02 Molecular Basis of Inheritance

01. NUCLEIC ACIDS

- Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are two types of nucleic acids, found in living systems.
- **F. Miescher** discovered nucleic acid in nucleus of pus cell and called it **'Nuclein'**. The term nucleic acid was coined by **'Altman'**.
- Nucleic acids are polymer of nucleotides.

  Example : DNA and RNA.

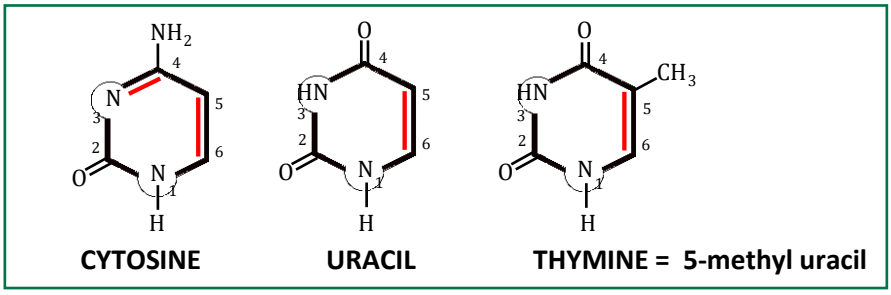
  Nucleotide = Nitrogen base + pentose sugar + phosphate.

  Nucleoside = Nitrogen base + pentose sugar
- DNA acts as the genetic material in most of the organisms. Though RNA also acts as a genetic material in some viruses but mostly functions as a messenger. RNA has additional roles as well. It functions as adapter, structural, and in some cases as a catalytic molecule.

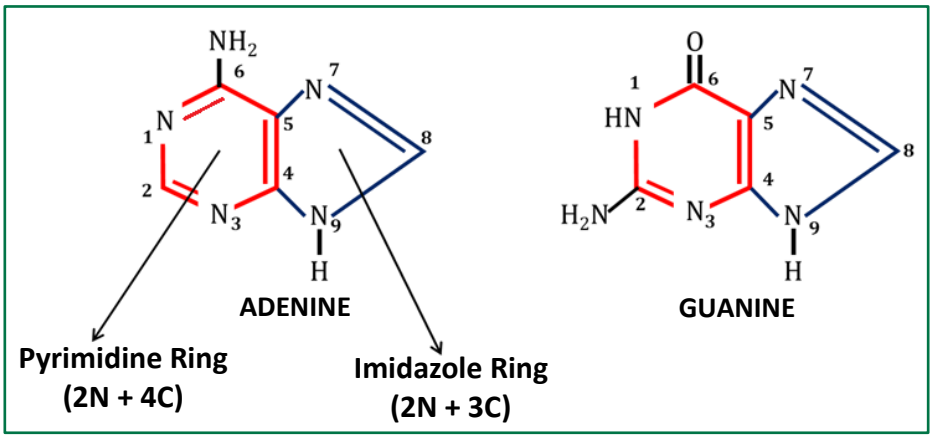
(1) NITROGENOUS BASES :

On the basis of structure Nitrogenous bases are broadly of two types :

(A) Pyrimidines : Consist of one pyrimidine ring. Skeleton of ring composed of two nitrogen and four carbon atoms. ***e.g.*** Cytosine, Thymine and Uracil.

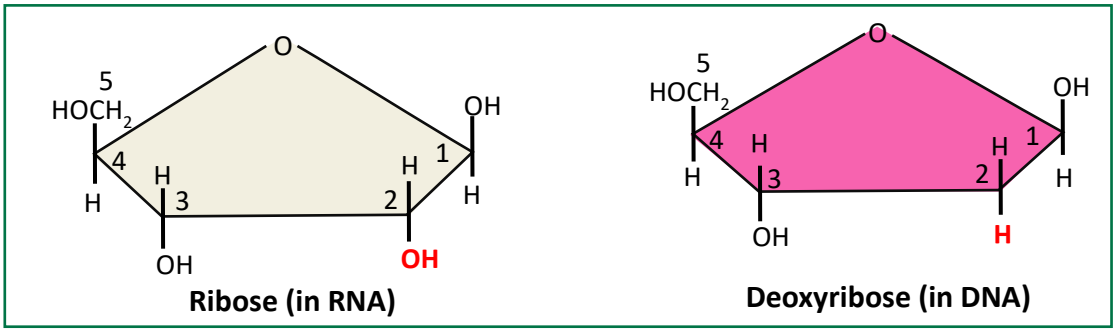


(B) Purines : Consist of two rings i.e. one pyrimidine ring (2N + 4C) and one imidazole ring (2N + 3C) ***e.g.*** Adenine & Guanine.



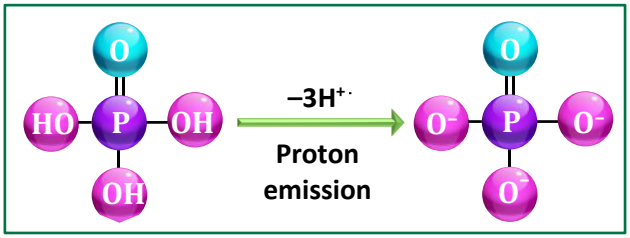
(2) PENTOSE SUGAR :

Number of carbon atoms is five. DNA has deoxy ribose sugar. RNA has ribose sugar.



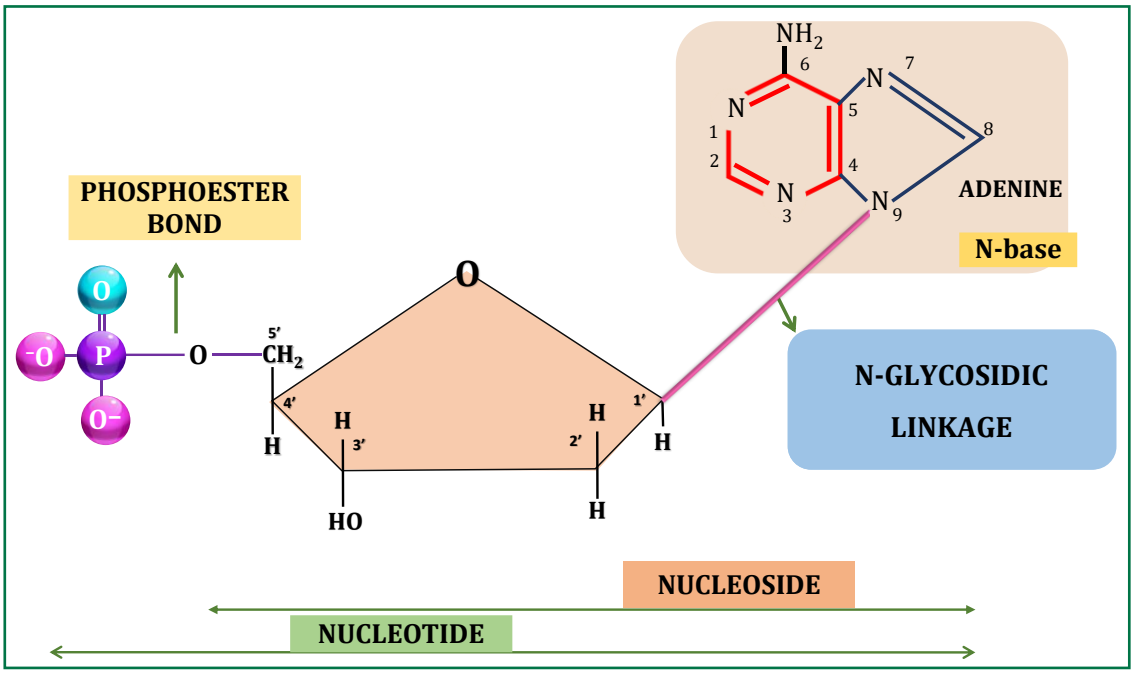
(3) PHOSPHATE :

These are acidic and negatively charged.



(4) NUCLEOSIDE AND NUCLEOTIDE :

- Nitrogenous 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 phosphoester bond (covalent bond) with fifth Carbon of sugar to form a complete nucleotide.



NCERT Question :

Group the following as nitrogenous bases and nucleosides:

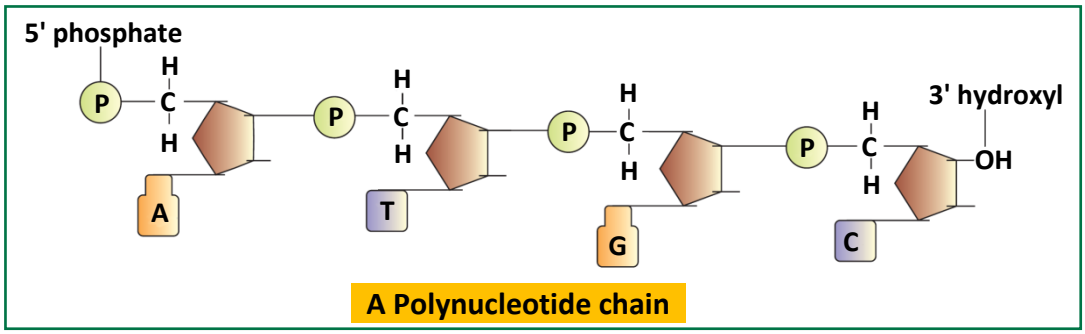
Adenine, Cytidine, Thymine, Guanosine, Uracil and Cytosine.

Types of Nucleotides : For RNA

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

Types of Nucleotides : For DNA

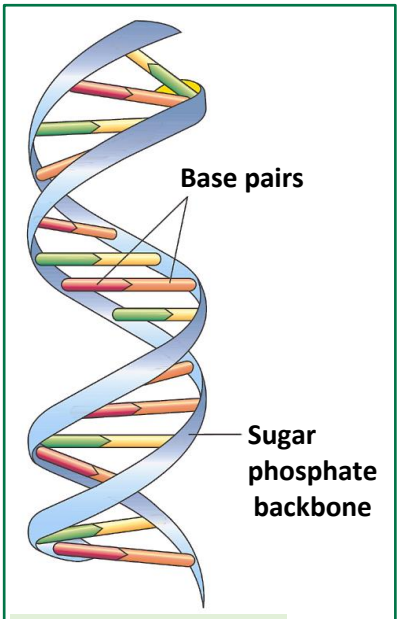
N-base	Nucleosides = N base + Deoxy Ribose sugar	Nucleotides = Nucleoside + P
Adenine	Deoxyadenosine	Deoxyadenylic acid (dAMP)
Guanine	Deoxyguanosine	Deoxyguanylic acid (dGMP)
Cytosine	Deoxycytidine	Deoxycytidylic acid (dCMP)
Thymine	Deoxythymidine	Deoxythymidylic acid (dTMP)



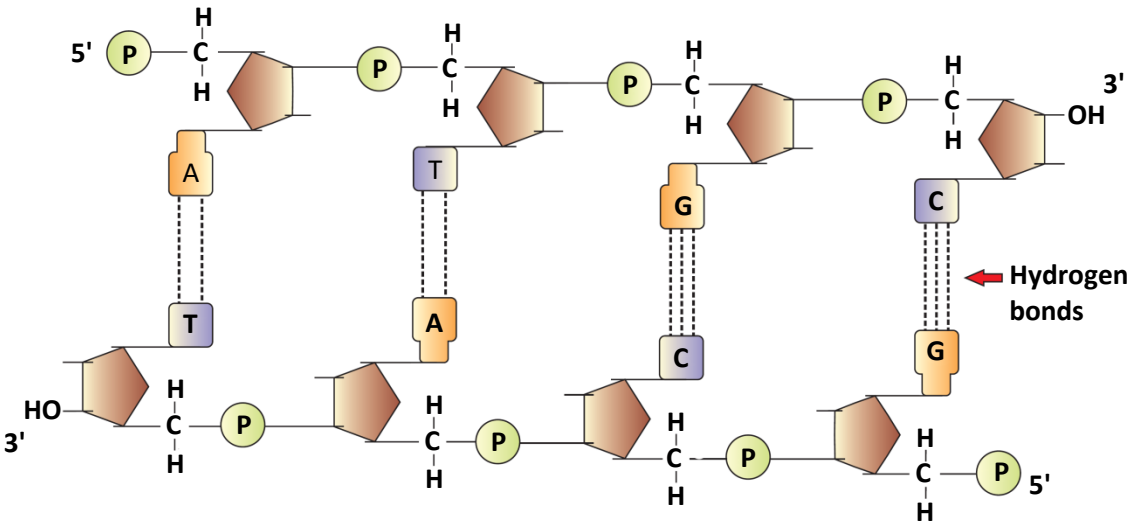
A Polynucleotide chain

02. DNA (DEOXYRIBONUCLEIC ACID)

- DNA as an acidic substance, present in nucleus, was first identified by **Friedrich Miescher** in 1869.
- DNA term was given by – **Zacharias.**
- DNA is long polymer of deoxyribonucleotides.
- DNA is negatively charged (due to presence of Phosphate).
- In DNA pentose sugar is deoxyribose sugar and four types of Nitrogenous bases A, T, G, C
- **Wilkins** and **Franklin** studied DNA molecule with the help of **X-ray crystallography**.
- **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 hallmark (main point) of double helix model is complementary base pairing between purine and pyrimidine.
- According to this model, DNA is composed of two polynucleotide chains.
- Both polynucleotide chains are complementary and antiparallel to each other.
- Both strands of DNA are held together by hydrogen bonds. These hydrogen bonds are present between nitrogenous bases of both 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.
- In DNA plane of one base pair stack over the other in double helix. This, in addition to H-bonds, confers stability of the helical structure of DNA.



DNA Double Helix



Double Stranded Polynucleotide Chain

NCERT Question :

If the sequence of one strand of DNA is written as follows:

5'-ATGCATGCATGCATGCATGCATGCATGC-3'

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

Chargaff's Equivalency rule :

In a double stranded DNA amount of purine nucleotides is equals to amount of pyrimidine nucleotides.

Purine = Pyrimidine

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

$$\frac{[A]+[G]}{[T]+[C]} = 1$$

Also, A = T & G = C

The ratio between adenine and thymine, guanine and cytosine are constant and equals one.

NCERT Question :

If a double stranded DNA has 20 per cent of cytosine, calculate the percent of adenine in the DNA.

Ans. A = 30%

Sol. According to Chargaff's Rule A + G = T + C

If, C = 20% then G = 20%

So, G + C = 40%

Rest A + T = 100 – 40 = 60%

Therefore, T = A = 60/2 = 30%

Base ratio :-

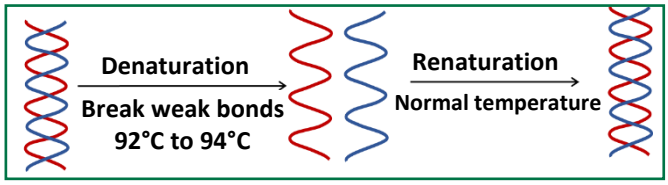
- $\frac{A+T}{G+C}$ = constant for a given species. i.e. species specific.
- In a DNA A + T > G + C $\Rightarrow$ A – T type DNA. Base ratio of A – T type of DNA is more than one.
  e.g. Eukaryotic DNA
- In a DNA G + C > A + T $\Rightarrow$ G – C type DNA. Base ratio of G – C type of DNA is less than one.
  e.g. Prokaryotic DNA

Melting point of DNA depends on G – C content.

- More G – C content 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 (92°C - 94°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 attached and recoil to each other. It is called renaturation of DNA.



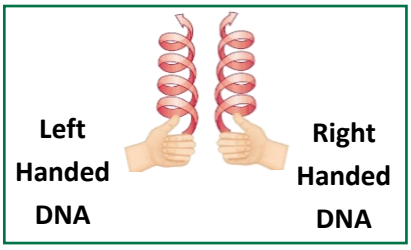
Φ × 174 bacteriophage (Single stranded and circular DNA)	5386 Nucleotides
λ bacteriophage (double stranded and circular DNA)	48502 base pairs
E. coli (double stranded and circular DNA)	4.6×10^6 base pairs
Human -Haploid cell (double stranded linear DNA)	3.3×10^9 base pairs
Human -Diploid cell (double stranded linear 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 : The DNA for which Watson and Crick proposed model was 'B' DNA. It was a right-handed DNA. It shows Clockwise twisting.

Left Handed DNA : Anticlockwise twisting ***e.g.*** Z-DNA. Phosphate and sugar backbone is *zig-zag*.



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

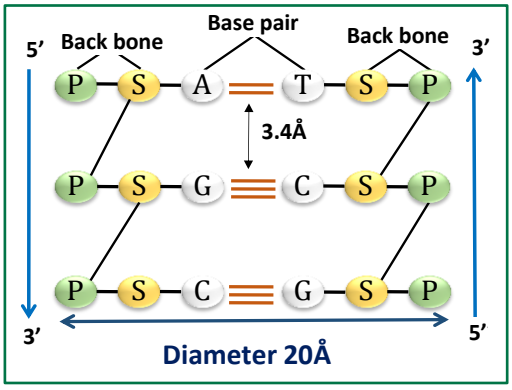
Palindromic DNA :

- Wilson and Thomas (1974) used this term. Sequence of nucleotides same from both ends. (In double stranded DNA where in reading 5' to 3' forward on one strand matches the sequence reading backward 5' to 3' on the complementary strand).

$$5' \xrightarrow{\quad} GAATTC \quad 3'$$
$$3' \quad CTTAAG \xleftarrow{\quad} 5'$$

Structure and Configuration of DNA :

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



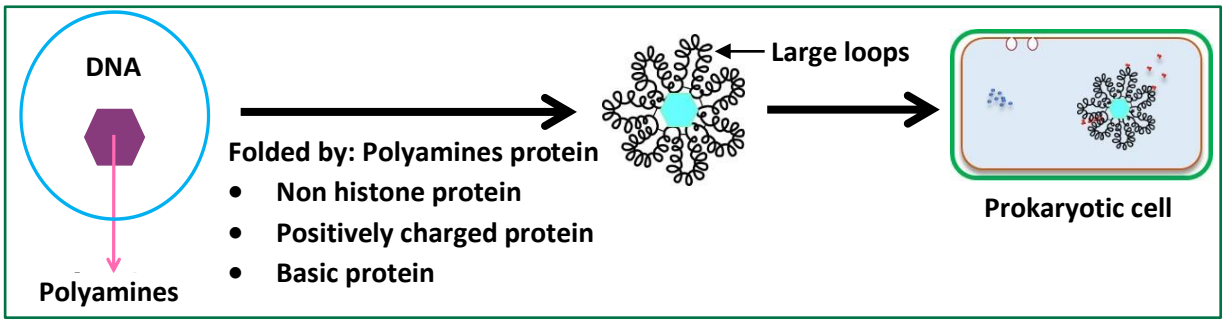
- Two chains have anti-parallel polarity. It means, if one chain has the polarity 5' $\rightarrow$ 3', the other has 3' $\rightarrow$ 5'.
- Distance between two successive steps is 3.4 Å. In one complete turn of DNA molecule there are such 10 steps (10 pairs of Nitrogenous bases). So, the length of one complete turn is 34 Å. This is called helix length.
- Diameter of DNA molecule i.e. distance between phosphates of two strands is 20 Å.
- Each step of ascent is represented by a pair of bases. At each step of ascent, the strands turns 36°.
- In nucleus of eukaryotes the DNA is associated with histone protein to form nucleoprotein.
- Bond between DNA and Histone is salt linkage (Mg^{2+}).

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 (approximately 10^{-6} m).
- The number of base pairs in ***Escherichia coli*** is 4.6×10^6. The total length is 1.36 mm. The long-sized DNA accommodated in small area (about 1 μm in ***Escherichia coli***) only through packing or compaction.
- DNA is acidic due to presence of large number of phosphate groups. Compaction occurs by folding acid attachment of DNA with basic proteins, **polyamine** in prokaryotes and histone in eukaryotes.

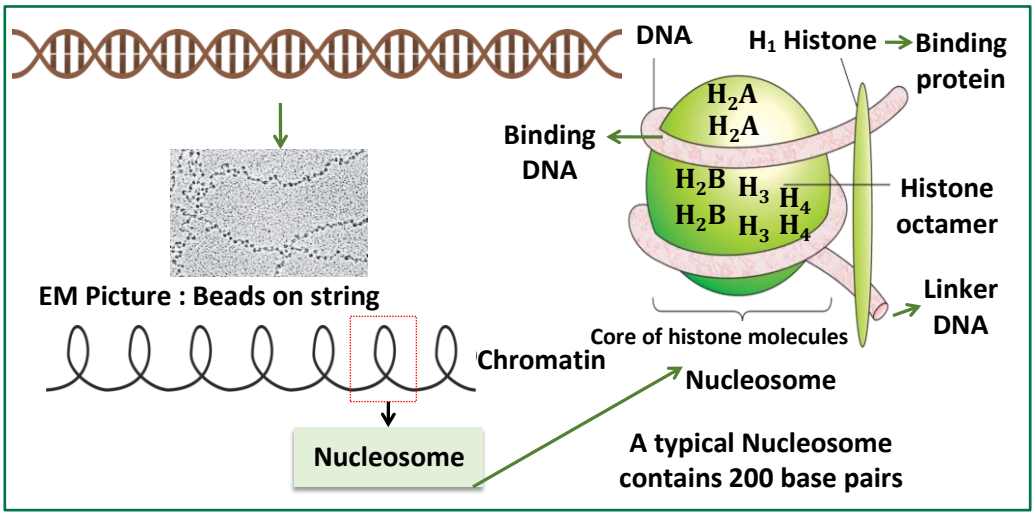
(1) DNA PACKAGING IN PROKARYOTES :

- DNA is found in cytoplasm in supercoiled state. The coils are maintained by non-histone basic protein like polyamines. This compact structure of DNA is called **nucleoid** or **genophore**.

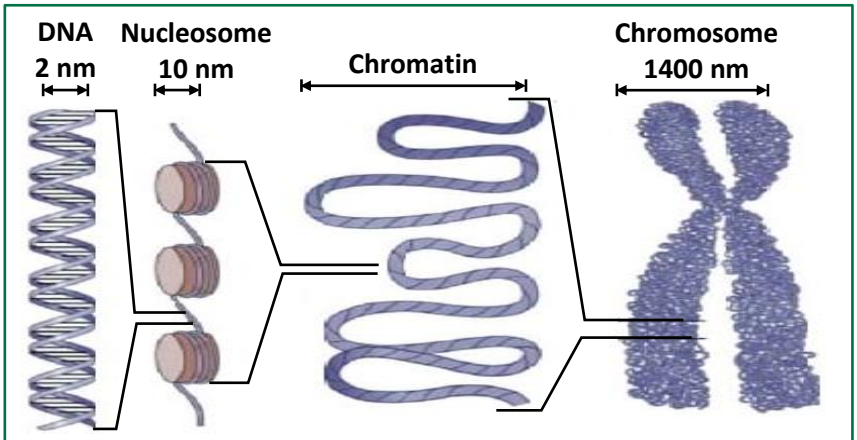


(2) DNA PACKAGING IN EUKARYOTES :

- It is carried out with the help of **lysine** and **arginine** rich basic proteins called **histone**.
- The unit of compaction is **nucleosome**.
- There are five types of histone proteins H_1 H_2A, H_2B, H_3 and H_4. Four of them occur in pairs to produce histone octamer (2 copies of each – H_2A, H_2B, H_3 and H_4), called **nu body** or core of nucleosome.
- Their (nu body) positively charged ends are directed outside. The negatively charged DNA is wrapped around the positively charged histone octamer to form a structure called **nucleosome**.

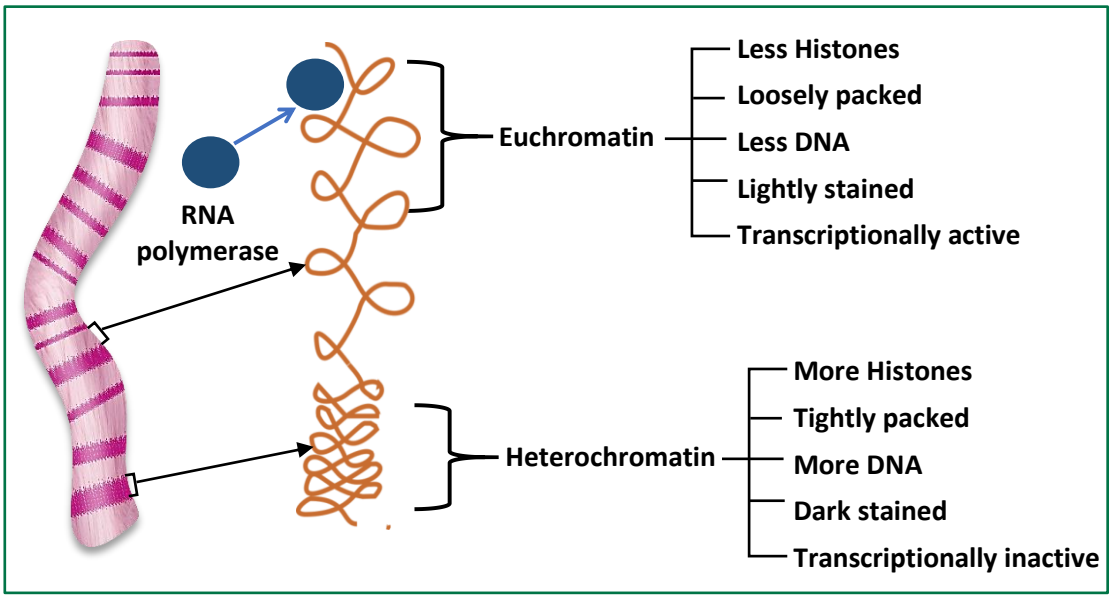


- A typical nucleosome contains 200 bp of DNA Helix.
- DNA present between two adjacent nucleosome is called **linker DNA**. It is attached to H_1 histone protein. Length of linker DNA varies from species to species.
- Nucleosome chain gives a **beads on string** appearance under electron microscope.
- The nucleosomes further coil to form **solenoid**. It has diameter of 30 nm as found in chromatin.
- 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 (acidic) that collectively are referred to as non-histone chromosomal (NHC) proteins.
- Non-Histone chromosomal proteins are of three types :
  (a) Structural NHC protein
  (b) Functional NHC protein e.g., DNA polymerase, RNA polymerase
  (c) Regulatory NHC protein



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

- **Euchromatin and Heterochromatin :-** In a typical nucleus, some region of chromatin is 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, specifically euchromatin is said to be transcriptionally active and heterochromatin is transcriptionally inactive.



04. 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., S-III and R-II.
- Smooth (S) or capsulated type which have a mucous coat and produce shiny colonies. These bacteria are virulent and cause pneumonia in mice.
- 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 in mice.
- The experiment can be described in following four steps:
  (a) Smooth type bacteria were injected into mice. The mice died as a result of pneumonia caused by bacteria.
  Live S strain → Injected into mice → Mice died

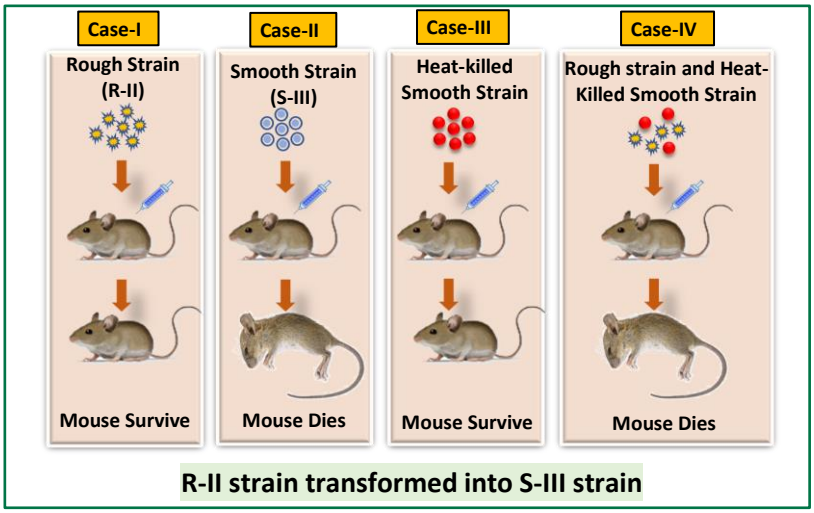
(b) Rough type bacteria were injected into mice. The mice lived and pneumonia was not produced. **Live R strain → Injected into mice → Mice survived**

(c) Smooth type bacteria which normally cause disease were heat killed and then injected into the mice. The mice lived and pneumonia was not caused.

S strain (heat killed) → Injected into mice → Mice survived

(d) Rough type bacteria (living) and smooth type heat-killed bacteria (both do not cause disease) were injected together into mice. The mice died due to pneumonia and virulent smooth type living bacteria could also be recovered from their bodies.

S strain (heat killed) + R strain (living) → Injected into mice → Mice died



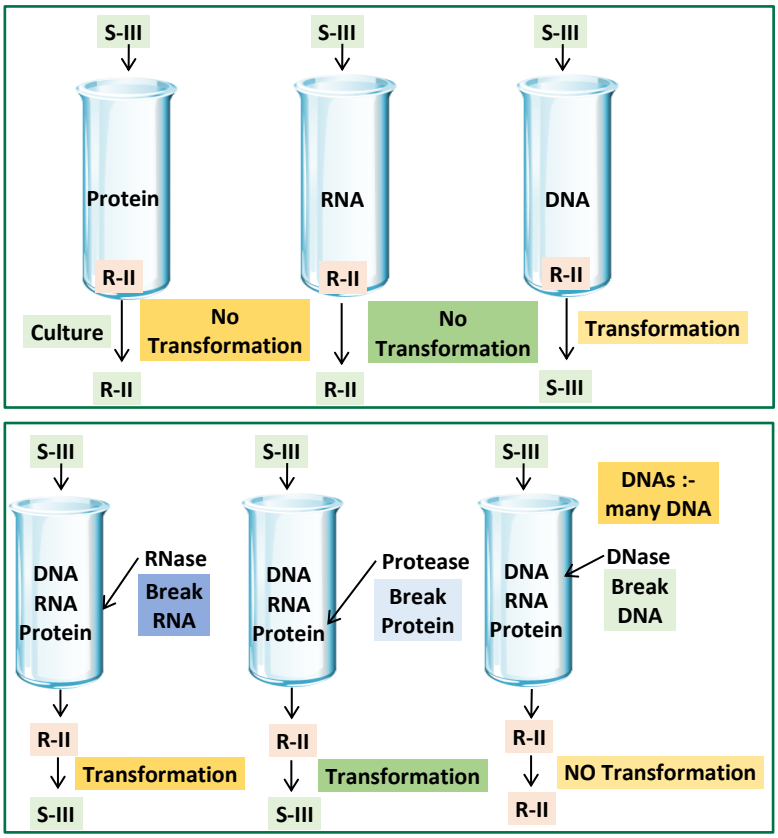
R-II strain transformed into S-III strain

- He concluded from fourth step of the experiment that some rough bacteria (nonvirulent) were transformed into smooth type of bacteria (virulent). This occurred perhaps due to absorption of some transforming substance by rough type bacteria from heat killed smooth type bacteria.
- This transforming substance from smooth type bacteria caused the synthesis of capsule which resulted in production of pneumonia and death of mice. Therefore, transforming principle appears to control genetic characters (for example, capsule as in this case).
- This must be due to the transfer of genetic material.
- However, the biochemical nature of genetic material was not defined from this experiment.

(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.

(3) EVIDENCE FROM EXPERIMENTS WITH BACTERIOPHAGE :

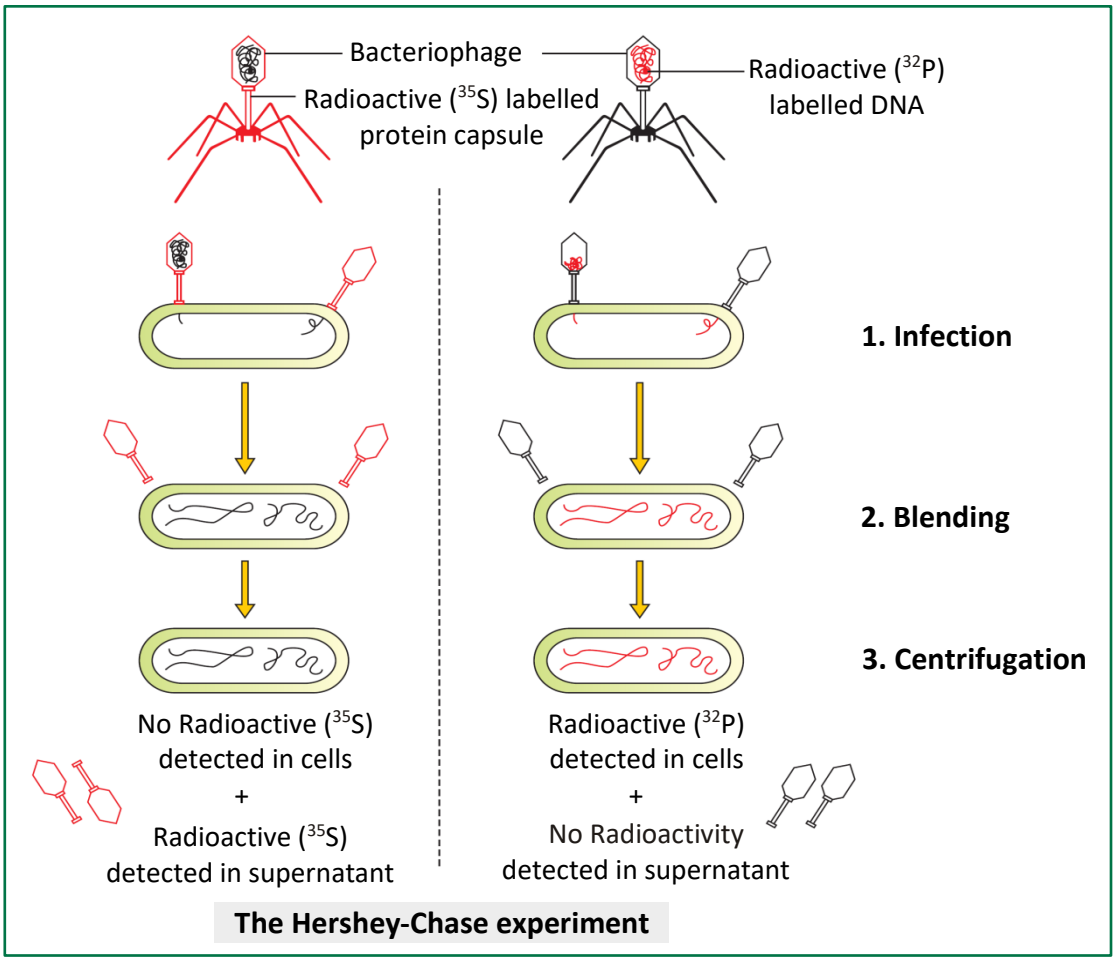
- T_2 bacteriophage is a virus that infects bacterium ***Escherichia coli*** and multiplies inside it.
- T_2 phage is made up of DNA and protein coat. Thus, it is the most suitable material to determine whether DNA or protein contains information for the production of new virus (phage) particles. **Hershey and Chase** (1952) demonstrated that only DNA of the phage enters the bacterial cell and, therefore, contains necessary genetic information for the assembly of new phage particle.
- The functions of DNA and proteins could be found out by labelling them with radioactive tracers. DNA contains phosphorus but not sulphur. Therefore, phage DNA was labelled with P^{32} by growing bacteria infected with phages in culture medium containing $^{32}PO_4$.

- Similarly, protein of phage contains sulphur but no phosphorus. Thus, the phage protein coat was labelled with S^{35} by growing bacteria infected with phages in another culture medium containing $^{35}SO_4$.

NCERT Question :

How did Hershey and Chase differentiate between DNA and protein in their experiment while proving that DNA is the genetic material?

- After the formation of labelled phages. Three steps were followed, i.e., infection, blending, centrifugation.
  - **(i)** **Infection :** Both type of labelled phages were allowed to infect normally cultured bacteria in separate experiments.
  - **(ii)** **Blending :** These bacteria cells were agitated in a blender to break the contact between virus and bacteria.
  - **(iii)** **Centrifugation :** The virus particles were separated from the bacterium by spinning them in a centrifuge.

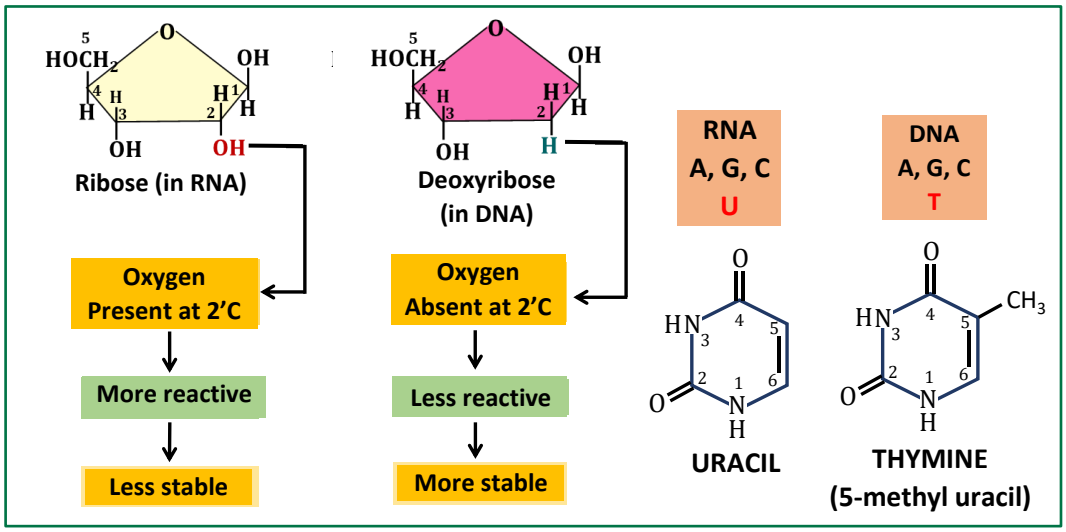


The Hershey-Chase experiment

- 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.
- DNA, therefore, is the infective part of virus and also carries all the genetic information. This provided the **unequivocal proof** that DNA is the genetic material.

Properties of Genetic Material :

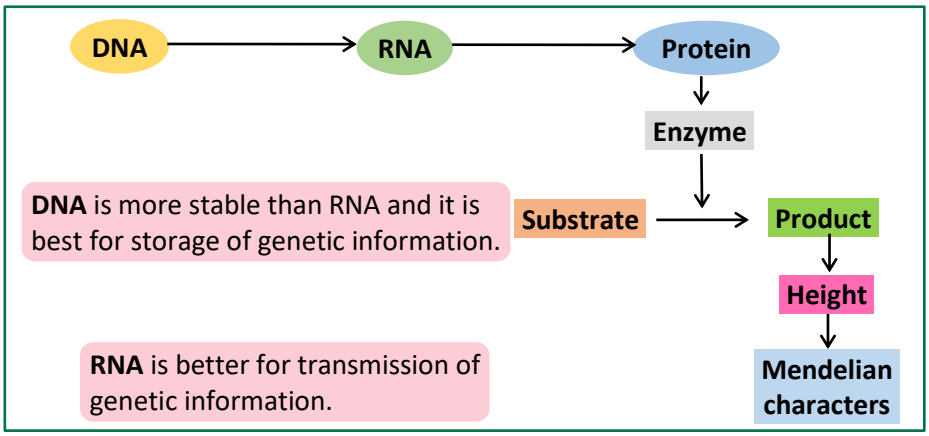
- It became an established fact that it is DNA that acts as genetic material. However, it subsequently became clear that in some viruses, RNA is the genetic material (for example, Tobacco Mosaic viruses, QB bacteriophage, etc.).
- Following are the properties and functions which should be fulfilled by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

DNA	RNA
More Stable	Less stable
Slow mutation	Fast mutation
Less variation	More variation
Slow evolution	Fast evolution

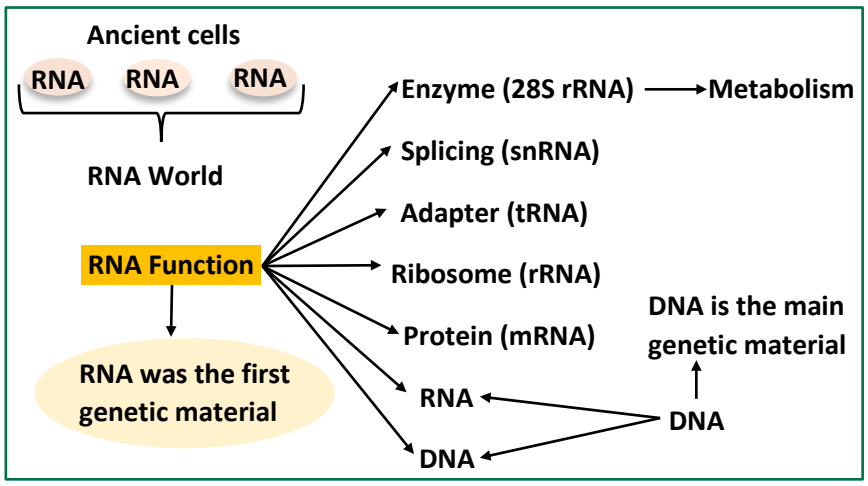
(d) The genetic material should be able to transmit faithfully to the next generation, as Mendelian characters. RNA can directly code for the synthesis of proteins, hence, can easily express the characters. DNA, however, is dependent on RNA for synthesis of proteins. The protein synthesising machinery has evolved around RNA.



- The above discussion indicate that both RNA and DNA can function as genetic material, but DNA being more stable is preferred for storage of genetic information. For the transmission of genetic information, RNA is better.

05. RNA WORLD

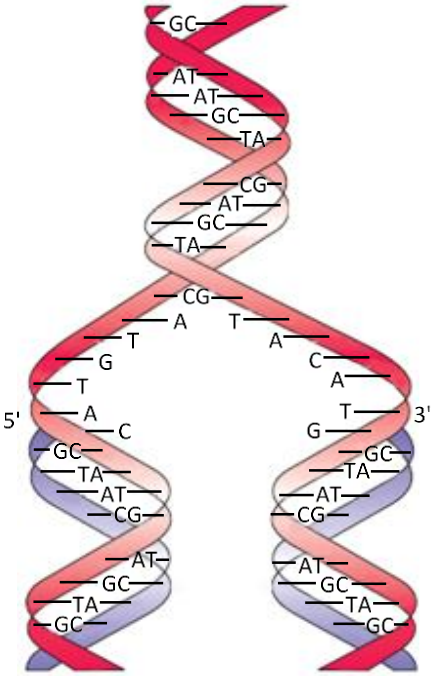
- RNA was the first genetic material. There are evidences to suggest that essential life processes, such as metabolism, translation, splicing etc. evolved around RNA.
- RNA used to act as a genetic material as well as a catalyst, there are some important biochemical reactions in living systems that are 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.



Watson-crick model for semiconservative DNA replication

- DNA capable of self-duplication.
- All living beings have the capacity to reproduce because of this characteristic of DNA.
- DNA replication takes place in **"S - Phase"** of the cell cycle. At the time of cell division, it divides in equal parts in the daughter cells.

NCERT Question :

Which property of DNA double helix led Watson and Crick to hypothesise semi-conservative mode of DNA replication? Explain.

07. SEMI CONSERVATIVE MODE OF DNA REPLICATION

- Semi conservative mode of DNA replication was first proposed by Watson & Crick. Later on it was experimentally proved by **Meselson & Stahl (1958)** ***in E. coli*** and **Taylor** in ***Vicia faba*** **(1958)**. To prove this method, Taylor used Radiotracer Technique in which Radioisotopes (tritiated thymidine = $_1H^3$) were used. Meselson and Stahl used heavy isotope of Nitrogen (N^{15}).
- Matthew Meselson and Franklin Stahl performed the following experiment in 1958:
  (i) 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 Caesium 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 densities).
  (ii) Then they transferred the cells into a medium with normal $^{14}NH_4Cl$ and took samples at various definite time intervals as the cells multiplied and extracted the DNA that remained as double-stranded helices.
  The various samples were separated independently on CsCl gradients to measure the densities of DNA.
  (iii) Thus, the DNA that was extracted from the culture 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.

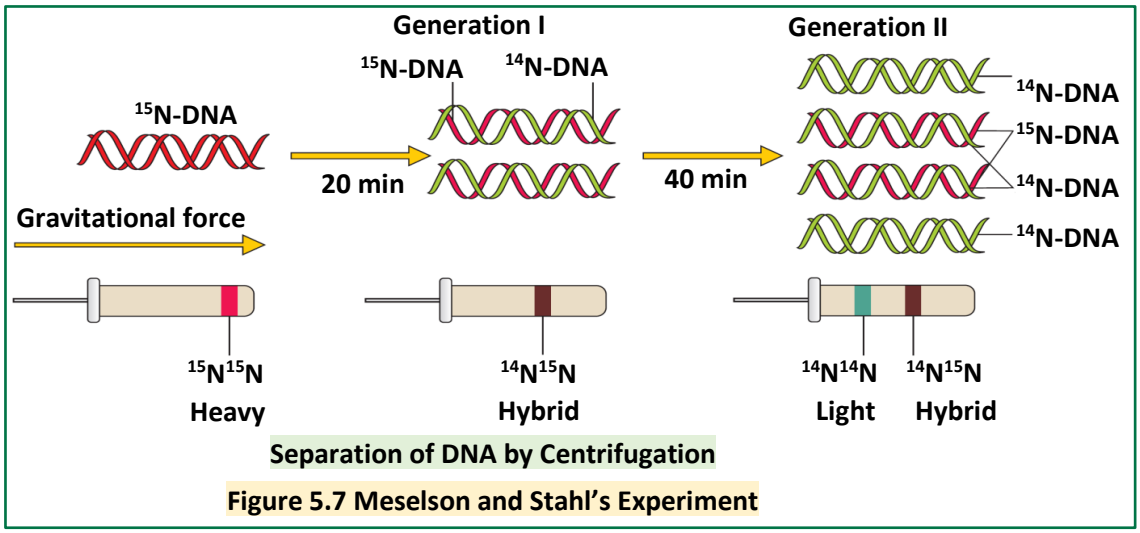


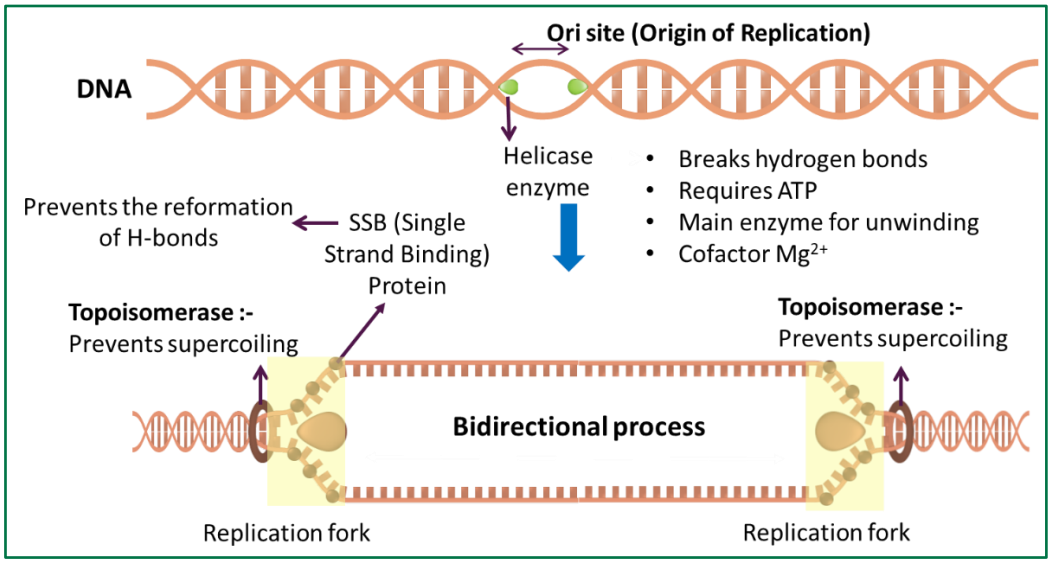
Figure 5.7 Meselson and Stahl's Experiment

08. MECHANISM OF DNA REPLICATION

The following steps are included in DNA replication :-

(1) UNZIPPING (UNWINDING) :

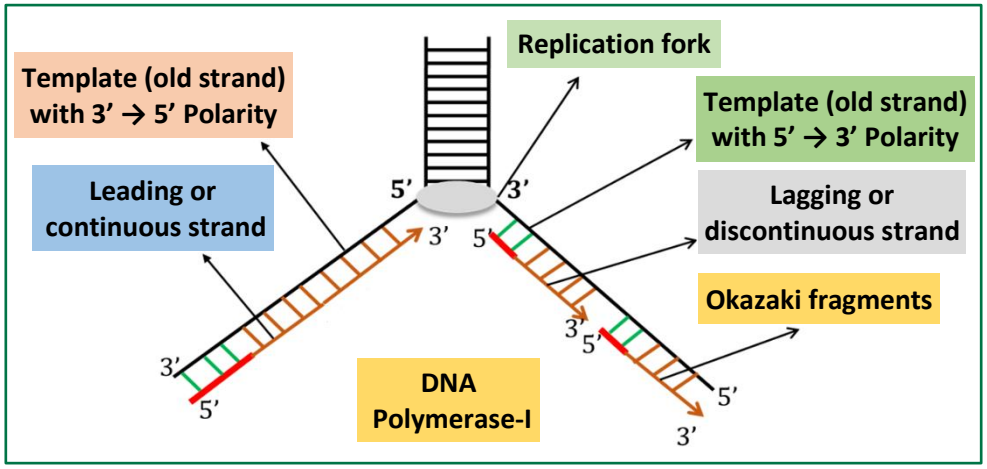
- The separation of 2 chains of DNA is termed as unzipping. And it takes place due to the breaking of Hydrogen bonds. The process of unzipping starts at a certain specific point which is termed as **initiation point** or **origin of replication**.
- In prokaryotes there occurs only one origin of replication but in eukaryotes there occur many origins of replication i.e. unzipping starts at many points simultaneously. At the place of origin, the enzyme responsible for unzipping (breaking the hydrogen bonds) is **Helicase**. In the process of unzipping Mg^{2+} acts as cofactor.
- **SSB** (single stranded DNA binding protein) prevents the reformation of H-bonds.
- **Topoisomerase** (in prokaryotes also called as DNA gyrase) releases 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 :

- 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' $\rightarrow$ 3' direction.
- After the formation of new chain, this RNA is removed. For the formation of new chain nucleotides are obtained from nucleoplasm. In the nucleoplasm, nucleotides are present in the form of triphosphates like dATP, dGTP, dCTP, dTTP etc.

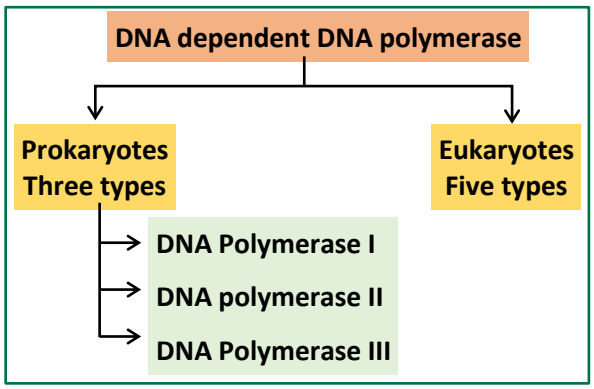


- During replication, the 2 phosphate groups of all nucleotides are separated. In this process energy is yielded which is consumed in DNA replication.
- Energetically replication is a very expensive process. Deoxyribonucleoside triphosphates serve **dual purposes** in addition to **acting as substrates** they provide **energy for polymerisation**.
- The formation of new chain always takes place in 5' - 3' direction. As a result of this, one chain of DNA is continuously formed, and it is termed as **Leading strand**. The formation of second chain begins from the centre and not from the terminal points, so this chain is discontinuous and is made up of small segments called **Okazaki Fragments**. This discontinuous chain is termed as **Lagging strand**. Ultimately all these segments are joined together, and a completely new chain is formed.
- The Okazaki fragments are joined together by an enzyme **DNA Ligase**.
- The formation of new chains is catalysed by an enzyme DNA Polymerase. In prokaryotes it is of 3 types :

(a) DNA - Polymerase I : This was discovered by KORNBERG (1957). So, it is also called as 'Kornberg's enzyme'. Kornberg also synthesized DNA first of all, in the laboratory. This enzyme functions as exonuclease. It separates RNA - primer from DNA and also fills the gap. It is also known as **DNA-repair enzyme**.

(b) DNA - Polymerase II : It is least reactive in replication process. It is also helpful in DNA-repairing in absence of DNA-polymerase-I and DNA polymerase-III

(c) DNA - Polymerase III : This is the main enzyme in DNA - Replication. It is most important. The larger chains are formed by this enzyme. This is also known as Replicase.



- In the semi conservative mode of replication each daughter DNA molecule receives one chain of polynucleotides from the mother DNA - molecule and the second chain is synthesized.

Special Point :

- All DNA polymerase I, II and III enzymes have 5'-3' polymerisation activity and 3'-5' exonuclease activity.
- DNA polymerase I has an additional 5'-3' exonuclease activity for removal of RNA primer.
- Any failure in cell division after DNA replication result into polyploidy.
- Difference between DNAs and DNase is that DNAs means many DNA and DNase means DNA digestive enzymes.
- In living cells, such as *E. coli*, the process of replication requires a set of catalysts (enzymes). The main enzyme is referred to as DNA-dependent DNA polymerase, since it uses a DNA template to catalyse the polymerisation of deoxynucleotides.
- These enzymes are highly efficient enzymes as they have to catalyse polymerisation of a large number of nucleotides in a very short time. *E. coli* that has only 4.6×10^6 bp (compare it with human whose diploid content is 6.6×10^9 bp), completes the process of replication within 18 minutes; that means the average rate of polymerisation has to be approximately 2000 bp per second. Not only do these polymerases have to be fast, but they also have to catalyse the reaction with high degree of accuracy. Any mistake during replication would result into mutations.

BEGINNER'S BOX-1

NUCLEIC ACIDS TO MECHANISM OF DNA REPLICATION

1. Nu body 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)

BMI395

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}

BMI396

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

BMI397

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

BMI398

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

BMI399

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) $^{15}NH_4NO_3$

BMI400

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

BMI401

8. The enzyme which shows polymerising activity in 5'$\rightarrow$ 3' direction is :

(1) DNA polymerase-III (2) DNA polymerase-II

(3) DNA polymerase-I (4) All of these

BMI402

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)

BMI403

10. DNA replication in lagging strand of most of the eukaryotic organisms is :-

(1) Conservative and continuous

(2) Semi-conservative but discontinuous

(3) Conservative and semi-discontinuous

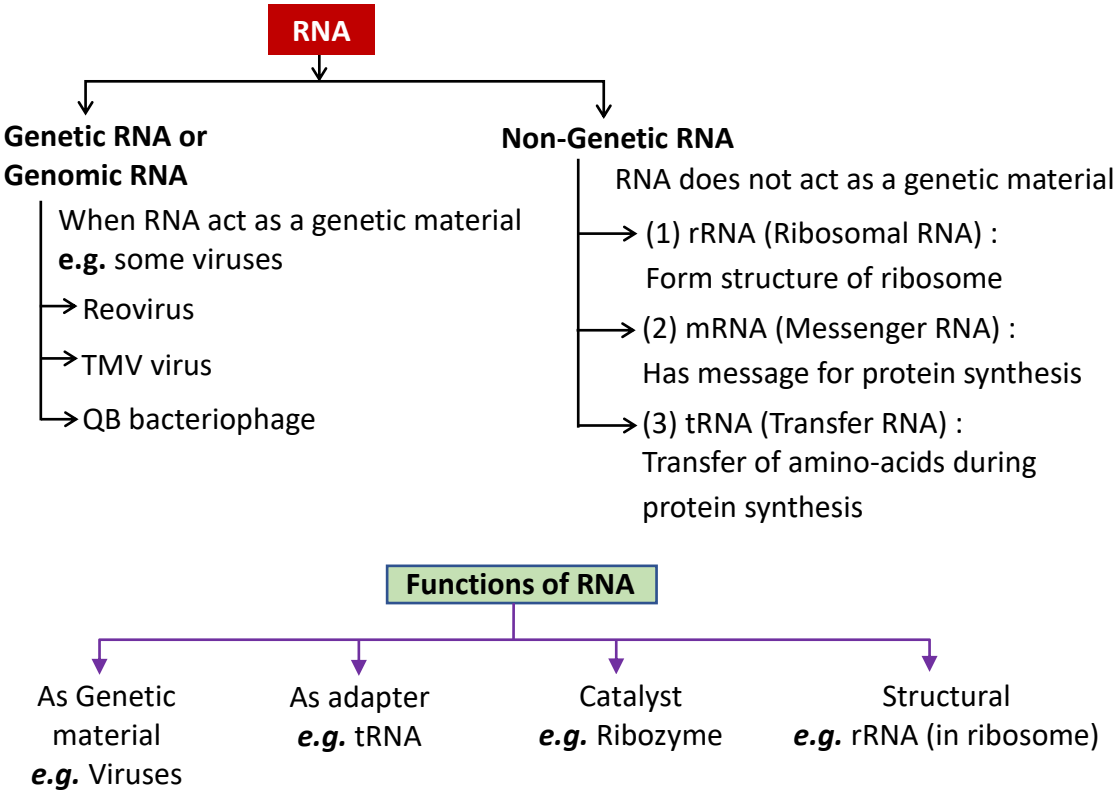
(4) Semi-conservative but continuous

BMI404

09. STRUCTURE AND TYPE OF RIBONUCLEIC ACID (RNA)

- Structure of RNA is fundamentally same as DNA, but there are some differences. The differences are as follows:-
  - (a) In place of Deoxyribose sugar in DNA, there is present Ribose sugar in RNA.
  - (b) In place of Nitrogenous base thymine in DNA, there is uracil in RNA.
  - (c) RNA is made up of only one polynucleotide chain i.e. RNA is single stranded.
  - (d) RNA functions as adapter, structural and in some cases as a catalyst (Ribozyme).

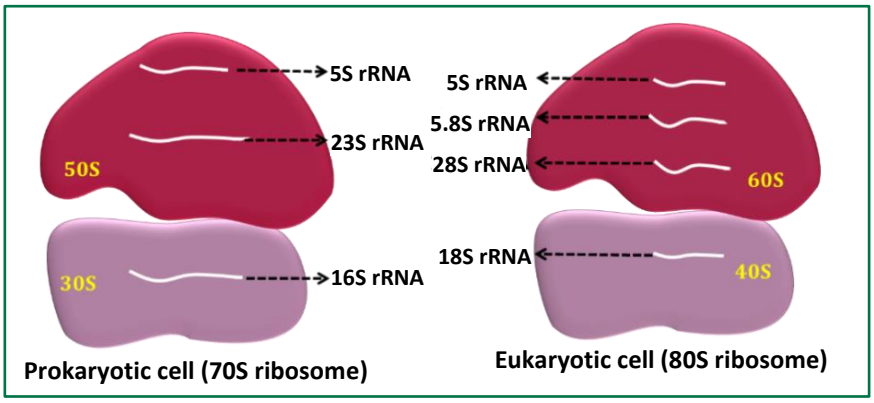
  Exception : RNA found in Reovirus is double stranded, i.e. it has two polynucleotide chains.



(1) RIBOSOMAL RNA (r-RNA) :

- This RNA is 80% of the cell's total RNA
- It is found in ribosomes.
- It is the most stable form of RNA.

- These are present in **80S type** of ribosomes in Eukaryotic cells. Their subunits are **60S** and **40S**. In 60S sub unit of ribosome three types of r-RNA are found – 5S, 5.8S, 28S. 40S sub unit of ribosome has only one type of rRNA i.e. 18S.
- In eukaryotic cell, maximum rRNA are produced in nucleolus except 5S rRNA. 5S rRNA is produced in nucleoplasm.
- Prokaryotic cells have 70S type of ribosomes and its subunits are 50S and 30S.
- 50S sub unit of ribosome contains 2 types of rRNA i.e. 5S and 23S.
- 30S sub unit of ribosome has 16S type of rRNA. So, 70S ribosome has total 3 types of r–RNA.



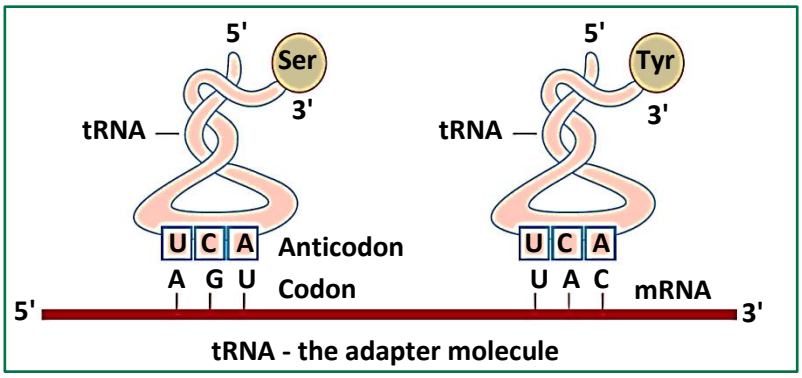
Function :

- At the time of protein synthesis, rRNA provides attachment site to tRNA and mRNA and attaches them on the ribosome.
- It attaches tRNA to the larger subunit of ribosome and mRNA to smaller subunit of ribosome.

(2) TRANSFER RNA (tRNA) :

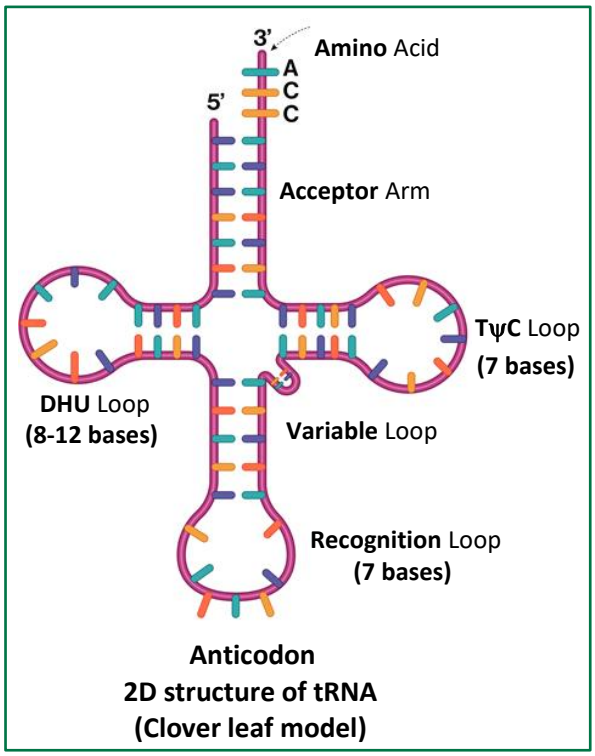
- It is 10-15% of total RNA.
- It is synthesized in the nucleus by DNA.
- It is also known as soluble RNA (sRNA)
- It is also known as Adapter RNA.
- It is the smallest RNA (4S).

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



Structure :

- The structure of tRNA is most complicated.
- Holley presented **Clover leaf model** of its structure. In two dimensional structure the tRNA appears clover leaf like but in three dimensional structure **(by Kim)** it appears inverted **L–shaped**.
- There are three nucleotides present in a particular sequence at 3' end of tRNA and that sequence is CCA. Whereas at 5' end G (guanine) is present.
- 3' end is known as Acceptor end.
- tRNA accepts amino acids at acceptor end. Amino acid binds to 3' end by its – COOH group.
- The molecule of tRNA 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**.



2D structure of tRNA (Clover leaf model)

Loops :

- There are some abnormal Nitrogenous bases in the loops, that is why hydrogen bonds are not formed.

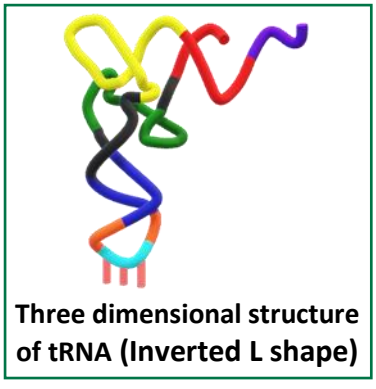
  e.g. Inosine (I), Pseudouracil (Ψ), Dihydrouridine (DHU)

TΨC Loop or Attachment loop :-

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

Recognition Loop (Anticodon loop) :-

- This is the most specific loop of tRNA and different types of tRNA are different due to this loop. There is a specific sequence of three nucleotides called anticodon, present at the end of this loop.
- tRNA recognizes its place on mRNA with the help of anticodon.


Three dimensional structure of tRNA (Inverted L shape)

- The anticodon of tRNA recognises its complementary sequence on mRNA. This complementary sequence is known as codon.

DHU Loop :

- It is also known as Aminoacyl synthetase recognition loop. Aminoacyl synthetase is a specific type of enzyme. The function of this enzyme is to activate a specific type of amino acid. After activation this enzyme attaches the amino acid to the 3' end of tRNA.
- There are 20 types of enzymes for 20 types of amino acids.
- The function of DHU loop is to recognize this specific Aminoacyl- tRNA synthetase enzyme.

(3) MESSENGER RNA (m-RNA) :

- The mRNA is 1 - 5% of the cell's total RNA.
- The name mRNA was given by **Jacob** and **Monod**.
- The mRNA is produced by genetic DNA in the nucleus.
- It is least stable RNA.

10. TRANSCRIPTION

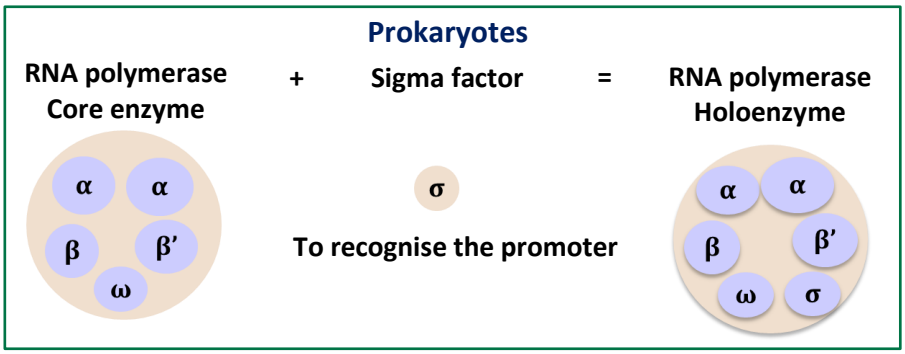
- The process of copying genetic information from one strand of the DNA into RNA is termed as transcription. Out of two strands of DNA only one strand participates in transcription and called **‘‘Antisense strand’’** or **"Template strand".**
- If both strands act as a template during transcription they would code for RNA molecule with different sequence and if they code for proteins the sequence of amino acid in these proteins would be different and another reason that if the two RNA molecules are produced they would be complementary to each other and form a dsRNA which prevents translation of RNA.
- A gene is defined as the functional unit of inheritance. It is difficult to literally define a gene in terms of DNA sequence, because the DNA sequence coding for tRNA or rRNA molecule is also define as gene (But information of protein is present on the DNA segment which codes mRNA).
- The segment of DNA which contains signal for the synthesis of one polypeptide is known as **‘‘Cistron’’**.
- RNA polymerase enzyme is involved in transcription. In eukaryotes there are three types of RNA polymerases.

  RNA polymerase–I for 28S rRNA, 18S rRNA, 5.8S rRNA synthesis.

  RNA polymerase–II for hnRNA synthesis (Precursor of m-RNA).

  RNA polymerase–III for tRNA, 5S rRNA, snRNA synthesis.

- Prokaryotes have only one type of RNA polymerase which synthesizes all types of RNAs.
- RNA polymerase (Core enzyme) of ***E. coli*** has five polypeptide chains.
- σ polypeptide chain is also known as **σ factor** (sigma factor).



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

(i) A promoter

(ii) A structural gene

(iii) A terminator

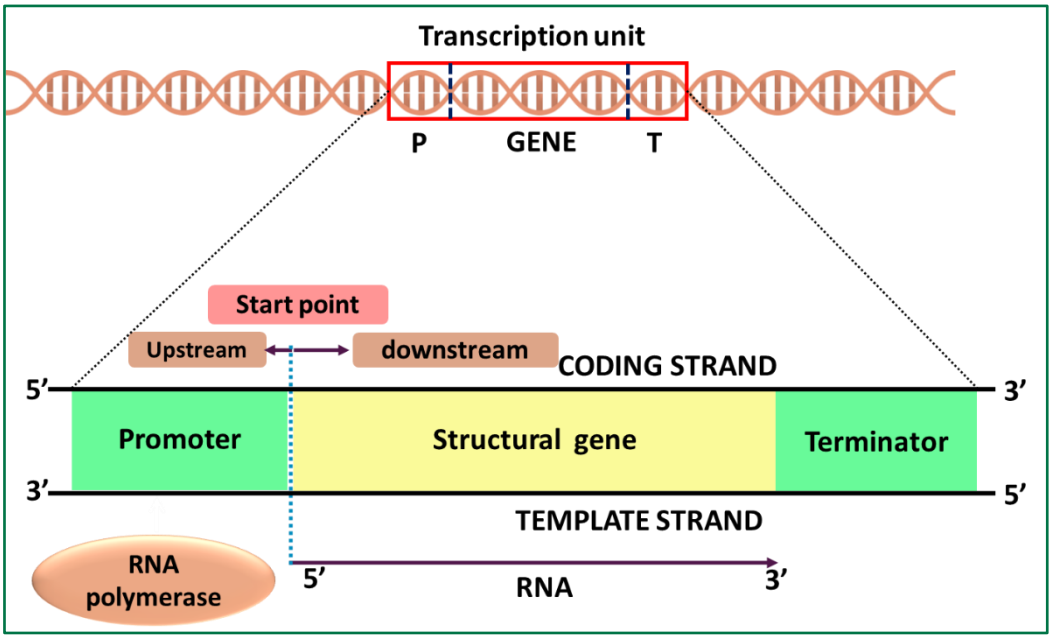


Figure 5.9 Schematic structure of a transcription unit

- The promoter and terminator flank the structural gene in a transcription unit. 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 a DNA sequence that provides binding site for RNA polymerase, and 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.

- The terminator is located towards 3′-end (downstream) of the coding strand and it usually defines the end of the process of transcription. There are additional regulatory sequences that may be present further upstream or downstream to the promoter. Some of the properties of these sequences shall be discussed while dealing with regulation of gene expression.

NCERT Question :

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

5'-ATGCATGCATGCATGCATGCATGCATGC-3'

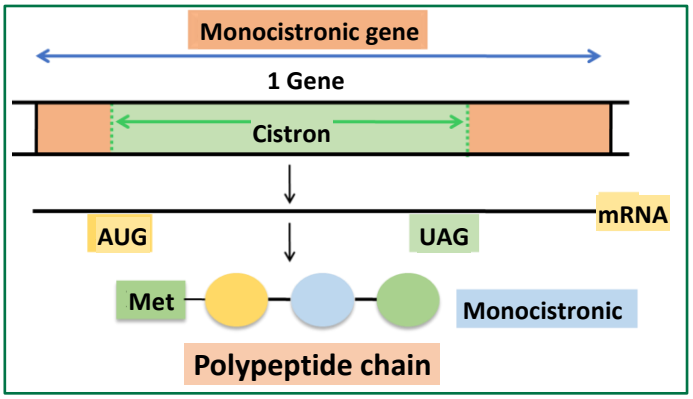
Write down the sequence of mRNA.

The structural gene in a transcription unit could be said as **monocistronic** (mostly in eukaryotes) or **polycistronic** (mostly in bacteria or prokaryotes).

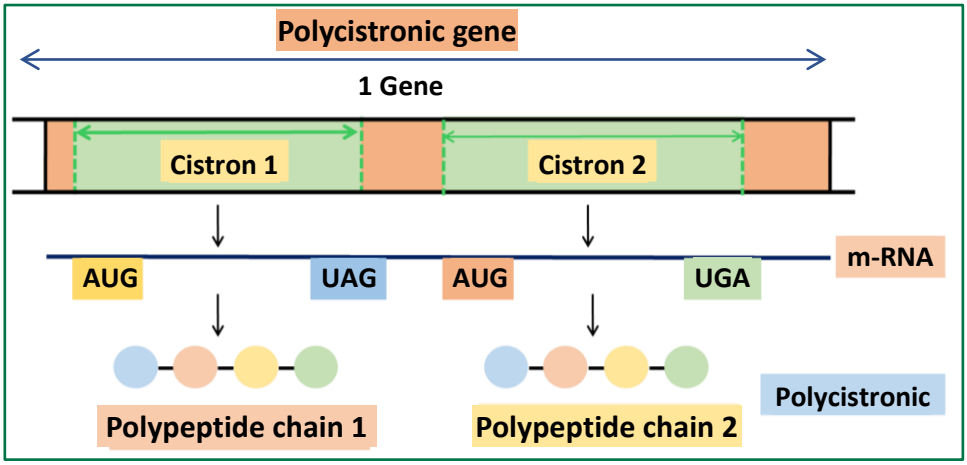
Structural genes and mRNA :-

(a) Monocistronic : The mRNA in which genetic signal is present for the formation of only one polypeptide chain.

eg. Eukaryotes.



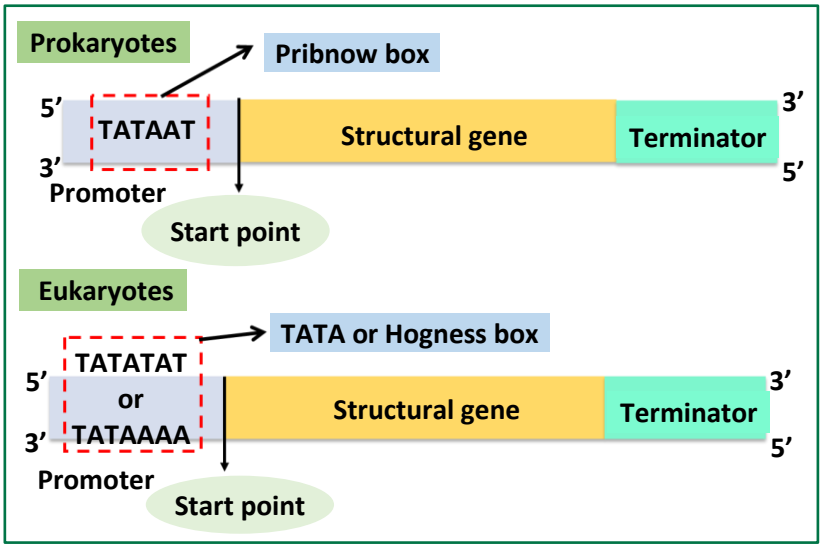
(b) Polycistronic : The mRNA, in which genetic signal is present for the formation of more than one polypeptide chains ***e.g.*** Prokaryotes.



Following Steps are Present in Transcription :-

(i) Initiation :

- DNA has a **"Promoter site"** where RNA polymerase binds and a **"Terminator site"** where transcription stops.
- Sigma factor (σ) recognises the promoter site of DNA.
- With the help of sigma factor RNA polymerase enzyme attaches to a specific site of DNA called **"Promoter site"**.
- In prokaryotes before the 10 N_2 base 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_2 base from **"Starting point"** a sequence of 7 base pairs (TATAAAA) or (TATATAT) is present on DNA which is called **"TATA box** or **Hogness box".**



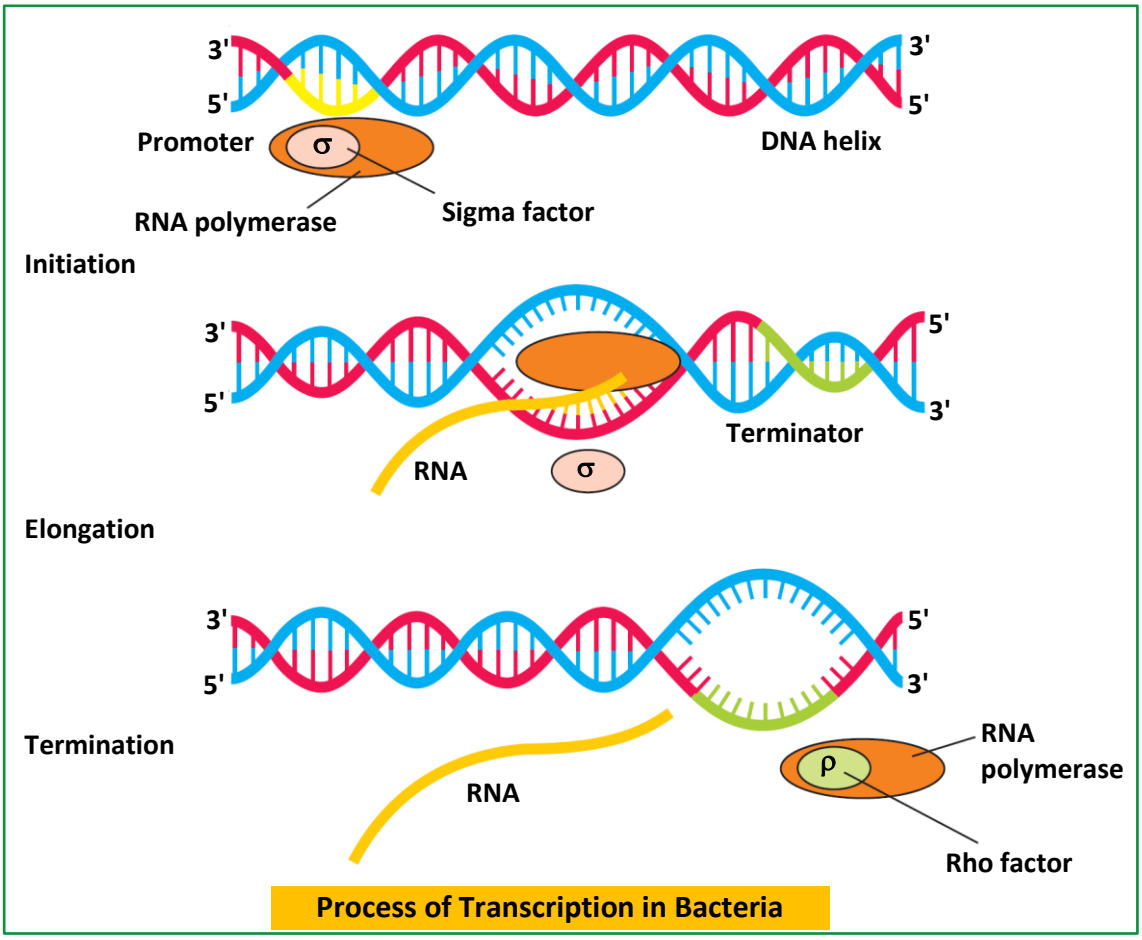
- At start point RNA polymerase enzyme breaks H–bonds between two DNA strands and separates them.
- One of these strands takes part in transcription. Transcription proceeds in 5' $\rightarrow$ 3' direction.
- Ribonucleoside triphosphates come to lie opposite complementary nitrogenous bases of anti-sense strand.
- These Ribonucleotides present in the form of triphosphate ATP, GTP, UTP and CTP. When they are used in transcription, pyrophosphatase hydrolyses two phosphates from each activated nucleotide. This releases energy. This energy is used in the process of transcription.

(ii) Elongation :

- RNA polymerase enzyme establishes phosphodiester bond between adjacent ribonucleotides.
- Sigma factor separates and RNA polymerase moves along the anti-sense strand till it reaches terminator site.

(iii) Termination :

- When RNA polymerase enzyme reaches at terminator site, it separates from DNA template.
- In most cases RNA polymerase enzyme can recognise the 'Terminator site' and stop the synthesis of RNA chain, but in prokaryotes, it recognises the terminator site with the help of **Rho factor (ρ factor)**.
- Rho (ρ) factor is a specific protein which helps RNA polymerase enzyme to recognise the terminator site.

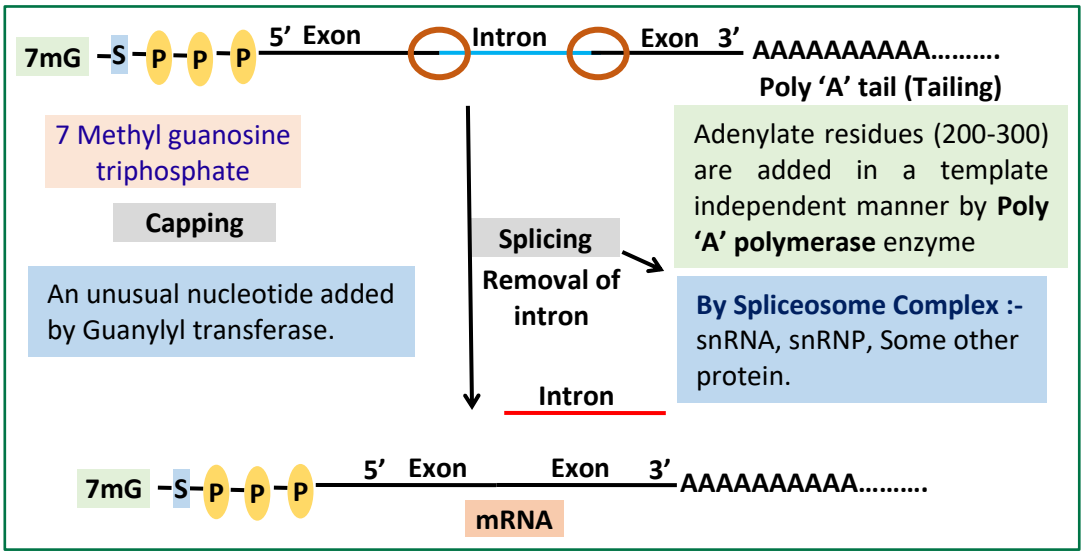


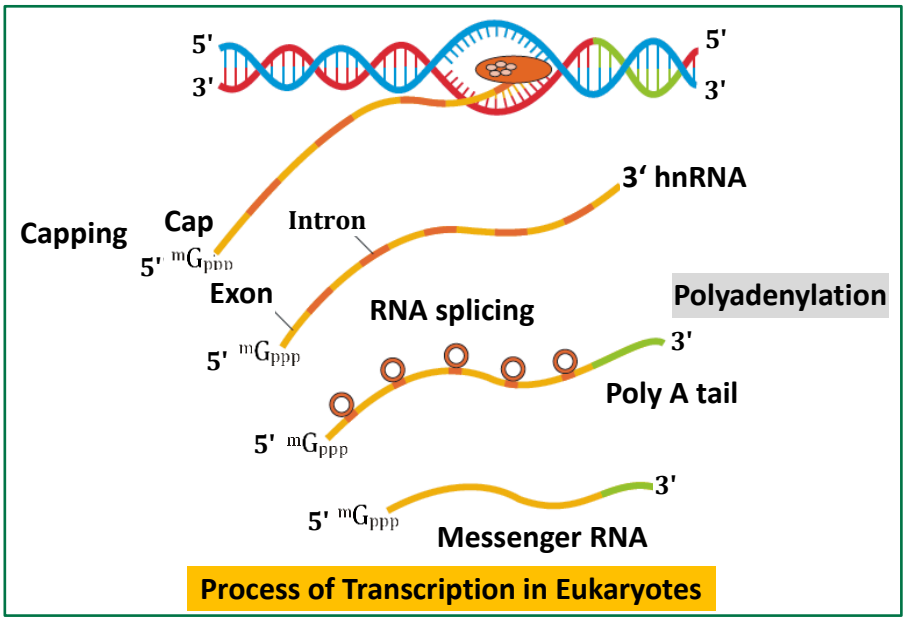
Process of Transcription in Bacteria

11. SPLIT GENE

- Discovered by **Sharp** and **Roberts**. They were awarded Nobel Prize in 1993. Gene which contains non-functional part along with functional part is known as split gene. Non-functional part is called **intron** and functional 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. The split-gene arrangement further complicates the definition of a gene in terms of a DNA segment.
- 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.
- By transcription split gene produces an RNA which contains coding and non-coding sequence and called hnRNA (Heterogenous nuclear RNA). This hnRNA is unstable. Now 7 methyl guanosine is added to its 5' end, and a cap like structure is formed. It is called **capping** and 200-300 nucleotides of adenylic acid are added in template independent manner to its 3' end, which is called **poly 'A' tail**. Now it becomes stable. By the process of **RNA splicing** hnRNA produces functional m-RNA that is exonic RNA.
- In RNA splicing non-coding parts is removed with the help of spliceosome complex and coding parts join together with the help of RNA ligase. Some specific proteins are also helpful in RNA - splicing called **'Small nuclear ribonucleoprotein' or 'snRNP' or 'Snurps'**. These snRNP proteins combine with some other proteins and snRNA to form **spliceosome complex**. This spliceosome complex uses energy of ATP to cut the RNA, releases the non-coding part and joins the coding-part to produce functional RNA. Non-coding part of hnRNA remained inside the nucleus and not translated into protein. Only coding part moves from nucleus to cytoplasm and gets translated into protein.
- Mostly Eukaryotic genes are example of split gene, **but gene which forms histone and interferon protein are non-split gene**.

  Mostly prokaryotic genes are example of non-split gene.
  - In eukaryotes after transcription, splicing process also occurred.
  - The split gene represent an ancient (primitive) feature of gene.
  - Presence of intron is probably **reminiscent of antiquity** (primitive character).
  - The splicing process represent the dominance of RNA world.





12. GENETIC CODE

- Term Given by **George Gamow**.
- Discovered by **Nirenberg, Matthaei and Khorana**.
- The relationship between the sequence of amino acids in a polypeptide chain and nucleotide sequence of DNA or m-RNA is called ***genetic code***.
- There are 20 types of amino acids which participate in protein synthesis. DNA contains information for the synthesis of many types of polypeptide chain. In the process of transcription, information is transferred from DNA to mRNA in the form of complementary N_2–base sequences.

- m-RNA contains code for each amino acid and it is called as codon. A codon is the nucleotide sequence on mRNA which codes for particular amino acid; whereas the genetic code is the sequence of nucleotides on mRNA molecule, which contains information for the synthesis of polypeptide chain.

(1) TRIPLET CODE :

- The main problem of genetic code was to determine the exact number of nucleotides in a codon which codes for one amino acid.
- There are four types of N_2–bases in mRNA (A, U, G, C) for 20 types of amino acids.
- If genetic code is **singlet** i.e. codon is the combination of only one Nitrogenous base, then only four codons are possible A, C, G and U. These are insufficient to code for 20 types of amino acids.

Nitrogenous bases = 4 Amino acids = 20	Total codon	Coded amino acids	Remaining amino acids
Single (one N-base code one amino acid)	$4^1 = 4$	4	16
Doublet (two N-base code one amino acid)	$4^2 = 16$	16	4
Triplet (three N-bases code one amino acid)	$4^3 = 64$	20	0
Tetraplate (four N-base code one amino acid)	$4^4 = 256$	20	0

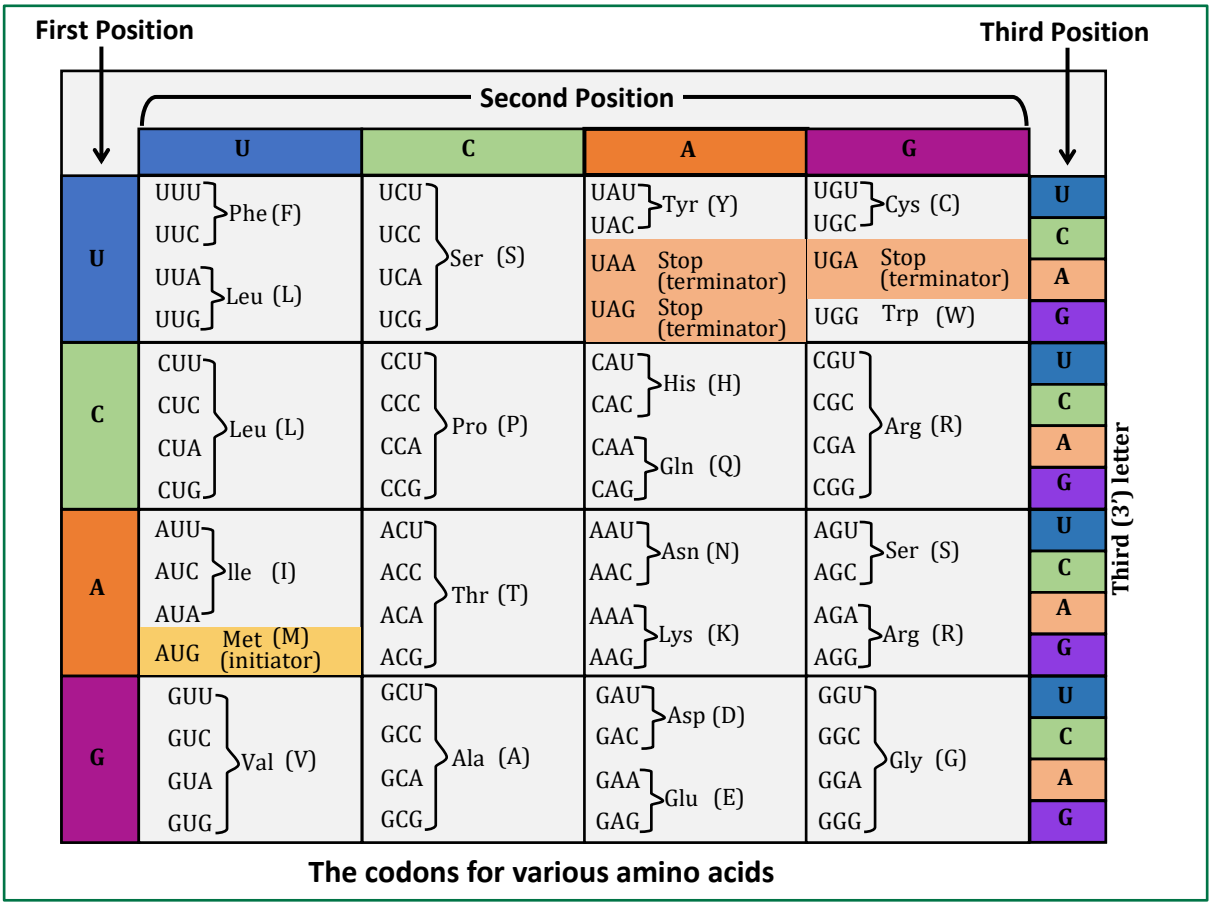
- Singlet code = $4^1 = 4 \times 1 = 4$ codons
- If genetic code is ***doublet*** (i.e. codon is the combination of two Nitrogenous bases) then 16 codons are formed.
- Doublet code = $4^2 = 4 \times 4 = 16$ codons.
- 16 codons are insufficient for 20 amino acid

UU	**UC**	**UA**	**UG**
CU	**CC**	**CA**	**CG**
AU	**AC**	**AA**	**AG**
GU	**GC**	**GA**	**GG**

Doublet Code: 4 × 4 = 16 codons

- Gamow (1954) pointed out the possibility of three letters code (Triplet code).
- Genetic code is triplet i.e. one codon consists of three Nitrogenous bases.
  Triplet code = $4^3 = 4 \times 4 \times 4 = 64$ codons
- In this case there occurs 64 codons in dictionary of genetic code.

- 64 codons are sufficient to code 20 types of amino acids.
- The chemical method developed by Har Gobind Khorana was instrumental in synthesising RNA molecules with defined combinations of bases (homopolymers and copolymers). Marshall Nirenberg's cell-free system for protein synthesis finally helped the code to be deciphered. **Severo Ochoa enzyme (polynucleotide phosphorylase)** was also helpful in polymerising RNA with defined sequences in a template independent manner (enzymatic synthesis of RNA).



The codons for various amino acids

(2) CHARACTERISTICS OF GENETIC CODE :

(i) Triplet in Nature :- A codon is composed of three adjacent Nitrogenous bases which specifies one amino acid in polypeptide chain.

For *e.g.* :-

- In mRNA if there are total 90 N_2 – bases.
- Then this mRNA determines 30 amino acids in polypeptide chain.
- In above example, number of Nitrogenous bases are 90, so codons $\Rightarrow$ 30.
- And 30 codons decide 30 amino acids in polypeptide chain.

(ii) Universality (Nearly Universal) :- The genetic code is applicable universally. The same genetic code is present in all kinds of living organism including viruses, bacteria, unicellular and multicellular organisms. **Exceptions:** codons in mitochondria and some protozoans.

(iii) Unambiguous and Specific :

- Genetic code is unambiguous **i.e.** one codon specifies only one amino acid and not any other.
- In this case one codon never, codes for two different amino acids. Exception GUG codon which codes both valine and methionine amino acids.

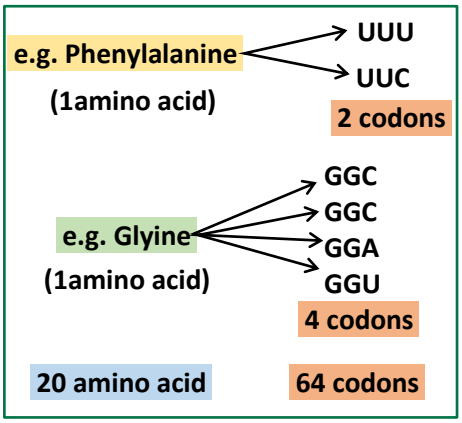
(iv) Non–Overlapping :- A Nitrogenous base is a constituent of only one codon.

(v) Comma Less :

- There is no punctuation (comma) between the adjacent codon **i.e.** each codon is immediately followed by the next codon.
- The codon is read in mRNA in a contiguous fashion.
- If a nucleotide is deleted or added, the whole genetic code read differently.
- A polypeptide chain having 50 amino acids shall be specialized by a linear sequence of 150 nucleotides. If a nucleotide is added in the middle of this sequence, the first 25 amino acids of polypeptide will be same but next 25 amino acids will be different.

(vi) Degeneracy 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. Single amino acid coded by more than one codon is called ''Degeneracy of genetic code''. This incident was discovered by Baurnfield and Nirenberg.



- Only two amino acids **Tryptophan** and **Methionine** are specified by single codon.
  All the other amino acids are specified or coded by 2 to 6 codons.
- Leucine, serine and arginine are coded or specified by 6–codons.
  Leucine = CUU, CUC, CUA, CUG, UUA & UUG
  Serine = UCU, UCC, UCA, UCG, AGU, AGC
  Arginine = CGU, CGC, CGA, CGG, AGA, AGG

- Degeneracy of genetic code is related to third position (3' – end of triplet codon) of codon. The third base is described as **"Wobbly base"**.

(vii) Chain Initiation and Chain Termination Codon :

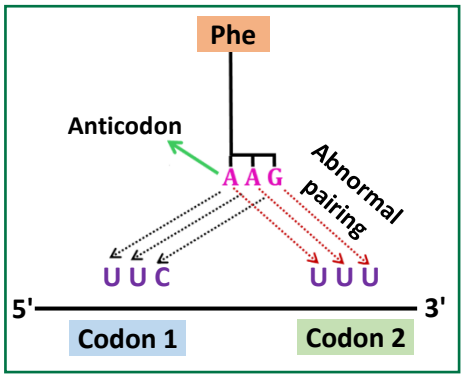
- Polypeptide chain synthesis is signalled by two initiation codons AUG or GUG.
- AUG codes for methionine amino acid in eukaryotes and in prokaryotes AUG codes for N–formyl methionine.
- Sometimes GUG also functions as start codon its codes for valine amino acid normally but when it is present at starting position its codes for methionine amino acid.
- **Out of 64 codons 3–codons are stopping or nonsense or termination codon.**
- **Nonsense codons do not specify 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 acid.

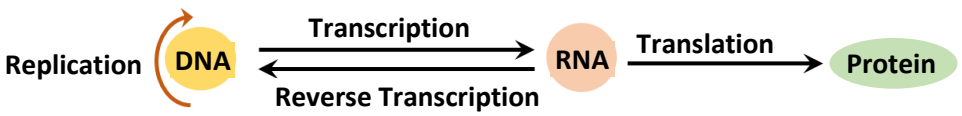
(3) WOBBLE HYPOTHESIS :

- 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.**
- For ***e.g.*** anticodon AAG can recognise two codons i.e. UUU and UUC, both code for phenyl alanine (Phe).



13. CENTRAL DOGMA

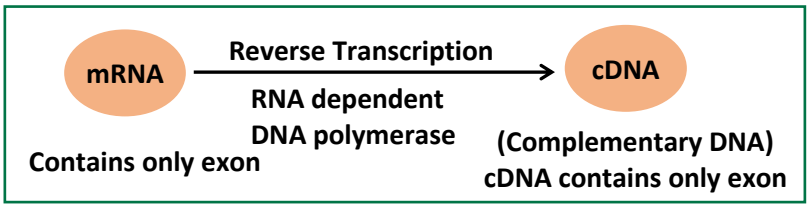
- Central dogma was given by **Crick**.
- The formation (production) of mRNA from DNA and then synthesis of protein from it, is known as Central Dogma.



- 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**.
- ss–RNA of Rous–Sarcoma virus (Retro virus) produces dsDNA in host's cell with the help of enzyme reverse transcriptase (DNA–polymerase). This DNA is called cDNA (Complementary DNA).

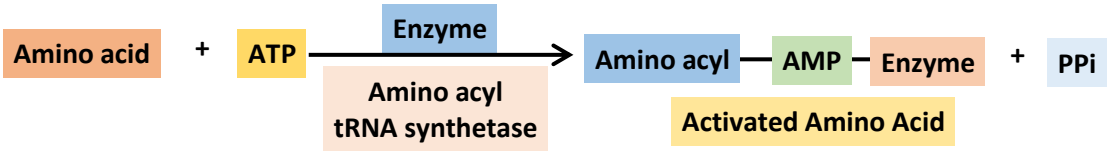


14. TRANSLATION (PROTEIN SYNTHESIS)

- 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. The amino acids are joined by a bond which is known as peptide bond. Formation of a peptide bond require energy. Therefore, in the first phase itself amino acids are activated in presence of ATP and linked to their cognate tRNA a process commonly known as charging of tRNA or aminoacylation of tRNA.

(1) ACTIVATION OF AMINO ACID :

- 20 types of amino acids participate in protein synthesis.
- Amino acid reacts with ATP to form **''Amino acyl AMP enzyme complex''**, which is also known as **'Activated Amino acid'**.



- This reaction is catalyzed by a specific '**Amino acyl tRNA synthetase**' enzyme.
- There is a separate 'Amino acyl tRNA synthetase' enzyme for each kind of amino acid.

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

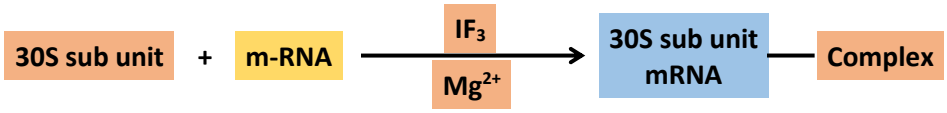
- Specific activated amino acid is recognised by its specific tRNA.
- Now amino acid attaches to the 'Amino acid attachment site' of its specific tRNA and AMP and enzyme are separated from it.

- Amino acyl tRNA complex is also called 'Charged tRNA'.
- Now Amino acyl tRNA moves to the ribosome for protein synthesis.

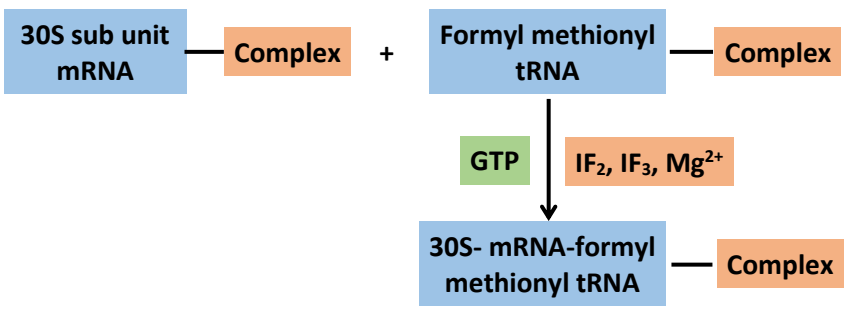
(3) TRANSLATION : 3 steps :–

(a) Initiation of polypeptide chain:

- In this step 30S and 50S sub units of ribosome, GTP, Mg^{2+}, charged tRNA, mRNA and some initiation factors are required.
- In prokaryotes there are three initiation factors present – **IF_1, IF_2, IF_3.**
- Initiation factors are specific proteins.
- GTP and initiation factors promote the initiation process.
- In prokaryotes with the help of **'SD sequence' (Shine–Dalgarno sequence)** mRNA recognises the smaller subunit of ribosome. A sequence of 8 N_2 base is present before the 4-12 N_2 base of initiation codon on mRNA, called **'SD sequence'**. In smaller subunit of ribosome, a complementary sequence of **'SD sequence'** is present on 16S rRNA, which is called **'Anti Shine-Dalgarno sequence'** (ASD sequence)
- With the help of 'SD' and 'ASD sequence' mRNA recognises the smaller sub unit of ribosome.
- While in eukaryotes, smaller subunit of ribosome is recognised by **'7mG cap'**.
- In eukaryotes, 18S rRNA of smaller sub unit has a complementary sequence of **'7mG cap'**.



- This '30S mRNA – complex' reacts with 'Formyl methionyl tRNA – complex' and '30S-mRNA – formyl methionyl tRNA – complex' is formed. This tRNA attaches with codon part of mRNA. A GTP molecule is required.

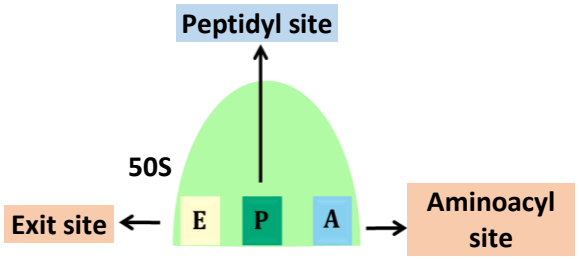


- Now larger sub unit of ribosome (50S subunit) joins this complex.
- In larger subunit of ribosome there are three sites for tRNA –

  'P' site = Peptidyl site.

  'A' site = Amino acyl site.

  'E' site = Exit site



- Starting codon of mRNA is near to 'P' site of ribosome, so tRNA with formyl methionine amino acid first attaches to **'P' site** of ribosome and next codon of mRNA is near to 'A' site of ribosome. So next new tRNA with new amino acid always attach at **'A' site** of ribosome but in initiation step **'A' site** is empty.

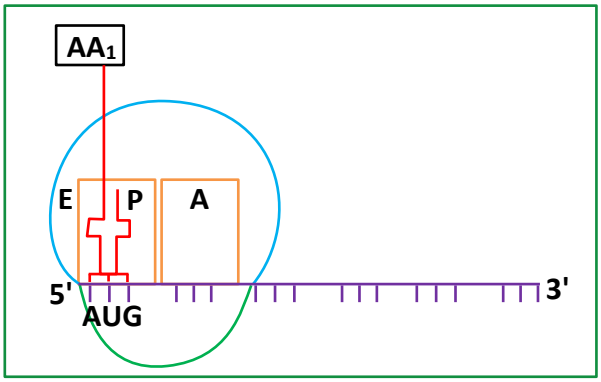


Fig. (1)

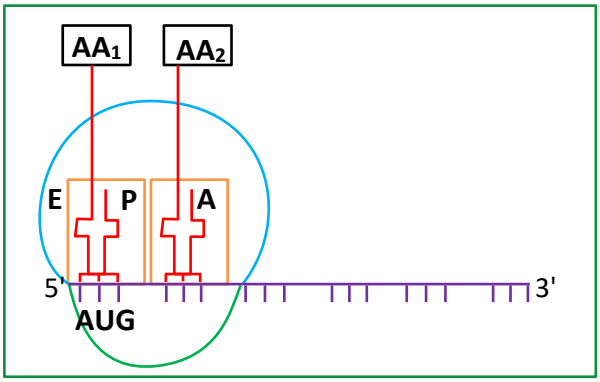


Fig. (2)

(b) Chain – Elongation :

- New tRNA with new amino acid is attached at **'A' site** of ribosome.
- The link between amino acid of **'P' site** of tRNA is broken and tRNA of **'P' site** is discharged so –COOH of **'P' site** amino acid becomes free.
- Now peptide bond takes place between **–COOH group** of **P site** amino acid and **$-NH_2$** group of **A site** amino acid.
- 23S rRNA induces the formation of peptide bond. This rRNA acts as an enzyme so it is called **''Ribozyme''**.
- After formation of peptide bond tRNA of **P site** is released from ribosome via **E site** and dipeptide attaches with **A site**.
- Now, tRNA of **A site** is transferred to **P site** and hence **A site** becomes empty.
- Now, ribosome slides over mRNA strand in **5' → 3'** direction. Due to sliding of ribosome on mRNA, new codon of mRNA is continuously available at **A site** of ribosome and according to new codon of mRNA new amino acid attaches in polypeptide chain.

- **Translocase enzyme** is helpful in movement of ribosome (translocation). **GTP** provides energy for sliding of ribosome.
- In elongation process some protein factors are also helpful, which are known as **'Elongation factors**.
- In prokaryotes three **'Elongation factors'** are present – **EF–Tu, EF–Ts, EF–G**.

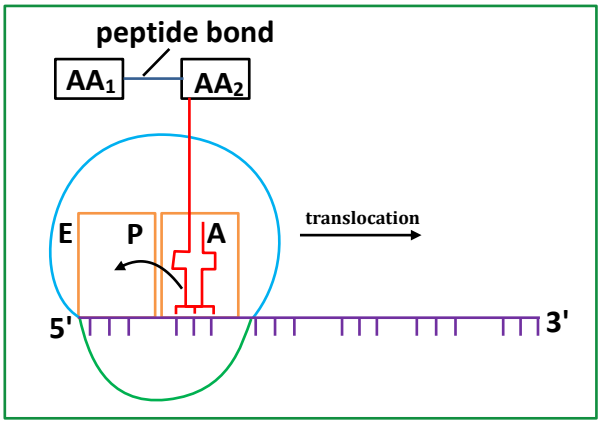


Fig. (3)

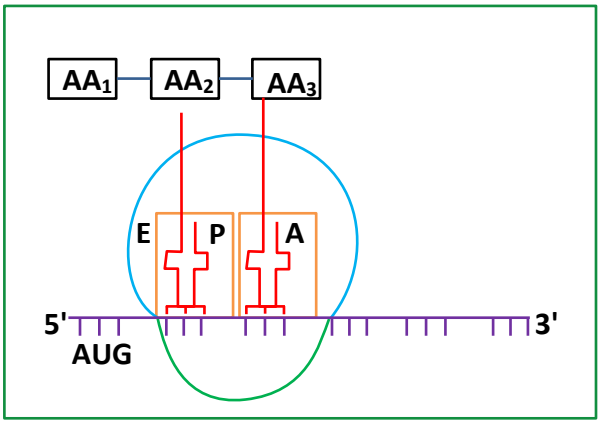


Fig. (4)

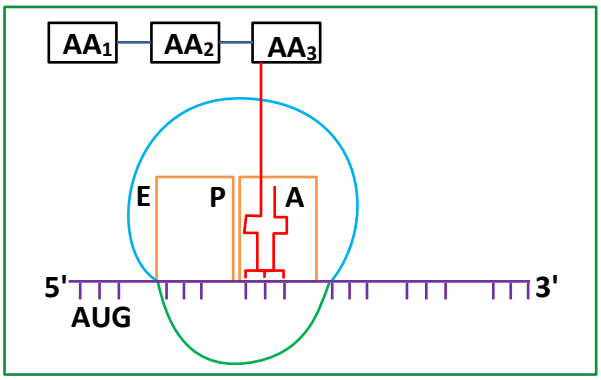


Fig. (5)

(c) Chain – Termination :

- Due to sliding of ribosome over mRNA when any **Nonsense codon** (UAA, UAG, UGA) available at **A site** of ribosome, then polypeptide chain terminates.
- The linkage between the last tRNA and the polypeptide chain is broken by three **release factor** called **RF_1, RF_2, RF_3** with the help of **GTP**.

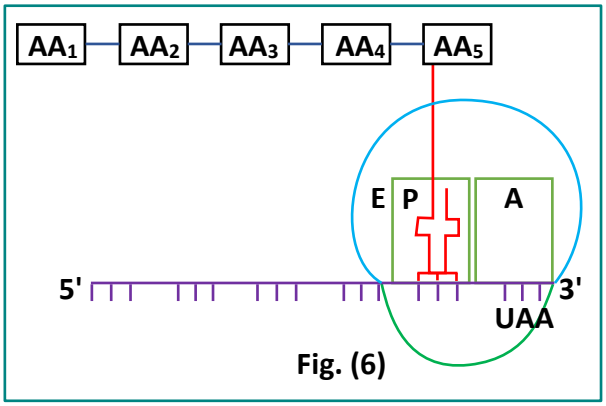


Fig. (6)

- At the end, a release factor binds to the stop codon, terminating translation and releasing the complete polypeptide from the ribosome.
- An mRNA also has some additional sequences that are not translated and are referred as untranslated regions (UTR). The UTRs are present at both 5'end (before start codon) and at 3'end (after stop codon).
- The UTR (untranslated regions) present on mRNA are required for efficient translation process (for recognition of smaller subunit of ribosome by mRNA.)

BEGINNER'S BOX-2 RIBONUCLEIC 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

BMI405

2. In eukaryotes, RNA Pol. III 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

BMI406

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

BMI407

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

BMI408

5. During protein synthesis, amino acid gets attached to tRNA with the help of :-

(1) mRNA (2) Aminoacyl synthetase

(3) Ribosome (4) rRNA

BMI409

6. The first amino acid in any polypeptide chain of prokaryote is always :-

(1) Formylated methionine (2) Formylated arginine

(3) Lysine (4) Methionine

BMI410

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

BMI411

8. To code 50 amino acids in a polypeptide chain, what will be the minimum number of nucleotides in its cistron?

(1) 50 (2) 153 (3) 306 (4) 309

BMI412

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

BMI413

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

BMI414

15. REGULATION OF GENE

- The mechanism which stimulates the expression of certain genes and inhibits that of others is called **regulation of gene expression**.

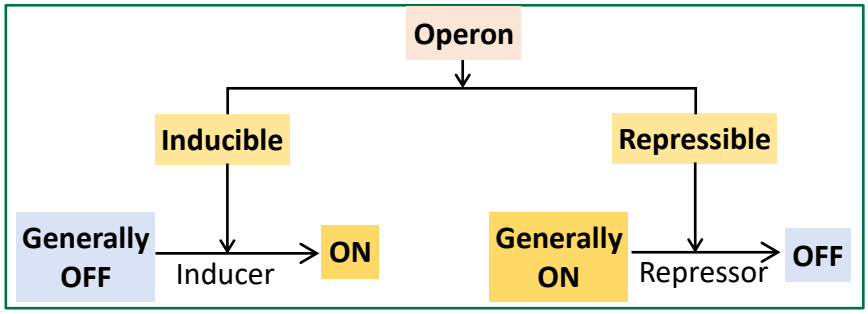
(1) CONSTITUTIVE GENES (House-Keeping Genes) :

- These genes are expressed constantly, because their products are constantly needed for cellular activity. **e.g.** genes for glycolysis, gene of ATPase enzyme.

(2) NON-CONSTITUTIVE GENES (Smart Gene or Luxury Gene) :

- These genes remain silent and are expressed only when the gene product is needed. They are switched 'on' or 'off' according to the requirement of cellular activities. Non-constitutive genes are of two types; **inducible** and **repressible**. The inducible genes are switched-on in presence of a chemical substance called inducer, required for the functioning of gene activity.

- The repressible genes continue to express themselves till a chemical, often an end product of the metabolism inhibits or represses their activity. Such type of inhibition is called **feedback inhibition or feedback repression**.
- A set of genes is 'switched on' when enzymes are required to metabolise a new substrate. The enzymes produced by these genes metabolise the substrate.



- The molecules of metabolite that come to switch on of the genes are termed as inducers and the phenomenon is called **induction**.
- Regulation of gene expression refers to a very broad term that may occur at various levels. Considering that gene expression results in the formation of a polypeptide, it can be regulated at several levels.
- In eukaryotes, the regulation could be exerted at :

  (i) transcriptional level (formation of primary transcript).

  (ii) processing level (regulation of splicing).

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

  (iv) translational level.
- The genes in a cell are expressed to perform a particular function or a set of functions. For example, if an enzyme called beta-galactosidase is synthesised by *E. coli*, it is used to catalyse the hydrolysis of a disaccharide, lactose into galactose and glucose; the bacteria use them as a source of energy.
- Hence, if the bacteria do not have lactose around them to be utilised for energy source, they would no longer require the synthesis of the enzyme beta-galactosidase. Therefore, in simple terms, it is the metabolic, physiological or environmental conditions that regulate the expression of genes.

- The development and differentiation of embryo into adult organisms are also a result of the coordinated regulation of expression of several sets of genes.
- In prokaryotes, control of the rate of transcriptional initiation is the predominant site for control of gene expression.
- In a transcription unit, the activity of RNA polymerase at a given promoter is in turn regulated by interaction with accessory proteins, which affect its ability to recognise start sites. These regulatory proteins can act both positively (activators) and negatively (repressors).
- The accessibility of promoter regions of prokaryotic DNA is in many cases regulated by the interaction of proteins with sequences termed operators.
- 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. 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.

Operon Concept :

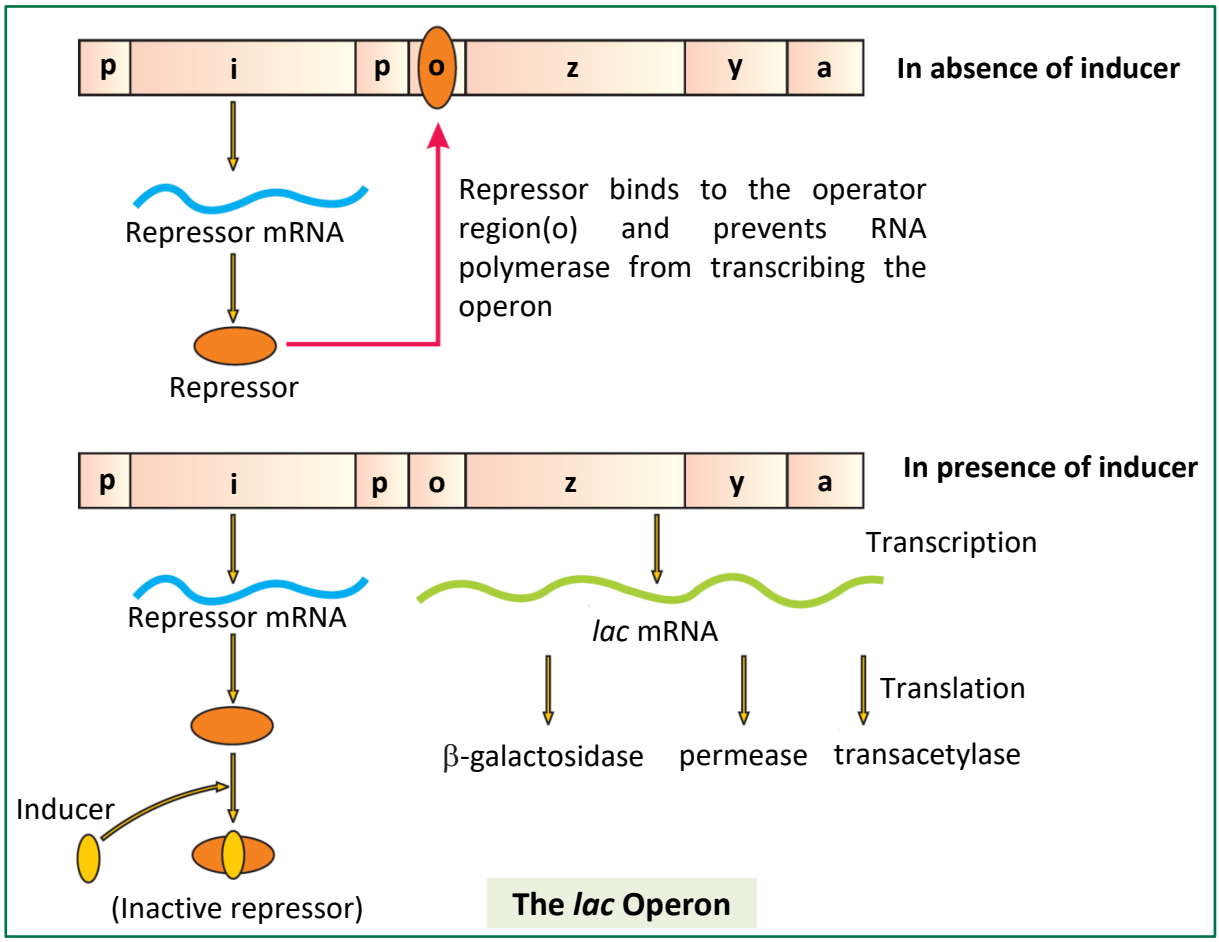
- In 1961, two French microbiologist **Francois Jacob** and **Jacques Monod** at the Pasteur Institute in Paris, proposed a mechanism called operon model for the regulation of gene action in ***E. coli.***
- An operon is a part of genetic material or DNA, which acts as a single regulated unit having one or more structural genes-an operator gene, a promoter gene, a regulator gene.
- Operons are of two types : **(i) Inducible (ii) Repressible**.

Inducible System (*Lac* Operon of *E. coli*) :

- An inducible operon system normally remains in switched off condition and begins to work only when the substance to be metabolised by it is present in the cell. Inducible operon system generally occurs in catabolic pathways. ***e.g.*** *lac* operon of ***E. coli.***
- Regulation of Lac operon by repressor is referred to as **negative regulation**.

  Active repressor + Inducer = Inactive repressor

  An inducible operon system consists of four types of genes :



(i) Structural genes :

- These genes synthesise mRNAs, which in turn synthesise polypeptide or enzyme over the ribosomes. An operon may have one or more structural genes. Each structural gene of an operon is called **cistron**. The *lac* operon (lactose operon) of ***Escherichia coli*** contains three structural genes (*z, y* and *a*). These genes occur adjacent to each other and thus are linked. They transcribe a polycistronic mRNA molecule (a single stretch of mRNA covering all the three genes), that helps in the synthesis of three enzymes **β-galactosidase** (breaks lactose into glucose and galactose), **lactose permease** (helps in entry of β-galactosides (lactose) in cell from outside) and **transacetylase** (transfers an acetyl group from acetyl Co A to β-galactosides).
- The *lac* operon consists of one regulatory gene (the *i* gene here the term does not refer to inducer, rather it is derived from the word inhibitor) and three structural genes (*z, y,* and *a*).
- The i *gene* codes for the repressor of the *lac* operon. The *z gene* codes for β - galactosidase, which is primarily responsible for the hydrolysis of the disaccharide, lactose into its monomeric units, galactose and glucose. The *y gene* codes for permease, which increases permeability of the cell to β-galactosides (lactose).

- *lac a* gene encodes a 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.

(ii) Operator gene :

It lies adjacent to the structural genes and directly controls the synthesis of mRNA over the structural genes. It is switched off by the presence of a repressor. An inducer can inactivate the repressor and switch on the gene that directs the structural genes to transcribe.

(iii) Promoter gene :

This gene is the site for initial binding of RNA polymerase. When the operator gene is turned on, the enzyme RNA polymerase moves over it and reaches the structural genes to perform transcription.

(iv) Regulator gene (*i gene*) :

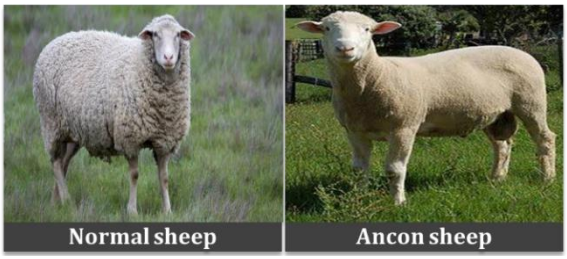
- It produces a repressor that binds to operator gene and prevents binding of RNA polymerase at the promoter. It is synthesised (all the time - constitutively) from the *i gene.*

Repressor - It is a protein, produced by the regulator gene. It binds to the operator gene so that the transcription of structural gene stops. Repressor has two binding sites (1) operator gene (2) effective molecule (inducer/corepressor)
Inducer - It is a chemical (substrate, hormone or some other metabolite) which after coming in contact with the repressor, forms an inducer-repressor complex. This complex cannot bind with the operator gene, which is thus switched on. The free operator gene allows the structural gene to transcribe mRNA to synthesise the enzymes.

- The inducer for lac operon of ***Escherichia coli*** is lactose (in fact allolactose an isomer of lactose).
- When the sugar lactose is added to the culture of ***E. coli***, a few molecules of lactose gets into the bacterial cells by the action of the enzyme permease, a small amount of this enzyme is present in the cell even when the operon is not working. These few lactose molecules are then converted into an active form which acts as an inducer and binds to the repressor protein.
- The inducer-repressor complex fails to join with the operator, which is turned on. The three genes are expressed as three enzymes to metabolise lactose. **Allolactose is real inducer of *lac* operon**.
- Regulation of *lac* operon by repressor is referred to as negative regulation. *Lac* operon is under control of positive regulation as well, but it is beyond the scope of discussion at this level.

16. MUTATION

- Sudden heritable change in genetic material of an organism is called as **Mutation**.
- Mutations are source of discontinuous variation.
- Only those mutation are heritable which occur in **germinal cell** of an organism. While somatic mutations are non-heritable. **Somatic mutations are also heritable in vegetatively propagated plants.**
- Mutation word was given by **Hugo De Vries**.
- De Vries studied mutations in the plant ***Oenothera lamarckiana*** (evening primrose).
- Mutation was first observed by **Seth Wright**. who observed some short-legged sheep (Ancon) variety in a population of long-legged sheep.



- **Beadle and Tatum** induced mutations in *Neurospora* by the help of U. V. rays or X-rays.

Normal (wild) *Neurospora* Prototroph $\xrightarrow{\text{U.V. rays}}$ **Mutant *Neurospora* Auxotroph**

- Normal-*Neurospora* can be grown in **minimal medium** because *Neurospora* can make all essential nutrients required for it. This is known as **Prototroph**.
- Mutant *Neurospora* doesn't have capability to grow in minimal medium because due to mutation it loses those genes which codes for the enzyme that helps to prepare some special nutrients for it. This form is known as **Auxotroph**. They gave "one gene-one enzyme" concept.
- M.S. Swaminathan induced mutations in wheat by the help of γ-rays to obtain good varieties for ***e.g.*** Sharbati Sonora, Pusa Lerma. Swaminathan established γ-garden in IARI-New Delhi (Pusa Institute).

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 :

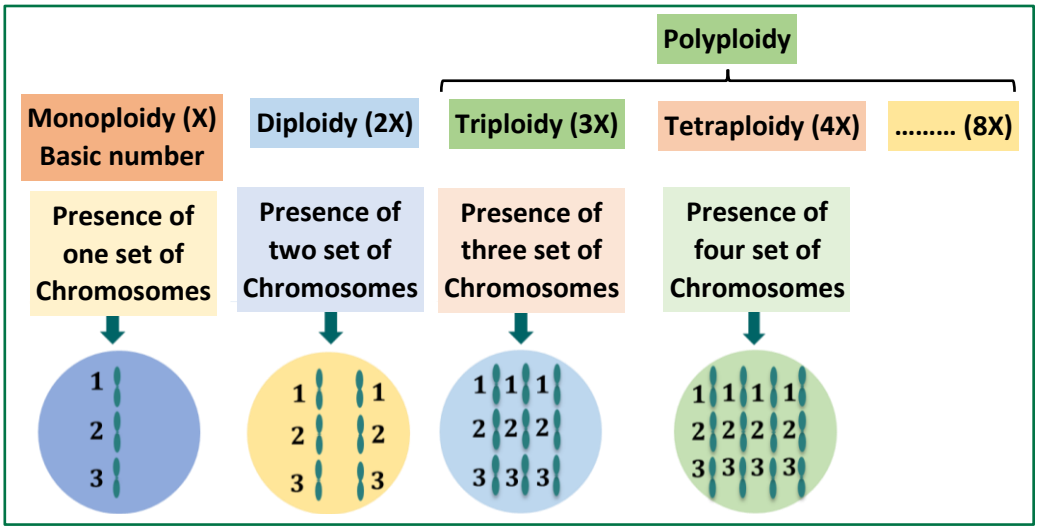
  (i) Euploidy → Change in number of chromosome sets.

  (ii) Aneuploidy → 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)



- Polyploid plants with even number of sets are always fertile, reproduce sexually and form seeds.
- Polyploid plants with odd number of sets are always sterile don't reproduce by sexual reproduction, they don't produce seeds but they may produce seedless fruits by parthenocarpy. ***e.g.*** Banana and seedless grapes.

Example :

Triticale : ***Triticum aestivum*** (6x = 42, Sterile) X Rye **(*Secale cereale*)** (2x = 14)

↓

Triticale

4x = 28

↓ Colchicine

Fertile Triticale

8x = 56

(ii) Aneuploidy :- Loss or addition of chromosomes in a set of chromosomes.

Types of Aneuploidy :

(a) Hypoaneuploidy (loss)

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

(b) Hyperaneuploidy (addition)

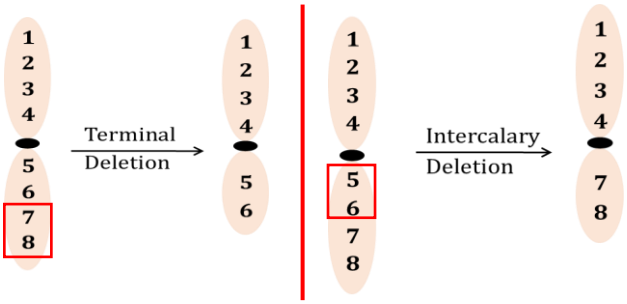
- **2n + 1 = Trisomy :-** addition of one chromosome in one set.
- **2n + 1 + 1 = Double Trisomy :-** addition of one chromosome in each set.
- **2n + 2 = Tetrasomy :-** addition of two homologous chromosome in set.
- Cause of aneuploidy is chromosomal non-disjunction means chromosomes fail to separate during meiosis.
- Chances of aneuploidy are more in higher aged female due to less activity of oocyte, so chances of syndrome increase in children who are born from higher aged female.

(B) Chromosomal Aberrations :- Change in structure of chromosome.

(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. ***e.g.*** The **cry -du-chat** syndrome is an example of terminal deletion in short arm of 5^{th} chromosome.



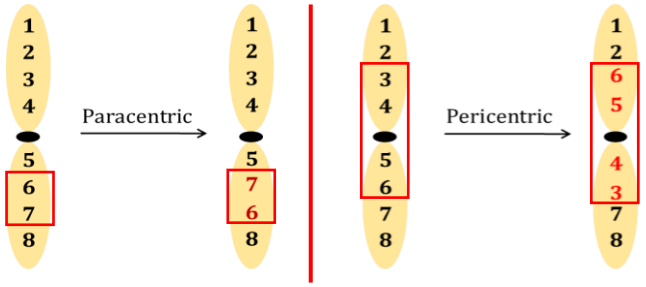
(b) Intercalary deletion - Loss of chromosomal part between the ends.

(ii) Inversion : Breakage of chromosomal segment but reunion on same chromosome in reverse orders. It leads to change in distance between genes on chromosome or sequence of genes on chromosome so crossing over is affected.

It is of 2 types :-

(a) Paracentric - If inversion occurs only in one arm and inverted segment does not include centromere.

(b) Pericentric - In this type of inversion inverted segment includes centromere.

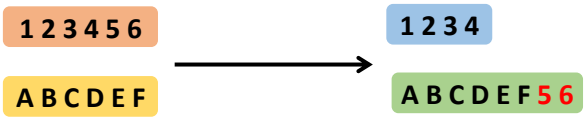


(iii) Duplication :- Occurrence of a chromosomal segment twice on a chromosome.

Example : In *Drosophila* **"Bar eye character"** is observed due to duplication in X-chromosome. Bar eye is a character where eyes are narrower as compared to normal eye shape.

Homologous chromosomes

(iv) Translocation :- In this, a part of the chromosome is broken and may be joined with non-homologous chromosome.



Non - Homologous chromosomes

Types of translocation –

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

(b) Reciprocal Translocation → Exchange of segments between two non-homologous chromosomes. This is also known as illegitimate crossing over (illegal crossing over).

Non - Homologous chromosomes

e.g. **Chronic Myeloid Leukaemia [CML]** is a type of blood cancer. This disease is a result of reciprocal translocation between 22 and 9 chromosome.

Note : If exchange of segments takes place in between homologous chromosomes then it is **called crossing over**.

(2) GENE MUTATION OR 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. ***e.g.*** Sickle cell anaemia.
- **Change may not occur sometimes because for one amino 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 :

- 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 : 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 all amino acids in structure of protein so a new protein is formed which is completely different from previous protein.
- So, frame shift mutations are more harmful as compared to substitution.
  eg. : **Thalassemia (lethal genetic disorder).**

Mutagens :

- Mutagens are those substances which cause mutations.
  Ex. Non-ionising radiation :- U. V. rays.

★ Golden Key Points ★

- Mostly mutations are harmful.
- Sometimes they are lethal which leads to death of organisms. But sometimes they are beneficial which are used to obtain good varieties of plants and animals. It is called as **Mutation Breeding**.
- Mostly mutations are recessive and they never eliminate from a population.

Forward and Backward Mutation:-

Wild gene $\underset{\text{Backward}}{\overset{\text{Forward}}{\rightleftharpoons}}$ Mutant gene

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 polypeptide synthesis by producing non-sense codon.

Same sense codon: 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-3

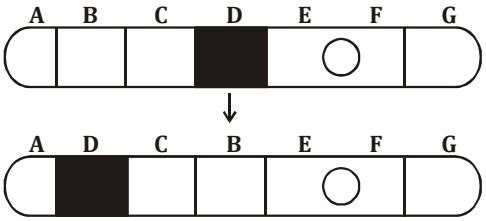
REGULATION OF GENE EXPRESSION TO MUTATION

1. (2n–1) condition of chromosomes is called :-

(1) tetrasomy (2) trisomy (3) monosomy (4) nullisomy

BMI415

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



(1) Deletion (2) Inversion

(3) Duplication (4) Reciprocal translocation

BMI416

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

BMI417

4. Addition or deletion of a nitrogenous base causes :-

(1) Frameshift mutation (2) Inversion

(3) Transformation (4) Translocation

BMI418

5. Trisomy of which chromosome is involved in Down's syndrome :-

(1) 8^{th} (2) 13^{th} (3) 21^{st} (4) 22^{nd}

BMI419

6. Match the Column-I with Column-II with respect to mutation :-

Column-I		Column-II	
(A)	Frame-shift	(i)	Change in structure of chromosome
(B)	Chromosomal aberrations	(ii)	Replacement of one Nitrogen base by another nitrogen base
(C)	Substitution	(iii)	Increase the set of chromosomes
(D)	Euploidy	(iv)	Deletions or insertions of nitrogen base of DNA

(1) A-iv, B-i, C-iii, D-ii (2) A-iv, B-i, C-ii, D-iii
(3) A-iv, B-ii, C-iii, D-i (4) A-iv, B-iii, C-ii, D-i

BMI420

7. Which of the following genetic mutations will never affect the protein produced?
(1) Transversions (2) Silent (3) Insertion (4) Frame shift

BMI421

8. Effect of point mutation that insert a base in structural gene :-
(1) Transition (2) Transversion
(3) Frame shift (4) Chromosomal aberration

BMI422

9. A monosomic (2n –1) abnormality in human is :-
(1) Klinefelter's syndrome (2) Turner's syndrome
(3) Down's syndrome (4) All the above

BMI423

10. *Lac*-operon codes for three types of enzymes. Which of the following enzyme is formed in *lac* operon?
(1) β-galactosidase (2) Permease (3) Amylase (4) Both (1) and (2)

BMI424

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

- It is a 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 Fingerprinting has been started by **Dr. V.K. Kashyap & Dr. Lal Ji Singh**.
- DNA of human is almost the same for all individuals but very small amount differs from person to person that forensic scientists analyze to identify people.
- These differences are called Polymorphism (many forms) and are the key of DNA typing.
- Allelic (again recall the definition of alleles from Chapter 5) 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.

- Polymorphisms are most useful to forensic scientist. It consists of variation in the length of DNA at specific loci. It is most important segment for DNA test made up of short repetitive nucleotide sequences, these are called VNTR (variable number of tandem repeat).
- VNTR also called minisatellites were discovered by **Alec Jeffreys**. These loci called restriction fragments consist of hypervariable repeat region of DNA having a basic repeat sequence of 11-60 bp and flanked on both sides by restriction site.
- The number and position of mini satellites or VNTR in restriction fragment is different for each DNA and length of restriction fragment depends on number of VNTR.
- Therefore, when the genome of two persons is cut using the same restriction enzyme, the length of fragments obtained is different for both the persons.
- These variations in length of restriction fragment is called RFLP or **Restriction fragment length polymorphism.**
- Restriction Fragment Length Polymorphism distributed throughout human genomes are useful for DNA Fingerprinting.
- DNA Fingerprint can be prepared from extremely minute amount of blood, semen, hair follicle or any other cell of the body.
- DNA content of 1 - Microgram is sufficient.

(1) TECHNIQUE OF DNA FINGERPRINTING INVOLVES THE FOLLOWING MAJOR STEPS :

(A) Extraction :

- DNA extracted from the cell (by cell lysis). If the content of DNA is limited, then DNA can be amplified by Polymerase chain reaction (PCR). This process is called amplification.

(B) Restriction Enzyme Digestion :

- Restriction enzyme cuts DNA at specific 4 or 6 base pair sequences called restriction site.
- Hae III (*Haemophilus aegyptius*) is most commonly used enzyme. It cuts the DNA, everywhere the bases are arranged in the sequence GGCC. These restriction fragments are transferred to Agarose Polymer gel.

(C) Gel Electrophoresis :

- Gel electrophoresis is a method that separates macromolecules-either nucleic acid or proteins-**on the basis of size and electric charge**.
- Gel electrophoresis refers to the technique in which molecules are forced across a span of gel, motivated by an electrical current. Activated electrodes at either end of the gel provides

the driving force. A molecule's properties determine, how rapidly an electric field can move the molecule through a gelatinous medium.

- Nowadays the most commonly used matrix is **agarose** which is a natural polymer extracted from sea weeds. The DNA fragments separate (resolve) according to their size through **sieving effect** provided by the agarose gel.
- Many important biological molecules such as amino acids, peptides, proteins, nucleotides, and nucleic acids possess ionisable groups and, therefore, at any given pH, exist as electrically charged species in solution, either as cation (+) or anion (–). Depending on the nature of the net charge, the charged particles will migrate either to the cathode or to the anode.
- By the gel electrophoresis these restriction fragments move towards the positive electrode (anode) because DNA has negative electric charge (PO_4^{3-}).
- Smaller Fragments move faster towards the positive pole due to less molecular weight. So, after the gel electrophoresis DNA fragments are arranged according to molecular weight.
- These separated fragments can be visualized by staining them with a dye that fluoresces in ultraviolet radiation.

(D) Southern Transfer/Southern Blotting :

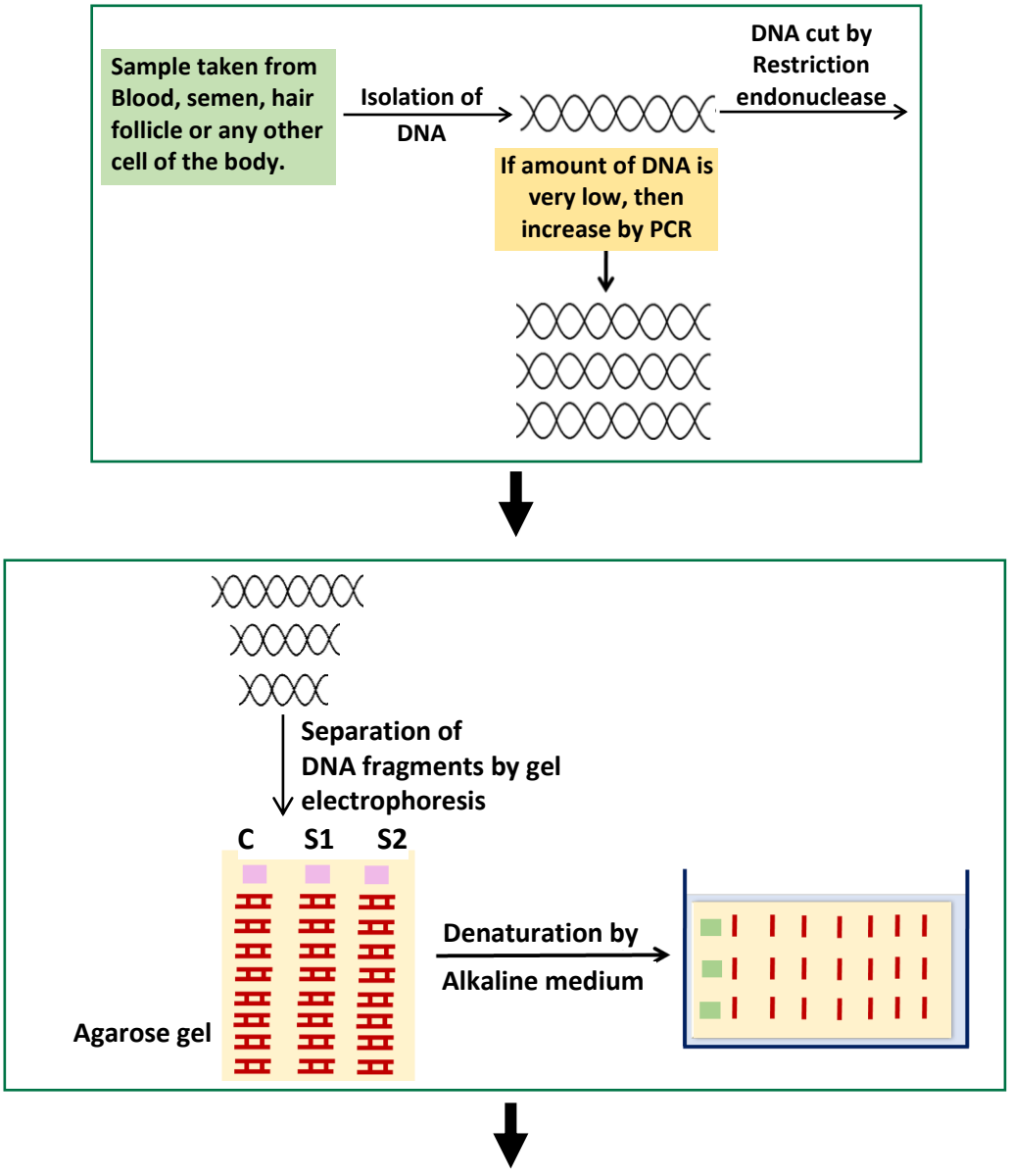
- The gel is **fragile**. It is necessary to remove the DNA from the gel and permanently attach it to a solid support. This is accomplished by the process of Southern blotting. The first step is to denature the DNA in the gel which means that the double-stranded restriction fragments are chemically separated into the single stranded form.
- The DNA then is transferred by the process of blotting to a sheet of nylon. The nylon acts like an ink blotter and "blots" up the separated DNA fragments, the restriction fragments, invisible at this stage are irreversibly attached to the nylon membrane.
- This process is called Southern blotting by the name of **Edward Southern (1970)**.

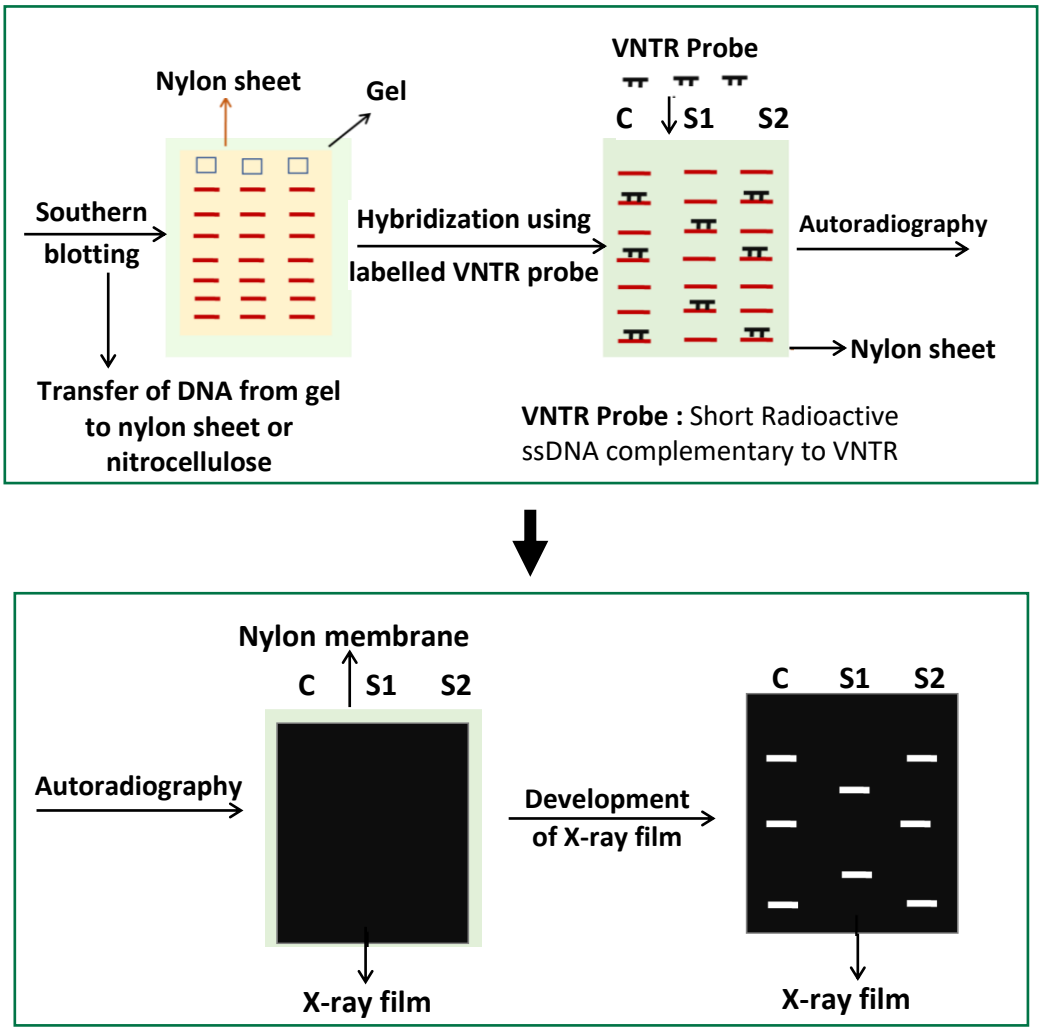
(E) Hybridization :

- To detect VNTR locus on restriction fragment, we use single stranded **Radioactive (P^{32}) DNA probes** which have the base pair sequences complementary to the DNA sequences at the VNTR locus. Commonly we use a combination of at least 4 to 6 separate DNA probes.
- Labelled VNTR probes are attached with the VNTR loci of restriction DNA fragments, this process is called **Hybridization**.

(F) Autoradiography :

- Nylon membrane containing radioactive probe is kept with x-ray film. Specific bands appear on X-ray film. These bands are the areas where the radioactive probe is bound to the VNTR.
- This represents the RFLP (Restricted Fragment Length Polymorphism) pattern. This pattern is different in different individuals. This is called DNA fingerprint.
- These allow analyser to identify a particular person DNA, the occurrence and frequency of a particular genetic pattern contained in this X-ray film. This x-ray film is called **DNA signature** of a person which is specific for each individual.
- The probability of two unrelated individual having same pattern of location and repeat number of minisatellite (VNTR) is one in ten billion (world's population 6.1 billion).
- In India, the centre for DNA fingerprinting and diagnostics (CDFD) is located at Hyderabad.





Steps of DNA fingerprinting

NCERT Question :

What is DNA fingerprinting? Mention its application.

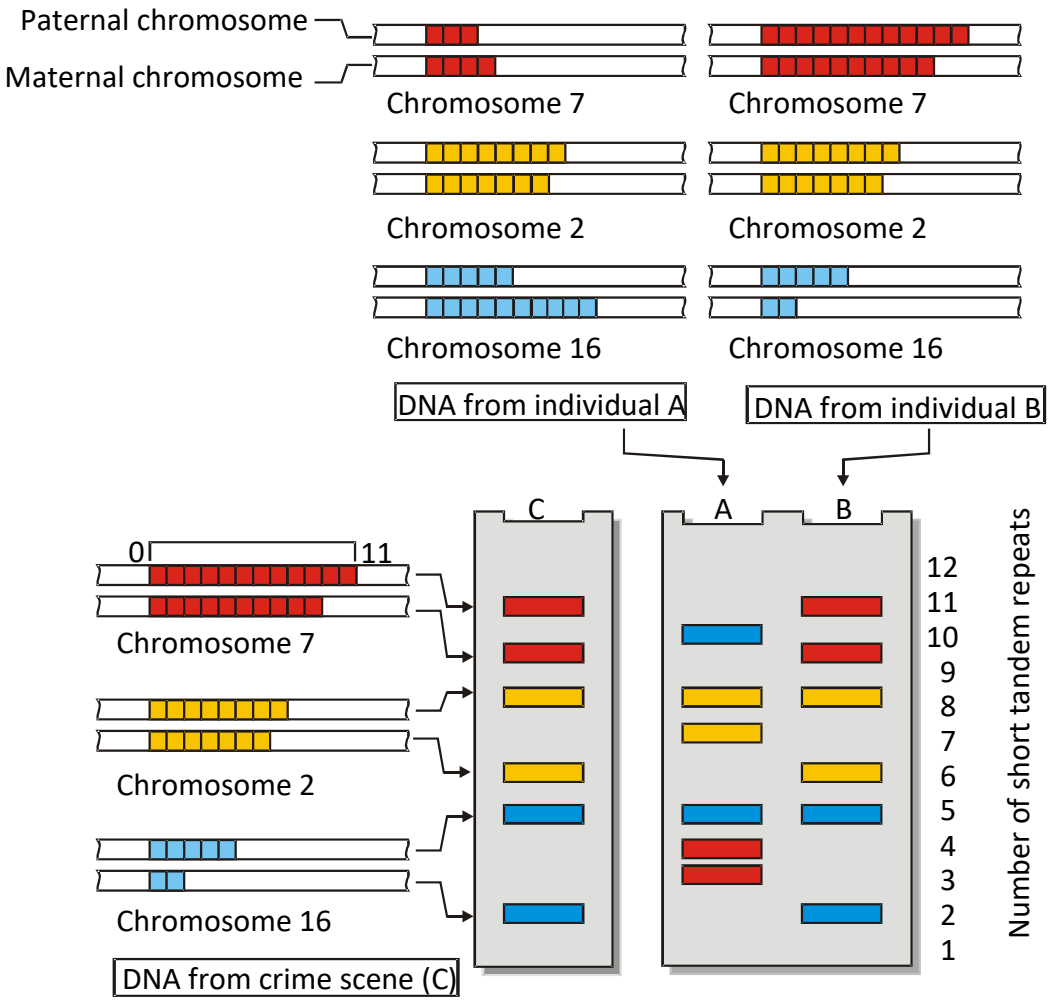
(2) APPLICATION OF DNA FINGERPRINTING :

(A) Paternity Tests :

- The major application of DNA fingerprinting is in determining family relationships. For identifying the true (biological) father, DNA samples of child, mother and possible fathers are taken and their DNA fingerprints are obtained. The prints of child DNA match to the prints of biological parents (approximately 50%). The RFLP pattern of child matches 50% with father and 50% with mother.
- In Above case father B is biological father.

(B) Identification of the criminal.

(C) Determination of population and genetic variability.



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.

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.

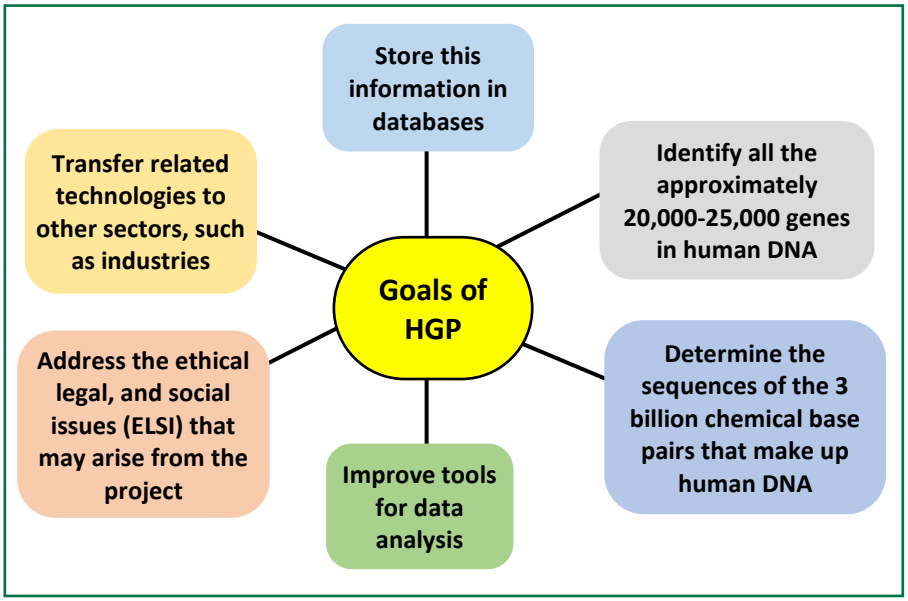
NCERT Question :

Why is the Human Genome project called a mega project?

- 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 approximately **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**.

(1) GOALS OF HGP :-

Some of the important goals of HGP are as follows :



- 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.
- 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 towards solving challenges in health care, agriculture, energy production, environmental remediation.
- Many non-human model organisms, such as bacteria, yeast, *Caenorhabditis elegans* (a free-living non-pathogenic nematode), *Drosophila* (the fruit fly), plants (rice and *Arabidopsis*), etc., have also been sequenced.

(2) METHODOLOGIES :

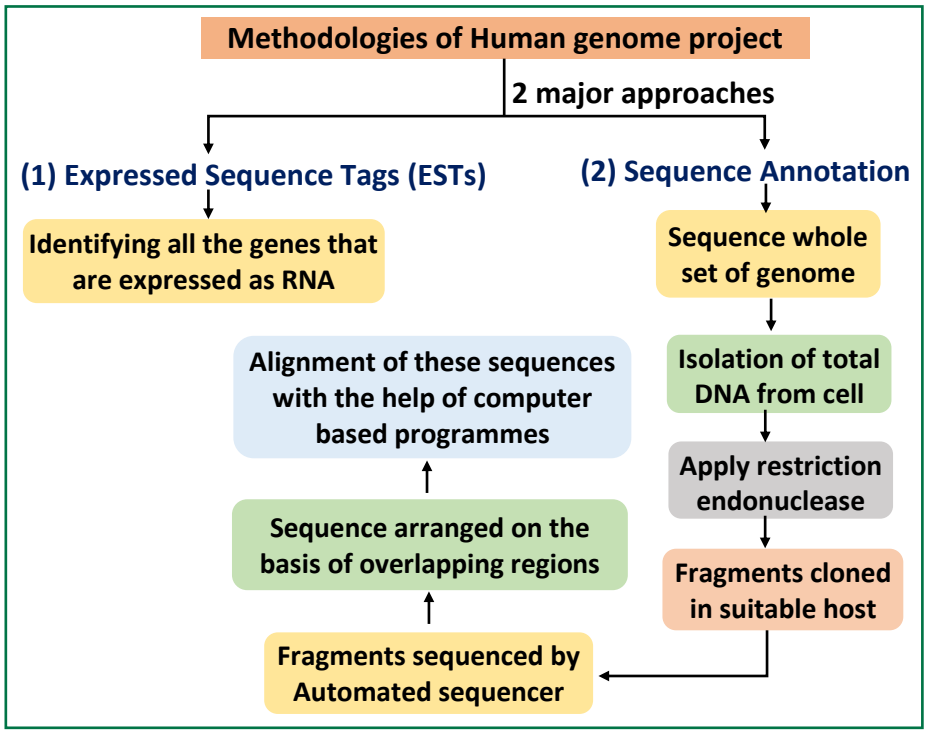
The methods involved two major approaches.

(a) Expressed Sequence Tags (ESTs) :

Identifying all the genes that are expressed as RNA.

(b) Sequence Annotation :

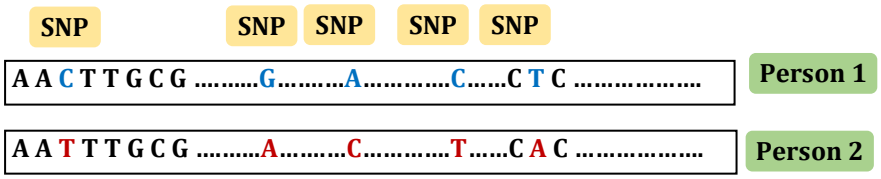
- The **blind approach** of simply sequencing the whole set of genomes that contained all the coding and non-coding sequence, and later assigning different regions in the sequence with functions.
- For sequencing, the total DNA from a cell is isolated and converted into random fragments of relatively 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 (multiplication) of each piece of DNA fragment so, that it subsequently could be sequenced with ease.
- The commonly used hosts were bacteria and yeast, and the vectors were called as **BAC (bacterial artificial chromosomes)**, and **YAC (yeast artificial chromosomes).**
- 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.

(3) SALIENT FEATURES OF HUMAN GENOME :

- Some of the salient observations drawn from human genome project are as follows:
  - (i) The human genome contains **3164.7 million** base pairs.
  - (ii) The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being **dystrophin at 2.4 million bases** (bp).
  - (iii) The total number of genes is estimated at 30,000-much lower than previous estimates of 80,000 to 1,40,000 genes. Almost all (99.9 per cent) nucleotide bases are exactly the same in all people.
  - (iv) The functions are unknown for over 50 per cent of discovered genes.
  - (v) Less than 2 per cent of the genome codes for proteins.
  - (vi) Repeated sequences make up very large portion of the human genome.
  - (vii) Repetitive sequences are stretches of DNA sequences that are repeated many times, sometimes hundred to thousand times. They are thought to have no direct coding functions, but they shed light on chromosome structure, dynamics and evolution.
  - (viii) Chromosome 1 has most genes (2968) and the Y has the fewest (231).
  - (ix) Scientists have identified about 1.4 million locations where single-base DNA differences **(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**.



- First prokaryote in which complete genome was sequenced is ***Haemophilus influenzae***.
- First eukaryote in which complete genome was sequenced is ***Saccharomyces cerevisiae (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)**.
- Genomes of rice and fruit fly have also been sequenced.

BEGINNER'S BOX-4

DNA FINGERPRINTING 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

BMI425

2. Which of the following is produced by *E. coli* in the lactose operon?

(1) β galactosidase (2) Transacetylase

(3) Permease (4) All of the above

BMI426

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

(1) 1^{st} (2) X (3) Y (4) 10^{th}

BMI427

4. In lac operon RNA polymerase binds with :-

(1) Promoter gene (2) Operator gene

(3) Structural gene (4) Regulator gene

BMI428

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

BMI429

6. In gel electrophoresis :-

(1) DNA fragments are arranged according to length on gel by applying electric field

(2) DNA fragments are arranged according to charge

(3) DNA fragments are treated by alcoholic solution

(4) DNA fragments are destroyed

BMI430

7. Satellite DNA :-

(A) Is classified in many categories such as micro-satellites, minisatellites, etc. on the basis of base composition, length of segments and number of repetitive units.

(B) Normally does not code for any protein

(C) Shows polymorphism

(D) Forms the basis of DNA finger printing

Choose the correct statements.

(1) A, C (2) A, B, C, D (3) B, C, D (4) A, B, C

BMI431

8. Allelic sequence variations, where more than one variant (allele) occurs at a locus in a human population with a frequency greater than 0.01 is referred to as :-

(1) incomplete dominance (2) EST

(3) snRNP (4) DNA polymorphism

BMI432

9. In human :-

(I) non-coding DNA is abundant.

(II) less than 2% of genome codes for protein.

(III) the function of more than 50% genes are unknown

(IV) total number of genes is 30000.

Correct options are :-

(1) I, II, III and IV (2) I and III (3) I, II and IV (4) I, II and III

BMI433

10. Which term is not associated with human genome project ?

(1) Expressed sequence tag

(2) Single nucleotide polymorphism

(3) Sequence Annotation

(4) Variable number of tandem repeats

BMI434

BEGINNER'S BOX ANSWERS KEY

BEGINNER'S BOX-1	Que.	1	2	3	4	5	6	7	8	9	10
	Ans.	4	2	1	2	4	2	2	4	4	2

BEGINNER'S BOX-2	Que.	1	2	3	4	5	6	7	8	9	10
	Ans.	1	3	2	3	2	1	2	3	3	2

BEGINNER'S BOX-3	Que.	1	2	3	4	5	6	7	8	9	10
	Ans.	3	2	1	1	3	2	2	3	2	4

BEGINNER'S BOX-4	Que.	1	2	3	4	5	6	7	8	9	10
	Ans.	1	4	1	1	3	1	2	4	1	4

Mutation (Sudden heritable change in genetic material)

Chromosomal mutation
[Change in number of chromosome & change in structure of chromosome]

Gene mutation
(Change in Nucleotide)

Heteroploidy mutation
(Change in number of chromosome)

Chromosomal aberration
(Change in structure of chromosome)

Substitution mutation
(Replacement of one nitrogenous base by another nitrogenous base
e.g. : Sickle cell anaemia)

Deletion

Terminal

Intercalary or interstitial

Duplication

Inversion

Paracentric

Pericentric

Translocation

Simple Translocation

1 2 3 4 5 6 7 8
A B C D E F G H
1 2 3 4 5 6
A B C D E F G H 7 8

Reciprocal Translocation

1 2 3 4 5 6 7 8
A B C D E F G H
1 2 3 4 5 6 G H
A B C D E F 7 8

Frame shift mutation
(Loss or addition of one or rarely more than one nitrogenous bases in structure of DNA
e.g. : Thalassemia)

Euploidy
(change in number of sets of chromosomes)

Monoploidy
[Loss of set of chromosome]
(2n–n)

n
n → n

Polyploidy
(Addition of sets of chromosomes)

e.g. 2x + x = 3x triploid
2x + 2x = 4x tetraploid
2x + 3x = 5x pentaploid
2x + 4x = 6x hexaploid
2x + 5x = 7x heptaploid
2x + 6x = 8x octaploid

Aneuploidy
(change in number of chromosomes in sets)

Hypoaneuploidy
(Loss of chromosomes in sets)

Monosomy [2n – 1]
Double monosomy [2n – 1 –1]
Nullisomy [2n – 2]

Hyperaneuploidy
(Add. of chromosomes in sets)

Trisomy [2n + 1]
Double trisomy [2n + 1 +1]
Tetrasomy [2n + 2]

Golden Key Points

- Miescher (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 Gobind 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 Jacques Monod – *Lac* operon
- Frederick Sanger – Developed method for determination of amino acid sequence in proteins.

NCERT SUMMARY

Flow Nucleic acids are long polymers of nucleotides. While DNA stores genetic information, RNA mostly helps in transfer and expression of information. Though DNA and RNA both function as genetic material, but DNA being chemically and structurally more stable is a better genetic material. However, RNA is the first to evolve and DNA was derived from RNA. The hallmark of the double stranded helical structure of DNA is the hydrogen bonding between the bases from opposite strands. The rule is that Adenine pairs with Thymine through two H-bonds, and Guanine with Cytosine through three H-bonds. This makes one strand complementary to the other. The DNA replicates semiconservatively, the process is guided by the complementary H-bonding. A segment of DNA that codes for RNA may in a simplistic term can be referred as gene. During transcription also, one of the strands of DNA acts a template to direct the synthesis of complementary RNA. In bacteria, the transcribed mRNA is functional, hence can directly be translated. In eukaryotes, the gene is split. The coding sequences, exons, are interrupted by non-coding sequences, introns. Introns are removed and exons are joined to produce functional RNA by splicing. The messenger RNA contains the base sequences that are read in a combination of three (to make triplet genetic code) to code for an amino acid. The genetic code is read again on the principle of complementarity by tRNA that acts as an adapter molecule. There are specific tRNAs for every amino acid. The tRNA binds to specific amino acid at one end and pairs through H-bonding with codes on mRNA through its anticodons. The site of translation (protein synthesis) is ribosomes, which bind to mRNA and provide platform for joining of amino acids. One of the rRNA acts as a catalyst for peptide bond formation, which is an example of RNA enzyme (ribozyme). Translation is a process that has evolved around RNA, indicating that life began around RNA. Since, transcription and translation are energetically very expensive processes, these have to be tightly regulated. Regulation of transcription is the primary step for regulation of gene expression. In bacteria, more than one gene is arranged together and regulated in units called as operons. Lac operon is the prototype operon in bacteria, which codes for genes responsible for metabolism of lactose. The operon is regulated by the amount of lactose in the medium where the bacteria are grown. Therefore, this regulation can also be viewed as regulation of enzyme synthesis by its substrate.

Human genome project was a mega project that aimed to sequence every base in human genome. This project has yielded much new information. Many new areas and avenues have opened up as a consequence of the project. DNA Fingerprinting is a technique to find out variations in individuals of a population at DNA level. It works on the principle of polymorphism in DNA sequences. It has immense applications in the field of forensic science, genetic biodiversity and evolutionary biology.