CHAPTER -6

MOLECULAR BASIS OF INHERITANCE

Introduction to nucleic acid:

- The inheritance and variation in living organism solely depends upon their genetic material which is basically DNA that is deoxyribonucleic acid.
- Most of the living beings except retroviruses (RNA) possess DNA as their genetic material which are the polymers of nucleotide.
- The new era of genomics not only studying the structure and function of DNA /RNA but also helping the mankind in many more fields like drug design, vaccines, medicines ,diagnosis etc.

Searching for genetic material: Griffith's experiment:

- Mendel proposed the principles of inheritance in year 1866, but it took a long to be discovered DNA as the genetic material. By 1926 the quest to determine the mechanism for genetic inheritance had reached the molecular level.
- In the year 1928 Frederick Griffith conducted a series of experiment the bacterium causing pneumonia and witnessed a miraculous transformation in the bacterium. During the course of his experiment the bacterium had changed in physical form.

Griffith's experiment:

- Griffith performed series of experiment by selecting two strains of bacterium ***Streptococcus pneumonia*** (pneumococcus) namely **S- strain** and **R- strain**.
- When the bacteria are grown on the culture plate some produce smooth mucous coat with shiny colonies (S- strain), while others produce rough non mucous coat with rough colonies (R- Strain).
- The S- strain bacterias are **virulent** and cause pneumonia. The R- strain bacterias are **non virulent** and do not cause pneumonia.
- Mice infected with S-strain bacteria died with pneumonia where as mice infected with R- strains do not develop pneumonia.
- Griffith was able to kill bacteria by heating them and the heat killed strain (both R and S) injected to Mice.
- The experiment can be described in following four steps:

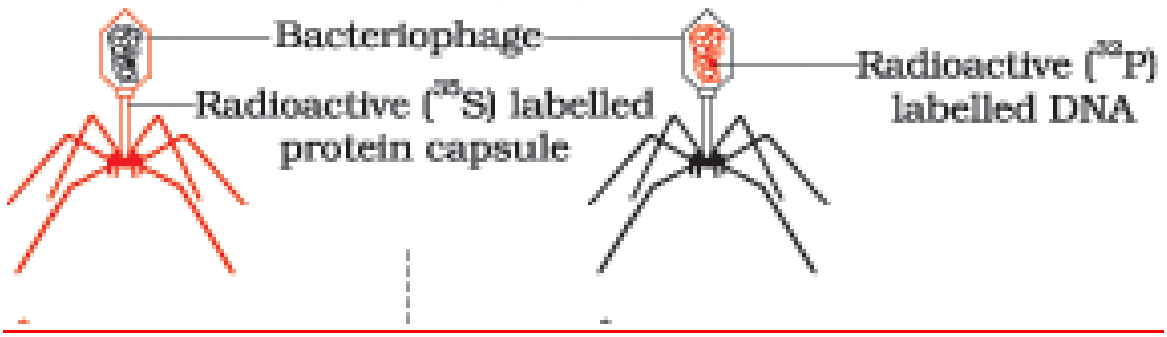
S-strain bacteria(live)	Injected into mice	Mice died
R-strain bacteria(live)	Injected into mice	Mice survived
S-strain bacteria (heat killed)	Injected into mice	Mice survived
S-strain bacteria (heat killed) +R-strain bacteria (live)	Injected into mice	Mice died

- From the above series of experiments he concluded that R- strain bacteria somehow **transformed** by the heat killed S- strain bacteria.
- That means heat killed S- strain bacteria had enabled the R- strain to synthesize a smooth mucous coat and became virulent.
- However he knew that there was something that could cause bacteria to transform from one type to another type and called it **transforming principle**.

Biochemical characterization of Transforming Principle:

- The nature of genetic material was not defined from Griffith's experiment. So **Oswald Avery, Colin MacLeod, Maclyn Mc Carty** (1933-44) worked to determine the biochemical nature of transforming principle.
- They purified biochemicals like proteins, DNA and RNA from heat killed S- strain to see which one could transform live R- strains into S- strains.
- They also discovered enzymes like **proteases, DNases, RNases**. Out of them proteases and RNases did not affect transformation, so the transforming substance was not protein or RNA.
- But digestion with **DNases inhibited the transformation**, suggesting that the DNA caused transformation and is the genetic material.

The genetic material DNA: Hershy & Chase experiment:



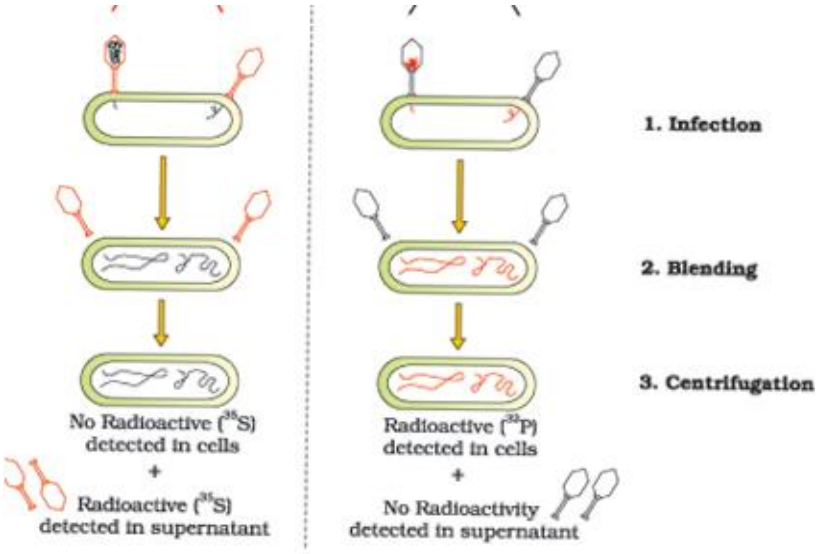


Figure 6.5 The Hershey-Chase experiment

- Not all biologists were convinced that DNA is the hereditary material. So the unequivocal proof that DNA is the genetic material came from transduction experiment of Alfred Hershy and Martha Chase in year 1952.
- They worked with T_2 Bacteriophage which infects *E.coli*. The bacteriophage attaches to the bacterial cell wall and their genetic material enters inside bacterial cell. Subsequently the viral genetic material integrates with bacterial genome and manufactures more viral particle.
- So it was the most suitable material for Hershy and Chase to determine whether DNA or protein entered inside the bacterial cell and is the genetic material.
- They grew some viruses on the media containing **radioactive phosphorous** and some others on the media containing **radioactive sulphur**. Viruses grown on radioactive phosphorous media contained only **radioactive DNA (no protein)** whereas the other grown on radioactive sulphur media contained only **radioactive protein (no DNA).**
- Radioactive viruses were allowed to attach to *E.coli* and infection proceeded. Therefore bacteria of both the cultures became labeled.
- After labeling three steps were followed: (i) Infection (ii) Blending (iii) Centrifugation

(i) **Infection**: After attachment of radioactively labeled virus upon E.coli, the genetic material of the phage enters into bacterial cell through cell wall.

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

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

- After centrifugation the heavier bacteria settled down in form of pellet and the lighter viral coats remain suspended as supernatant. Both the pellet and supernatant were analysed.
- It was found that the virus with labeled protein did not make the bacteria labeled. Instead radioactivity was restricted to the supernatant which was found to contain only viral capsid.
- But the Bacteriophage labeled with radioactive DNA made the bacteria labeled which became heavier indicating that **DNA was the genetic material that passed from virus to bacteria.**

The DNA: Length of DNA :

- DNA is a long polymer of deoxyribonucleotides. The length of DNA usually defined as number of nucleotides or a pair of nucleotide referred to as base pairs (bp) present in it which is the characteristic of an organism. For example:
- Bacteriophage $\emptyset * 174$ has 5386 nucleotides.
- Bacteriophage lambda has 48502 base pairs.
- E.coli has $4.6*10^6$ bp.
- Haploid content of human DNA is $3.3*10^9$ bp.

Structure of polynucleotide chain:

- The chemical structure of polynucleotide chain is made up of its monomer nucleotide which is composed of nucleoside and a phosphate group. The nucleoside composed of nitrogenous base and pentose sugar. Overall the nucleotide is composed of three components like Po_4 group, pentose sugar and nitrogenous base.

Pentose sugar: It is a 5 carbon sugar. In DNA the sugar is deoxyribose but in RNA it is ribose sugar.

Nitrogenous base: A heterocyclic nitrogen containing called nitrogenous base is a major component of nucleotide. There are two types of nitrogenous bases:

(a) **Purine** - It is a double ring structure with N at position 1 , 3, 7 & 9 . e.g. Adenine and Guanine.

(b) **Pyrimidine**- It is a single ring structure with N at position 1& 3. e. g. Cytosine, Thymine, Uracil. Cytosine is common in both DNA and RNA; Thymine is present in DNA and Uracil is present in RNA at the place of thymine.

Phosphate group: A phosphate group is present.

A polynucleotide chain shows following bonds or linkage:

Glycosidic linkage: A linkage between a nitrogenous base and a pentose sugar to form a nucleoside is called glycosidic bond. The nucleosides are deoxyadenosine, deoxyguanosine, deoxycytidine, deoxythymidine. Purine nucleosides have **1-9 glycosidic linkage** and Pyrimidine nucleosides have **1-1 glycosidic linkage**.

Phosphodiester bond: Adjacent nucleosides are connected together to form a long polymer chain. For any one nucleotide the phosphate attached to the OH group at the 5' carbon of pentose sugar is in turn bonded to the 3' carbon of pentose sugar of the next nucleotide. Since each PO_4 –OH bond is an ester bond ,hence the linkage between two nucleotides is called 3'-5' Phosphodiester bond.

Hydrogen bond: The opposite and complementary nitrogen bases are held together by hydrogen bond two between A-T & three between C-G.

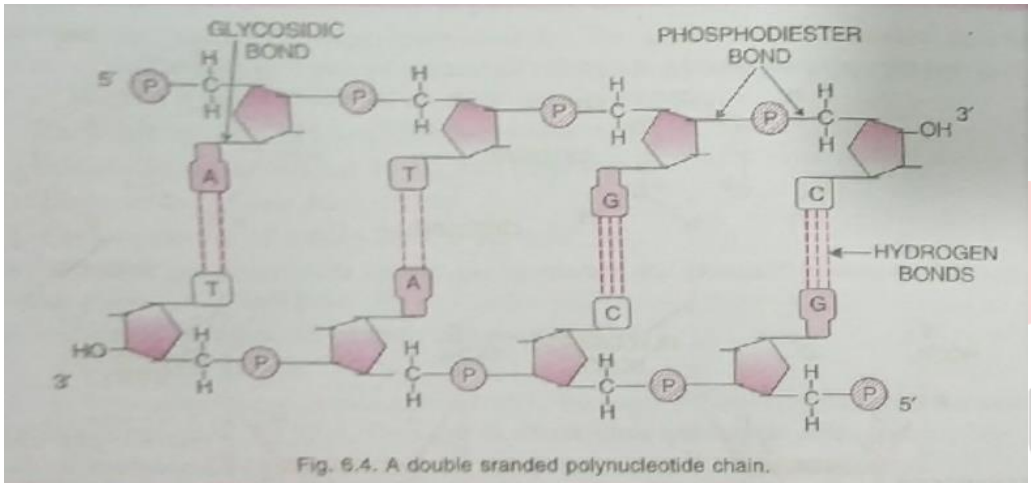


Fig. 6.4. A double sranded polynucleotide chain.

Properties of genetic material

A molecule that can act as genetic material must have the following properties:

- It should be able to generate its replica.
- It should be chemically and structurally stable.
- It should provide the scope of slow change (mutation) that is necessary for evolution.
- It should be able to express itself in form of Mendelian traits.

Structure of polynucleotide chain (RNA):

- In RNA also each nucleotide has three components. Bothe Purine and Pyrimidine are present in RNA.
- Only Uracil replaces thymine in RNA. So the nucleotides in RNA are **A, G, C, U**.
- The sugar in case of RNA is ribose sugars which possess a reactive **additional OH group** on the **2' position**.

DNA versus RNA:

DNA	RNA
It is double stranded.	It is single stranded.
A, T, C, G are the nitrogenous bases.	A, U, C, G are the nitrogenous bases.
Presence of thymine in place of Uracil provides much more stability	Comparatively less stable due to absence of thymine.
The rate of mutation is slow	The rate of mutation is high
DNA is much more capable of storing genetic information.	Presence of addition 2'OH group makes it more reactive and less stable.

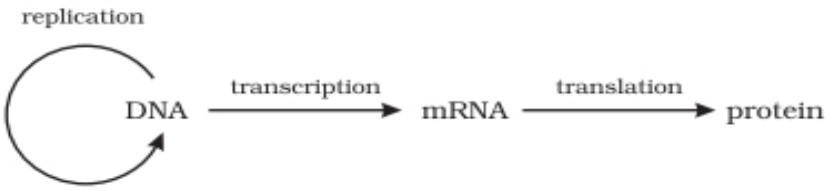
Discoveries related to the structure and composition of DNA:

- Nucleic acids were first isolated in 1869 by **Friedrich Meischer** from the nuclei of pus cell and named them as Nuclein.
- The correct structure of DNA was first worked out by **James Watson** and **Francis Crick** in year 1953. But their **double helix** model was based on two major investigations:
- **Maurice Wilkins and Rosalind Franklin's data of X-ray diffraction** pattern of DNA in year 1953 and **Erwin Chargaff's complementary base pairing rule** in year 1950.
- The X-ray diffraction pattern of DNA determined the **three dimensional structure** and Chargaff's rule summarized that the purines and pyrimidines are always in equal amount *i.e.* **A+G=T+C or A=T & G=C.**

The salient features of DNA:

- DNA has two polynucleotide chains with antiparallel polarity that is from 5' -3' and 3'-5'.
- Backbone of polynucleotide chains made up of alternate sugar phosphate group and the nitrogen base projects inwardly.
- The bases in two strands are paired through hydrogen bond forming base pairs; Adenine forms two H- bonds with Thymine from opposite strand and vice versa. Guanine is bonded with Cytosine with three H-bonds .As a result always a Purine comes opposite to Pyrimidine. This generated approximately distance between the two strands of the helix.
- The two strands are coiled in a helical, right handed fashion. This coiling produces minor and major grooves alternately.
- The pitch length of helix is 3.4nm ($34A^0$) with roughly 10 base pairs in each turn. The average distance between two adjacent base pairs comes to about 0.34nm.
- Planes of adjacent base pairs are stacked over one another. Along with hydrogen bonding, the stacking confers stability to the helical structure.

Central Dogma:



- It is the flow of information from DNA to mRNA (transcription) ten decoding the information present in mRNA in the formation of polypeptide chain.
- The concept of central dogma was proposed by Crick. It proposes unidirectional flow of information from DNA to RNA and then protein.
- But in some viruses the flow of information is in the reverse direction as RNA of these viruses first synthesizes DNA through reverse transcriptase and the process is called reverse transcription.

Packaging of prokaryotic and eukaryotic DNA:

Packaging of DNA helix:

- The average distance between the two adjacent base pairs is 0.34nm ($0.34*10^{-9}$). The number of base pairs in E.coli is $4.6*10^{6}$. The total length of its DNA is 1.36mm.
- Similarly $6.6*10^{9}$ b.p of the two human genomes or any other mammalian cell i.e. Such a long sized DNA got accommodated inside small area in a cell which is possible only through packaging or compaction.

Packaging of Prokaryotic DNA:

- In prokaryotes such as E.coli though they do not have a defined nucleus, the DNA (being negatively) is held with some proteins (positive charge) in a region termed as **Nucleoid.**
- The DNA in nucleoid is organized in large loops held by proteins.

Packaging of Eukaryotic DNA:

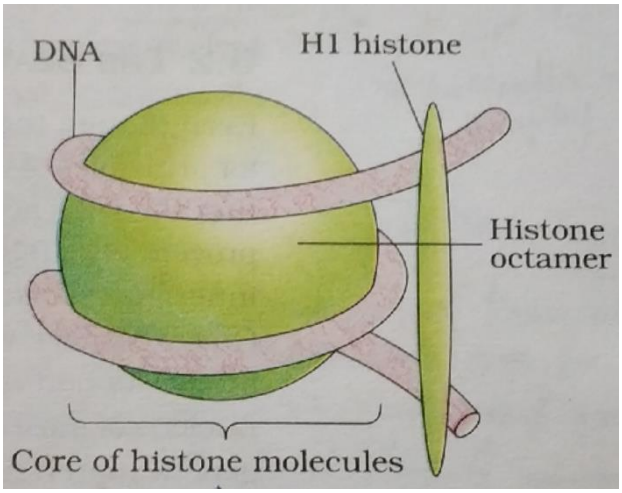


Figure Nucleosome

- In eukaryotes the organization is much more complex. The packaging is carried out by a set of positively charged proteins called as Histones. Histones and DNA are organized to form Nucleosome.

Nucleosome:

- The **negatively charged DNA** is wrapped around the **positively charged histone octamer** form the structure called nucleosome.
- A typical nucleosome contains **200 bp** of DNA helix. It constitutes the repeating unit of a structure in nucleus called **chromatin**. The chromatin is a thread like stained bodies seen in the nucleus.
- The nucleosomes in chromatin are seen as beads on string structure when viewed under electron microscope.
- Small segment of DNA connecting two adjacent nucleosomes is called **linker DNA.**
- The beads on string structure in chromatin is packaged to form chromatin fibers that are further coiled and **condensed at metaphase stage** of cell division to form chromosomes.

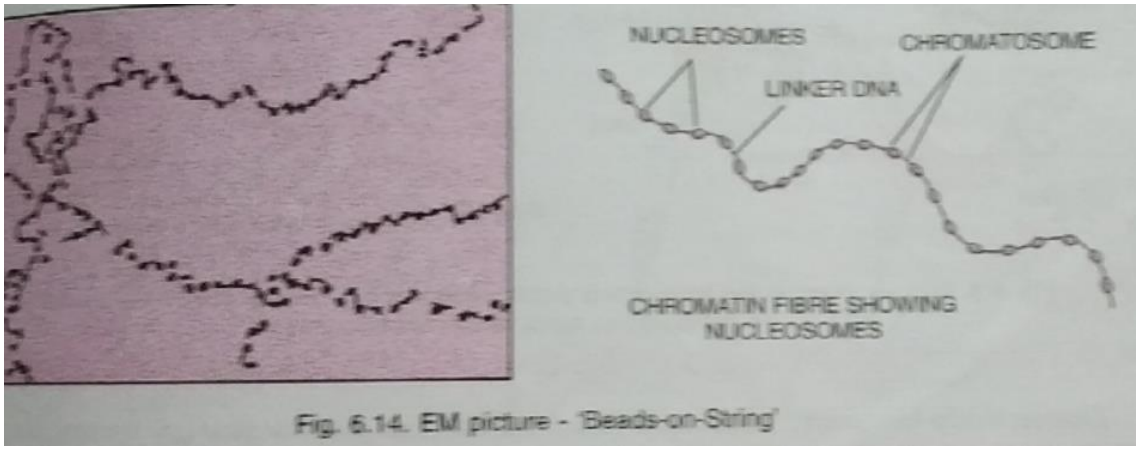


Fig. 6.14. EM picture - 'Beads-on-String'

Histones:

- Histones are positively charged basic proteins which are rich in basic amino acid residues, lysines and arginines. Both the amino acids carry positively charges side chains.
- There are 5 types of histone proteins – **H_1, H_2A, H_2B, H_3, and H_4**. Four of them (H_2A, H_2B, H_3, and H4) occurs in pair to produce histone octamer.
- A fifth type of histone called **H_1** is attached over the linker DNA.

NHC proteins:

- The packaging of chromatins at higher level requires additional set of proteins that collectively called as **non histone chromosomal proteins (NHC).**
- On the basis of **staining behavior** in a typical nucleus, chromatin is of two types: **Euchromatin & Heterochromatin** .

Difference between Euchromatin and Heterochromatin:

Euchromatin	Heterochromatin
It stains lightly.	It stains darkly.
This chromatin is loosely packed.	This chromatin is densely packed.
It is transcriptionally active.	It is transcriptionally inactive
Replication takes place at early S phase as it takes less time to unwind.	Replication takes place at late S phase as it takes longer time to unwind.

RNA world:

- There is a little doubt that early life was RNA centric with very important function being controlled by it. The **first genetic material was RNA**.

- There is now enough evidence to suggest that essential life processes such as **metabolism, translation, splicing evolved around RNA**. The first biocatalysts were RNAs e.g. **ribozyme**. But RNA being a catalyst was reactive and unstable.
- RNAs worked well in early unstable environmental condition. Therefore DNA has evolved from RNA with chemical modifications that make it more stable. DNA being double stranded and having complementary stand further resist changes by evolving a process of repair.

DNA replication: semi conservative mode of DNA replication:

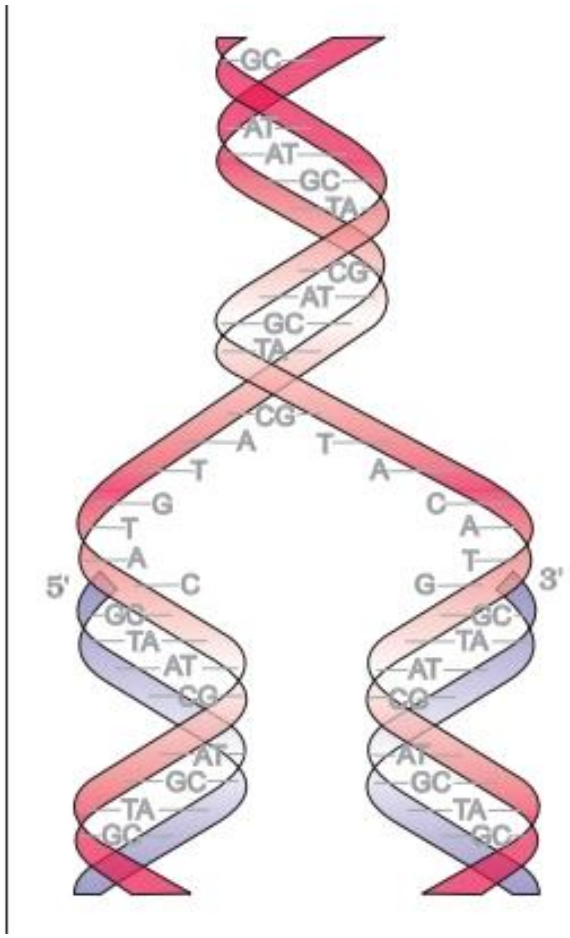


Figure6.6Watson-Crickmodel for semiconservativeDNAreplication

Replication is a duplication process requiring copying from a template. While proposing the double helical structure of DNA, Watson and Crick had immediately proposed a scheme for replication of DNA.

- The scheme suggested that the two strands would separate and act as a template for the synthesis new complementary strands. After the completion of replication, each DNA molecule would have one parental and one newly synthesized strand.
- This scheme is termed as **semi conservative DNA replication.**

Meselson & Stahl's experiment in support of DNA replication:

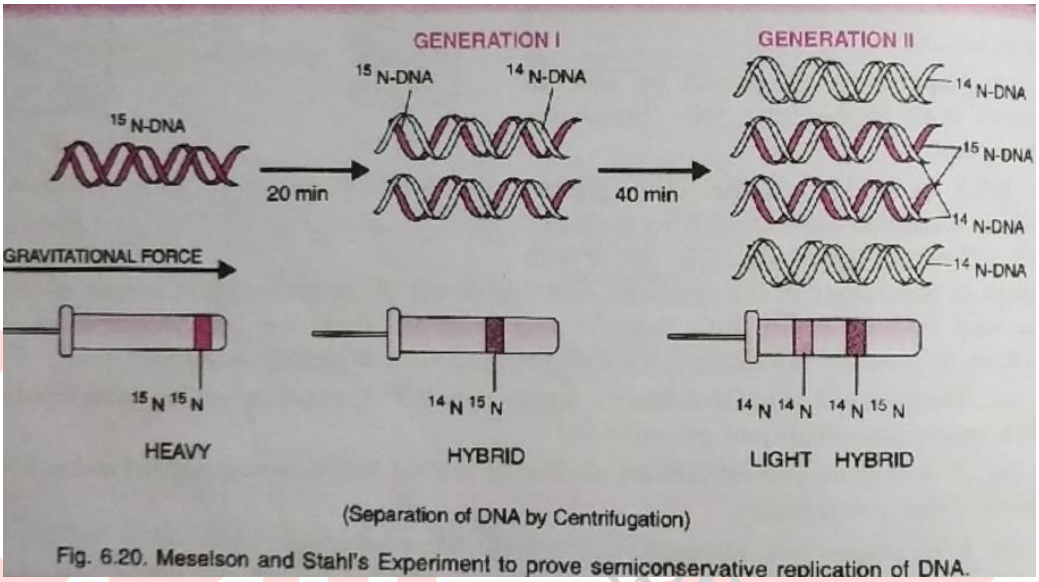


Fig. 6.20. Meselson and Stahl's Experiment to prove semiconservative replication of DNA.

- Matthew Meselson and Franklin Stahl performed the experiment in E.coli and proved that DNA replicates semi conservatively.
- They grew E.coli in a medium containing **$^{15}NH_4Cl$** as the only nitrogen source for many generations. ^{15}N is the heavy isotope of nitrogen and was incorporated into newly synthesized DNA as well as other nitrogen-containing compounds.
- This heavy DNA molecule could be distinguished from normal DNA molecule by centrifugation in **CsCl density gradient method**.
- A dense solution of CsCl, on centrifugation, forms density gradient bands of a solution of lower density at the top that increases gradually towards bottom with highest density.
- Then they transferred the cells into the medium with normal **$^{14}NH_4Cl$** and took samples at various definite time intervals as the cells multiplied and extracted the DNA that remained as double stranded helix. The samples were separated independently on CsCl gradients to measure the densities of DNA.
- Thus the DNA that was extracted from the culture after 1st generation i.e. just after 20 minutes had a hybrid or intermediate density. DNA extracted from the culture another generation i.e. 2nd generation or 40 minutes was composed of equal amounts of this **hybrid DNA (N^{14} N^{15})** and **light DNA (N^{14} N^{14}).**

- Increase in the amount of light DNA and decrease in hybrid DNA amount can be possible due to semi conservative mode of replication.

Taylor's experiment:

- Very similar experiments were performed on ***Vicia faba*** by **Taylor and colleagues** in 1958.
- They used **radioactive thymidine** to detect distribution of newly synthesized DNA in the chromosomes of faba beans.
- The experiments proved that the DNA in chromosomes also replicate semi conservatively.

Enzymes involved in DNA replication and their role:

DNA dependent DNA polymerase: It is the main enzyme which takes part in combining deoxyribose nucleotides to form new DNA strands. This was discovered by Kornberg in 1957. In prokaryotes DNA polymerase-III and in eukaryotes DNA polymerase-delta possesses replication activity.

DNA ligase: This enzymes joins newly synthesized fragments of DNA.

Helicase: This enzyme helps in unwinding the DNA helix during replication.

Topoisomerase: This enzyme helps in making a swivel cut above the replication fork to prevent knotting of DNA.

Process of DNA Replication:

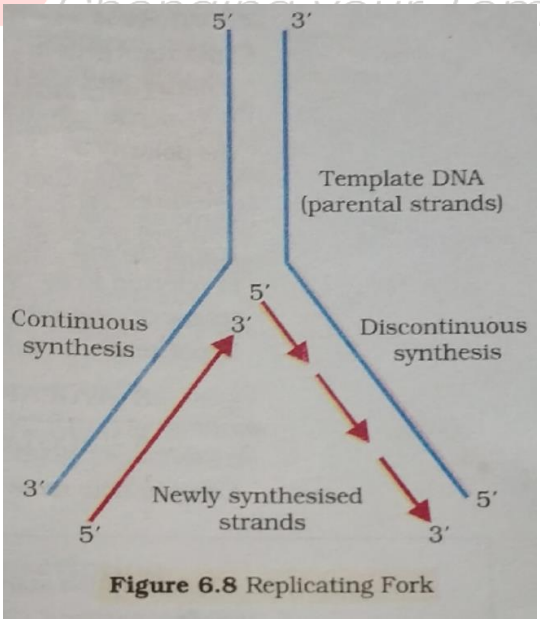


Figure 6.8 Replicating Fork

- The main enzyme involved in replication is DNA dependent DNA polymerase. These enzymes are highly efficient enzymes as they have to catalyse polymerization of a large number of nucleotides in a very short time.

Rate of DNA replication:

- E.coli which has only 4.6 *10^6 bp completes the process of replication within **38 minutes**; that means the average rate of polymerization has to be approximately **2000 bp per second.**
- The catalyst not only adds the nucleotide fast but also completes the reaction with **high degree of accuracy**. Any mistake during the process would result in **mutation**.
- In eukaryotes the replication takes place in S –phase of cell cycle. Both replication and cell cycle should be highly coordinated. A failure in cell division after DNA replication results in **polyploidy.**
- DNA replication is a complex multistep process that requires a number of enzymes and protein factors:

Origin of Replication:

- The DNA polymerases on their own cannot initiate the process of replication and also it does not initiate randomly at any place in DNA.
- There is a definite region in prokaryotes (E.coli) where the replication originates such regions are termed as **origin of replication (ORI).** It is because of the requirement of the origin of replication that a piece of DNA if needed to be propagated during **r-DNA technology** requires a **vector**. The vectors provide the origin of replication.
- In eukaryotes there is a number of origins of replications called as **replicon.**

dNTPs:

- The deoxyribonucleotides (deoxyribo nucleoside triphosphate) occur freely inside nucleoplasm.
- It serves dual role. It acts as the substrate for polymerization. As replication is energetically very expensive so dNTPs provide energy for polymerization (the two terminal phosphates in a dNTPs are high energy phosphate bonds).

Replication fork:

- The replication occurs at a small opening of the DNA helix called as **Replication fork** which is **Y shaped** in structure.
- The enzyme **helicase** unwinds the DNA at ORI site by destroying the H-bond. This exposes the single strand for further enzymatic activity. The template strand with 3'-5'

polarity synthesizes daughter DNA **continuously** and the strand with polarity 5'-3' polarity synthesizes daughter DNA **discontinuously**.

- The replicated strand which grows continuously is called **leading strand** and the one which grows discontinuously called **lagging strand.**
- During unwinding a super coiling gets developed on the end of DNA which is released by **topoisomerase**.
- The discontinuous strand synthesized in fragments and they are known as Okazaki fragments. The fragments are later joined by **DNA ligase**.

Transcription:

- The process of **copying genetic information from one strand of the DNA into RNA** is termed as transcription. The process follows complementarity like that of replication. In transcription only a segment of DNA and only one of the strands is copied into RNA.
- The process is mediated by enzyme **RNA polymerase**.
- There are two reasons **why both the strands are not copied during transcription**.

(1) If both strands act as template they would code for RNA molecule with different sequences and in turn if will code for proteins, the sequence of amino acid will be different. This would complicate the genetic information transfer machinery.

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

Transcription unit:

Transcription unit in DNA molecule comprises mainly of three regions:

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

- The strand of the DNA with polarity from **3'-5'** that directs the synthesis of RNA is called **template strand or antisense strand**.
- The other strand of the DNA with polarity from **5'-3'** complementary to the template strand is called as **coding strand or sense strand**.
- During this process Uracil is replace by Thymine complementary to Adenine.
- The coding strand has same polarity and same nucleotides (except 'U' in place of 'T') as the RNA transcript.

(i) **Promoter:** It is the region on DNA molecule to which RNA polymerase binds and initiates the process of transcription. It is located on the **upstream (to the left) of the structural gene i.e. towards 5' end of the coding strand, 3' end of the template strand**.

(ii) **Structural gene:** It is the area of template strand that is involved in transcription or formation of RNA..

(iii) **Terminator:** The region on the DNA molecule is **present downstream (to the right) of the structural gene, i.e., towards 3' end of coding strand, 5'end of template strand.** It usually defines the end of the process transcription.

Transcription unit and gene:

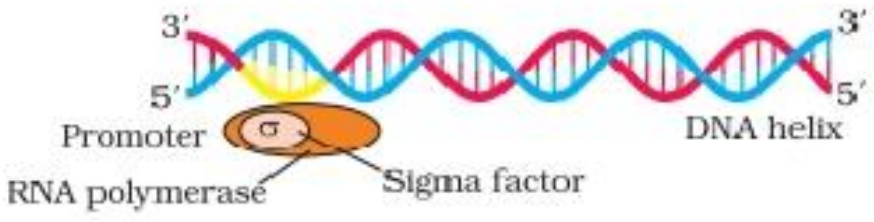
- DNA not only specifies the synthesis of polypeptides through mRNAs but also synthesis other RNAs. Therefore the term gene is now-a-days being replaced by Cistron. It is defined as a functional unit (structural gene) of gene, it is a segment of DNA coding for a polypeptide.
- Structural gene is of two types: Monocistronic and Polycistronic. **Monocistronic** gene carries information for synthesis of only one polypeptide and is mostly found in eukaryotes**. Polycistronic** gene information for synthesis of more than one polypeptide and is mostly found in prokaryotes
- The Monocistronic structural genes have interrupted coding sequences i.e. the genes in eukaryotes are **split genes**.
- The coding or expressed sequences are known as **Exons** and the non coding or intervening sequences are known as **Introns.**

Types of RNAs and process of transcription:

- In prokaryotes there are three major types of RNAs: mRNA, tRNA and rRNA. All three RNAs are needed to synthesize a protein in a cell.
- The mRNA provides the template, tRNA brings the amino acid and reads the genetic code, and rRNAs play structural and catalytic role during translation.

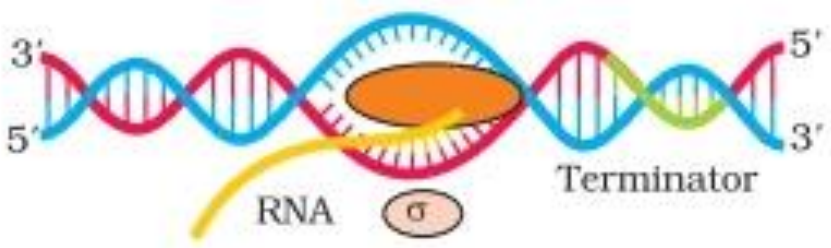
Transcription in prokaryotes: It occurs in cytoplasm with the help RNA polymerase that catalyses transcription of all RNAs. It involves three steps:

Initiation:



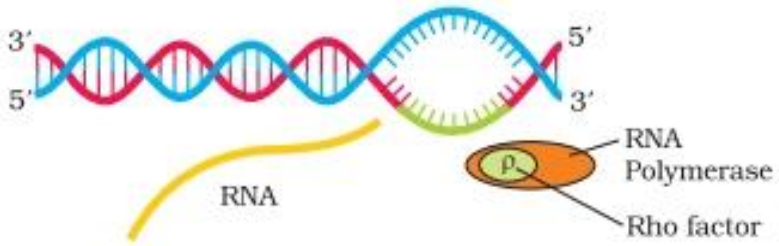
RNA polymerase reaches the promoter region and bids to it. RNA polymerase has a factor called as **initiator factor** or **sigma factor** (**σ**). The enzyme recognizes the promoter by the sigma factor and then initiates the process. It uses the dNTPs as substrates and polymerases in a template dependent fashion from 5' -3' direction.

Elongation:



After synthesis of small chain of nascent RNA the **σ factor leaves the complex and the RNA polymerase continues the process. As the enzyme moves along the DNA, RNA chain becomes longer.**

Termination:

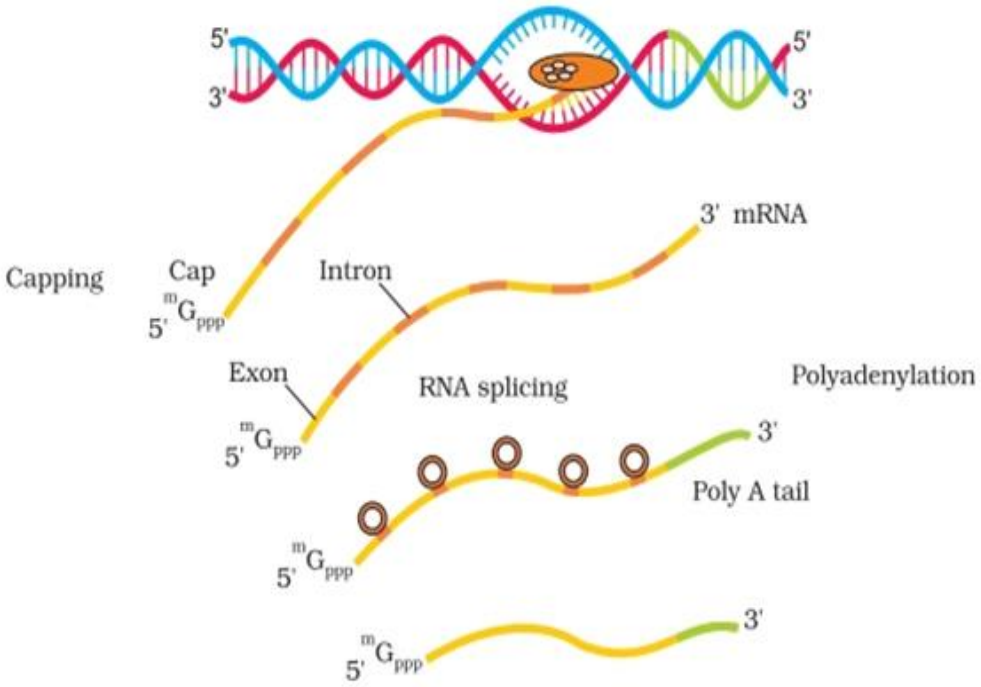


When the RNA polymerase reaches the terminator region a specific chain **terminating protein** called **rho factor** (**ρ**) stops the synthesis of RNA chain. It separates RNA Polymerases as well as the newly formed RNA strand. The RNA and RNA polymerases fall off and it results in the termination of transcription.

- In bacteria mRNA does not require any processing to become active and both transcription and translation take place in the same compartments as there is no separation of cytosol and nucleus.

Transcription in eukaryotes: There are at least 3polymerases in the nucleus which show a clear cut division of labour. The two additional complexities in eukaryotes:

- RNA pol-I transcribes rRNAs (28S, 18S, and5.8S), RNA pol-II transcribes hnRNA (precursor of mRNA), and RNA pol-III transcribes tRNA, 5sRNA and snRNA (small nuclear RNAs).
- The primary transcript contains both exons and Introns and it is subjected to processes like capping, splicing, tailing (**post transcriptional change**).



Messenger RNA

Figure:process of Transcription in Eukaryotes.

Capping:

- The hnRNA undergoes a change as an unusual nucleotide 7 methyl guanosine is added at 5' end. The methylated cap protects the mRNA from degradation by nuclease.

Splicing:

- It is the process of removal of Introns through cutting and joining of exons in a defined order.

Tailing:

- At the 3'end of hnRNA about 200-300 adenine residues are added and the process is called as polyadenylation which protects the Transcript from enzymatic activities.

Genetic codon:

- There is an intimate connection between genes and synthesis of polypeptide. **Genetic code** refers to the relationship between the sequence of nucleotides on mRNA and the sequence of amino acids in the polypeptide.
- DNA is made up of only four types of nucleotides and the number of amino acids is 20. Arrangement of nucleotides is connected with the synthesis of proteins by influencing the incorporation of amino acids between them.
- It was therefore, hypothesized by **George Gamow,** a physicist, that triplet code is co operative. He argued that since there are only 4 bases and if they have to code for 20amino acids, the code should constitute a combination of bases.
- He suggested that in order to code for all the 20 amino acids, the code should be made up of three nucleotides.
- This was a very bold proposition, because a permutation combination of 4^3 would generate 64 codons.
- The chemical method developed by Har Govind Khorana proved that the codon was triplet.
- He introduced the instrumental method in synthesizing RNA molecules with defined combinations of bases (homopolymers and copolymers).
- Marshell Nirenberg's cell free system for protein synthesis finally helped the code to be deciphered severo Ochoa enzyme (polynucleotide phosphorylase) was also helpful in polymerizing RNA with defined sequences in a template dependent manner.

The salient features of genetic codon:

Triplet codon: Three adjacent nitrogenous bases constitute a codon which specifies the placement of one amino acid in a polypeptide.

Start signal: Polypeptide synthesis is signaled by a initiation codon-AUG. This codon possesses dual function as it acts as start codon and it also codes for amino acid methionine.

Stop codon: Polypeptide chain termination is signaled by three termination codon-UAA, UAG, UGA. The do not specify for any amino acid and are hence known as nonsense codon.

Universal codon: The genetic codon is applicable universally i.e. a codon specifies the same amino acid from virus to a tree or human beings.

Nonambiguous codon: One codon specifies only one amino acid and not any other.

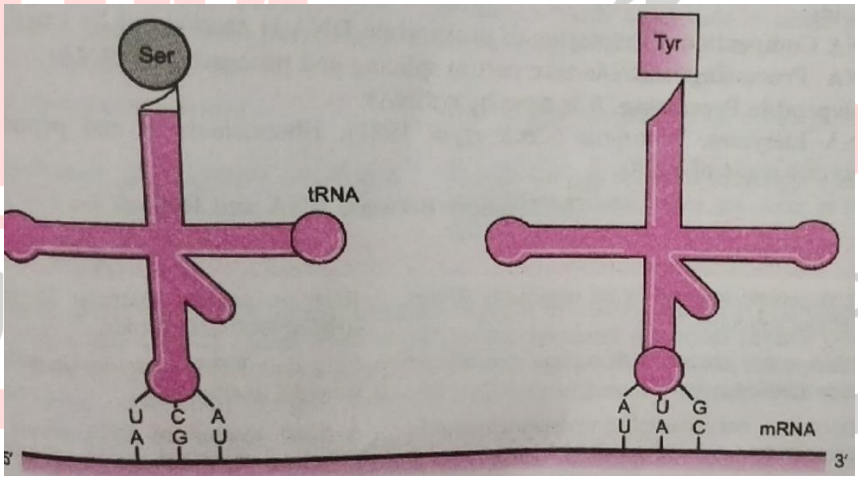
Degeneracy of code: As there are 64 codons and only 20 amino acids, the incorporation of some amino acids must be influenced by more than one codon.

Comma less: The genetic codon is continuous and does not possess any punctuation mark after the triplets.

Mutation and Genetic code:

- The relationship between genes and DNA are best understood by mutation studies. Effects of large deletion and rearrangements in a segment of DNA are easy to comprehend. It may result in loss or gain of a gene and so a function.
- **Point mutation**: It is the change of single base pair in the gene. e.g. Sickle cell anemia. In this case a single base pair change replaced the amino acid sequence from the polypeptide chain and causes the disease.
- **Frame shift mutation**: It is the type of mutation where addition/insertion or deletion of one or two bases changes the reading frame from the site of mutation, resulting in a protein with a different set of amino acid.

t-RNA –the adaptor molecule:



- Francis Crick was clear that there has a mechanism to read the code and to link it with amino acids. He postulated that there is a adapter molecule which on one hand will read the codon and on the other hand will bind to specific amino acids. Hence the tRNA is called as sRNA (soluble RNA).The secondary structure of tRNA looks like clover-leaf structure but actually it is a compact molecule which looks like inverted L.,
- t-RNA possess four different loops like: Anticodon loop, T C loop, DHU loop and Amino acid acceptor loop.
- Anticodon loop: This loop possess the bases complementary to that of codon upon the mRNA.
- Amino acid acceptor loop: This loop binds to a specific amino acid coded by the codon.
- T C loop: This loop upon the tRNA provides binding site for ribosome.
- DHU loop: This loop upon the tRNA holds the enzyme amino acyl tRNA synthetase.

- For the initiation of translation there is a specific tRNA called as initiator tRNA and there is no tRNAs for stop codons.

Translation:

- Translation refers to the process of polymerization of amino acids to form a polypeptide as per the codon of mRNA. The order and the sequence of the amino acids are defined by the sequence of bases of mRNA and the information given by bases of DNA.

Translation machinery:

- It consists of tRNA, amino acids, ribosome and mRNA. mRNA functions as a template having genetic information.

Ribosome:

- Ribosomes are known as protein factories as protein synthesis occurs over the ribosomes. It consists of structural RNAs and 80 different types of proteins. Each ribosome has equal parts, **small and large**. The Smaller and larger subunit fits over each other and leaves a tunnel for mRNA. The subunits only come together during the time of protein formation and this association requires Mg^{2+} ion.
- Ribosome possesses three different reactive parts –**A site** (aminoacyl site), **P site** (peptidyl site), **E site** (exit site).

t-RNA:

- They are transfer or soluble RNAs which pick up the particular amino acids from cellular pool.
- This process is called charging. The enzyme aminoacyl tRNA synthetase helps in combining the amino acid to its particular tRNA.

Amino acid:

- There are many different amino acids and amides which constitute the building blocks or monomers of the proteins.

Translation mechanism:

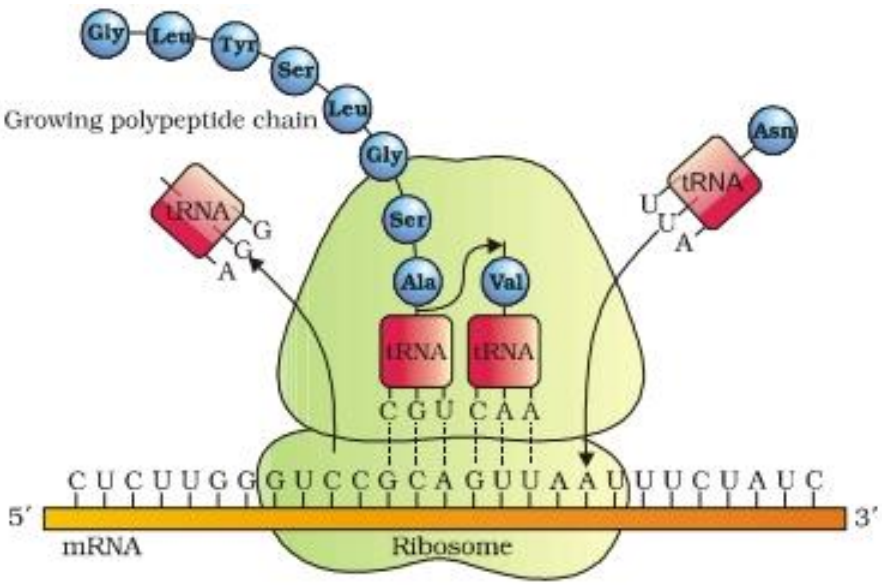


Figure :Translation

Charging of tRNA or aminoacylation of tRNA:

- In the presence of **ATP an amino acids combines with its specific aminoacyl tRNA synthetase** and produces substrate enzyme complex.
- The complex then reacts with the specific tRNA to form **aminoacyl-tRNA-complex**. Then enzyme and AMP are released.

Initiation:

- mRNA attaches itself upon the smaller subunit of ribosomes in such a way that the **codon of mRNA 'AUG' comes and lie at P site of ribosome**.
- Then the initiator tRNA gets charged with methionine in the cytoplasm. By that time the larger subunit of ribosome binds to the complex.
- The charged tRNA comes and attaches to the P site of the ribosome making mRNA-tRNA complex at P site but keeps the A site exposed.

Elongation:

- An aminoacyl tRNA complex teaches the A-site and attaches to mRNA codon next to the initiation codon with the help of its Anticodon. The subsequent amino acids being close enough to each other **forms peptide bond**.
- The ribosome acts as catalyst (**23sRNA in bacteria is the enzyme-ribozyme**) for the formation of peptide bond. This enzyme is called as **peptidyl transferase**.

Termination:

- Polypeptide synthesis is terminated when a nonsense codon of mRNA reaches the A site. There are three stop codon like **UAG, UGA, and UAA**.
- These **codons are not recognized by any t-RNAs**. Therefore no more aminoacyl t-RNAs reaches the A site.
- At the end, a **release factor** binds to the stop codon terminating translation and releasing the complete polypeptide from the ribosome.

A translation nit in mRNA is the sequence of RNA that is flanked by the start codon (AUG) and the stop codon and codes for a polypeptide.

An mRNA also has some additional sequences that are not translated are as referred as **untranslated region (UTR)**. These UTRs are present at both5' and 3' ends of mRNA and are required for efficient translation process.

Regulation of gene expression:

- Gene expression is the mechanism at the molecular level by which a gene is able to express itself in the phenotype of an organism.
- In prokaryotes the gene expression is controlled by the initiation of transcription.

In eukaryotes the regulation can be regulated at four different levels:

- Transcription level (formation of primary transcript)
- Processing level (splicing)
- Transport of mRNA from nucleus to cytoplasm
- Translational level.

Operon concept:

Two French scientists **Jacob** and **Monad** in 1961 proposed a model of gene regulation on the metabolism of bacterium *E.coli*. The model was known as operon model.

Operon is co ordinate group of genes such as structural gene, regulator gene, promoter gene, operator gene etc which functions together and regulate a metabolic pathway as a unit.

Structural gene: Transcribes mRNA for polypeptide synthesis.

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

Promoter gene: Provides attachment site for RNA polymerase.

Regulator gene: It synthesizes a regulator protein which can act positively as activator and negatively as repressor. Overall it controls the activity of operator gene.

Lac- Operon:

- The lac-operon refers to lactose and in *E.coli* breakdown of lactose requires three enzymes. These enzymes are synthesized together in a co-ordinate manner by functional unit of DNA.

Structural genes:

- Three structural genes are:
- ***Lac z*:** The z gene codes for **β- galactosidase** which is primarily responsible for the hydrolysis of the disaccharide, lactose into its monomeric units galactose and glucose.
- ***Lac y*:** The y gene codes for **permease**, which increases permeability of the cell to β-galactosidase.
- ***Lac a***: The a gene codes for **transacetylase** which can transfer acetyl group to β-galactosidase.

Regulator gene:

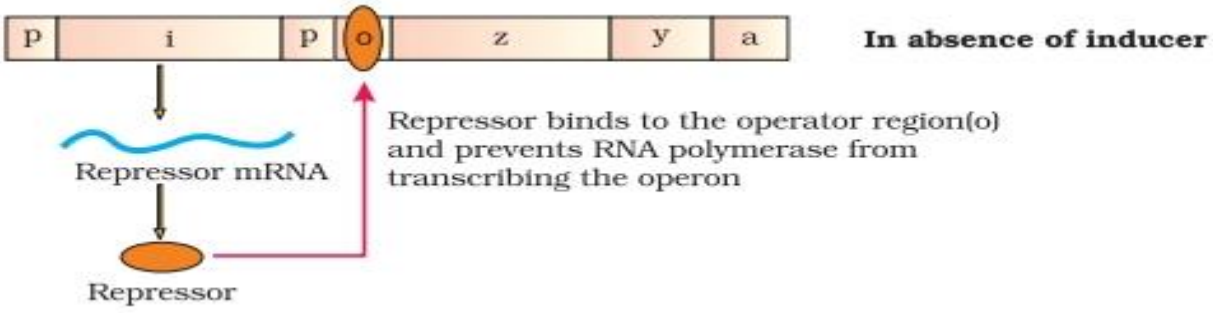
- The gene codes for a protein known as repressor protein, it is synthesizes all the time from **i-gene** and it is functional always.

Promoter gene: The gene possesses site for RNA polymerase.

Operator site: It interacts with a protein molecule or regulator molecule, which prevents the transcription of structural genes.

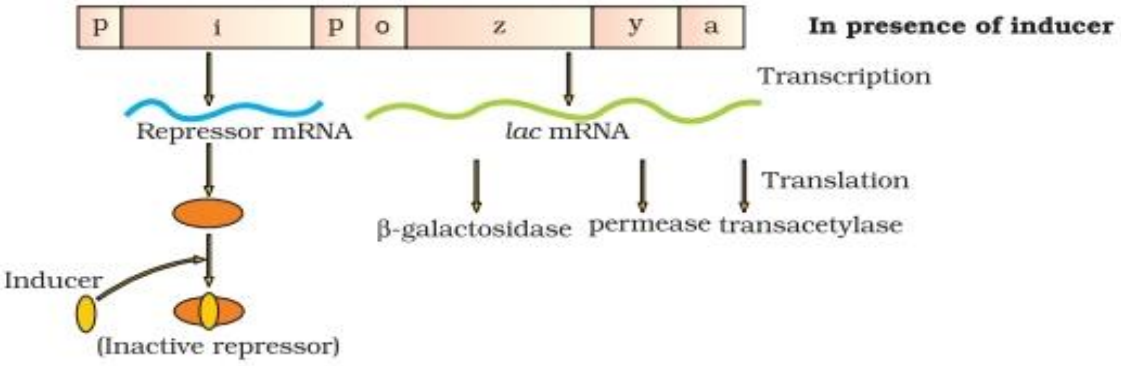
- The operon is switched off when repressor protein produced by regulatory or inhibitor gene binds to the operator gene and when lactose is present in the medium the operon is switched on as repressor gets inactivated.

Switch off:



- When lactose is absent in the medium then the repressor protein synthesized by regulatory gene (constitutive) binds to the operator gene.
- RNA polymerase binds to the promoter but cannot move forward and is unable to read the structural gene. So there will be no transcription and no translation.

Switch on:



If lactose is provided in the growth medium, it is transported inside to the bacterial cell through the action of permease.

- A very low level of expression of lac-operon has to be present in the cell all the time; otherwise lactose cannot enter inside the cell.
- In the presence of inducer lactose or allolactose, the repressor is inactivated by interaction with inducer. This allows the RNA polymerase access to the promoter and transcription proceeds. Then the structural gene synthesizes the mRNA and respective proteins are formed. The proteins modified into enzymes to carry out further function.
- Regulation of lac- operon by repressor is referred to as **negative control**.
- Lac-operon is under control of **positive regulation for glucose**.

Human Genome Project (HGP):

- Each individual has an identity that is due to one's genetic make-up. No two individuals are similar because they differ in their genetic make-up. Differences in genetic make-up are due to differences in nucleotide sequences of their DNA.
- It was therefore always an ambition of the scientists to map human genome. So in the year **1990, U.S Department of Energy and National Institute of Health** embarked and coordinated on the project of sequencing human genome called HGP.
- **Welcome Trust (UK)** joined the project as a major partner. Later on Japan, France, Germany, China and some other countries joined.
- The HGP was launched on **October 1, 1990** and was completed in **2003**. Thus it was a **13 year project**. John Craig Venter was one of the first scientists to sequence human genome.
- HGP has been called as a **mega project** as:
- Due to its huge cost which estimated near about **9 billion US dollar**.
- There are a large number of base pairs to be identified i.e.**$3*10^9$ b.p**.
- A large number of scientists, technicians and supporting staffs are involved.
- Storage of data generated which requires some **3300 books, each book with 1000 pages and each page having 1000 typed letters.**
- However, high speed computational devices for storage, retrieval and analysis of data was made this easier due to introduction of bioinformatics.
- **Bioinformatics** is the branch of science in which computer hardware and software technologies are developed and use to gather, store, analyze and disseminate information, biological data and images.

Goals of HGP:

- Some major goals of HGP are to:
- Identify all the genes (approximately 20000-25000) in human DNA.
- To determine the sequence of the 3 billion chemical base pairs that makes up human DNA.
- To store this information in databases.
- To improve tools for data analysis.
- Transfer related technologies to other sectors, such as industries, medicines etc.
- To solve ethical legal and social issues (**ELSI**) that may arise from the project.

Sequencing of model organisms:

Non human organisms DNA can lead to an understanding of their natural capabilities that can be applied for betterment.

Organisms	Base pairs	No.of genes
E.coli	4.7 million	4000
S.cerevisiae	12 million	6000
C.elegans	97 million	18000
D.melanogaster	180 million	13000
Arabidopsis	130 million	25000
O.sativa	430 million	32000-50000

Methodologies of HGP:

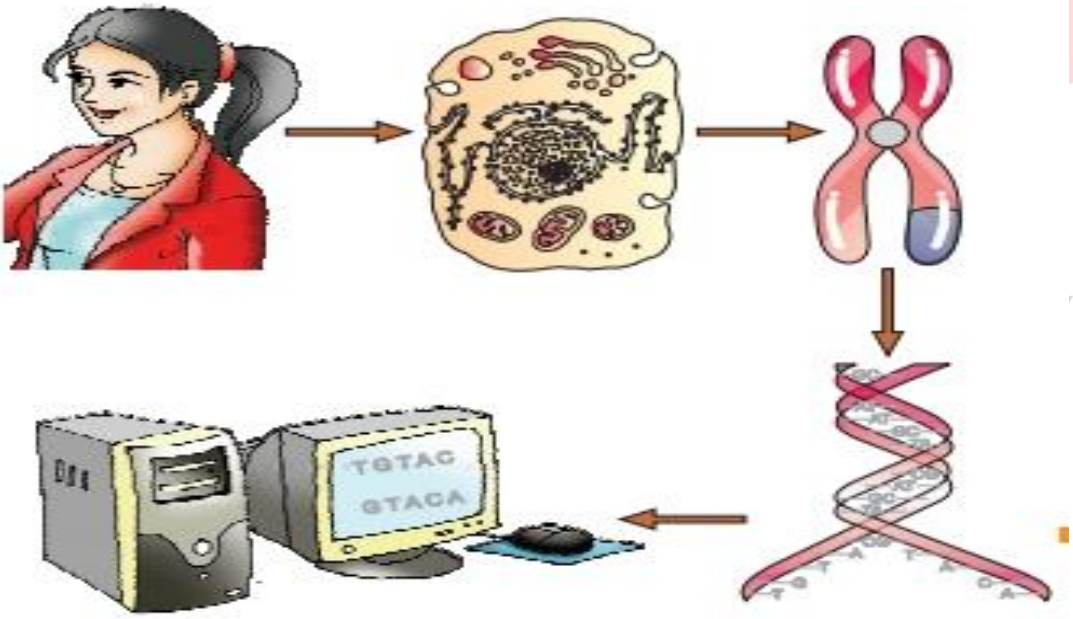


Figure Arepresentativediagramofhuman genomeproject

- There are two approaches for sequencing HGP- (1) Expressed sequence tags (2) Sequence annotation.

Expressed sequence tags:

- It was a blind approach for simple sequencing of whole genome. All the genes that are **expressed as RNAs** were tagged and sequenced.

Sequence annotation:

- The whole DNA was isolated from the cell and broken into fragments by restriction enzymes.
- Those fragments were inserted into specific vectors like **YAC** (Yeast Artificial Chromosome) and **BAC** (Bacterial Artificial chromosome).
- The fragments were cloned in specific hosts or by PCR cloning was conducted. The fragments were sequenced using automated DNA sequencers, the method developed by **Frederick Sanger**.
- The sequences were then arranged on the basis of some overlapping regions. It necessitated the generation of overlapping fragments for sequencing.
- Computer based programmes were used to align the sequences.
- The sequences were then annotated and assigned to different chromosomes.
- All the human chromosomes have been sequenced,22 autosomes, X and Y chromosomes. Chromosome 1was to be sequenced in May, 2006.

Salient features of human genome:

- Human genome contains 3164.7 million nucleotide bases.
- The average gene consists of 3000 bases, but sizes vary greatly. The total number of genes is about 30,000, much lower than previous estimates of 80,000 to 140,000 genes. All most all (99.9%) nucleotide bases are exactly the same in all people.
- The functions are unknown for over 50%of the discovered genes.
- Less than 2% of the genome codes for proteins.
- Repeated sequences make up very large portion of the human genome.
- Repeated sequences are stretches of DNA sequences that are repeated many times, sometimes 100 to 1000 times. They are thought to have no direct coding functions, but they shed light on chromosome structure, dynamics and evolution.
- Chromosome 1 has most genes (2968), and the Y has fewest (231).
- Scientists have identified about 1.4 million locations where single base DNA differences (SNPs-single nucleotide polymorphism pronounced as snips) occur in humans. The information promises to revolutionize the processes of finding chromosomal location for disease associated sequences and tracing human history.

Application and future challenges of HGP:

- More than 1200 genes are responsible for common human cardiovascular, endocrine, neurological diseases. The HGP is providing the scope for studying that group of genes and for finding out solution for the same.
- Efforts are on progress to determine the genes that will change the cancerous cells to normal.
- It is helping the industries to produce protein, medicines in large sector for public benefits

DNA fingerprinting:

- As the finger, palm prints of every individual is unique so as the also the DNA. DNA fingerprinting is a technique of determining nucleotide sequences of certain areas of DNA which are unique to each individual. Each person has a unique DNA fingerprint.

Principle:

- There is 99.9% of similarity in every individual genome with only 0.1% of difference and this difference provides uniqueness to each individual.
- Human genomes possess numerous small non coding but inheritable sequences of bases which are repeated many times.
- These sequences occur near telomere, centromere and Y chromosomes or heterochromatinised area.
- The area with small sequence of bases repeated several times is called repetitive DNA and is also known as Satellite DNA.
- The satellite DNA is separated from the bulk DNA by density gradient centrifugation. The bulk DNA during this method forms major peak whereas satellite DNA forms small peaks.
- In satellite DNA the repetition of bases is tandem and depending upon the length, base composition and numbers of tandemly repeated units these satellite DNAs are categorized into microsatellite and minisatellite.
- These satellite shows polymorphism and the term polymorphism is used when a variant at a locus is present with a frequency of more than 0.01 populations.
- Variation occurs due to mutation. While mutation in genes produces alleles with different expressions, mutations in non coding repetitive DNA have no immediate impact.
- These mutations in the non coding sequences have piled up with time and for the basis of DNA polymorphism (variation at genetic level due to mutation).
- This DNA polymorphism is the basis of genetic mapping and DNA fingerprinting.

- The minisatelites include VNTR (variable number of tandem repeats) which is used in the DNA fingerprinting technique as probe.
- The size of the VNTR varies from 0.1 to 20kb.
- The sensitivity of the technique has been increased by use of PCR.

Technique:

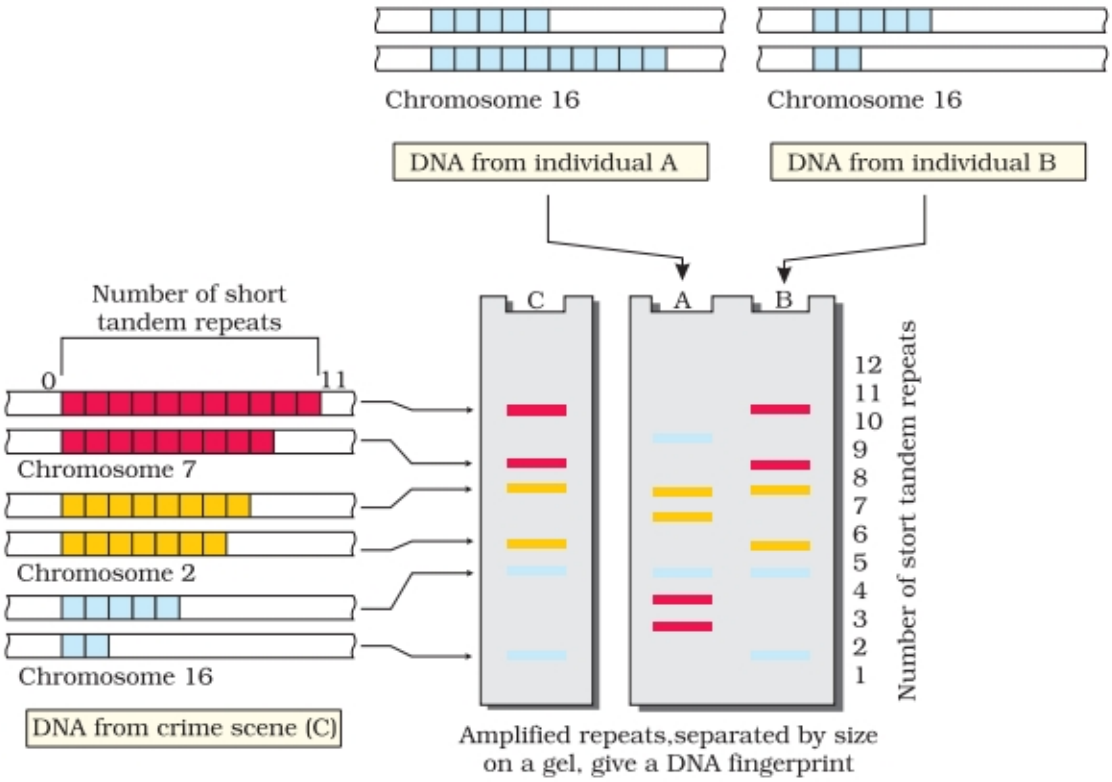


Figure- SchematicrepresentationofDNAfingerprinting:Fewrepresentativechromosomes havebeenshowntocontaindifferentcopynumberofVNTR.Forthesakeofunderstandingdifferentcolourschemeshavebeenusedtotracetheoriginofeachbandinthegel.Thetwoalleles(paternalandmaternal)ofachromosomealsocontaindifferentcopynumbersofVNTR.Itisclearthatthebandingpatternof DNAfromcrimescenematcheswithindividualB,andnotwithA.

- The DNA is separated from the nuclei of sample collected from crime spot.
- The DNA molecule is digested by restriction endonucleases.
- The fragments are separated by gel-electrophoresis.
- The separated DNA fragments are transferred to synthetic membrane such as nylon membrane or nitrocellulose membrane. The technique is known as blotting.
- The separated DNA fragments are now hybridized by VNTR probes.
- The hybridized fragments are detected by autoradiography under X-ray film.

Application of DNA-fingerprinting:

- It can identify the real genetic mother, father and off springs
- It is very helpful in detection of crime and legal pursuits.
- It helps to distinguish one individual from another except monozygotic twins.
- It has much wider application, such as in determining population and genetic diversities. It can identify the racial groups, their origin, historical migration and invasion

Rice genome project:

- An international Rice Genome sequencing project (**IRGSP**) was conceived in September 1997 at a workshop held in conjunction with the International Symposium on Plant Molecular biology in Singapore. Some 32 institution of 10 countries participated in it. The Indian participants were IARI and university of Delhi.
- The IRGSP sequenced the genome of the single inbreed cultivar *Oryza sativa ssp. Japonica* and adopted the tools like **BAC (Bacterial Artificial Chromosome) and PAC (P1-Phage derived Artificial Chromosome**).
- The map based draft sequence was released in December 2002.
- Rice has the smallest genome among the major cereals with only **389 million base pair.**
- The number of genes is, however, high some **37544**. Many of them occur in clustered gene families.
- Nearly 0.4% of nuclear genes contain organelle DNA segments.
- Some 35% of transposons are present.
- **18.9 million SNPs** and 80,127 polymorphic sites occur in genome.
- There are 5 varietal groups in Rice: **Basmati, India, Aus, Tropical japonica and Temperate japonica**

IMPORTANT TERMS

Transcription: copying of genetic material from DNA to RNA.

Purine: Dicyclic nitrogenous bases which includes adenine and guanine.

Pyrimidine: 6 membered cyclic ringed nitrogenous bases which includes cytosine, thymine and Uracil.

Nucleotide: It consists of pentose sugar, nitrogenous base and phosphate group.

Nucleoside: It consists of nitrogenous base and pentose sugar.

DNA ligase: It joins the discontinuously synthesized okazaki fragments.

Cistron: The DNA sequence which synthesizes structural protein.

Exon: The coding sequences in cistron.

Intron: The non coding sequences in cistron.

Splicing: The process in which Introns are removed.

Capping: The process in which an unusual nucleotide is added at 5' end of nascent mRNA.

Tailing: The process in which 3' end of mRNA is polyadenylated.

Codon: It is formed by 3 nitrogenous bases on mRNA that codes for an amino acid.

Anticodon: It has bases complementary to the code.

Polymorphism: Variation at genetic level is called polymorphism.

Repetitive DNA: Small stretches of DNA present at telomeric and centromeric region of chromosomes.

HGP: Human Genome Project.

ESTs: Expressed sequence tags.

BAC: Bacterial artificial chromosomes.

YAC: Yeast artificial chromosomes.

SNPs: Single nucleotide polymorphism.

VNTR: Variable number of tandem repeats.

UTR: Untranslated regions

snRNA: small nuclear RNA.

hnRNA: Heterogeneous nuclear RNA.

NHC: Non histone chromosome.