MOLECULAR BASIS OF INHERITANCE

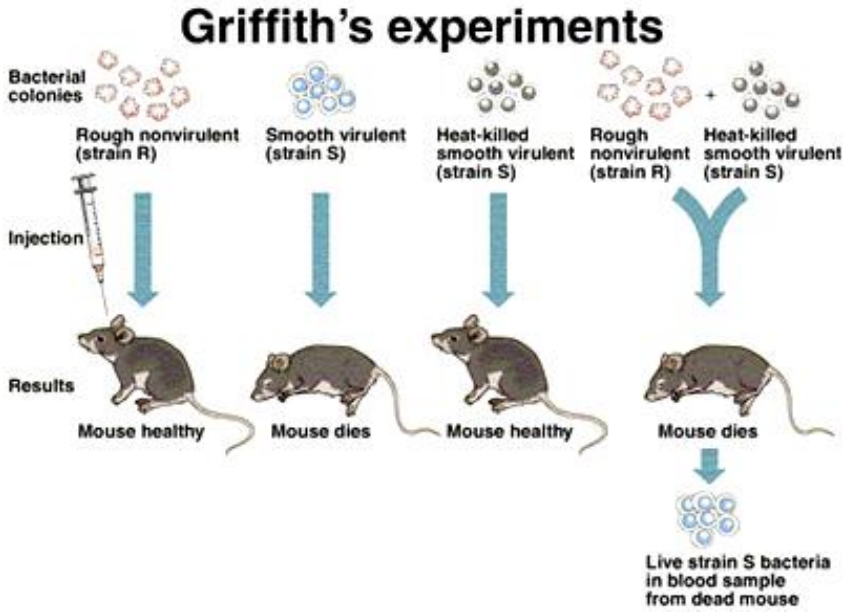
Genetic Material:

- It is the substance that controls the inheritance of traits from one generation to the next generation and is also able to express its effect through the formation and functioning of the traits.

DNA as genetic material:

(i) Transformation:

- It is the conversion in the genetic constitution of an organism by picking up genes present in the remains of its dead relatives.
- The transformation experiments conducted by **Frederick Griffith** in 1928. He used two strains of bacterium ***Diplococcus pneumoniae*** **(*Streptococcus pneumoniae*).**



(a) Smooth (S) or capsulated type which have a capsule. These bacteria are of **virulent** strain and cause **pneumonia.**

(b) Rough (R) or non-capsulated in which capsule is absent. These bacteria are of **non virulent strain** and do not cause pneumonia.

The experiment described in following four parts.

(a) When S-type bacteria injected into mice. The latter died as a result of pneumonia caused by bacteria.

(b) When R- type bacteria injected into mice. The latter lived and pneumonia was not produced.

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

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

- The fourth part of the experiment indicates that some R-type bacteria (non-virulent) were transformed into S- type of bacteria (virulent). The phenomenon is called **Griffith effect or transformation.**
- Later, **Avery, Macleod and McCarty (l944)** repeated the experiment ***in vitro*** to identify transforming substance. They proved that this substance is infact DNA.
- They purified biochemicals from the **killed S-type bacteria** into three components – **DNA, carbohydrate and protein**.
- DNA fraction was further divided into two parts: one with deoxyribonuclease or **DNase** and the **other without it.** The four components were then added to separate culture tubes containing R-type bacteria. After some time they were then analysed for bacteria.
- Only DNA of S-type can changed R-type of bacteria into S-type. Therefore, the character or gene of virulence is located in DNA. Thus they proved that the chemical to be inherited is DNA and it forms the **chemical or molecular basis of heredity.**

(ii) Multiplication of Bacteriophage (Transduction):

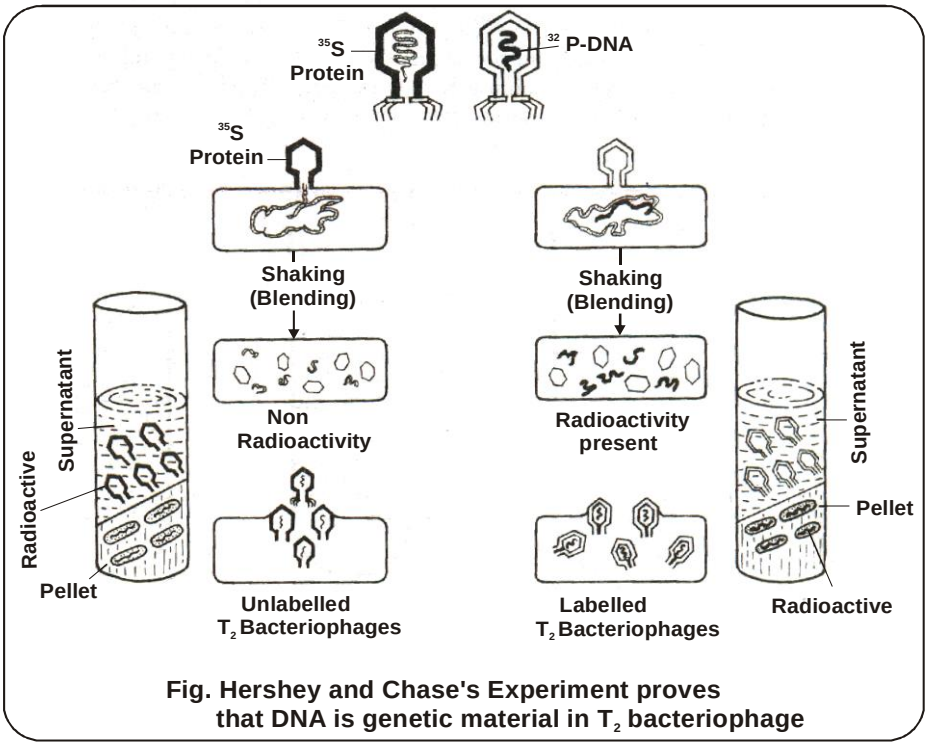


Fig. Hershey and Chase's Experiment proves that DNA is genetic material in T_2 bacteriophage

- The transfer of genetic material from one bacterium to another through bacteriophage is called **transduction**. **T_2** is a **Bacteriophage** which infects **E. coli.**
- **Hershey and Chase (1952)** used radioactive phosphorus **^{32}P** & radio-active sulphur **^{35}S** for their experiment and proved that DNA is a genetic matarial.

Nucleic acids:

- Nucleic acids were first discovered by a Swiss biochemist, **Friedrich Miescher (1869)** who called them **nuclein** due to their acidic nature.
- Chemical analysis of chromosomes shows presence of two nucleic acids-DNA (Deoxyribo nucleic acid) and RNA (Ribo nucleic acid).
- Nucleic acid is a macromolecule & consists of many (polymer) monomeric units, called nucleotides. Each nucleotide is composed of a nucleoside and a phosphate group. Thus nucleotide is a phosphoric ester of nucleoside.

- Each nucleoside consists of sugar molecule and a nitrogenous base. The relationship can be shown as given below.

  Nucleic acid = Many Nucleotides

  Nucleotide = Nucleosides + Phosphate

  Nucleoside = Sugar + Nitrogenous base

  Thus nucleotide = Phosphate + Sugar+ Nitrogenous base

  Nucleic acids bear different components that are briefly discussed below.

  1. **Phosphoric acid:** The acidic nature of nucleic acids is due to the presence of phosphoric acid. Sugar of nucleoside combines with phosphoric acid by a phosphodiester bond formed at 5th or 3rd carbon of the sugar.
  2. **Sugar:** It is a five carbon (pentose) sugar. There are two types of pentose sugars-ribose and deoxyribose. Deoxyribose sugar has one oxygen atom less at second carbon. Ribose sugar is present in RNA while deoxyribose sugar occurs in DNA.
  3. **Nitrogenous bases:** They have two catagories.
     - **(a) Pyrimidine:** It includes Cytosine, Thymine and Uracil. Pyrimidine bases are made of only one ring of carbon.
     - **(b) Purine:** It includes Adenine and Guanine. Purine bases are made of two ring of carbon and nitrogen bases of DNA contains adenine, guanine, cytosine and thymine. In RNA, **uracil** is present in place of **thymine.**

- DNA is a long polymer of deoxyribonucleotides. The length of DNA is usually defined as number of nucleotides (or a pair of nucleotide referred to as base pairs) present in it. This also is the characteristic of an organism. For example, a bacteriophage known as ϕ ×174 has 5386 nucleotides, Bacteriophage lambda has 48502 base pairs (bp), *Escherichia coli* has 4.6×10^6 bp, and haploid content of human DNA is 3.3×10^9 bp.
- Many nucleotide monomer units join one another to give rise to polynucleotide chain.
- The two adjacent nucleotides are joined by formation of **phosphodiester bond** (a bond that involves two ester bonds). A polynucleotide chain is often written as 5'p 3'OH. This indicates that it is a dinucleotide with phosphate group (p) attached to the 5th carbon of terminal nucleotide and hydroxyl group (OH) is present at 3rd carbon of basal nucleotide.

Structure of DNA (Deoxyribonucleic acid):

- **J.D. Watson and F.H.C. Crick (1953)** proposed double helical structure of DNA based on the results of **M.H.F.Wilkins and co-workers.** All these three persons were awarded Nobel Prize in 1962 for this work.

  The following are some of the characteristic features of double helical structure of DNA.

  (1) **Each nucleotide** consists of **sugar, phosphate and a nitrogenous base**. Many such nucleotides are linked by phosphodiester bonds to form a polynucleotide chain or strand.

  (2) **Phospho diester bonds** are formed between 5'carbon of sugar of one nucleotide and 3' carbon of sugar of the next nucleotide.

  (3) **Nitrogenous base** is attached to **1' carbon of sugar**. At this place **purine base** is attached by its **9 position** and **pyrimidine** by its **1 position**.

  (4) Polynucleotide strand is made of backbone of sugar and phosphate forming its long axis and bases at right angles to it.

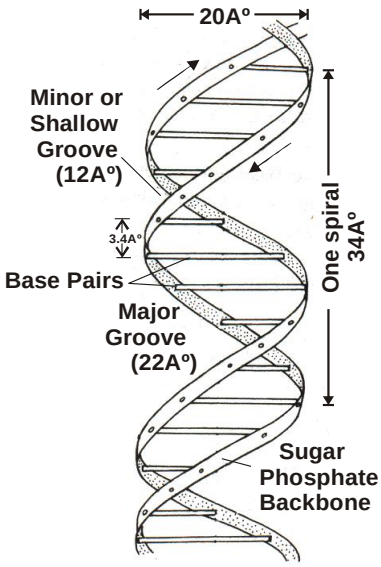


Fig. Structure of DNA- coiling in double helix of DNA

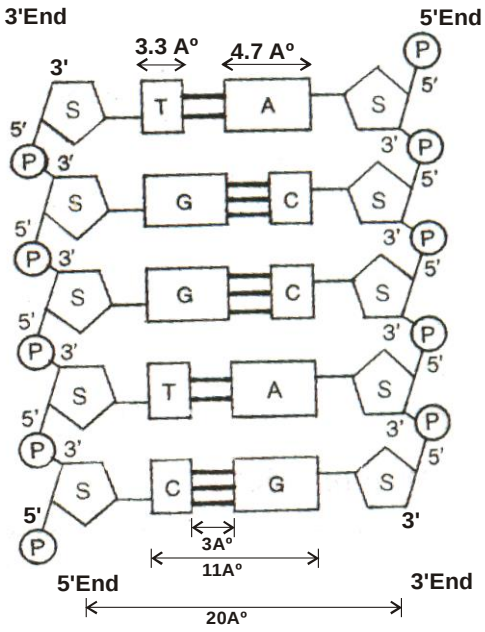


Fig. Arrangement of various Components of DNA

Chargaff (1950) made observations on the base and other contents of DNA. These observations or generalizations are called Chargaffs rule.

(i) Purine and pyrimidine base pairs are in equal amount, that is, adenine + guanine =thymine + cytosine.

(ii) Molar amount of purine-adenine is always equal to the molar amount of pyrimidine thymine. Similarly, guanine is equalled by cytosine.

(iii) Sugar deoxyribose and phosphate occur in equimolar proportions.

(iv) **The ratio of A + T/G + C is constant for a species.**

(v) Chargaffs rule states that in natural DNAs the base ratio AT is always close to unity and the GC ratio also to always close to unity indicated that A always pairs with T and G pairs with C. A and T, G and C, therefore, are complementary base pairs.

(vi) Thus, if one DNA strand has A, the other would have T and if it has G, the other, would have C. Therefore, if the base sequence of one strand is CAT TAG GAC, the base sequence of other strand would be GTA ATC CTG. Hence, the two poly nucleotide strands are called complementary to one another.

(vii) Two such complementary strands are joined with one another by hydrogen bonds between their complementary nitrogenous bases. There are three hydrogen bonds between cytosine and guanine and two hydrogen bonds between adenine and thymine.

(viii) The two polynucleotide chains are helically coiled around the same axis in such a way that these can separate from one another only by uncoiling. Helical coiling is supposed to be right handed. Such a form of DNA is now called B-DNA

(ix) The two chains or strands are antiparallel, i.e., they run in opposite directions in relation to their sugar molecules. Their 5'p 3' OH phosphodiester linkages are in opposite directions

(x) **Double standed DNA molecule has a diameter of 20A°.**

(xi) **The helix makes one complete turn every 34 A° (3.4 nm) along its length.**

(xii) There are 10 nucleotides per turn of helix. Thus the distance between two neighbouring base pairs is 3.4 Aº. Since the discovery of DNA structure, some other forms of DNA have also been recognised. These forms have been classified considering (i) the number of base pairs per turn of helix and (ii) the distance of base pairs along the helical axis. Accordingly, besides commonly known B-DNA, other forms are A, C (sometimes D and E) and Z DNA. Some important similarities and differences among different types of DNA are given in.

S.No.	Property	B	Z
1	Handedness of helix	Right handed	Left handed
2	Pitch of helix per turn	34 Å	46 Å
3	Diameter of helix	20 Å	18 Å (thinnest)
4	Stability	Stable and Physiologically active form	Unstable
5	Base pairs per turn of helix	10	12 (6 dimers)
6	Distance between two base pairs	3.4Å	3.8 Å
7	Repeating unit	Mononucleotide	Dinucleotide

Packaging of DNA Helix:

- The distance between two base pairs of DNA duplex is 0.34 nm. (0.34×10^{-9}m) and the length of DNA duplex in a typical mammalian cell is calculate (by multiplying the total number of bp with distance between two consecutive bp, that is, 6.6×10^9 bp × 0.34 nm/bp). It is about 2.2 metres.

DNA Packing in Prokaryotes:

- In prokaryotes, such as E.coli, though they do not have a defined nucleus, the DNA is not scattered throughout the cell. DNA (being negatively charged) is held with some proteins (that have positive charges) in a region known as **'nucleoid'**. The DNA in nucleoid is organised in large loops held by proteins.

DNA Packing in Eukaryotes:

- It is much more complex. The DNA is folded with the help of positively charged basic proteins called histones. The histones are of low molecular weight, acting as repressor of genes to prevent transcription. These are **rich in lysine and arginine**. These are of **5 types (H_1, H_{2A}, H_{2B}, H_3, H_4)** depending upon the ratio of lysine, arginine and histidine.

- Histones are organised to form a unit of eight molecules known as histone octamer **(octomer of histone- Each of H_{2A}, H_{2B}, H_3 and H_4 - Oudet 1975)**. The negatively charged DNA (It has **$1\frac{3}{4}$ turns** of DNA **containing 146 nucleotides)** is wrapped around the positively charged histone octomer to form a structure called **nucleosome or nu particle.** The nucleosomes are structural sub units of chromatin. A nucleosome has a size of 110 × 60 Å is ellipsoidal and bead like. The fifth histone H_1 is found on the part of DNA between 2 nucleosomes called **linker DNA (size 145 Å with 70 bp). Nucleosome + DNA linker is collectively called chromatosome.**

- Nucleosomes give beaded appearance to chromatin like **string on a bead**. H_1 Protein helps in grouping of nucleosomes while nucleosomes help in packing of DNA. A chain of nucleosomes is once again coiled with 6 nucleosomes per turn to form **solenoid (Klug 1982)**. Each solenoid has 1200 bp of DNA and 300 Å size. H_1 is most species specific, most divergent and act as **marker protein.** It is called plugging protein. It is rich in lysine and is non conservative. It is not found in pairs and not form nucleosome. H_3 and H_4 are most conserved proteins.

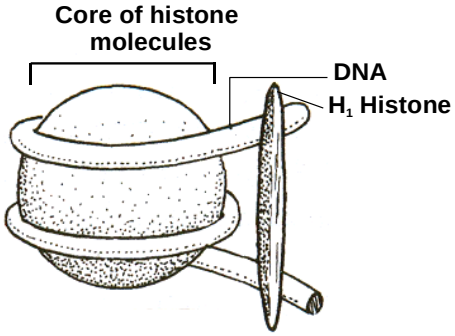


Fig. Components of nucleosome

DNA (2 nm wide) → Nucleosome (10 nm wide) —further coiling→ Solenoid (30 nm wide) —further coiling→ chromatin fibre (300 nm wide) —further coiling→ Chromatid (700 nm) → Chromosome (1400 nm)

Properties of Genetic Material (DNA versus RNA)

- 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.).
- A molecule that can act as a genetic material must fulfill the following criteria:
  (i) It should be able to generate its replica (Replication).
  (ii) It should chemically and structurally be stable.
  (iii) It should provide the scope for slow changes (mutation) that are required for evolution.
  (iv) It should be able to express itself in the form of 'Mendelian Characters'.
- If one examines each requirement one by one, because of rule of base pairing and complementarity, both the nucleic acids (DNA and RNA) have the ability to direct their duplications. The other molecules in the living system, such as proteins fail to fulfill first criteria itself.
- 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.
- 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.
- 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.

RNA WORLD

RNA was the first genetic material. There is now enough evidence 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). But, RNA being a catalyst was reactive and hence unstable. Therefore, DNA has evolved from RNA with chemical modifications that make it more stable. DNA being double stranded and having complementary strand further resists changes by evolving a process of repair.

Structure of RNA (Ribonucleic acid):

❖ RNA or ribonucleic acid is present in all the living cells. It is found in the cytoplasm as well as nucleus. Sugar in RNA is ribose sugar. Phosphoric acid is similar to that present in DNA. Purine bases are adenine and guanine but pyrimidine bases are cytosine and uracil (thymine being replaced by uracil).

Types of RNA:

❖ RNA is generally involved in protein synthesis but in some viruses, it also serves as a genetic material. Therefore two major types of RNA are as follows.

(a) Genetic RNA (b) Non-genetic RNA.

(a) Genetic RNA: H. Frankle-Conrat (1957) showed that RNA present in TMV (Tabacco Mosaic Virus) is genetic material. RNA acts as a genetic material in most plant viruses.

(b) Non-genetic RNA: This type of RNA is present in cells where DNA is genetic material. Non-genetic RNA is synthesized on DNA template. It is of following three types.

(i) Messenger RNA (mRNA): It carries genetic information present in DNA. mRNA constitutes about 5-10% of the total RNA present in the cell. The molecular weight varies from 25,000 to 20,00,000.

(ii) Ribosomal RNA (rRNA): It is most stable type of RNA and is found associated with ribosomes. It forms about 80% of the total cell RNA. The molecular weight varies from 35,000 to 8,00,000.

(iii) Transfer RNA (tRNA): It is also known as soluble RNA (sRNA).

These are the smallest molecules which carry amino acids to the site of protein synthesis.There are approximately 73-93 bases. These constitute about 10-15% of the total cell RNA. The molecular weight of tRNA varies from 23,000 to 30,000.

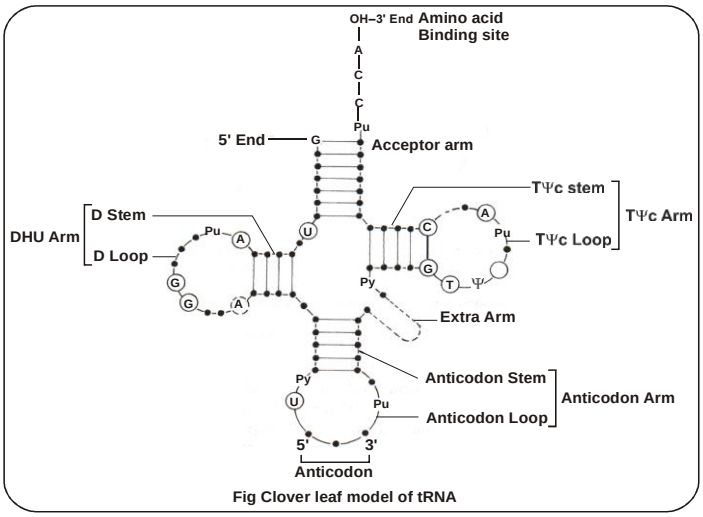


Fig Clover leaf model of tRNA

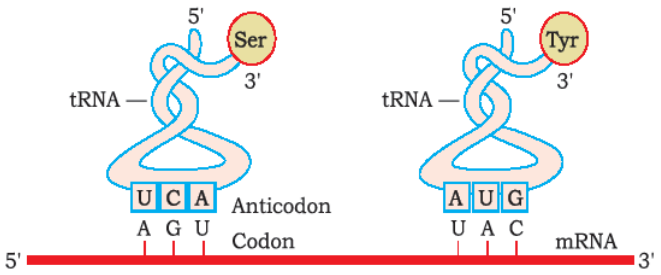


Figure: tRNA - the adapter molecule

FOR AIIMS

Structure of t-RNA:

- **2-D clover leaf model of t-RNA was proposed by Holley, (1965).** tRNA molecule appears like a clover leaf being folded with three or more double helical regions, each having loop.
  - **(i) Anticodons loop :** It has 7 bases out of which three bases form anticodon (nodoc) for recognising and attaching to the codon of mRNA.
  - **(ii) AA-Binding Site :** It is amino acid binding site. The site lies at the 3' end opposite the anticodon and has CCA–OH group. The 5' end bears G. Amino acid or AA-binding site and anticodon are the two recognition sites of tRNA.
  - **(iii) T Ψ C Loop :** It has 7 bases out of which Ψ (Pseudouridine) and rT (ribothymidine) are unusual bases.The loop is the site for attaching to ribosome.
  - **(iv) DHU Loop :** The loop contains 8– 12 bases. It is largest loop and has dihydroxyuridine. It is binding site for aminoacyl synthetase enzyme.
  - **(v) Extra Arm :** It is a variable side arm or loop which lies between T Ψ C loop and anticodon. It is not present in all tRNAs. The exact role of arm is not known.
- The **3-D structure** of this **tRNA** was, however, found to be characteristic **L-shaped** by **Kim (1973).**

Differences Between DNA and RNA		
S.No	**DNA**	**RNA**
1	It usually occurs inside nucleus and some cell organelles.	Very little RNA occurs inside nucleus. Most of it is found in the cytoplasm.
2	DNA is the genetic material.	RNA is not the genetic material except in certain viruses, e.g. Reovirus.
3	It is double stranded with the exception of some viruses (e.g. ϕ x 174)	RNA is single stranded except in certain viruses, double stranded Reovirus.
4	DNA contains over a million nucleotides.	Depending upon the type, RNA contains 70-12000 nucleotides.
5	DNA is of only two types; intra-nuclear and extra-nuclear.	There are at least three types of RNAs– mRNA, rRNA and tRNA.
6	It contains deoxyribose sugar.	It contains ribose sugar.

7	Nitrogen base thymine occurs in DNA alongwith three others–adenine, cytosine and guanine.	Thymine is replaced by uracil in RNA The other three are similar – adenine, cytosine and guanine.
8	It replicates to form new DNA molecules. DNA transcribes genetic information to RNA.	It cannot normally replicate itself. RNA translates the transcribed message for forming polypeptides.
9	DNA control metabolism and genetics. Including variation.	It only controls metabolism under instruction from DNA.
10	Purine and pyrimidine bases are in equal number.	There is no proportionality between number of purine and pyrimidine bases.

Replication of DNA:

- The synthesis of DNA from DNA is called Replication.
- It usually occurs during **S-phase** of cell cycle.
- DNA performs two types of functions
  - **(A) Autocatalytic:** DNA synthesizes new DNA by replication,
  - **(B) Heterocatalyic:** DNA helps in the synthesis of other substances like RNA, protein.
- According to **Delbruck**, DNA replication is of three types.
  - **(1) Dispersive:** Old structure undergoes fragmentation. Fragments synthesise complementary structures, both of which assemble randomly to form two replica.
  - **(2) Conservative:** A New structure is formed over the template of old structure (conserved)
  - **(3) Semiconservative:** One half is parent structure and one half new structure in each replica. It was firstly suggested by **Watson and Crick (1953).**
- **It is now proven that DNA replicates semiconservatively. It was shown first in** *Escherichia coli* and subsequently in higher organisms, such as plants **and human cells. Matthew Meselson and Franklin Stahl performed thefollowing experiment in 1958:**
- They cultured E. coli bacteria in culture medium containing $N^{15}H_4Cl$ (Heavy isotope of N as N^{15}) for several generations.
- Now they introduced labelled Bacteria in another culture medium contain N^{14} H_4 Cl (Normal N^{14}).
- Now they used **density gradient centrifugation** method with **CsCl (Cesium chloride)** to examine DNA of its offsprings.
- They found that **DNA** was **intermediate type** in **first generation** in which one strand was heavy (containing N^{15}) and other strand was light (containing N^{14}).
- **Second generation** of bacteria contained two types of DNA, **50% light ($N^{14}N^{14}$)** and **50% intermediate ($N^{15}N^{14}$).**
- In **third generation** of bacteria contained **25% intermediate ($N^{15}N^{14}$)** and **75% light ($N^{14}N^{14}$)** in **1 : 3 ratio** and **fourth generation** bacteria contained **12.5% $N^{15}N^{14}$** and **87.5% $N^{14}N^{14}$ DNA** in **1 : 7 ratio.**

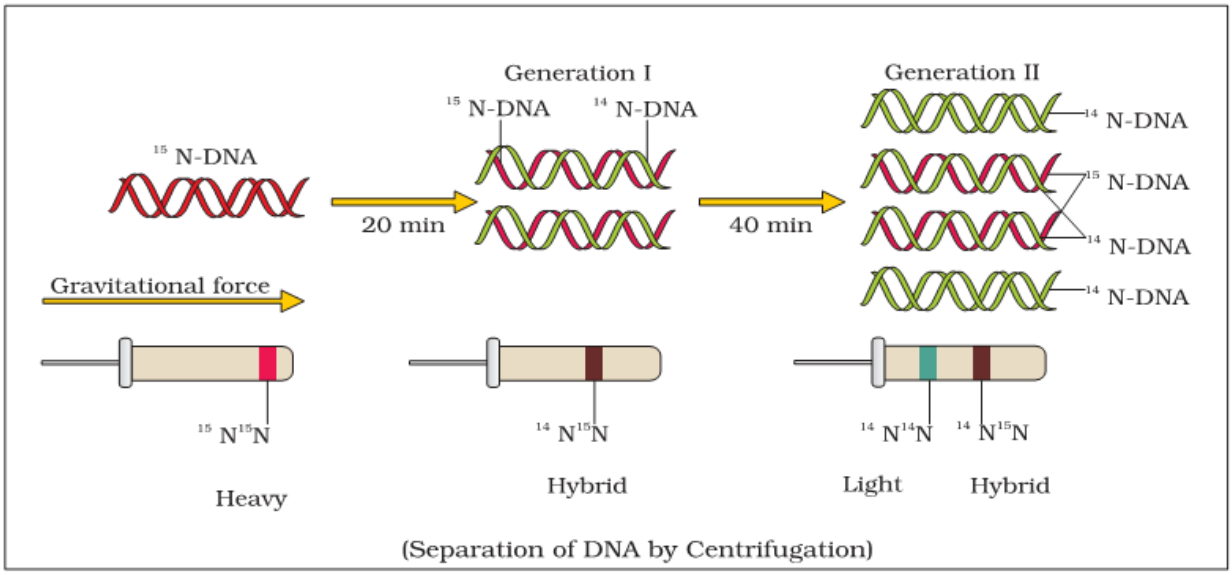


Figure: Semi-conservative replication of DNA (Results of experiment of Meselson & Stahl)

- Very similar experiments involving use of radioactive thymidine to detect distribution of newly synthesised DNA in the chromosomes was performed on Vicia faba (faba beans) by Taylor and colleagues in 1958. The experiments proved that the DNA in chromosomes also replicate semiconservatively.

Mechanism of DNA Replication:

(1) Origin of Replication: It starts at a particular place called **origin of replication** or **Ori**. In prokaryotes replication starts at one point & entire DNA strand takes part in replication thus it contains **single replicon** while in Eukaryotes **several replicons** present.

- DNA replication is **bidirectional, semidiscontinuous and semiconservative** in nature.

(2) Activation of Deoxyribonucleotides: The phosphorylated nucleotides (dAMP, dGMP, dCMP, dTMP) are found in inactivated form. They react with ATP in the presence of **phosphorylase enzyme** & converted in to active dATP, dGTP, dCTP, dTTP.

(3) Exposure of DNA helix: Helicase enzyme acts over the ori site of DNA template and unwinds the two strands of DNA.

- **SSB (single stranded binding) Protein** prevents the recoiling of uncoiled DNA strands.
- **Topoisomerase** cause nicking of one strand of DNA (for removing coils) and resealing the same. Along with **Topoisomerase**, **bacteria** possess another enzyme called **DNA Gyrase** which can introduce negative supercoils.
- Whole of the DNA does not open in one stretch due to very high energy requirement but the point of separation proceeds slowly from one end to other. It gives the appearance of Y-shaped structure called **replication fork.**

(4) RNA Primer: It is small strand of RNA (5–10 nucleotide). It is synthesized at 5'end of new strand with help of enzyme **Primase**. Formation of RNA primer constitutes the initiation phase of synthesis because without the presence of RNA primer, DNA polymerase can not add nucleotides.

Resonate the Concept

- In **eukaryotes**, the **function** of **primase** is carried out by **enzyme DNA polymerase α.**

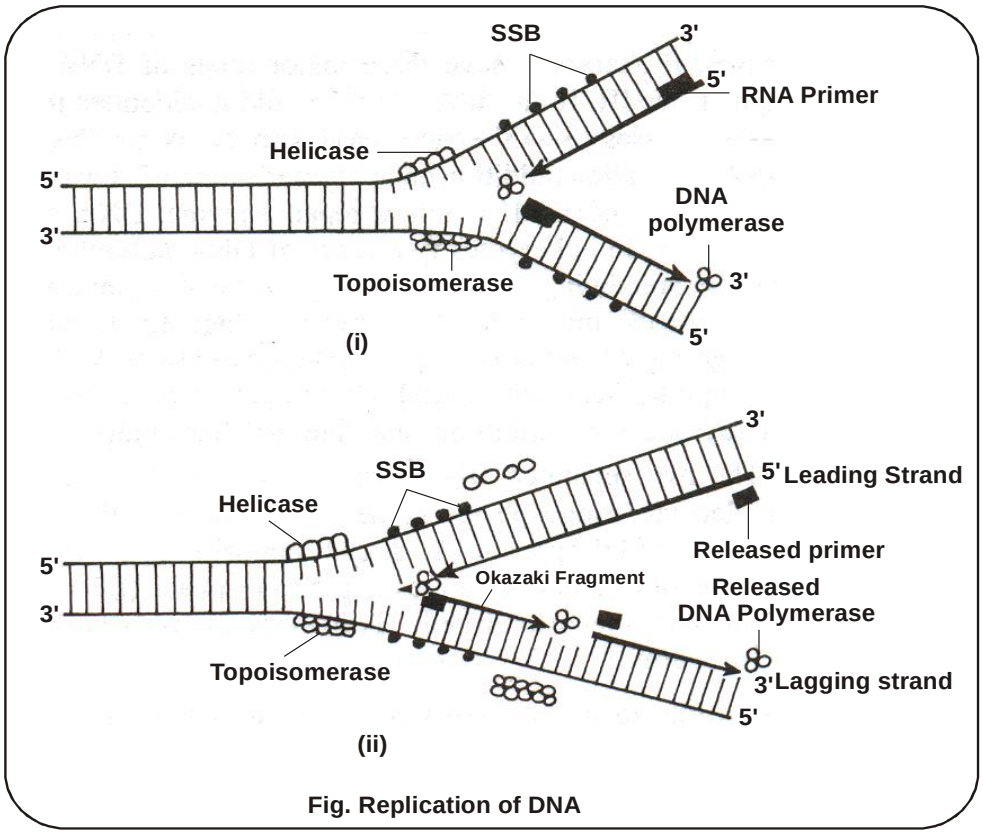


Fig. Replication of DNA

(5) DNA Polymerase: Prokaryotes possess three types of DNA synthesising enzymes called **DNA polymerases III, II and I** they add nucleotides in **5' → 3' direction** on **3' → 5' strand.** DNA replication is mainly performed by DNA polymerase III. DNA polymerase I is major repair enzyme where as polymerase II is minor repair enzyme.

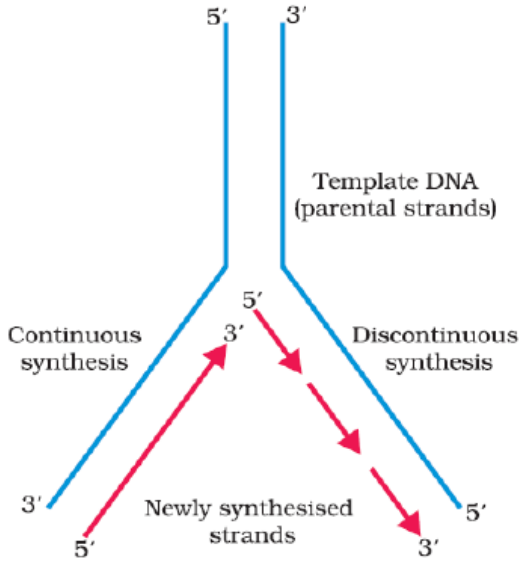
- In **eukaryotes five types of DNA polymerases (α, β, γ, δ, ε)** have been reported. out of them α, δ, ε are major enzymes. According to **sugino et al, DNA polymerase α** acts at both the leading and lagging strands and initiates DNA synthesis along with primase activity while **DNA polymerase δ, ε** are involved in **elongation** of the **leading and lagging strands** respectively.

(6) Base Pairing: Two separated strands of DNA in the replication fork function as template.

- Deoxyribnucleoside triphosphates come to lie opposite the nitrogen bases of exposed DNA templates – dTTP opposite-A, dCTP opposite G, dATP opposite T and dGTP opposite C.
- With the help of **pyrophosphatase** enzyme the two extra phosphates present on the dexyribonucleotides separate. Energy is released in this process that is utillized for base pairing.
- Energy is used in establishing hydrogen bonds between the free nucleotides and nitrogen bases of templates.

(7) Chain formation: It requires **DNA polymerase III** in **prokaryotes** and **polymerase δ/∈** in **eukaryotes**. **DNA polymerase III** is a complex enzyme having **seven subunits (α, β, δ, γ, ∈, θ, τ).**

- In the presence of Mg^{++}, ATP/ GTP, TPP and DNA polymerase -III, the adjacent nucleotides attached to nitrogen bases of each template DNA strand establish **phosphodiester bonds** and get linked to form replicated DNA strand. Two strands of DNA run antiparalled to each other.



- Replication on one DNA template is continuous in **5' → 3' direction** due to opening of its 3' end this newly formed strand is called **leading strand**.
- On the second DNA template the replication of DNA is **discontinuous** due to opening of small stretch of fork at a time. Small fragments deposite with the help of RNA primer. These fragements are called **okazaki fragements (1000 - 2000 nucleotides in prokaryotes and 100–200 in eukaryotes).**
- After deposition of each Okazaki fragment RNA primer is released and gap is filled by the activity of DNA polymerase thus the new strand is formed called **Lagging strand**.

  After deposition of bases **DNA Ligase** enzyme seals these bases.
- Thus one strand grows continuously while the other strand is formed discontinuously hence DNA replication is **semidiscontiuous**.

(8) Proof reading and DNA repair: Sometimes wrong base is deposited in the strand. **DNA poymerase III** is able to check this error and removes the wrong base. It allows addition of proper base but DNA polymerase III can not distinguish **uracil** from **thymine** such an error is corrected by number of enzymes.

- **DNA polymerase I** removes the wrong base and attaches the correct base in the strand in Prokaryotes where as **DNA polymerase β** in eukaryotes.

Transcription:

- Synthesis of RNA from DNA is called transcription.
- One of the two strands of DNA takes part in transcription. According to **Lewin (2000)** transcription takes place at **anti sense strand or – strand.**

Transcription Unit: The part of DNA which takes part in transcription is called **transcription unit.** The latter has three components.

(i) Promoter (ii) Structural gene (iii) Terminator

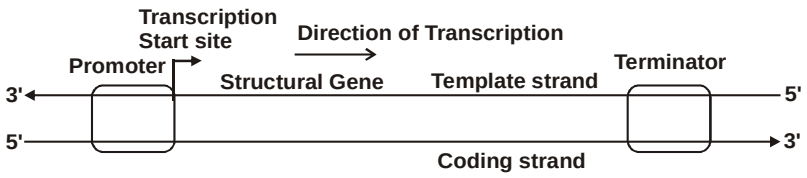


Fig. Components of a transcription unit.

- Promoter is situated upstream of structural gene at 5' end of coding strand where as terminator at downstream of structural gene at 3' end of coding strand. Promoter bears different parts for attachment to various **transcription factors**. In most of the cases, the promoter contains **AT rich** regions called **TATA box (TATAAAA)**.The latter has groove for the attachment of **specific protein.** In prokaryotes it is called **Pribnow box (TATAAT)**.
- **Structural gene** is a part of strand of DNA having **3'→ 5'** polarity on which transcription proceeds only in **5'→3' direction** of new strand. This strand of DNA is called **template strand or master strand or antisense, or (–) strand**. The other strand is **nontemplate strand** that does not take part in transcription is also also called **sense or coding strand or plus (+) strand**.

Mechanism of Transcription:

- It takes place in **G_1 and G_2 phases** of interphase of cell cycle in the nucleus in eukaryotes where as in prokaryotes it occurs in the cytoplasm.

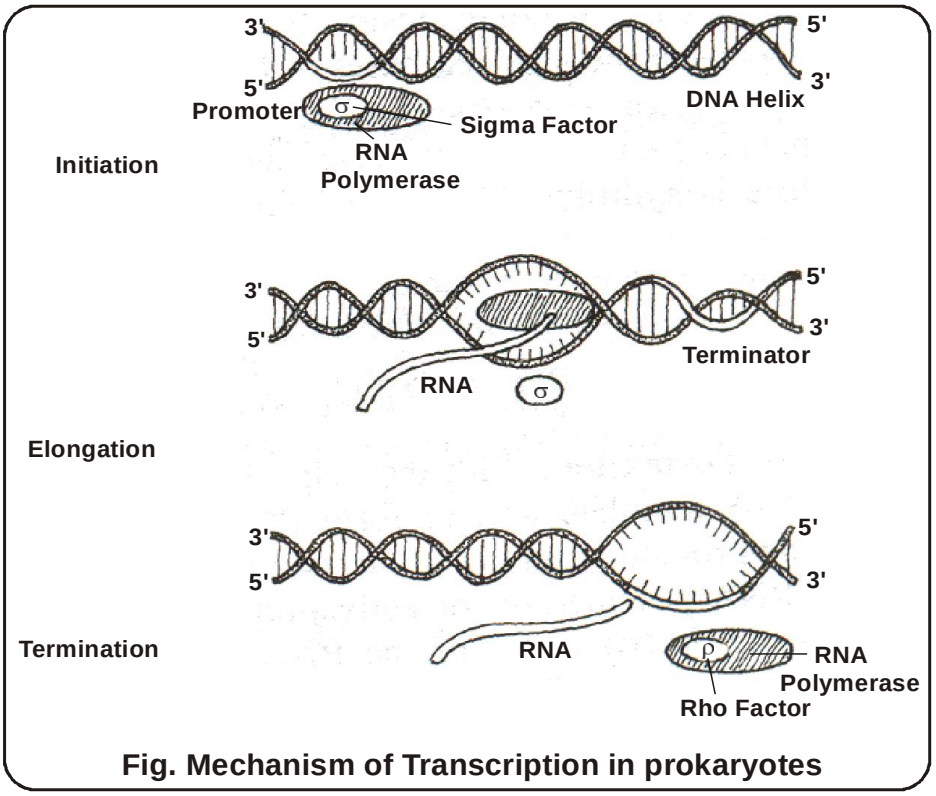


Fig. Mechanism of Transcription in prokaryotes

- Single **RNA polymerase** enzyme performs transcription in Prokaryotes. The former consists of α,α' β , β', σ and ω **factors**.Out of them σ **(sigma) factor** recognizes the site of transcription on the promoter region of DNA template and resting part of enzyme is called **core enzyme.**
- In Eukaryotes, Three types of RNA polymerase involve in the synthesis of three different types of RNA.

(i) RNA polymerase : I - synthesizes - rRNA (28S, 18S, 5·8S).

(ii) RNA polymerase : II - synthesizes - hn RNA (Precursor of mRNA)

(iii) RNA polymerase : III - synthesizes - tRNA, 5S rRNA & SnRNA.

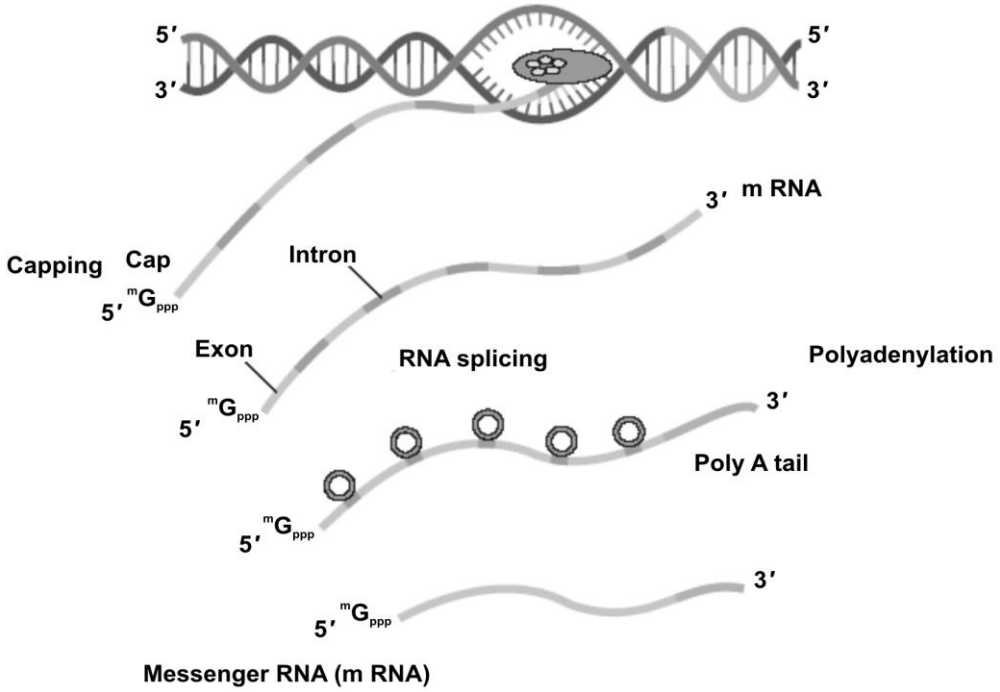


Fig. Process of Transcription in Eukaryotes

- Ribonucleoside mono Phosphates - UMP, AMP, CMP and GMP react with ATP in the presence of enzyme phosphorylase to form active UTP, ATP, CTP and GTP.

Cistron: Prokaryotes bear **polycistronic RNA** and **Eukaryotes** bear **monocistronic RNA**.

- Helicase enzyme performs uncoiling of DNA strands. SSB prevents their recoiling.
- **Sigma factor** binds on TATA box on the promoter region of master strand.
- UTP comes to lie opposite to the A, GTP opposite C, CTP opposite to G, ATP opposite to T. At this time pyrophosphatase separates two groups from each Ribonucleoside triphosphate resulting they are converted into UMP, GMP, CMP, AMP during deposition on master strand.
- Core enzyme proceeds transcription from promotor region towards terminator region. Sigma factor is released.
- When the chain of RNA reaches at terminator region**, rho factor (ρ factor)** prevents its synthesis by ATPase activity resulting newly synthesized RNA becomes separated. Which is called Primary transcript or Hn RNA (Heterogenous RNA).
- After separation of RNA, Both strands of DNA are again coiled for duplex formation.

Processing of RNA or Primary Transcript:

It involves following methods.

(i) Cleavage : Ribonuclease – P enzyme cleaves 5–7 bases of primary transcript & the latter may form t–RNA precursors.

(ii) Splicing : In eukaryotic primary transcript, Introns have no information about protein synthesis.

SnRNA (Small nuclear RNA) combines with some peptides to form small nuclear Ribonucleo protein or **snurp**. The latter combines with some peptides to form **spliceosome.**

Spliceosomes cut the introns of primary transcript and the exons of primary transcripts are joined by **RNA ligase** & thus active m–RNA is formed.

(iii) Terminal addition: It involves addition of some nucleotides at the terminal part of Primary ranscript for example – CCA sequence is added on **3' end of t-RNA, Poly A** is added on **3' end of hn–RNA in tailing or polyadinylation**, in **capping**, **7– methyl guanosine** (formed by modification of GTP) is added on **5' end of hn–RNA**.

(iv) Nucleotide modification: Nucleotides are methylated, ethylated or deaminated **Ex: Inosine, methylcytosine, Methylguanosine, dihydrouracil, Pseudouracil.**

Genetic code:

- It represents relationship of sequence of Amino acids in polypeptide and sequences of nucleotides of mRNA/DNA.
- Genetic code was discovered by **Nirenberg and Matthaei**.
- **Crick (1961)** stated that deletion or addition of one or two bases in DNA disturbs the DNA functioning. It was George Gamow, a physicist, who argued that since there are only 4 bases and if they have to code for 20 amino acids **(4^1 = 4) Double codon** can specify **4^2 = 16 aminoacids** that are not sufficient for the coding of essential 20 amino acid. **Triplet codon** can specify **4^3 = 64 aminoacids**. That are sufficient for 20 amino acids.
- **George Gamow** gave concept of **Triplet codon**. He also **coined** the term **Genetic code.**

Table Assignment of mRNA codons to Amino Acids

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

Features of genetic code:

(i) Triplet codon: Genetic code is Triplet codon composed of three adjacent nitrogen bases.
Codon - A sequence of three nucleotides specifying an amino acid

(ii) Start signal or Initiation codon: It is mostly **AUG (Methionine codon)**. But in prokaryotes it can be GUG and UUG (Lewin 2000), in all cases they specify Methionine. GUG and UUG specify different amino acids inside the polypeptide chain (GUG - Valine, UUG- Leucine).

(iii) Stop signal or Termination codon: Polypeptide chain termination is signalled by three termination codon **UAA (ochre), UAG (Amber) and UGA (opal)**. They do not specify any amino acid and hence are called **non sense codons**.

(iv) Non ambiguous codon: Normally one codon specifies only one amino acid and not any other.

(v) Non overlapping code: A nitrogen base is a constituent of only one codon.

(vi) Universal code: A codon specifies the same amino acid in all organisms from virus to human.

(vii) Commaless: There are no pauses so that genetic code reads continuously. If a nucleotide is deleted or added, the whole genetic code will read differently.

(viii) Colinearity: The sequence of codons of DNA/mRNA correspond to the sequence of amino acids in a polypeptide.

(ix) Related codons: Amino acids with similar properties have related codons Eg: aromatic amino acids tryptophan (UGG), Phenylalanine (UUC, UUU), and tyrosine (UAC, UAU).

(x) Degeneracy of codons: Since there are 64 triplet codons and only 20 amino acids, the incorporation of some amino acids is influenced by more than one codon only **Tryptophan (UGG) and Methionine (AUG)** are specified by **single codons**. All other amino acids are specified by 2–6 codons. The latter are called degenerated codons.

Wobble hypothesis (Crick, 1966): In degenerated codons the first two nitrogen bases are similar while the third one is different. The third nitrogen base has no effect on coding actually 5' end base of t-RNA anticodon is able to wobble and get paired with even noncomplementary base of m-RNA **Eg: CCA, CCC, CCG, and CCU all specify amino acid proline.**

One gene–One enzyme hypothesis:

- **Archibald garrod (1909):** He stated that diseases inherited from parents are inborn error of metabolism.These are due to defect in a single enzyme that catalyzes a particular reaction.
- **Beadle and Tatum, (1948)** Proposed **One gene–One enzyme hypothesis**. They conducted experiments on the nutrition of **pink mould (Neurospora crassa).** This fungus grows on simple nutrient medium and has the ability to synthesize all its cellular components. Such an organism is called **prototroph**.
- An organism that is unable to synthesize a particular cellular component such as an amino acid or coenzyme is called **auxotroph**.

Beadle and tatum produced **arginine** (an amino acid) auxotrophs (mutants of Neurospora unable to synthesize arginine).

Arginine synthesis passes through the following path-

Ammonia + Sugar $\xrightarrow{\text{Enz A}}$ **Ornithine** $\xrightarrow{\text{Enz B}}$ **Citrulline** $\xrightarrow{\text{Enz C}}$ **Arginine**

They found that any step of this metabolic chain could be blocked by a mutation in a specific enzyme catalyzing the reaction, each enzyme representing a different gene product. Thus Beadle and Tatum reached a conclusion that each gene regulates the synthesis of a single enzyme. This laid the foundation of biochemical genetics. Beadle and Tatum were awarded Nobel Prize in 1958.

One gene–one polypeptide Hypothesis:

- **Yanofsky et al (1965):** proposed one gene–one polypeptide hypothesis He stated that a structural gene regulates the synthesis of a single polypeptide. Haemoglobin is composed of two α and two β polypeptides synthesised by two separate genes.

One gene–one Function Hypothesis:

- A **gene or cistron** (It is a part of DNA composed of stretch of deoxyribonucleotides that codes for a biochemical controlling other cistrons, rRNA, tRNA or polypeptide through mRNA) performs one function, **structural or regulatory**.

Central dogma:

- It is the unidirectional flow of information that proceeds from DNA to mRNA and then decoding information present in m-RNA in the formation of polypetptide chain or protein (translation).

DNA $\xrightarrow{\text{transcription}}$ **mRNA** $\xrightarrow{\text{translation}}$ **Polypeptide (Protein)**

- The concept of **central dogma** was proposed by **crick** in 1958.
- **Commoner** (1968) propounded concept of **circular flow of information (from DNA → RNA → Protein → RNA → DNA).**

Reverse transcription:

- **Temin and Baltimore (1970)** reported that **retroviruses** operate a **central dogma reverse** (Inverse flow of information). RNA of these viruses first synthesises DNA through **reverse transcriptase or RNA dependent DNA polymerase.** This DNA synthesized on RNA template is called **c-DNA or Retroposon.**

RNA → c-DNA → RNA → Protein

Protein synthesis:

Mechanism of Protein Synthesis (Translation): According to the coded information of m-RNA the formation of polypeptide over ribosome is called translation.

It involves following steps

(1) Activation of Amino acid: Specific Amino acid reacts with ATP in the presence of Amino acyl t-RNA synthetase enzyme to form Amino acyl adenylate enzyme complex.

AA + ATP + E $\xrightarrow{Mg^{2+}}$ **AA ~ AMP – E + PPi**

amino acid **amino acyl tRNA synthetase** **Pyrophosphate**

AA ~ AMP – E + tRNA $\longrightarrow$ **AA - tRNA + AMP + E.**

(2) Synthesis of polypeptide chain : It involves following steps.

(i) Initiation of polypeptide chain :

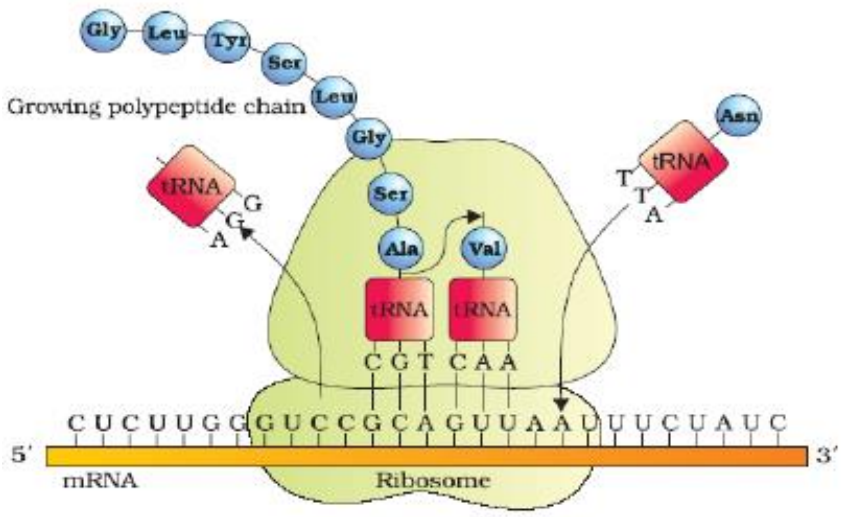
- In **prokaryotes IF_3, IF_2, IF_1,** - **Initiation factors** are required for initiation of polypeptide chain, whereas in **eukaryotes eIF_2, eIF_3, $eIF_{1,}$ eIF_{4A}, eIF_{4B}, eIF_{4C}, eIF_{4D}, eIf_5, eIF_6** are required..
- m-RNA is fused with P-site of small subunit 40S of ribosome (30 S in procaryote) in the presence of **eIF_2** to form **40 S mRNA complex** (**30 S mRNA complex** forms in the presence of **IF_3**).
- Now 40 S mRNA complex attracts Amino acyl t- RNA Complex both fuse to form **40 S m RNA + t RNA_m^{met} complex** (**30 S m RNA + t RNA_f^{met} complex** in **prokaryotes** in the presence of **IF_2 & GTP**) in the presence of **eIF_3 and GTP**.
- Now this complex is fused with large subunit 60 S of ribosome (50S in prokaryotes) on RER to form **80 S mRNA t- RNA complex** in the presence of **eIF_1, eIF_{4A}, eIF_{4B}, eIF_{4C}, (70 s m RNA tRNA complex in the presence of IF_1).**
- The **0·001M concentration of Mg^{++} ions** is required for the formation of **intact ribosome.**

  At this time P site and complex are covered by Ribosome now its A site is exposed at the front of next codon of mRNA.

(ii) Elongation of Polypeptide chain:

- Now **Amino acyl tRNA complex** reaches on A side due to stimulation of **Elongation factors (eEFα1 & eEFα2** in Eukaryotes and **EF Tu** & **EF Ts** in Prokaryotes)**.**

  Now **Peptidyl transferase (Ribozyme) enzyme** stimulates the fusion of Amino acids of both complexes.
- In this process - **COOH (carboxylic group) group** of **amino acid** of **complex of P-site** and – **NH_2 group (Amino group)** of **amino acid** of **complex of A-site** are fused to form **CO – NH bond or peptide bond** in the presence of **peptidyl transferase and GTP.**
- After this process **tRNA of P site** breaks and slips away and the dipeptide complex moves on A- site due to **translocase enzyme (EF-G in prokaryotes and eEFβ in eukaryotes)**. Ribosome or m-RNA rotates slightly this movement is called **translocation**. As a result dipeptide complex is shifted on p-site and the A-site is again exposed at the front of next codon of mRNA.
- This process repeats again & again as result a long polypeptide chain is formed.



(iii) Termination of polypeptide chain:

- When termination codons **UAA or UAG or UGA** is exposed on m-RNA. It has no information about AA t-RNA complex therefore elongation of polypeptide chain is stopped.
- GTP based releasing factors require for the activation of termination codons. **RF (eRF) in eukaryotes** and **RF_1 (specific for UAG and UAA) and RF_2 (specific for UAA and UGA) in prokaryotes.**

Important Point: 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). They are required for efficient translation process.

Types of Genes:

(1) House Keeping Genes (Constitutive Genes): These genes are constantly expressing themselves in a cell because their products are required for the normal cellular activities, **Eg: genes for glycolysis.**

(2) Non-constitutive Genes: (Luxury Genes): The genes are not always expressing themselves in a cell. They are switched on or off according to the requirement of cellular activities **Eg: gene for nitrate reductase in plants**, lactose system in Escherichia coli. Non-constitutive genes are of further two types, inducible and repressible.

(3) Pseudogenes: They have homology to functional genes but are unable to produce functional products due to intervening nonsense codons, insertions, deletions and inactivation of promoter regions, **Eg: Several of snRNA genes**.

(4) Split Genes: They discovered by **Sharp and Roberts (1977).** Split genes possess extra or nonessential regions interspersed with essential or coding parts. The noessential parts are called **introns or spacer DNA or interveining sequences (IVS)**. Essential coding parts are called **exons.**

(5) Transposons (Jumping Genes): They are segments of DNA that can jump or move from one place in the genome to another, Transposons were first descovered by **Mc Clintock (1951**) in case of **Maize.** The term transposon coined by **Hedges and Jacob (1974).**

(6) Overlapping Genes: In **ϕ X 174,** genes E, B and K overlap other genes.

(7) Processed Genes: These are eukaryotic genes that lack introns. Processed genes have been formed probably due to reverse transcription by retroviruses. Processed genes are generally nonfunctional due to lack of promoters.

Regulation of Gene Expression:

- Regulation over the functioning of genes is called regulation of gene expression. It can be exerted at four levels.
  - **(i)** Transcriptional level during formation of primary transcript.
  - **(ii)** Processing like splicing, terminal additions or modifications.
  - **(iii)** Transport of RNAs from nucleus to cytoplasm.
  - **(iv)** Translation level.

Genes expression is of three types

- **(1) Inducible (2) Constitutive (3) Repressible.**
  - **(1) Inducible :** In this type the Gene is switched on in response to the presence of substrate (Inducer).
  - **(2) Constitutive :** Genes and their enzymes remain operational throughout.

(3) Repressible : It is of two types

(a) Positive Control: The product of regulatory gene initiates expression of genes under of its control.

(b) Negative control: The product of a regulatory gene switches off the expression of genes under its control.

Operon Model:

- An operon is a segment of DNA that functions as single regulated unit comprising a regulator gene, a promoter gene, an operator gene, one or more structural genes, a repressor and an inducer or corepressor, these systems are common in prokaryotes.
- **First operon lac–Operon** was discovered by **Jacob and Monod (1961) in E.coli**.

Operons involve two types.

(1) Inducible operon model

(2) Repressible operon model.

(1) Inducible Operon Model:

- It is found in **catabolic pathway Eg: Lactose operon or Lac operon.** Lac operon consists of following components.

(i) Structural genes: They actually control the synthesis of m-RNA through transcription. They determine primary structure of polypeptide chain. In Lac operon three **structural gene z, y, a** take part in the formation of **polycistronic mRNA** that regulates the synthesis of **β galactosidase, permease and transacetylase enzymes.** β-galactosidase hydrolyses lactose in glucose & galactose. Permease allows the entry of Lactose in the cell. Transacetylase performs metabolism of toxic thiogalactosides.

(ii) Operator gene: It controls the activity of structural genes. When repressor of regulator gene binds to the operator gene. The latter becomes switched off.

(iii) Promoter gene: It acts as initiation signal. It bears **RNA polymerase enzyme**. When operator gene is functional, its RNA polymerase travels on structural gene and perform transcription.

(iv) Regulator gene: It regulates the synthesis of **repressor**. It is also called **inhibitor gene or i gene.**

(v) Repressor: It is proteinaceous substance formed by Regulator gene. It has two allosteric sites, one for the attachment of operator gene and other for the attachment of inducer.

(vi) Inducer: It is chemical substance (Hormone, enzyme etc). When inducer is present in the medium, inducer combines with repressor resulting some conformational changes occur in the repressor in such a way that it becomes unable to bind on operator gene. Therefore the latter continuously operative. When inducer is completely consumed. Repressor is again activated. In lac-operon lactose acts as inducer **(Actual allolactose acts as inducer)** & substrate.

(viii) CAP (catoabolic Activator Protein): It exerts positive control on Lac operon but in its absence RNA polymerase is unable to recognize Promoter gene.

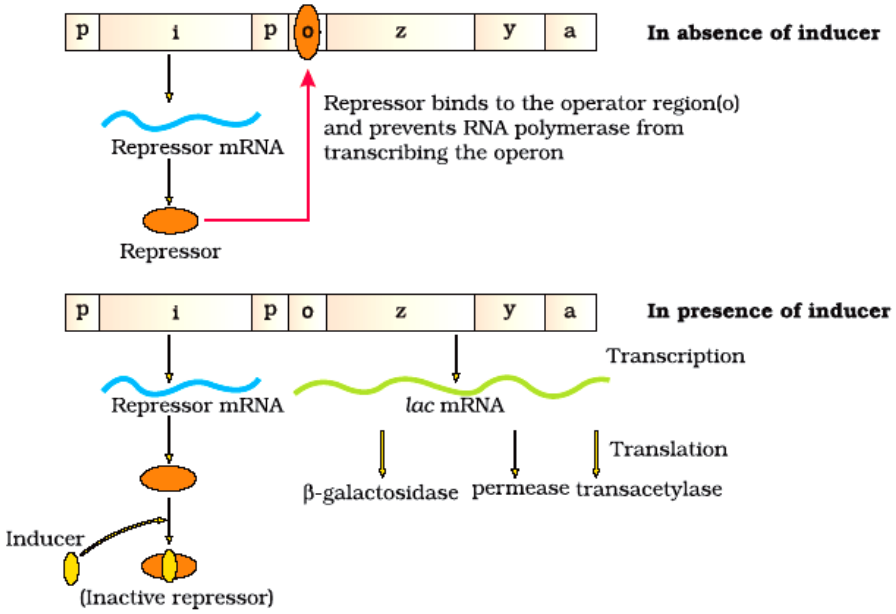


Figure: The lac Operon

(2) Repressible Operon Model:

❖ Repressible operon System is found in **anabolic pathways. Eg : Tryptophan or trp operon of E.coli**

Trp operon consists of the following components.

(i) Structural Genes: The genes are connected to transcription of mRNAs. Tryptophan operon has **five structural genes trp E, D, C, B, A.** They form enzymes for five steps of tryptophan synthesis.

(ii) Operator Gene: It regulates the activity of structural genes usually. Aporepressor produced by regulator gene is unable to completely block operator gene.

(iii) Promoter Gene: It acts as initiation signal. It bears **RNA polymerase enzyme**. When operator gene is functional, its RNA polymerase travels on structural gene and perform transcription.

(iv) Other Regulatory Components: It involves two components that lie between **operator gene** and **structural gene E.**

(a) Leader sequence (L): It is controller of attenuator.

(b) Attenuator (A): It helps in reducing tryptophan synthesis when it is available in sufficient amount without switching off the operon.

(v) Regulator Gene: It produces proteinaceous component (such as Aporepressor) for possible blocking the activity of operator gene.

(vi) Aporepressor: It is a proteinaceous substance formed by regulator gene. Independently it is unable to block the activity of operator gene. For this purpose it requires a corepressor.

(vii) Corepressor: It is a nonproteinaceous component of repressor that may be an end product of reactions. In trp operon, when end product Tryptophan is accumulated in sufficient amount its some molecules act as corepressor. The latter combines with aporepressor. And forms repressor that block the operator gene resulting structural genes become switched off. The phenomenon is called **feed-back repression** that exerts a negative control.

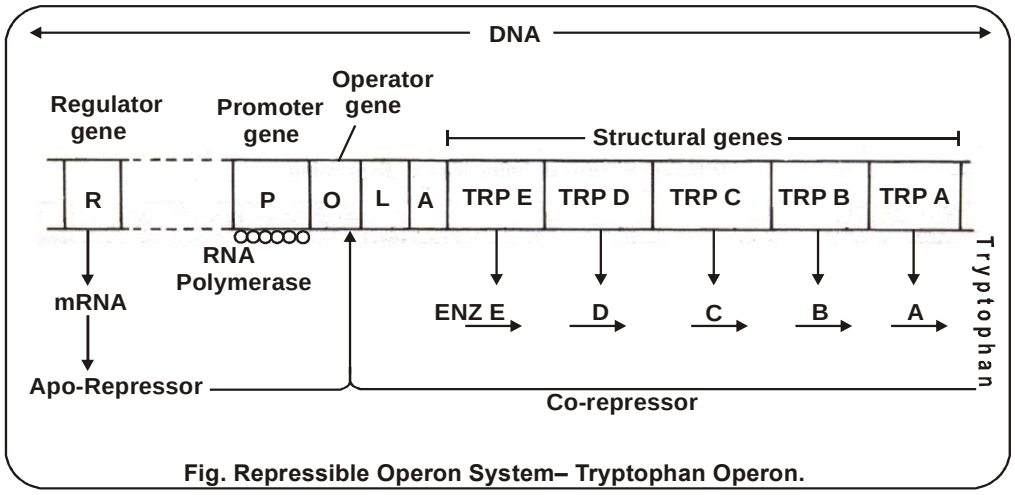


Fig. Repressible Operon System– Tryptophan Operon.

Human Genome project (HGP):

❖ The Human Genome Project was a 13-year project.started by **U.S. Department of Energy and National Institute of Health** for sequencing human genome in 1990. **Welcome Trust (UK)** joined the project as a **major partner.** Later on Japan, France, Germany, China and some other countries also joined it.

Human Genome Project (HGP) was called a mega project. Because human genome is said to have approximately 3 x 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. The enormous amount of data expected to be generated also necessitated the use of high speed computational devices for data storage and retrieval, and analysis. HGP was closely associated with the rapid development of a new area in biology called as **Bioinformatics.**

Aims of HGP:

(i) To determine the sequence of 3.2 billion base pairs of human genome

(ii) To Identify all the genes of human genome and determined their functions

(iii) To identify those genes that are responsible for genetic disorders.

(iv) To Store this information in data bases.

(v) The project may result in many ethical, legal and social issue which must be addressed and solved.

Methodology: Two approaches have been recognized for analysing the human genome.

(i) ESTs or expressed sequence tags: To identify all the genes that are expresed as RNA.

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

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 programs were developed. These sequences were subsequently annotated and were assigned to each chromosome.

The project was completed for sequencing in 2003.

However, final report of **chromosome I** was published at last in **May 2006.**

Salient features of human Genome:

(1) Human genome contains **30,000 genes** which are much lower than previous estimate of 80,000 – 10,0000.

(2) Human genome contains **3.1647 billion** nucleotide base.

(3) Less than 2% of the genome shows structural genes that code for proteins.

(4) Average gene size is 3000 base pairs but size vary greatly, with the largest known human gene being dystrophin at 2·4 million bases.

(5) **Chromosome I** contains **2968 genes (maximum gene)** while **Y-chromosome** bears **231 genes (minimum genes)** in human chromosome.

(6) **99.9% of the nucleotide bases** are **exactly similar** in all human beings.

(7) **About 1.4 million locations** have been reported where single nucleotide differences or **SNPs (snips) or single nucleotide polymorphism** are found. They have the potential to helping and finding chromosomal locations for disease associated sequences and tracing human history.

(8) Repeated sequences makeup very large portion of the human genome.

(9) Repeated sequences are stretches of DNA sequnces that are repeated hundred to thousand times. They are thought to have no direct coding functions but they shed light on chromosome structure. Dynamics & evolution. Many non human organisms like, Bacteria *(E. coli),* yeast, *Caenorhabditis elegans, Drosophila*, plants (Rice and *Arabidopsis*) etc have also been sequenced.

DNA Fingerprinting (DNA Profiling):

Historical Aspect:

❖ **Sir Alec Jeffreys** (1984) discovered the DNA fingerprinting technique. Dr. K. Kashyap and Dr. Lalji Singh started the fingerprinting technology in India.

What is DNA-fingerprinting: It is a technique to identify a person on the basis of his//her DNA specificity.

Principle of DNA Fingerprinting: Jeffreys observed that DNA of each individual contains some noncistronic hyper-variable repeat minisatellite sequences. These repeat minisatellite sequences are called **variable number of tandem repeats or VNTRs (These are also called minisatellites).** The numbers of repeat show very high degree of polymorphism. As a result the size of VNTR varies in size from 0.1 to 20 kb.

❖ These sequence show high degree of polymorphism and form the basis of DNA fingerprinting. Since DNA from every tissue (such as blood, hair-follicle, skin, bone, saliva, sperm etc.), from an individual show the same degree of polymorphism.

❖ What is DNA polymorphism?

DNA Polymorphism (variation at genetic level) arises due to mutations OR if an **inheritable mutation** is observed in a population at high frequency, it is referred to as **DNA polymorphism** (frequency greater than 0.01)

The latter are similar only in monozygotic twins and vary in number from person to person. These repeats are inherited in the offsprings by their parents.These are used as genetic markers in a personal identity test. One half of VNTR alleles of the offspring resemble that of the mother and other half that of the father.

*** Inheritable mutation is observed in a population at high frequency, it is referred to as DNA polymorphism.**

Technique of DNA fingerprinting:

(i) The DNA is isolated from the nuclei of white blood cells or spermatozoa or the hair follicle cells.

(ii) Restriction endonuclease enzyme performs digestion of DNA molecules. The former cuts DNA in to fragements. The fragments of DNA also contain the VNTRs.

(iii) **Gel electrophoresis** is used to separate these fragments according to their size.

(iv) VNTRs are multiplied through **PCR technique**. The VNTRs are treated with alkaline chemicals to split them into single stranded DNAs.

(v) SS DNA fragments of the gel are shifted onto a **nylon paper/nitrocellulose membrane** by **southern blotting technique**.

(vi) Radioactive SS-DNA-probes are used for hybridization with VNTR on the nylon membrane.

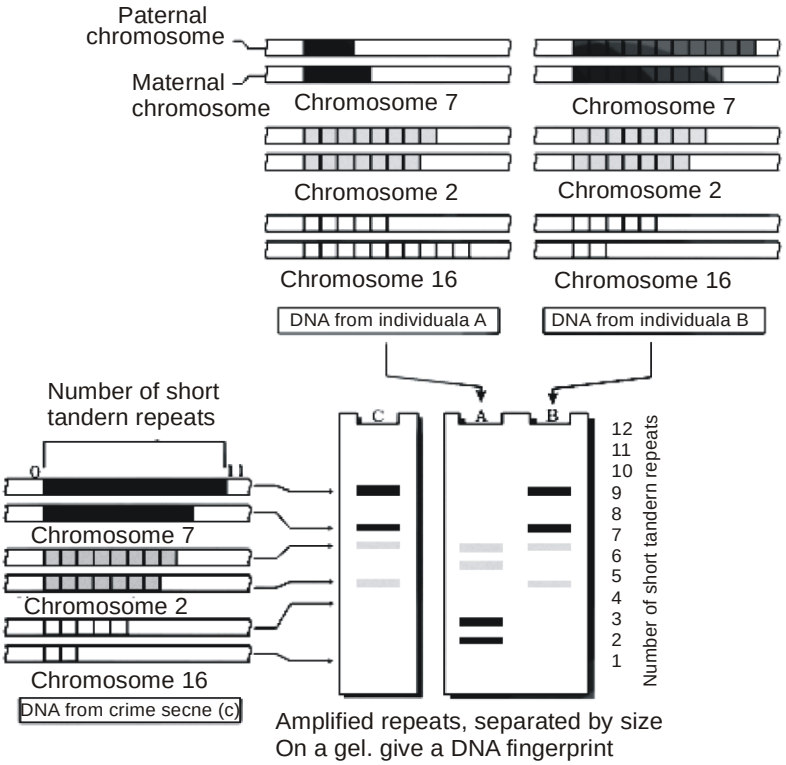


Figure: Schematic representation of DNA fingerprinting: Few representative chromosomes have been shown to contain different copy number of VNTR. For the sake of understanding different colour schemes have been used to trace the origin of each band in the gel. The two alleles (paternal and maternal) of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

(vii) Now the nylon membrane is exposed at the front of **X-ray film** and mark the places where the radioactive DNA probes have bound to the DNA fragements. These places are marked as dark bands when X-ray film is developed. This is known as **autoradiography**.

(viii) The dark bands on X-ray film show the DNA fingerprints.

Applications of DNA Fingerprinting:

(i) It is very useful in the detection of crime and illegal pursuits.

(ii) Paternity maternity disputes can be solved by DNA fingerprinting.

(iii) It can be used to study the breeding patterns of animals facing the danger of extinction.

(iv) It provides information about relationship of man with apes.

(v) It determining population and genetic diversities.

Resonate the Concept

(i) **Palindromic DNA:** DNA duplex have areas where sequence of nucleotides is the same but opposite in the two strands.

(ii) **Repetitive DNA:** Type of DNA containing multiple copies of identical squences of nitrogen bases. Repeated sequences account for more than 50% of the human genome. Repetitive DNA is inert in the synthesis of RNA, i.e it is heritable but does not transcribe and translate into any protein.

(iii) **Satellite DNA:** It is a part of repetitive DNA that has long repetitive nucleotide sequences in tanden that forms a separate fraction on density ultracentrifugation. Satellite DNA is of two types.

(a) **Microsatellite sequences** 1– 6 bp repeat units flanked by conserved sequences.

(b) **Minisatellite sequences** 11– 60 bp flanked by conserved restriction sites. These are hypervariable and are specific for each individual. They are used in DNA finger printing.

(iv) **Antisense Therapy:** Excess of anti-mRNA or antisense RNAs will not allow the pathogenic genes to express themselves. It is used for curing AIDS, cancers and other diseases.

(v) **Artificial Gene:** Khorana et al (1968) synthesised first artificial gene It was alanine It was alanine -t-RNA gene with 77 base pairs. It was non functional in living system. in 1979 he prepared tyrosine-tRNA gene with 207 base pairs. It was functional.

(vi) **Seymour Benzer:** He stated that gene involves three units **(1) cistron, (2) recon (3) muton**

(vii) **Ribozymes:** These are RNA molecules showing catalytic activity. **Cech et al** discovered in ciliated protozoan - **Tetrahymena thermophila**.

(viii) **SnRNA (Small Nuclear RNA):** Small sized RNA that is found in nucleus. Each RNA is combined with 7–8 molecules of proteins and form small nuclear ribonucleoprotein or snRNP (snurp). The latter takes part in splicing of (U_1 and U_2), rRNA processing (U_3) and mRNA processing (U_7).

(ix) **ScRNA (small Cytoplasmic RNA):** Small sized RNA that lies freely in the cytoplasm. SRP or signal recognition particle is formed by the fusion between 7S small cytoplasmic RNA and 6 protein molecules. The former helps in taking and binding a ribosome to endoplasmic reticulum for producing secretory proteins.

(x) **Denaturation and Renaturation:** In the presence of higher temperature (T_m 82–90ºC) or low or high pH, the H-bonds between nitrogen bases of two strands of DNA can break. It is called denaturation. The DNA strands separated by melting at T_m 82–90ºC can reassociate and form duplex on cooling to temperature at 65ºC. It is called renaturation or annealing.

(xi) **Hot spots:** At these place of DNA/chromosomes where frequency of mutation and recombination are very much increased.

(xii) **Cryptogram:** A method of expressing in a standard form of a collection of data used in classification

(xiii) **Punctuation Codons:** They are four in number, three represent termination of polypeptide synthesis, viz UAA, UAG and UGA. AUG is initiation codon.

(xiv) **C-Value:** Total amount of DNA present in single genome.