from a population.

Muton (unit of mutation) :

- Smallest part of DNA which undergoes mutation.
- It is one nucleotide.

Mis-sense mutation :-

- When a nucleotide change in genetic code causes the change of one amino acid of a polypeptide chain it is called mis-sense mutation.

Non-sense mutation :-

- When a nucleotide change in one codon causes termination of polypepetide synthesis by producing non-sense codon.

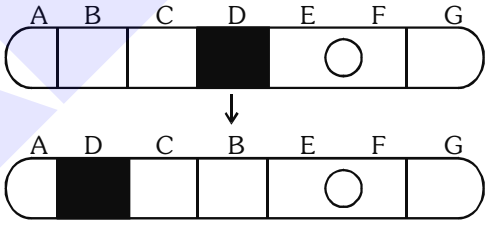
Same sense mutation :-

- A change in one nucleotide in a codon does not change amino acid in polypeptide chain, because both codons code same amino acid.

BEGINNER'S BOX — REGULATION OF GENE EXPRESSION TO MUTATION

1. (2n–1) condition of chromosomes is called :-
(1) tetrasomy (2) trisomy (3) monosomy (4) nullisomy

2. Given below is the representation of a kind of chromosomal mutation. It is :-



(1) Deletion (2) Inversion
(3) Duplication (4) Reciprocal translocation

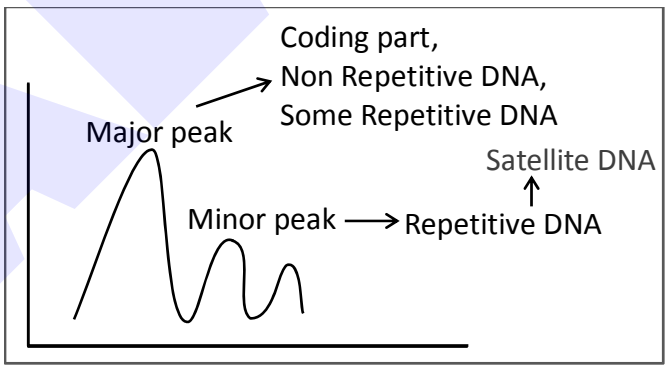
3. Which of the following has normal sex chromosome complements.
(1) Down's syndrome (2) Klinefelter's syndrome
(3) Super female (4) Turner's syndrome

4. Addition or deletion of a nitrogenous base causes
(1) Frameshift mutation (2) Inversion
(3) Transformation (4) Translocation

5. Trisomy of which chromosome is involved in Down's syndrome.
(1) 8^{th} (2) 13^{th} (3) 21^{st} (4) 22^{nd}

17. DNA FINGERPRINTING/DNA TYPING/DNA PROFILING/DNA TEST

- It is technique to identify a person on the basis of his/her DNA specificity.
- This technique was invented by sir Alec Jeffreys (1984)
- In India DNA Finger printing has been started by Dr. V.K. Kashyap and Dr. Lai Ji Singh.
- **Principle :-**
  - DNA of all humans are almost (99.9%) same but very small amount (0.1%) that differs from person to person.
  - Human genome = 3.3×10^9 bp
  - Only 0.1 % different, means 3.3×10^6 bp are different.
  - This differences are mainly at - Repetitive DNA (Satellite DNA).
  - DNA fingerprinting involves identifying differences in some specific regions in DNA sequence called as repetitive DNA, because in these sequences, a small stretch of DNA is repeated many times.
  - These repetitive DNA are separated from bulk genomic DNA as different peaks during density gradient centrifugation.
  - The bulk DNA forms a major peak and the other small peaks are referred to as satellite DNA.



- **Satellite DNA :-**

(1) During density gradient centrifugation maximum repetitive DNA appear as satellite, so this DNA referred to as satellite DNA.

(2) This DNA do not code for any protein.

(3) It makes up large portion of human genome.

(4) It show high degree of polymorphism (variation) in a population.

(5) It also show the high degree of polymorphism in homologus chromosomes of an individual (In homologus pair one chromosome come from father and other chromosome come from mother).

(6) So it form the basis of DNA fingerprinting.

(7) Since DNA from every tissue [such as blood cells, hair-follicle, skin, bone, saliva, sperm etc.] from an individual show the same degree of polymorphism.

(8) Depending on base composition (A : T rich or G : C rich), length of segment, and number of repetitive units, the satellite DNA is classified into many categories, such as micro-satellites, mini-satellites etc.

- **Micro-satellites :-**

(1) This satellites has 1 to 6 bp repetition

(2) Also called SSR (Simple Sequence Repeats)

- **Mini-satellites :-**

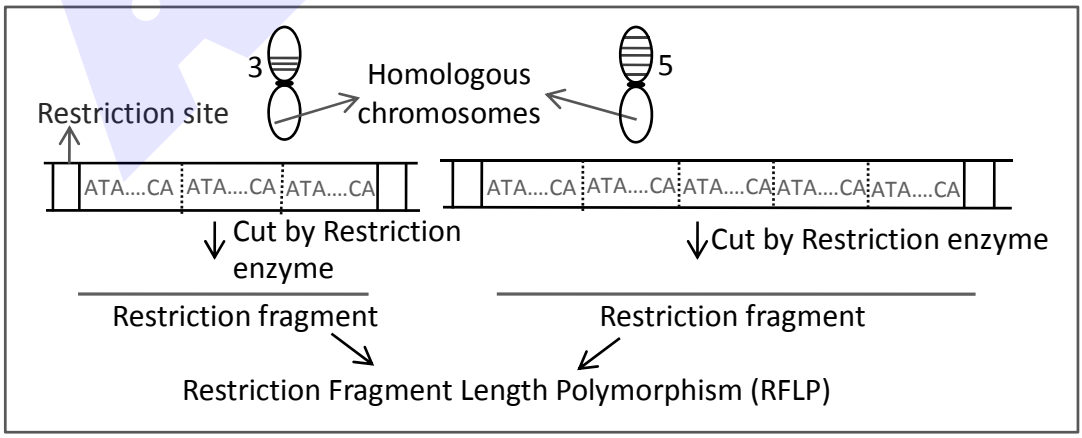
(1) This satellites has 11 to 60 bp repetition

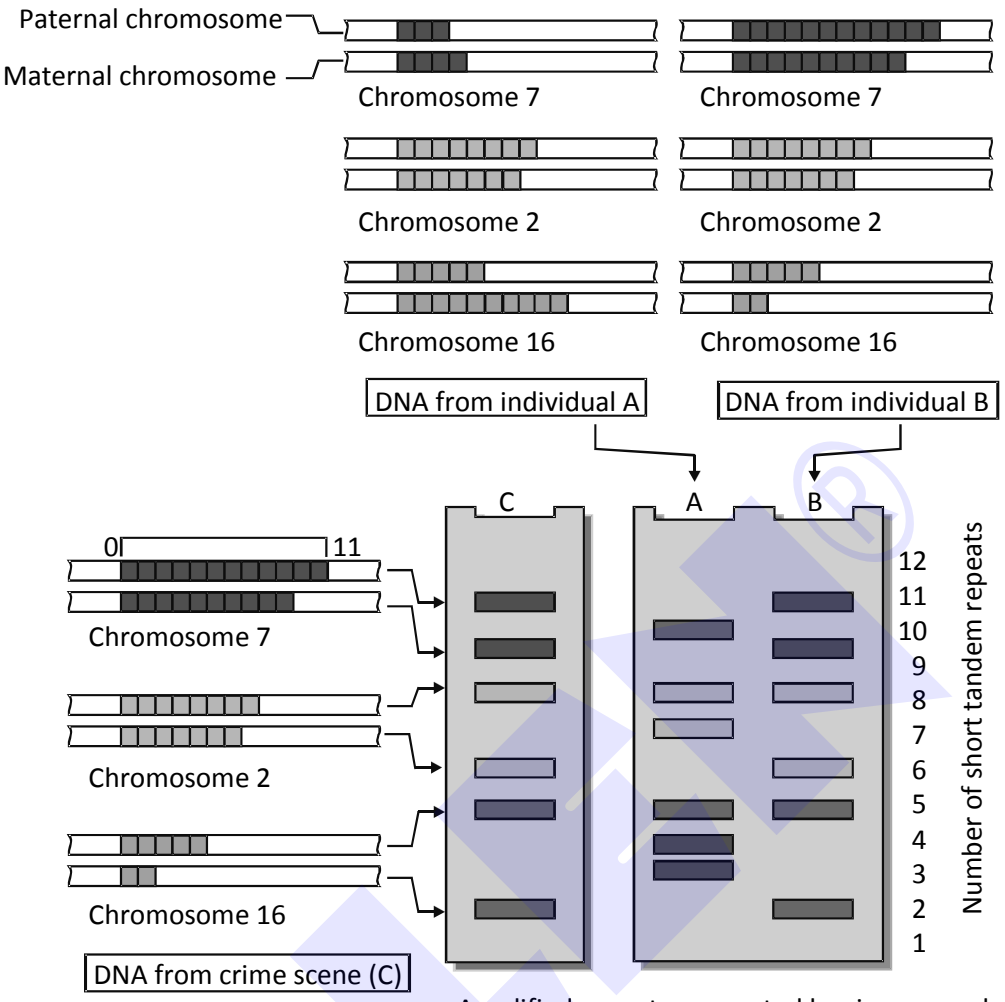
(2) Also called VNTR (Variable Number of Tandem Repeats)

(3) Size of VNTR varies in size from 0.1 to 20 kb

(4) At the end of VNTR palindromic sequence is present

(5) The repetition number of VNTR is vary in homologous chromosomes of an individual so DNA of all human beings are slightly different except monozygotic twins.



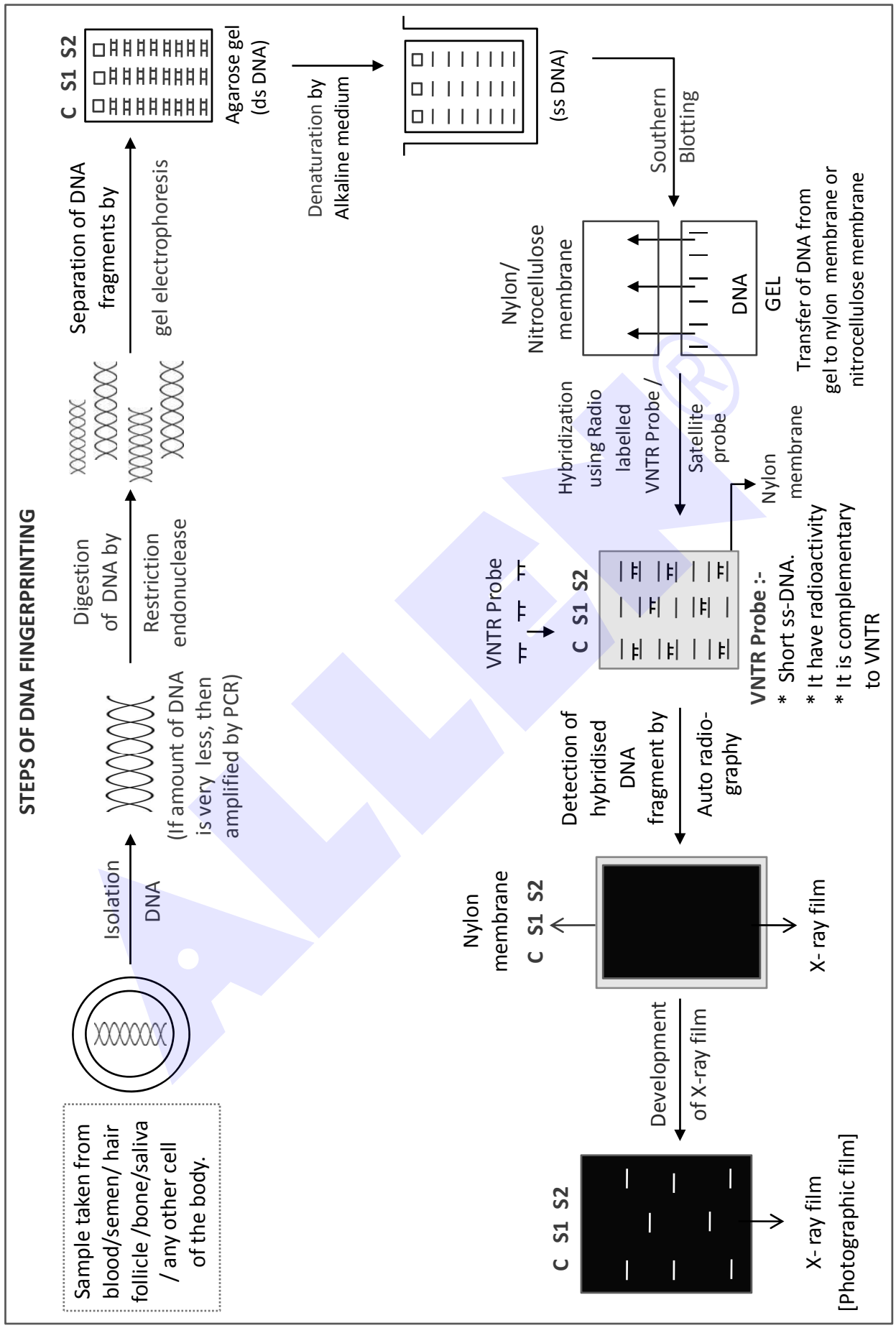


Schematic representation of DNA fingerprinting : Few representative chromosomes have been shown to contain different copy number of VNTR. For the sake of understanding colour schemes have been used to trace the origin of each band in the gel. The two alleles (paternal and maternal) of chromosome also contain different copy numbers of VNTR. It is clear that the banding pattern of DNA from crime sceme matches with individual B, and not with A.

NCERT XII Page No. 123, Figure No. 6.16

STEPS OF DNA FINGERPRINTING :-

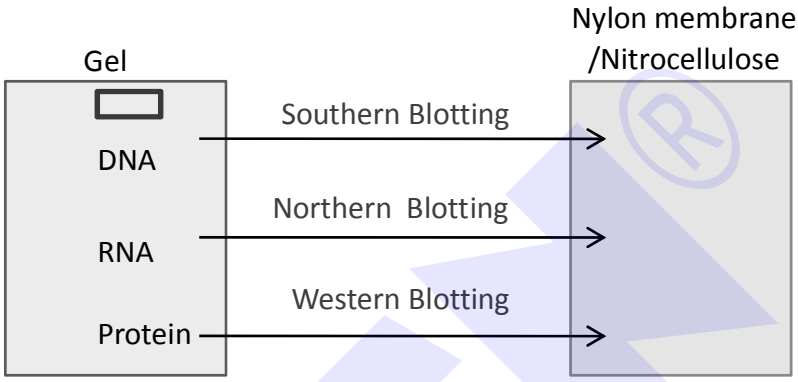
- DNA from a single cell is enough to perform DNA fingerprinting analysis.
- The sensitivity of the technique has been increased by use of polymerase chain reaction [PCR].


STEPS OF DNA FINGERPRINTING
Sample taken from blood/semen/ hair follicle /bone/saliva / any other cell of the body.
Isolation
DNA
(If amount of DNA is very less, then amplified by PCR)
Digestion of DNA by
Restriction endonuclease
Separation of DNA fragments by
gel electrophoresis
C S1 S2
Agarose gel (ds DNA)
Denaturation by Alkaline medium
(ss DNA)
Southern Blotting
Nylon/ Nitrocellulose membrane
DNA
GEL
Transfer of DNA from gel to nylon membrane or nitrocellulose membrane
Hybridization using Radio labelled VNTR Probe / Satellite probe
Nylon membrane
VNTR Probe :-
* Short ss-DNA.
* It have radioactivity
* It is complementary to VNTR
VNTR Probe
C S1 S2
Detection of hybridised DNA fragment by
Auto radio-graphy
Nylon membrane
C S1 S2
X- ray film
Development of X-ray film
C S1 S2
X- ray film [Photographic film]

- After hybridisation with VNTR probe, the autoradiogram gives many bands of different sizes.
- These bands give a characteristic pattern for an individual DNA.
- It differs from individual to individual in a population except in the case of monozygotic (identical) twins.

Blotting: Transferring of separated DNA / RNA / protein fragments from gel to synthetic membranes, such as nitrocellulose or nylon.

Three types :-



Application of DNA fingerprinting :-

1. Identification of the criminal (forensic science)
2. In determining population and genetic diversities
3. Paternity tests (The RFLP pattern of child 50% match with father and 50% with mother).

POLYMORPHISM:-

- In simple terms polymorphism (variation at genetic level) arises due to mutations.
- New mutations may arise in an individual either in somatic cells or in the germ cells (cells that generate gametes in sexually reproducing organisms).
- If a germ cell mutation does not seriously impair individual's ability to have offspring who can transmit the mutation, it can spread to the other members of population (through sexual reproduction).
- 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 a 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 noncoding DNA sequence would be higher as mutations in these sequences may not have any immediate effect/impact in an individual's 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.
- Polymorphisms play very important role in :-
  (i) DNA finger printing
  (ii) Genetic mapping
  (iii) Evolution
  (iv) Speciation

18. HUMAN GENOME PROJECT

- Genetic make-up of an organism or an individual lies in the DNA sequences. If two individuals differ, then their DNA sequences should also be different, at least at some places. These assumptions led to the quest of finding out the complete DNA sequence of human genome.
- 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.
- **Human Genome Project** (HGP) was called a mega project. You can imagine the magnitude and the requirements for the project if we simply define the **aims of the project as follows** :
- Human genome is said to have approximately 3×10^9 bp, and if the cost of sequencing required is US $ 3 per bp (the estimated cost in the beginning), the total estimated cost of the project would be aproximately 9 billion US dollars.
- Further, if the obtained sequences were to be stored in typed form in books, and if each page of the book contained 1000 letters and each book contained 1000 pages, then 3300 such books would be required to store the information of DNA sequence from a single human cell.
- HGP was closely associated with the rapid development of a new area in biology called as **Bioinformatics.**
- The Human Genome Project was a 13-year project coordinated by the U.S. Department of Energy and the National Institute of Health. During the early years of the HGP, the Wellcome Trust (U.K.) became a major partner; additional contributions came from Japan, France, Germany, China and others.

(A) Goals of HGP

Some of the important goals of HGP are as follows :

- Identify all the genes in human DNA.
- Determine the sequences of the 3 billion chemical base pairs that make up human DNA.
- Store this information in databases.
- Improve tools for data analysis.
- Transfer related technologies to other sectors, such as industries.
- Address the ethical, legal, and social issues (ELSI) that may arise from the project.

(B) Methodologies

The methods involved **two** major approaches.

(i) Expressed sequence tags (ESTs) -

Identifying all the genes that expressed as RNA.

(ii) Sequence annotation -

- The blind approach of simply sequencing the whole set of genome that contained all the coding and non-coding sequence, and later assigning different regions in the 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 is 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).
- The fragments were sequenced using automated DNA sequencers that worked on the principle of a method developed by **Frederick Sanger**. (Remember, Sanger is also credited for developing method for determination of amino acid sequences 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 programmes were developed.

- These sequences were subsequently annotated and were assigned to each chromosome. The sequence of chromosome I was completed only in May 2006 (this was the last of the 24 human chromosomes-22 autosomes and X and Y- to be sequenced).
- Another challenging task was assigning the genetic and physical maps on the genome. This was generated using information on polymorphism of restriction endonuclease recognition sites, and some repetitive DNA sequences known as microsatellites.

(C) Salient Features of Human Genome

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

- The human genome contains 3164.7 million bps.
- The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million bases.
- 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 per cent) nucleotide bases are exactly the same in all people.
- The functions are unknown for over 50 per cent of discovered genes.
- Less than 2 per cent of the genome codes for proteins.
- Repeated sequences make up very large portion of the human genome.
- 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.
- Chromosome 1 has most genes (2968). and the Y has the fewest (231).
- Scientists have identified about 1.4 million locations where single-base DNA differences **(SNPs- single nucleotide polymorphism,** pronounced as 'snips') occur in humans, this information promises to revolutionise the processes of finding chromosomal locations for disease-associated sequences and tracing human history.
- The project was completed in 2003. Knowledge about the effects of DNA variations among individuals can lead to revolutionary new ways to diagnose, treat and someday prevent the thousands of disorders that affect human beings.
- Besides providing clues to understanding human biology, learning about non-human organisms, DNA sequences can lead to an understanding of their natural capabilities that can be applied toward solving challenges in health care, agriculture, energy production, environmental remediation.

- Many non-human model organisms, such as bacteria, yeast, *Caenorhabditis elegans* (a freeliving non-pathogenic nematode), *Drosophila* (the fruit fly), plants (rice and *Arabidopsis*), etc., have also been sequenced.

- First prokaryotes in which complete genome was sequenced is ***Haemophilus influenzae.***
- First Eukaryote in which complete genome was sequenced is ***Saccharomyces cerviceae (Yeast).***
- First plant in which complete genome was sequenced is ***Arabidopsis thaliana (Small mustard plant).***
- First animal in which complete genome was sequenced is ***Caenorhabditis elegans (Nematode).***

BEGINNER'S BOX

DNA FINGER PRINTING TO HUMAN GENOME PROJECT

1. DNA finger printing involves identifying differences in some specific

(1) Repetitive DNA (2) Non repetitive DNA

(3) Selfish DNA (4) All of the above

2. Which of the following is produced by E-Coli in the lactose operon.

(1) B galactosidase (2) Transacetylase

(3) Permease (4) All of the above

3. Maximum number of gene present on which chromosome number in human.

(1) 1st (2) X (3) Y (4) 10th

4. In lac operon RNA polymerase binds with

(1) Promoter gene (2) Operator gene

(3) Structural gene (4) Regulator gene

5. Fill the gap in following statement

Human genome have approximately _____ and the cost of sequencing was _____ per base pair.

(1) 4×10^9 bp, 9 billion US dollars

(2) 9 billion US dollars, 4×19^9 bp

(3) 3×10^9 bp, 3 US dollars

(4) 4.7 million bp, 9 billion US dollars

NUCLEIC ACIDS TO MECHANISM OF DNA REPLICATION

Que.	1	2	3	4	5	6	7	8	9	10
Ans.	**4**	**2**	**1**	**2**	**4**	**2**	**2**	**4**	**4**	**2**

RIBO NUCLEIC ACID TO TRANSLATION

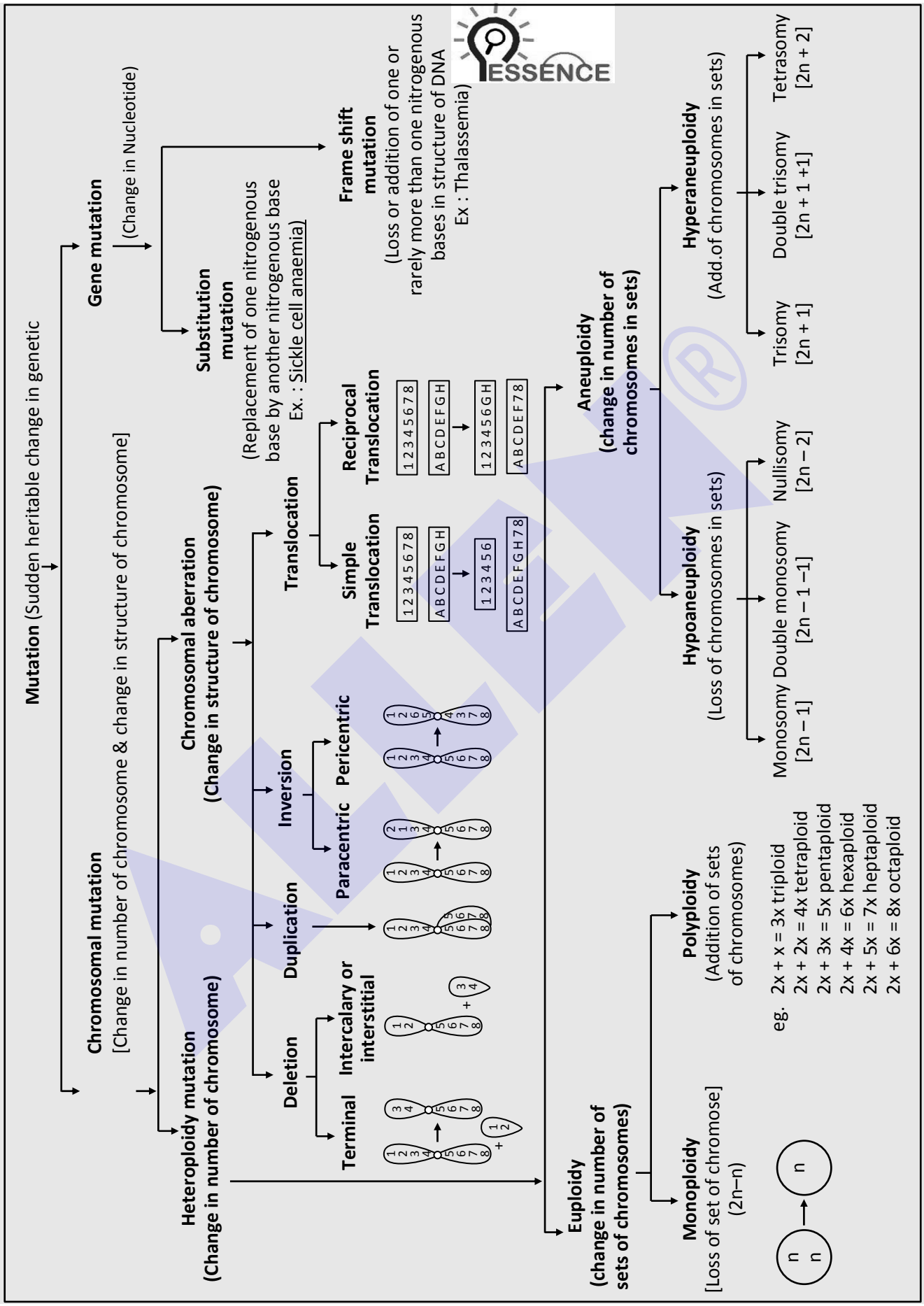
Que.	1	2	3	4	5	6	7	8	9	10
Ans.	**1**	**3**	**2**	**3**	**2**	**1**	**2**	**3**	**3**	**2**

REGULATION OF GENE EXPRESSION TO MUTATION

Que.	1	2	3	4	5
Ans.	**3**	**2**	**1**	**1**	**3**

DNA FINGER PRINTING TO HUMAN GENOME PROJECT

Que.	1	2	3	4	5
Ans.	**1**	**4**	**1**	**1**	**3**



- Meischer (1869) – Discovery of nucleic acid
- Frederick Griffith (1928) – Transforming principle
- Oswald Avery, Colin Macleod, Maclyn McCarty (1933-44) } Biochemical analysis of transforming principle i.e. DNA
- Alfred Hershey and Martha Chase (1952) – Unequivocal proof that DNA is the genetic material
- Maurice Wilkins and Rosalind Franklin – X-Ray diffraction data of DNA
- James Watson and Francis Crick (1953) – Double helix model for structure of DNA
- Watson and Crick – Scheme for DNA replication
- Mathew meselson and Franklin stahl (1958) – Experimental proof for semi-conservative replication of DNA (*E.coli*)
- Taylor and Colleagues (1958) – Experimental proof for semi-conservative replication of DNA in chromosome (*vicia faba*)
- George Gamow – Proposition of triplet code (made of 3 nucleotide)
- Har Govind Khorana – Synthesized RNA molecules with combination of bases (homopolymers and heteropolymers)
- Marshall Nirenberg's – Cell-free system of protein synthesis, helped in code to be deciphered.
- Francis Jacob and Jacque Monod – Lac operon
- Frederick Sanger – Developed method for determination of amino acid sequence in proteins.