Chapter 6

Molecular Basis of Inheritance

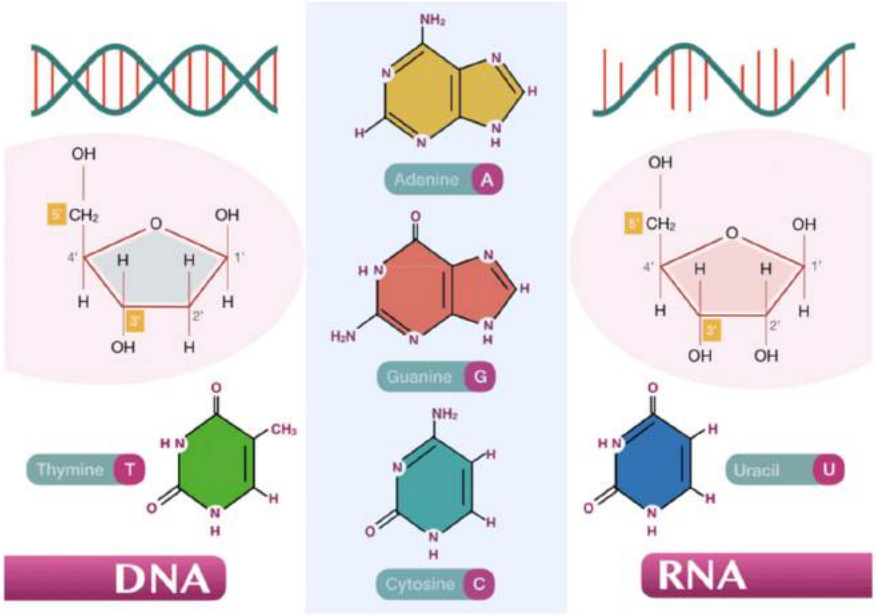
DNA: Structure of Polynucleotide Chain

What are Nucleic Acids?

Nucleic acids are long-chain polymeric molecules, the monomer (the repeating unit) is known as the nucleotides and hence sometimes nucleic acids are referred to as polynucleotides.
Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are two major types of nucleic acids. DNA and RNA are responsible for the inheritance and transmission of specific characteristics from one generation to the other.

Structure of Nucleic Acids

- Nucleic acids are the **biomolecules** that play a very important role in the process of Inheritance. They store all our **genetic information** that we pass down to future generations and they are able to perform their functions, due to the shape and structure they form.
- Just like in all biomolecules, nucleic acids are also made up of **repeating monomers**. Here these repeating units or monomers have a special name, nucleotides. These are the building blocks of nucleic acids.
- Two types of nucleic acids exist **DNA** (Deoxyribo Nucleic Acid) and **RNA** (Ribo Nucleic Acid)

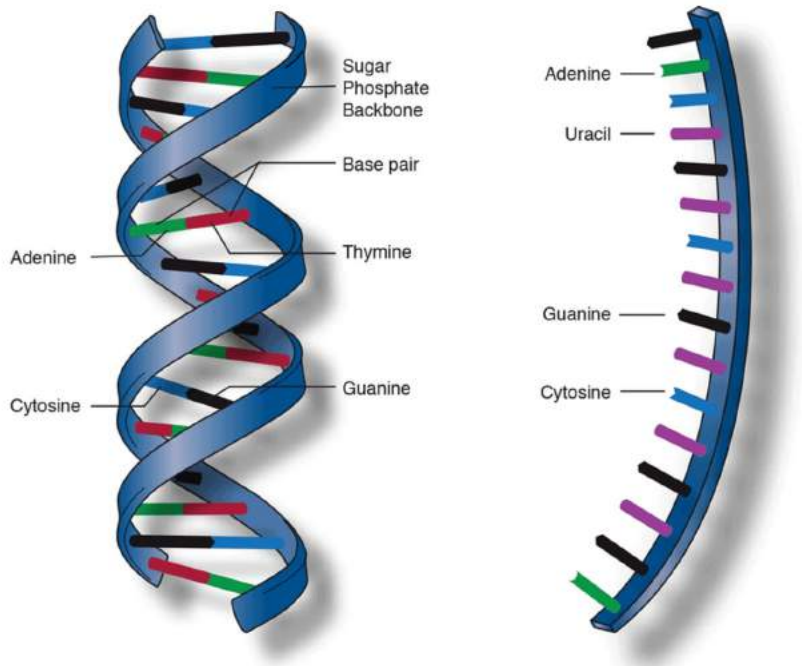


- **DNA** has a double-stranded structure. It is a polynucleotide whose monomer units are deoxyribonucleotide. The length of DNA is determined by the number of nucleotides in it.
- **RNA**, on the other hand, has a single-stranded structure. It is also a polymer whose monomer units are ribonucleotide.

Molecular Basis of Inheritance

The DNA and RNA World:

1. Over the years after Mendel, the nature of the genetic material was investigated, resulting in the realization that DNA is the genetic material in the majority of organisms.



Deoxyribonucleic acid (DNA)

Ribonucleic acid (RNA)

1. Deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA) are the two types of nucleic acid found in living systems. Nucleic acids are polymers of nucleotides.
2. **DNA** acts as **genetic material** in most organisms, whereas RNA acts as genetic material in some viruses.
3. **RNA** mostly functions as a **messenger**. RNA has other functions as an adapter, structural or as a catalytic molecule.

What is DNA?

DNA is a group of molecules that is responsible for carrying and transmitting the hereditary materials or the genetic instructions from parents to offsprings.

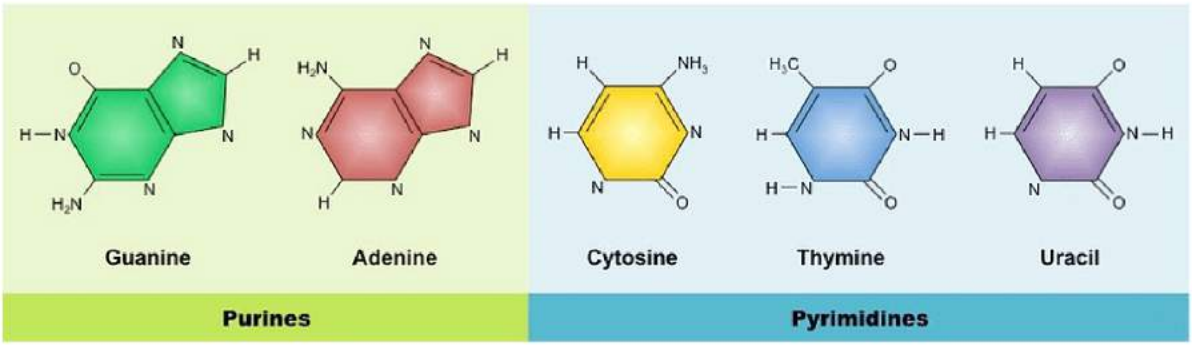
Structure of Polynucleotide Chain.

1. Each nucleotide is composed of three elements:

- **Nitrogenous base:**
  Purines- Adenine (A) and Guanine (G) present in DNA as well as RNA

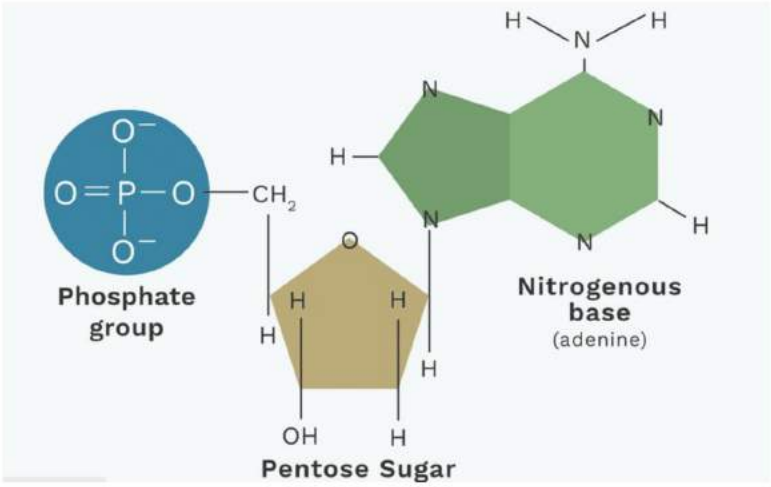
Pyrimidines- Cytosine (C) and Thymine (T) in DNA and Cytosine and Uracil in RNA. Thymine is also known as 5-methyl uracil and it is accounted for more stability of DNA molecule

- **Sugar:** Pentose sugar- Ribose in RNA (ribonucleic acid), deoxyribose in DNA
- **Phosphate group**



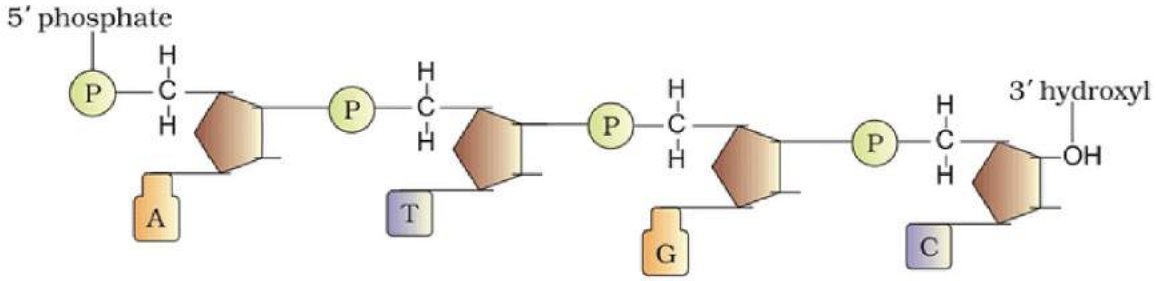
2. Nucleoside: a nitrogenous base linked to the hydroxyl group of 1' C of the pentose sugar by an N-glycosidic bond

3. Nucleotide: phosphate group is attached to the hydroxyl group present at 5' C of the nucleoside by a phospho-ester bond



4. When a phosphate group is linked to 5'-OH of a nucleoside through phosphodiester linkage, a corresponding nucleotide is formed. Two nucleotides are linked through 3'-5' phosphodiester linkage to form a dinucleotide.

5. More nucleotides can be joined in such a manner to form a polynucleotide chain. A polymer thus formed has at one end a free phosphate moiety at 5'-end of the ribose sugar. This end is referred to as the 5'-end of the polynucleotide chain. At the other end of the polymer, the ribose has a free 3'-OH group. This is referred to as the 3'-end of the polynucleotide chain.



Structure & Packaging of DNA Helix

Structure of DNA Double Helix

1. DNA is a long polymer of deoxyribonucleotides. It is made up of two polynucleotide chains, where the backbone is constituted by sugar-phosphate and the bases project inside.
2. The two chains have anti-parallel polarity, i.e. 5′ > 3′ for one, 3′ > 5′ for another

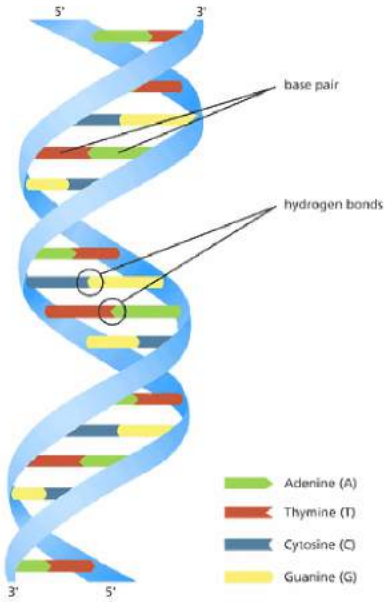


Fig: Double Helix Structure of DNA

1. The bases in two strands are paired through hydrogen bond (H—bonds) forming base pairs (bp). Adenine forms two hydrogen bonds with thymine from the opposite strand and vice-versa. Guanine bonds with cytosine by three H—bonds. Due to this, purine always comes opposite to pyrimidine. This forms a uniform distance between the two strands.
2. The two chains are coiled in a right-handed fashion. The pitch of the helix is 3.4 nm and there is roughly 10 bp in each turn. Due to this, the distance between a base pair in a helix is about 0.34 nm.
3. The plane of one base pair stacks over the other in a double helix. This confers stability to the helical structure in addition to H—bonds.

The length of a DNA double helix is about 2.2 meters (6.6×10^9 bp x 0.34×10^{-9} m/bp)

Therefore, it needs special packaging in a cell.

Packaging of DNA Helix

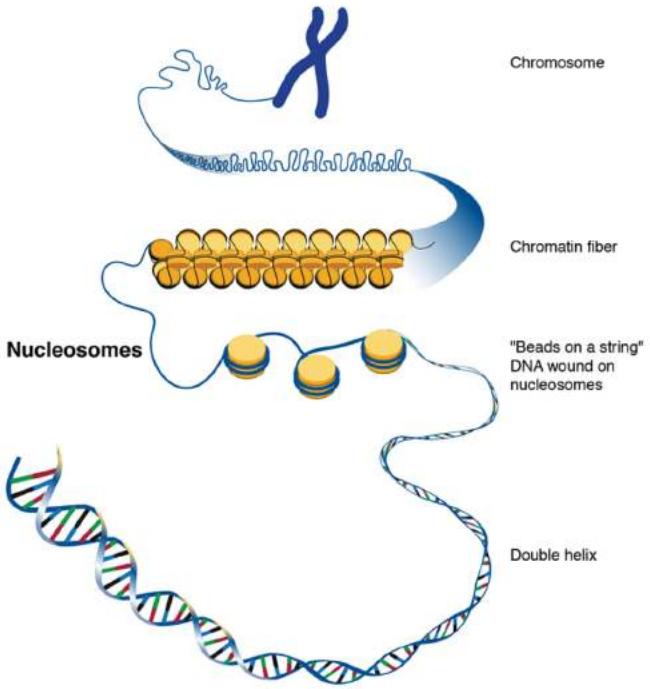
DNA packaging is the process of tightly packing up the DNA molecule to fit into the nucleus of a cell.

Have you ever wondered how DNA is present in a nucleus smaller than it?
The DNA is an organic, complex, molecular structure, found in both prokaryotic and eukaryotic cells and also in many viruses. It is a hereditary material that is found in the nucleus of the cell and is mainly involved in carrying genetic information.

The DNA structure has the following characteristics:

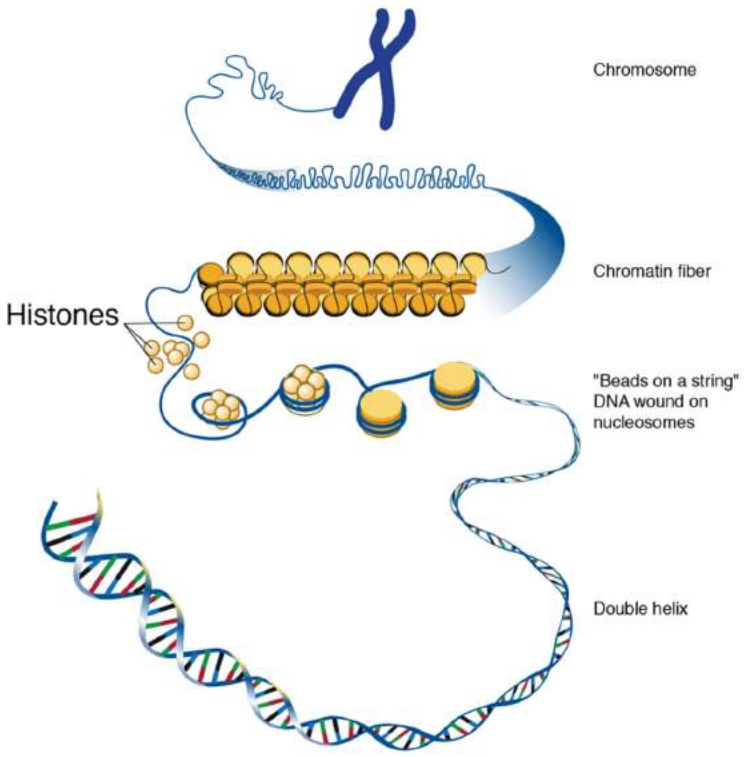
- The strands of the DNA are helically wounded, every single strand forms a right-handed coil.
- The pitch of each helix is 3.32 nm and about 10 nucleotides make up one turn.
- The distance between two succeeding base pairs is 0.34 nm
- The total length of a DNA is the distance between two succeeding base pairs and the product of a total number of base pairs.
- A typical DNA has an extent around 2.2 meters, which is much longer than a nucleus.
- Prokaryotic cells can be distinguished from eukaryotic cells by the presence of a well-defined nucleus. However, their negatively charged DNA is arranged in a region called the nucleoid. They appear as a loop wrapped around a protein molecule having a positive charge.

- All eukaryotes have a well-defined nucleus that contains DNA. DNA is a negatively charged polymer, packed compactly within the chromatin engirdling the histone proteins, a ball of positively charged proteins.
- The octamer of histone proteins is wrapped with a DNA helix, giving rise to a structure called a **nucleosome**. The nucleosomes are further coiled which results in the formation of chromatin fibres. Chromatin fibres are stained thread-like structures whereas nucleosomes are beads present over it. These chromatin fibres condense to form chromosomes during mitosis.



Histones

Histones are the proteins promoting the DNA packaging into chromatin fibres. Histone proteins are positively charged possessing several arginine and lysine amino acids binding to the negatively charged DNA.



There are two types of Histones:

1. Core Histones
2. Linker Histones

- H2A, H2B, H3 and H4 are the core histones. Two H3, H4 dimers and two H2A, H2B dimers form an octamer.
- Linker histones lock the DNA in place onto the nucleosome and can be removed for transcription.
- Histones can be modified to change the amount of packaging a DNA does. The addition of the methyl group increases the hydrophobicity of histones. This results in tight DNA packaging.
- Acetylation and phosphorylation make the DNA more negatively charged and loosens the DNA packaging.
- Enzymes that add methyl groups to histones are called histone methyltransferases. The enzymes that add acetyl groups to the histones are called histone acetyltransferases while the ones that remove the histones are called histone deacetylases.
- The packaging of chromatin at higher level requires additional set of proteins which are collectively called Non-Histone Chromosomal (NHC) proteins.

S-strain $\xrightarrow{\text{Injection}}$ Mice $\longrightarrow$ Mice die

R-strain $\xrightarrow{\text{Injection}}$ Mice $\longrightarrow$ Mice live

In a nucleus, some regions of chromatin are loosely packed (stains light) called euchromatin (transcriptionally active chromatin). In some regions, chromatin is densely packed (stains dark) called heterochromatin (inactive chromatin).

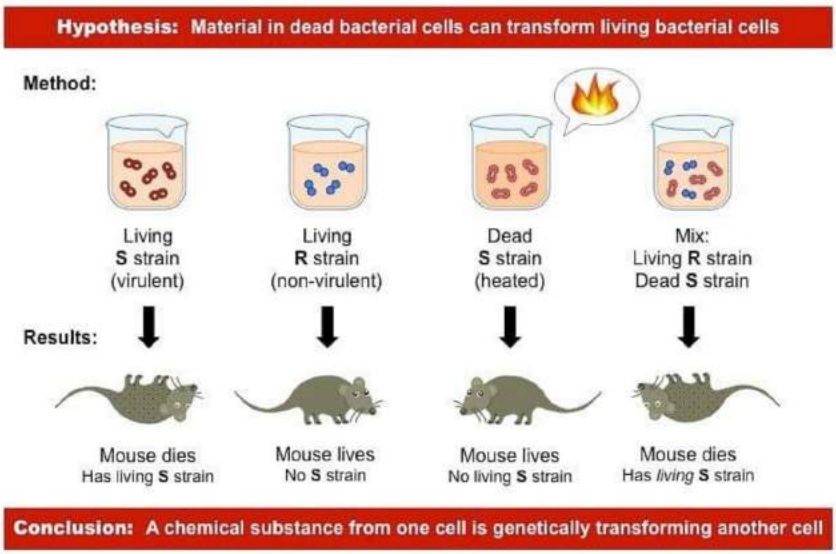
S-strain (Heat-killed) $\xrightarrow{\text{Injection}}$ Mice $\longrightarrow$ Mice live

S-strain (Heat-killed) + R-strain (live) $\xrightarrow{\text{Injection}}$ Mice $\longrightarrow$ Mice die

Search for Genetic Material

Transforming Principle

(i) Frederick Griffith (1928) carried out a series of experiments with Streptococcus pneumoniae (bacterium causing pneumonia).



(ii) According to him, when the bacteria are grown on a culture plate, some produce smooth shiny colonies (S), while others produce rough (R) colonies.

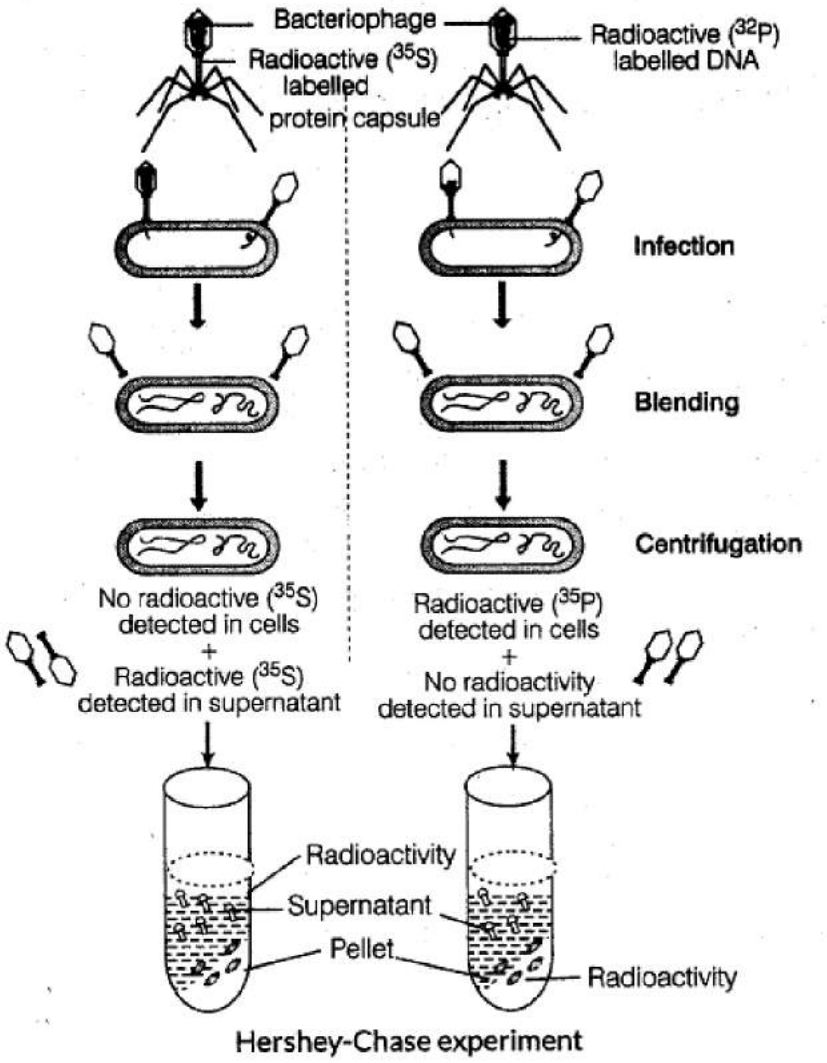
(iii) This is because the S-strain bacteria have a mucous (polysaccharide) coat, while R-strain does not.
(iv) Mice infected with S-strain (virulent) die from pneumonia but mice infected with R-strain do not develop pneumonia.
(v) Griffith killed bacteria by heating and observed that heat-killed S-strain bacteria injected into mice did not kill them. On injecting mixture of heat-killed S and live R bacteria, the mice died. He recovered living S-bacteria from dead mice.
(vi) From this experiment, he concluded that the 'R-strain bacteria' had been transformed by the heat-killed S-strain bacteria. Some transforming principle transferred from heat-killed S-strain, had enabled the R-strain to synthesize a smooth polysaccharide coat and become virulent. This must be due to transfer of the genetic material.

Biochemical Nature of Transforming Principle

(i) Oswald Avery, Colin MacLeod and Maclyn McCarty, worked to determine the biochemical nature of transforming principle in Griffith's experiment.
(ii) They purified biochemicals (proteins, RNA and DNA, etc) from heat-killed S-cells and discovered that DNA alone from S-bacteria caused R-bacteria to be transformed.
(iii) They also discovered that protease (protein digesting enzyme) and RNAases (RNA-digesting enzymes) did not affect transformation.
(iv) Digestion with DNAse did inhibit transformation, indicating that DNA caused transformation.
(v) They concluded that DNA is the hereditary material. But, still all the biologists were not convinced.

DNA is the Genetic Material

(i) Alfred Hershey and Martha Chase (1952) gave unequivocal proof that DNA is the genetic material.
(ii) In their experiments, bacteriophages (viruses that infect bacteria) were used.
(iii) They grew some viruses on a medium that contained radioactive phosphorus and some others on sulphur containing radioactive medium.
(iv) Viruses grown in the presence of radioactive phosphorus contained radioactive DNA but not radioactive protein because DNA contains phosphorus but protein does not. In the same way, viruses grown on radioactive sulphur contained radioactive protein, but not radioactive DNA because DNA does not contain sulphur.



Hershey-Chase experiment

(v) Radioactive phages were allowed to attach to E. coli bacteria. As the infection proceeded, viral coats were removed from the bacteria by agitating them in a blender. The virus particles were separated from the bacteria by spinning them in a centrifuge.
(vi) Bacteria which were infected with viruses that had radioactive DNA were radioactive, indicating that DNA was the material that passed from the virus to the bacteria.
(vii) Bacteria that were infected with viruses that had radioactive proteins were not radioactive. This indicated that the proteins did not enter the bacteria from viruses. It proved that DNA is a genetic material that is passed from virus to bacteria.

Properties of Genetic Material

(i) It became establised that DNA is the genetic material from the Hershey-Chase experiment.
(ii) In some viruses, RNA was also reported as genetic material, e.g. Tobacco mosaic viruses, QB bacteriophage, etc.
(iii) Characteristics of a Genetic Material
(a) It should be able to replicate.
(b) It should be chemically and structurally stable.
(c) It should provide scope for slow changes (mutation) that are required for evolution.
(d) It should be able to express itself in the form of 'Mendelian characters'.
(iv) According to the above mentioned rules, both the nucleic acids (DNA and RNA) have the ability to direct duplications.
Stability can be explained in DNA as the two strands being complementary if separated by heating come together in appropriate conditions.
(v) The 2′ — OH group present at every nucleotide in RNA is a reactive group and makes RNA labile and easily degradable, hence it is reactive.
(vi) DNA is chemically less reactive and structurally more stable when compared to RNA. Thymine also confers additional stability to DNA. So, among the two nucleic acids, the DNA is a predominant genetic material.
(vii) Both RNA and DNA are able to mutate. Viruses having RNA genome and having shorter life span mutate and evolve faster.
(viii) DNA is dependent on RNA for protein synthesis, while RNA can directly code for it. The protein synthesizing machinery has evolved around RNA. This concluded that the DNA being more stable is suitable for storage of genetic information, while for the transmission of genetic information, RNA is suitable.

Francis Crick proposed the central dogma in molecular biology, which states that the genetic information flows from

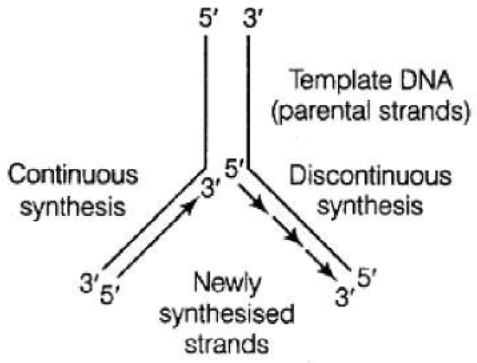
RNA World & DNA Replication

RNA world

RNA was the first genetic material. There are evidences to prove that essential life processes, such as metabolism, translation, splicing, etc., have evolved around RNA.
(i) There are some important biochemical reactions in living systems that are catalysed by RNA catalysts and not by protein enzyme.
(ii) DNA has evolved from RNA with chemical modifications that make it more stable because RNA being a catalyst was reactive and hence, unstable.
There are following three types of RNAs:

(i) mRNA (messenger RNA) provides the template for transcription.
(ii) tRNA (transfer RNA) brings amino acids and reads the genetic code.
(iii) rRNA (ribosomal RNA) plays structural and catalytic role during translation.
All the three RNAs are needed to synthesise a protein in a cell.

DNA replication machinery and enzymes - process of replication requires a set of catalysts (enzymes).
(i) The main enzyme is DNA-dependent DNA polymerase, since it uses a DNA template to catalyse the polymerisation of deoxynucleotides. The average rate of polymerisation by these enzymes is approximately 2000 bp/second.
(ii) These polymerases has to catalyse the reaction with high degree of accuracy because any mistake during replication would result into mutations.
DNA polymerisation is an energy demanding process, so deoxyribonucleoside triphosphates serve dual purposes, i.e. act as substrates and provide energy for polymerization reaction.
(iv) Many additional enzymes are also required in addition to DNA-dependent DNA polymerase.
(v) (a) Replication in DNA strand occurs within a small opening of the DNA helix, known as replication fork.



Replicating fork of DNA

(b) DNA-dependent DNA polymerases catalyse polymerisation only in one direction, i.e. 5′ -> 3. It creates additional complications at the replicating fork. Consequently, on one strand (template 3′-5′), the replication is continuous, while on the other strand (template 5′-3′), it is discontinuous. The discontinuously synthesised fragments called Okazaki fragments are later joined by DNA ligase.

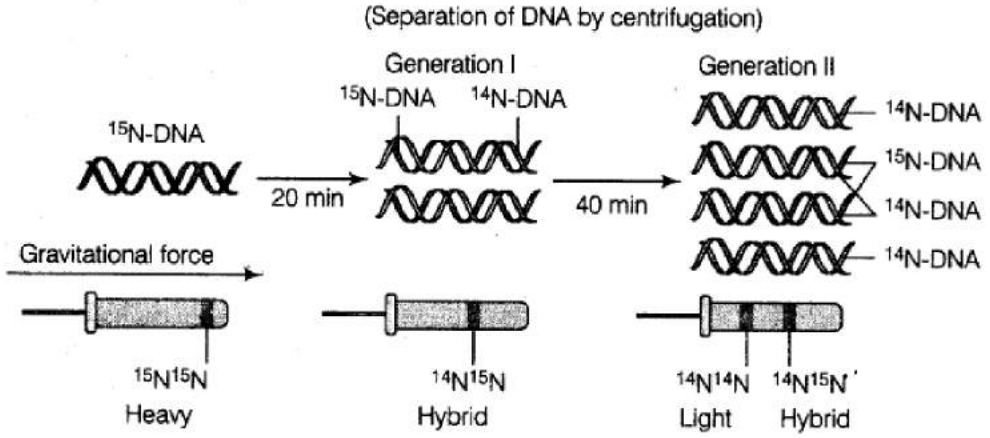
Origin of Replication

(i) DNA polymerases cannot initiate the process of replication on their own. Also,

replication does not initiate randomly at any place in DNA. So, there is a definite region in E.coli DNA where the replication originates. The region is termed as origin of replication.
(ii) Due to this requirement, a piece of DNA, if needed to be propagated during recombinant DNA procedures, requires a vector. The vectors provide the origin of replication.

Replication Scheme for replication of DNA termed as semi-conservative DNA replication was proposed by Watson and Crick (1953). According to it,
(i) The two strands would separate and act as a template for the synthesis of new complementary strands. .
(ii) After replication, each DNA molecule would have one parental and one newly synthesised strand.

Experimental proof that DNA replicates semi-conservatively, comes first from E. coli and later from higher organisms, such as plants and human cells.
Matthew Meselson and Franklin Stahl performed the following experiments to prove this in 1958.
(i) E. coli was grown in a medium containing $^{15}NH_4C1$ as the only nitrogen source for many generations. 15N got incorporated into newly synthesised DNA (and other nitrogen containing compounds). This heavy DNA molecule could be distinguished from the normal DNA by centrifugation in a cesium chloride (CsCl) density gradient.
(ii) They then transferred the cells into a medium with normal $^{14}NH_4Cl$ and took samples at various definite intervals as the cells multiplied and extracted the DNA that remained as double stranded helices. DNA samples were separated independently on CsCl gradients to measure DNA densities.
(iii) The DNA that was extracted from the culture, one generation (after 20 min) after the transfer from 15 N to 14N medium had a hybrid or intermediate density. DNA extracted from the culture after another generation (after 40 min) was composed of equal amounts of this hybrid DNA and of light DNA.
(iv) Very similar experiments were carried out by Taylor and Colleagues on Vicia faba (faba beans) using radioactive thymidine and the same results, i.e. DNA replicates semi-conservatively, were obtained as in earlier experiments.

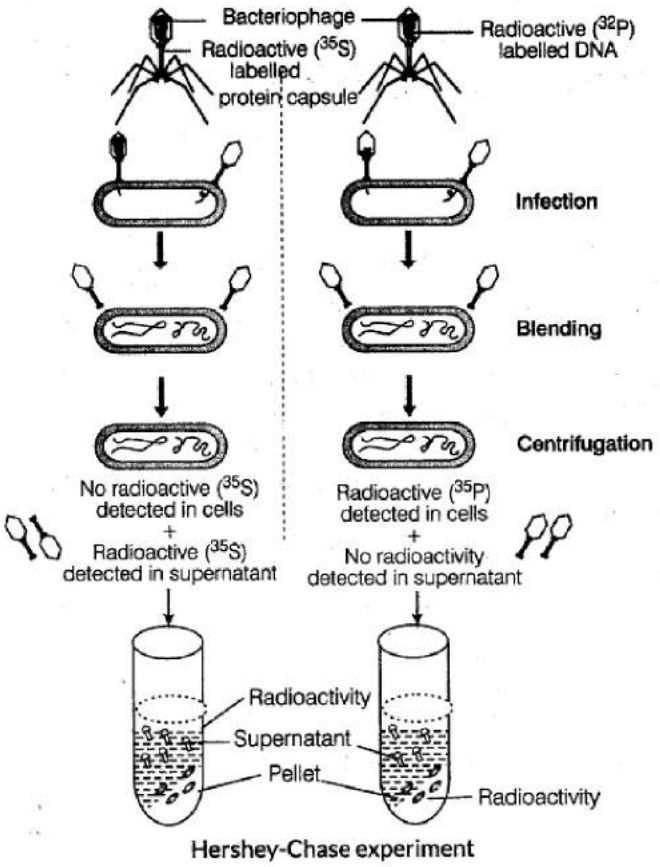


Meselson and Stahl's experiment

Genetic Material Criteria & RNA World

DNA is the Genetic Material

(i) Alfred Hershey and Martha Chase (1952) gave unequivocal proof that DNA is the genetic material.
(ii) In their experiments, bacteriophages (viruses that infect bacteria) were used.
(iii) They grew some viruses on a medium that contained radioactive phosphorus and some others on sulphur containing radioactive medium.
(iv) Viruses grown in the presence of radioactive phosphorus contained radioactive DNA but not radioactive protein because DNA contains phosphorus but protein does not. In the same way, viruses grown on radioactive sulphur contained radioactive protein, but not radioactive DNA because DNA does not contain sulphur.



Hershey-Chase experiment

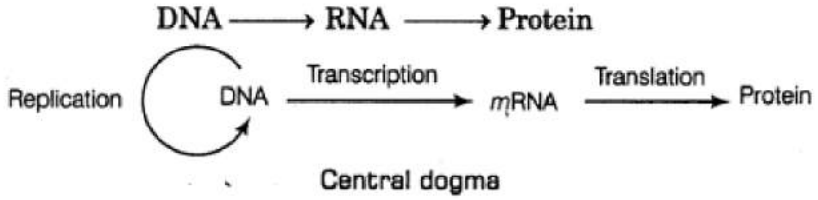
(v) Radioactive phages were allowed to attach to E. coli bacteria. As the infection proceeded, viral coats were removed from the bacteria by agitating them in a blender. The virus particles were separated from the bacteria by spinning them in a centrifuge.
(vi) Bacteria which were infected with viruses that had radioactive DNA were radioactive, indicating that DNA was the material that passed from the virus to the bacteria.
(vii) Bacteria that were infected with viruses that had radioactive proteins were not radioactive. This indicated that the proteins did not enter the bacteria from viruses. It proved that DNA is a genetic material that is passed from virus to bacteria.

Properties of Genetic Material

(i) It became establised that DNA is the genetic material from the Hershey-Chase experiment.
(ii) In some viruses, RNA was also reported as genetic material, e.g. Tobacco mosaic viruses, QB bacteriophage, etc.

(iii) Characteristics of a Genetic Material
(a) It should be able to replicate.
(b) It should be chemically and structurally stable.
(c) It should provide scope for slow changes (mutation) that are required for evolution.
(d) It should be able to express itself in the form of 'Mendelian characters'.
(iv) According to the above mentioned rules, both the nucleic acids (DNA and RNA) have the ability to direct duplications.
Stability can be explained in DNA as the two strands being complementary if separated by heating come together in appropriate conditions.
(v) The 2′ — OH group present at every nucleotide in RNA is a reactive group and makes RNA labile and easily degradable, hence it is reactive.
(vi) DNA is chemically less reactive and structurally more stable when compared to RNA. Thymine also confers additional stability to DNA. So, among the two nucleic acids, the DNA is a predominant genetic material.
(vii) Both RNA and DNA are able to mutate. Viruses having RNA genome and having shorter life span mutate and evolve faster.
(viii) DNA is dependent on RNA for protein synthesis, while RNA can directly code for it. The protein synthesizing machinery has evolved around RNA. This concluded that the DNA being more stable is suitable for storage of genetic information, while for the transmission of genetic information, RNA is suitable.

Francis Crick proposed the central dogma in molecular biology, which states that the genetic information flows from



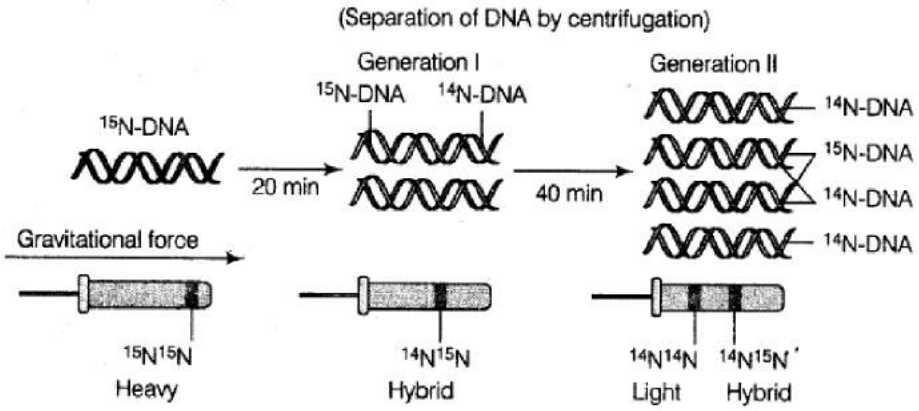
Replication Scheme for replication of DNA termed as semi-conservative DNA replication was proposed by Watson and Crick (1953). According to it,
(i) The two strands would separate and act as a template for the synthesis of new complementary strands. .
(ii) After replication, each DNA molecule would have one parental and one newly synthesised strand.
Experimental proof that DNA replicates semi-conservatively, comes first from E. coli and later from higher organisms, such as plants and human cells.
Matthew Meselson and Franklin Stahl performed the following experiments to prove this in 1958.

(i) E. coli was grown in a medium containing $15NH_4C1$ as the only nitrogen source for many generations. 15N got incorporated into newly synthesised DNA (and other nitrogen containing compounds). This heavy DNA molecule could be distinguished from the normal DNA by centrifugation in a cesium chloride (CsCl) density gradient.
(ii) They then transferred the cells into a medium with normal $14NH_4Cl$ and took samples at various definite intervals as the cells multiplied and extracted the DNA that remained as double stranded helices. DNA samples were separated independently on CsCl gradients to measure DNA densities.
(iii) The DNA that was extracted from the culture, one generation (after 20 min) after the transfer from 15 N to 14N medium had a hybrid or intermediate density. DNA extracted from the culture after another generation (after 40 min) was composed of equal amounts of this hybrid DNA and of light DNA.
(iv) Very similar experiments were carried out by Taylor and Colleagues on Vicia faba (faba beans) using radioactive thymidine and the same results, i.e. DNA replicates semi-conservatively, were obtained as in earlier experiments.

Replication Scheme for replication of DNA termed as semi-conservative DNA replication was proposed by Watson and Crick (1953). According to it,
(i) The two strands would separate and act as a template for the synthesis of new complementary strands. .
(ii) After replication, each DNA molecule would have one parental and one newly synthesised strand.

Experimental proof that DNA replicates semi-conservatively, comes first from E. coli and later from higher organisms, such as plants and human cells.
Matthew Meselson and Franklin Stahl performed the following experiments to prove this in 1958.
(i) E. coli was grown in a medium containing $15NH_4C1$ as the only nitrogen source for many generations. 15N got incorporated into newly synthesised DNA (and other nitrogen containing compounds). This heavy DNA molecule could be distinguished from the normal DNA by centrifugation in a cesium chloride (CsCl) density gradient.
(ii) They then transferred the cells into a medium with normal $14NH_4Cl$ and took samples at various definite intervals as the cells multiplied and extracted the DNA that remained as double stranded helices. DNA samples were separated independently on CsCl gradients to measure DNA densities.
(iii) The DNA that was extracted from the culture, one generation (after 20 min) after the transfer from 15 N to 14N medium had a hybrid or intermediate density. DNA extracted from the culture after another generation (after 40 min) was composed of equal amounts of this hybrid DNA and of light DNA.

(iv) Very similar experiments were carried out by Taylor and Colleagues on Vicia faba (faba beans) using radioactive thymidine and the same results, i.e. DNA replicates semi-conservatively, were obtained as in earlier experiments.



Meselson and Stahl's experiment

Genetic Code

Genetic Code

During replication and transcription a nucleic acid was copied to form another nucleic acid. Hence, these processes are easy to conceptualize on the basis of complementarity. The process of translation requires transfer of genetic information from a polymer of nucleotides to a polymer of amino acids. Neither does any complementarity exist between nucleotides and amino acids, nor could any be drawn theoretically. There existed ample evidences, though, to support the notion that change in nucleic acids (genetic material) were responsible for change in amino acids in proteins. This led to the proposition of a genetic code that could direct the sequence of amino acids during synthesis of proteins.

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

Providing proof that the codon was a triplet, was a more daunting task. The chemical method developed by Har Gobind Khorana was instrumental in synthesising RNA molecules with defined combinations of bases (homopolymers

and copolymers). Marshall Nirenberg's cell-free system for protein synthesis finally helped the code to be deciphered. Severo Ochoa enzyme (polynucleotide phosphorylase) was also helpful in polymerising RNA with defined sequences in a template independent manner (enzymatic synthesis of RNA). Finally a checker-board for genetic code was prepared which is given in Table .

Table: The Codons for the Various Amino Acids

First position	Second position U	Second position C	Second position A	Second position G	Third position
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA 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	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

The salient features of genetic code are as follows:
(i) The codon is triplet. 61 codons code for amino acids and 3 codons do not code for any amino acids, hence they function as stop codons.
(ii) One codon codes for only one amino acid, hence, it is unambiguous and specific.
(iii) Some amino acids are coded by more than one codon, hence the code is degenerate.
(iv) The codon is read in mRNA in a contiguous fashion. There are no punctuations.
(v) The code is nearly universal: for example, from bacteria to human UUU would code for Phenylalanine (phe). Some exceptions to this rule have been found in mitochondrial codons, and in some protozoans.
(vi) AUG has dual functions. It codes for Methionine (met) , and it also act as initiator codon.
If following is the sequence of nucleotides in mRNA, predict the sequence of amino acid coded by it (take help of the checkerboard):
-AUG UUU UUC UUC UUU UUU UUC-
Now try the opposite. Following is the sequence of amino acids coded by mRNA. Predict the sequence of nucleotide in the RNA:
Met-Phe-Phe-Phe-Phe-Phe-Phe Do you face any difficulty in predicting the opposite.

Mutations and Genetic Code

The relationships between genes and DNA are best understood by mutation studies. You have studied about mutation and its effect in Chapter 5. Effects of large deletions and rearrangements in a segment of DNA are easy to comprehend. It may result in loss or gain of a gene and so a function. The effect of point mutations will be explained here. A classical example of point mutation is a change of single base pair in the gene for beta-globin chain that results in the change of amino acid residue glutamate to valine. It results into a diseased condition called as sickle cell anemia. Effect of point mutations that inserts or deletes a base in structural gene can be better understood by following simple example. Consider a statement that is made up of the following words each having three letters like genetic code.

RAM HAS RED CAP

If we insert a letter B in between HAS and RED and rearrange the statement, it would read as follows:

RAM HAS B RED CAP

Similarly, if we now insert two letters at the same place, say BI'. Now it would read,

RAM HAS BI RED CAP

Now we insert three letters together, say BIG, the statement would read

RAM HAS BIG RED CAP

The same exercise can be repeated, by deleting the letters R, E and D, one by one and rearranging the statement to make a triplet word.

RAM HAS EDC AP

RAM HAS DCA P

RAM HAS CAP

The conclusion from the above exercise is very obvious. Insertion or deletion of one or two bases changes the reading frame from the point of insertion or deletion. Insertion or deletion of three or its multiple bases insert or delete one or multiple codon hence one or multiple amino acids, and reading frame remains unaltered from that point onwards. Such mutations are referred to as frame-shift insertion or

deletion mutations. This forms the genetic basis of proof that codon is a triplet and it is read in a contiguous manner.

Translation

Translation

Translation refers to the process of polymerisation of amino acids to form a polypeptide (Figure). The order and sequence of amino acids are defined by the sequence of bases in the mRNA. The amino acids are joined by a bond which is known as a peptide bond. Formation of a peptide bond requires energy. Therefore, in the first phase itself amino acids are activated in the presence of ATP and linked to their cognate tRNA–a process commonly called as charging of tRNA or amino acylation of tRNA to be more specific. If two such charged tRNAs are brought close enough, the formation of peptide bond between them would be favoured energetically. The presence of a catalyst would enhance the rate of peptide bond formation

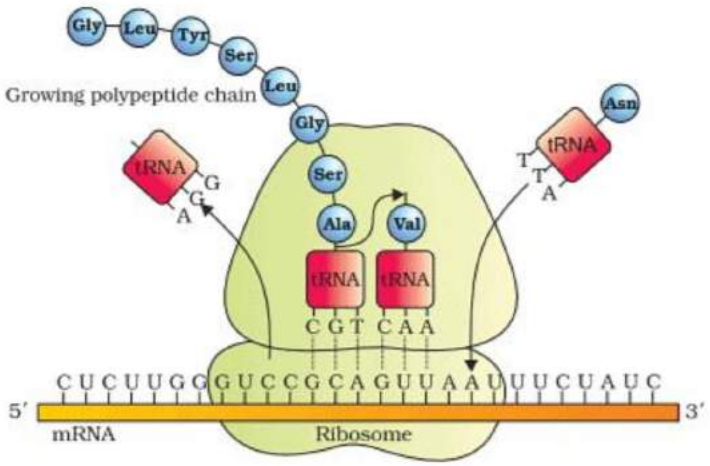


Figure. Translation

The cellular factory responsible for synthesising proteins is the ribosome. The ribosome consists of structural RNAs and about 80 different proteins. In its inactive state, it exists as two sub units; a large sub unit and a small sub unit. When the small sub unit encounters an mRNA, the process of translation of the mRNA to protein begins. There are two sites in the large sub unit, for subsequent amino acids to bind to and thus, be close enough to each other for the formation of a peptide bond. The ribosome also acts as a catalyst (23S rRNA in bacteria is the enzyme- ribozyme) for the formation of peptide bond.

A translational unit in mRNA is the sequence of RNA that is flanked by the start codon (AUG) and the stop codon and codes for a polypeptide. An mRNA also has some additional sequences that are not translated and are referred as untranslated regions (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. For initiation, the ribosome binds to the mRNA at the start codon (AUG) that is recognised only by the initiator tRNA. The ribosome proceeds to the elongation phase of protein synthesis. During this stage, complexes composed of an amino acid linked to tRNA, sequentially bind to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. The ribosome moves from codon to codon along the mRNA. Amino acids are added one by one, translated into Polypeptide sequences dictated by DNA and represented by mRNA. At the end, a release factor binds to the stop codon, terminating translation and releasing the complete polypeptide from the ribosome.

Human Genome Project

GENOMICS

Genomics is the study of genomes and genes based on DNA sequencing. **Genome** is the total gene complement of a haploid set of chromosomes and inherited as a unit from one parent through the gamete. A haploid (such as prokaryotic) cell contains a single genome, a diploid (such as a cell of higher plant or animal) has two genomes, one paternal and other maternal. Additional DNA is also present in mitochondria which is inherited from one's mother.

HUMAN GENOME PROJECT

Genetic make-up of an organism or an individual lies in the DNA sequences. If two individuals differ, then their DNA sequences should also be different, at least at some places. These assumptions led to the quest of finding out the complete DNA sequence of human genome. With the establishment of genetic engineering techniques where it was possible to isolate and clone any piece of DNA and availability of simple and fast techniques for determining DNA sequences, a very ambitious project of sequencing human genome was launched in the year 1990.

Huma n Genome Project (HGP) was called a mega project. You can imagine the magnitude and the requirements for the project if we simply define the aims of the project as follows : Human genome is said to have approximately 3 × 109 bp, and if the cost of sequencing required is US $ 3 per bp (the estimated cost in the beginning), the total estimated cost of the project would be approximately 9 billion US dollars. Further, if the obtained sequences were to be stored in typed form in

books, and if each page of the book contained 1000 letters and each book contained 1000 pages, then 3300 such books would be required to store the information of DNA sequence from a single human cell. HGP was closely associated with the rapid development of a new area in biology called as **Bio**

informatics .

Goals of HGP Some of the important goals of HGP are as follows :
(i) Identify all the genes in human DNA.
(ii) Determine the sequences of the 3 billion chemical base pairs that make up human DNA.
(iii) Store this information in databases.
(iv) Improve tools for data analysis.
(v) Transfer related technologie s to other sectors, such as industries.
(vi) Address the ethical, legal, and social issues (ELSI) that may arise from the project.
The project was completed in 2003. Knowledge about the effects of DNA variations among individuals can lead to revolutionary new ways to diagnose, treat and someday prevent the thousands of disorders that affect human beings. Besides providing clues to understanding human biology, learning about non-human organisms, DNA sequences can lead to an understanding of their natural capabilities that can be applied toward solving challenges in health care, agriculture, energy production, environmental remediation. Many non-human model organisms, such as bacteria, yeast, Caenorhabditis elegans (a free living non-pathogenic nematode), Drosophila (the fruit fly), plants (rice and Arabidopsis), etc., have also been sequenced.

Methodologies : The methods involved **two** major approache s.
(1) **Expressed Sequence Tags** (ESTs) -Identifying all the genes that expressed as RNA .
(2) **Sequence Annotation -** The blind approach of s imply sequencing the whole set of genome that contained all the coding and non-coding sequence, and later assigning different regions in the sequence with functions. For sequencing, the total DNA from a cell is isolated and converted into random fragments of relatively smaller sizes (recall DNA is a very long polymer, and there are technical limitations in sequencing very long pieces of DNA) and cloned in suitable host using specialised vectors.
The cloning resulted into amplification of each piece of DNA fragment so, that is subsequently could be sequenced with ease. The commonly used hosts were bacteria and yeast, and the vectors were called as **BAC** (bacterial artificial chromosome s), and **YAC** (yeast artificial chromosome s).

The fragments were sequenced using automated DNA sequencers that worked on the principle of a method developed by **Frederick Sanger** . (Remember, Sanger is also credited for developing method for determination of amino acid sequences in proteins). These sequences were then arranged based on some **overlapping regions** present in them. This required generation of overlapping fragments for sequencing. Alignment of these sequences was humanly not possible. Therefore, specialised computer based programmes were developed. These sequences were subsequently annotated and were assigned to each chromosome. The sequence of chromosome I was completed only in May 2006 (this was the last of the 24 human chromosomes -22 autosomes and X and Y- to be sequenced). Another challenging task was assigning the genetic and physical maps on the genome. This was generated using information on polymorphism of restriction endonuclease recognition sites, and some repetitive DNA sequences known as microsatellites.

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

(viii) Chromosome 1 has most genes (2968). and the Y has the fewest (231).
(ix) Scientists have identified about 1.4 million locations where single-base DNA differences **(SNPs- single nucleotide polymorphism,** pronounced as 'snips') occur in humans, This information promises to revolutionize the processes of finding chromosomal locations for disease-associated sequences and tracing human history.

Organisms	Base pair	Gene No.
Bateriophage	10,000	------/-----
Lily	106 Billion B.P.	
E.coli	4.7 million B.P.	4,000
S. cerviceae	12 Million B.P.	6,000
D. melangaster	180 Million B.P.	13,000
Caenorhabditis elegans	97 Million B.P.	18,000

Human	3 Billion B.P.	30,000

(a) First prokaryotes in which complete genome was sequenced is Haemophilus influenzae.
(b) First Eukaryote in which complete genome was sequenced is Saccharomyces cerviceae (Yeast).
(c) First plant in which complete genome was sequenced is Arabidopsis thaliana (Small mustard plant).
(d) First animal in which complete genome was sequenced is Caenorhabditis elegans (Nematode). – b – globin and insulin gene are less than 10 kilo base pair T.D.F. gene is the smallest gene (14 base pair) and Duchenne muscular Dystrophy gene is made up of 2400 kilo base pair.(Longest gene)

GENE LIBRARY (GENOMIC DNA LIBRARY) AND GENE BANKS

Agene libraryis a collection of many of the desired genes of DNA fragments maintained in clones of bacterial or some other cells. It is prepared by the following method.
DNA fragments containing one or few desired genes are obtained with the help of specific restriction endonucleases.
Each fragment is joined to a suitable vehicle DNA to form recombinant DNAs of different nature. These are then introduced into host (bacterial, yeast, plant or animal) cells. The cells containing recombinant DNAs are allowed to multiply in cultures. This will produce clones of cells where the daughter cells carry the same genes which are identical to those of parent cells. A collection of clones with recombinant
DNA containing desired genes is a gene library. Gene libraries are maintained through special techniques.
A gene bank is a store house of clones of known DNA fragments, genes, gene maps, seeds, spores, frozen sperms or eggs or embryos. These are stored for possible use in genetic engineering and breeding experiments where species have become extinct. The need of gene banks is being increasingly felt as the rate of extinction is increasing day by day. The human genome project is the most remarkable contribution in this field.

AFLP (AMPLIFIED FRAGMENT LENGTH POLYMORPHISMS)

This procedure was described by scientist Zabeau & Vos (1993). In this procedure DNA is cut by restriction enzyme then these restricted fragment are amplified by P.C.R. and band pattern of these restricted fragments is visualised after gel electrophoresis.
Procedure (3 steps)

(1) Digestion of total cellular DNA with one or more restriction enzymes.
(2) Selective amplication of some of these fragments by P.C.R. primer.
(3) Electrophoretic separation on gel matrix followed by visualisation of band pattern of these restricted fragment.

R.A.P.D. (RANDOM AMPLIFICATION OF POLYMORPHIC DNA)

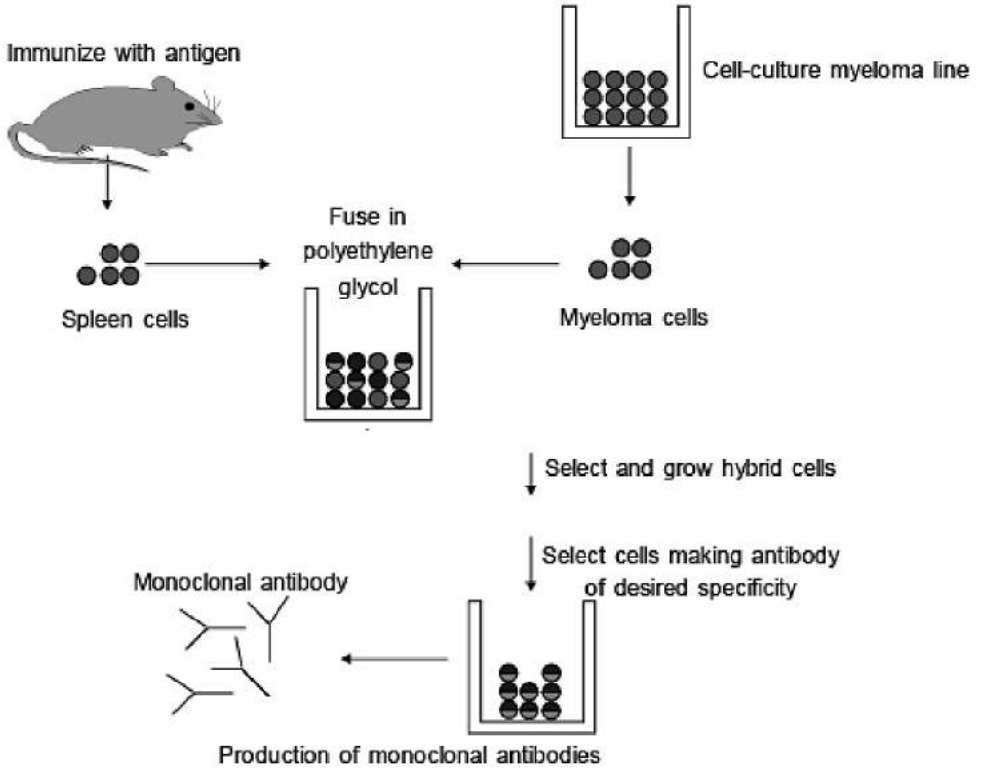
This is a type of P.C.R. in which random (unknown) DNA fragments are amplified.

BIOINFORMATICS

Definition : Bioinformatics is application of computer technology to the management of biological information.
Computer techniques are used to gather, store, analyze and integrate biological and genetic information which can be applied to gene based drug discovery and development.
Drug design based on bioinformatics Bioinformatics is a new approach for drug designing. In this procedure all the knowledge about the disease is collected, analysed and find out a 'target molecules' that cause the disease. Structure of these molecules is analysed by X-ray or N.M.R. technique, then drugs are developed which can bind and block the activity of these molecules.
Biological data base Biological data base is collection of genomic, proteomic and metabolic data associated with computer software, which include nucleotide sequence of gene or amino acid sequence of protein, information about structure, function, location of genes and protein.
Importance –
(1) Easy asses to the information.
(2) A method for extracting only required information.

MONOCLONAL ANTIBODY (MAB)

– Monoclonal antibodies are specific only for one antigen and synthesize outside the animal body .
– Monoclonal antibody produced by a specialized cells through a technique called as hybridoma technology.
– This technology was discovered by Georges Kohler and Milstein, were awarded with the 1984 nobel prize.
– Each hybrid clone grown in culture medium to produce monoclonal antibody.
– Monoclonal antibody which acts as an enzyme called abzyme.



DNA Fingerprinting

DNA FINGER PRINTING / DNA TYPING / DNA PROFILING/ DNA TEST

- It is technique to identify a person on the basis of his/her DNA specificity.
- This technique was invented by sirAlec. Jeffery(1984). In India DNA Finger printing has been started by Dr. V.K. Kashyap & Dr. Lal Ji Singh.
- DNA of human is almost the same for all individuals but very small amount that differs from person to person that forensic scientists analyze to identify people.

These differences are called Polymorphism (many forms) and are the key of DNA typing. Polymorphisms are most useful to forensic scientist. It is consist of variation in the length of DNA at specific loci is called Restricted fragment. It is most important segment for DNA test made up of short repetitive nucleotide sequences, these are called VNTRs (variable number of tandem repeat).

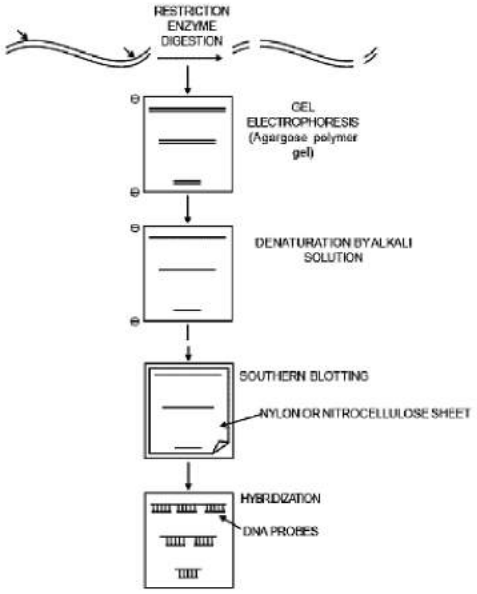
- VNTR's also called mini satellites were discovered by Alec Jeffery. Restricted fragment consist of hyper-variable repeat region of DNA having a basic repeat sequence of 11-60 bp and flanked on both sites by restriction site. The number and position of minisatellites or VNTR in

restriction fragment is different for each DNA and length of restricted fragment is depend on number of VNTR.

- Therefore, when the genome of two people are cut using the same restriction enzyme the length of fragments obtained is different for both the people.
- These variations in length of restricted fragment is called RFLP or Restriction fragment length polymorphism.
- Restriction Fragment Length Polymorphism distributed throughout human genomes are useful for DNA Finger printing.
- DNA Fingerprint can be prepared from extremely minute amount of blood, semen, hair bulb or any other cell of the body.
  DNA content of 1 - Microgram is sufficient.
  Technique of DNA Finger printing involves the following major stpes.

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

Steps of DNA Fingerprinting



Steps of DNA Fingerprinting

2. Restriction Enzyme Digestion : Restriction enzyme cuts DNA at specific 4 or 6 base pair sequences called restriction site.
Hae III (Haemophilus aegyptius) is most commonly used enzyme. It cuts the DNA, every where the bases are arranged in the sequence GGCC. These restricted fragment transferred to Agarose Polymer gel.

3. Gel Electrophoresis : – Gel electrophoresis is a method that separates macromolecules-either nucleic acid or proteins-on the basis of size, electric charge.

- A gel is a colloid in a solid form. The term electrophoresis describes the migration of charged particles under the influence of an electric field. Electro refers to the energy of electricity. Phoresis, from the Greek verb phoros, means "to carry across." Thus, gel electrophoresis refers to the technique in which molecules are forced across a span of gel, motivated by an electrical current. Activated electrodes at either end of the gel provides the driving force. A molecule's properties determine, how rapidly an electric field can move the molecule through a gelatinous medium.
- Many important biological molecules such as amino acids, peptides, proteins, nucleotides, and nucleic acids posses ionisable groups and, therefore, at any given pH, exist in solution as electrically charged species either as cation (+) or anions (–). Depending on the nature of the net charge, the charged particles will migrate either to the cathode or to the anode. – By the gel electrophoresis these restricted fragments move towards the positive electrode (anode) because DNA has –ve electric charge (PO_4^{-3}).
- Smaller Fragment more move towards the positive pole due to less molecular weight. So after the gel electrophoresis DNA fragment arranged according to molecular weight
- These separated fragments can be visualized by staining them with a dye that fluoresces ultraviolet radiation.
- This appears the specific restricted fragment length pattern. This length pattern is different in different individual.
- This is called Restricted Fragment length Polymorphism (RFLP).

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

5. Hybridization : To detect VNTR locus on restricted fragment, we use single stranded Radioactive (P^{32}) DNA probe which have the base pair sequences

complimentary to the DNA sequences at the VNTR locus. Commonly we use a combination of at least 4 to 6 separate DNA probes. Labelled Probes are attached with the VNTR loci of restricted DNA Fragments, this process is called Hybridization.

6. Autoradiography : Nylon membrane containing radio active probe exposed to X-ray. Specific bands appear on X-ray film. These bands are the areas where the radioactive probe bind with the VNTR.
These allow analyzer to identify a particular person DNA, the occurrence and frequency of a particular genetic pattern contained in this X-ray film. These x-ray film called DNA signature of a person which is specific for each individual.
The probability of two unrelated individual having same pattern of location and repeat number of minisatellite (VNTR) is one in ten billion (world population 6.1 billion) In India the centre for DNA finger printing and diagnosis (CDFD - center for DNA finger printing & diagnosis) located at Hyderabad.

Application of DNA Finger printing

1. Paternity tests. The major application of DNA finger printing is in determining family relationships. For identifying the true (biological) father, DNA samples of Child, mother and possible fathers are taken and their DNA finger prints are obtained. The prints of child DNA match to the prints of biological parents.

2. Identification of the criminal. DNA finger printing has now become useful technique in forensic (crime detecting) science, specially when serious crimes such as murders and rapes are involved. For identifying a criminal, the DNA fingerprints of the suspects from blood or hair or semen picked up from the scene of crime are prepared and compared. The DNA fingerprint of the person matching the one obtained from sample collected from scene of crime can give a clue to the actual criminal.

Transcription: Molecular Basis of Inheritance

Transcription Unit

The central dogma of molecular biology is as follows:
DNA → RNA → protein
Transcription is the copying of genetic information from one DNA strand into RNA by RNA polymerase. Like replication, it is also governed by the principle of complementarity. However, unlike replication, only one DNA strand is copied to RNA in transcription. Why is this?
Firstly, if both DNA strands act as templates, it will result in two different RNA sequences and in turn give rise to two different proteins. This complicates the

transfer of genetic information. Secondly, the two RNA sequences produced will be complementary to each other and produce a double-stranded RNA. This will prevent RNA from being translated into protein. Now, let's look at the different regions of a transcription unit in DNA.

1. Structural Gene

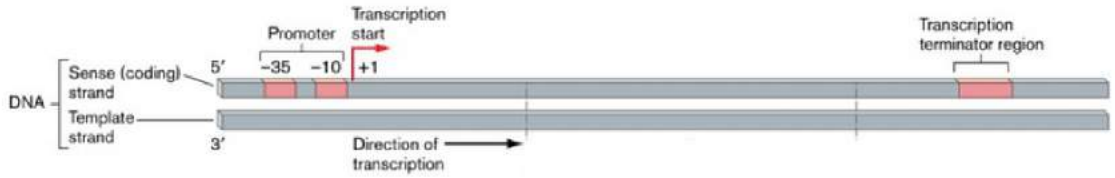
The two DNA strands within the structural gene have different names. Since RNA polymerase catalyzes polymerization in only one direction $5' \rightarrow 3'$, the strand with $3' \rightarrow 5'$ polarity becomes the template strand. The other strand with $5' \rightarrow 3'$ polarity is displaced during transcription and is called the coding strand even though it does not code for anything. In a transcription unit, the promoter and terminator regions lie on either side of the structural gene.

2. Promoter

It is a DNA sequence located towards the $5'$ end (upstream) of the coding strand. It is the binding site for RNA polymerase and is the site that tells the polymerase to start transcription. Additionally, the presence of the promoter defines the template and coding strand in a transcription unit.

3. Terminator

It is a DNA sequence located towards the $3'$ end (downstream) of the coding strand. It provides the stop signal and defines the end of transcription. Additional regulatory sequences may be present upstream or downstream of the promoter.



Transcription unit

The Gene

The functional unit of inheritance is a gene. Although genes are located on DNA, it is difficult to define a gene in terms of DNA sequence. A DNA sequence that codes for tRNA (transfer RNA) or rRNA (ribosomal RNA) is also a gene.
A cistron is a segment of DNA that codes for a polypeptide (a polymer of amino acids). A cistron can be polycistronic (mostly in prokaryotes and bacteria), i.e. it can code for several proteins. It can also be monocistronic (mostly in eukaryotes) i.e. it codes for a single protein. The monocistronic genes in

eukaryotes consist of coding sequences called exons and intervening sequences called introns. Exons appear in mature or processed RNA whereas introns do not.

Types of RNA

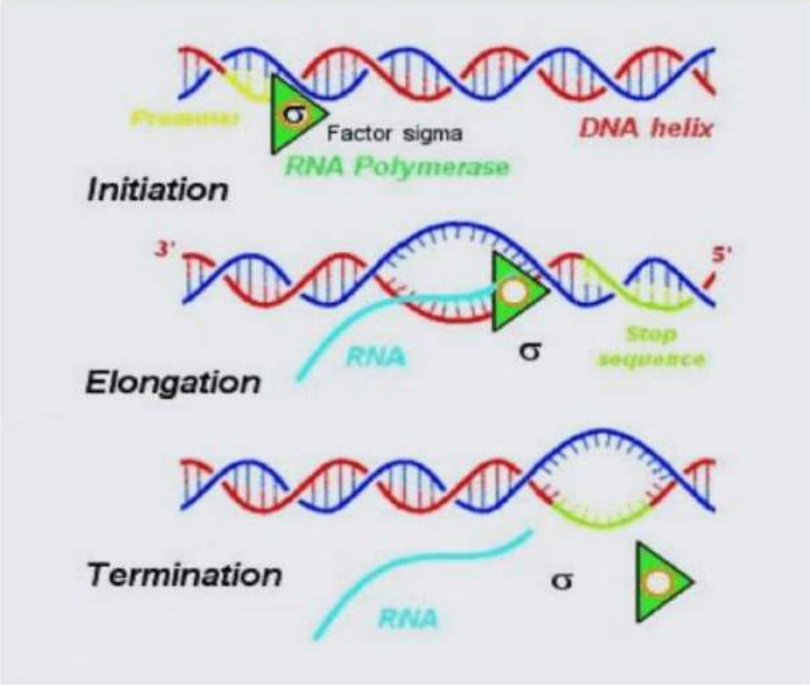
There are three major types of RNAs in bacteria:

- **Messenger RNA (mRNA)** – It provides the template to make protein.
- **Transfer RNA (tRNA)** – It reads the genetic code and transfers amino acids for protein synthesis.
- **Ribosomal RNA (rRNA)** – It has a structural and catalytic role in protein synthesis.

The Process of Transcription

Transcription has the following steps:

- **Initiation:** Here, RNA polymerase binds to the promoter region and transiently binds to the 'initiation factor' to initiate transcription.
- **Elongation:** This is the step where the RNA strand starts getting longer. RNA polymerase "walks" along one strand of DNA. For every nucleotide recognized on the DNA template, it adds a complementary RNA nucleotide to the growing RNA transcript.
- **Termination:** Transcription stops once the RNA polymerase reaches the terminator region. At this region, the RNA transcript and the RNA polymerase, both fall off. RNA polymerase transiently associates with the 'termination factor' to stop transcription.



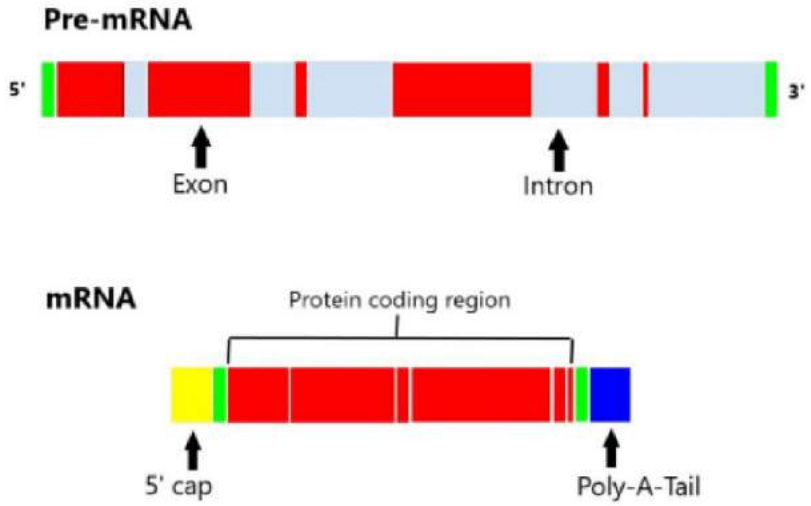
Transcription in bacteria

Complexity In Eukaryotic Transcription

In bacteria, since the mRNA does not need to be processed and since transcription and translation occur in the same cell compartment, the two processes can occur simultaneously. Also, the RNA Polymerase catalyzes transcription of all kinds of RNA. Eukaryotes, however, differ and show two main complexities. There are 3 types of RNA polymerases –

- RNA Polymerase I that transcribes rRNA.
- Type II that transcribes a precursor of mRNA – heterogenous nuclear RNA (hnRNA).
- RNA Polymerase III that transcribes tRNA and small nuclear RNAs (snRNA).

The primary transcript in eukaryotes is non-functional since it contains exons and introns. It undergoes splicing, a process that removes introns and joins the exons together in a specific order. The precursor hnRNA undergoes additional processing called capping and tailing.
An unusual nucleotide is added to the 5′ end of hnRNA during capping. In tailing, 200-300 adenylate residues are added to the 3′ end of hnRNA. This fully processed hnRNA, called mRNA is now transported out of the nucleus for translation.



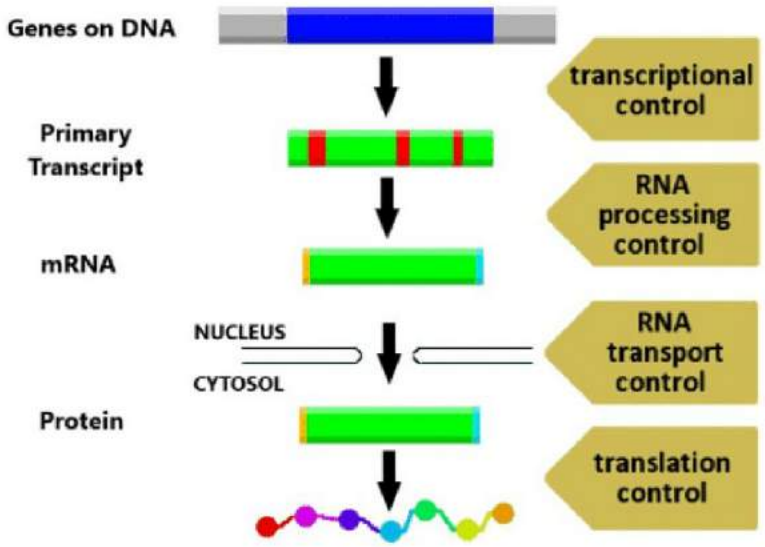
Post-transcriptional modifications in Eukaryotes

Regulation of Gene Expression

Regulation of Gene Expression

Protein synthesis begins at transcription, ends at translation and involves multiple steps. Therefore, regulation of gene expression can happen at any of these steps. In eukaryotes, gene regulation occurs at any of the following steps:

- Transcriptional level i.e. during the formation of the primary transcript.
- Processing level i.e. at the stage of splicing.
- During transport of mRNA from the nucleus to the cytoplasm.
- Translational level.



A great example of coordinated gene regulation is the development and differentiation of embryo into adult organisms. Metabolic, physiological and environmental conditions govern the regulation of gene expression. For example, E. coli uses lactose as a source of energy.
To do so, it synthesizes an enzyme called beta-galactosidase which hydrolyzes lactose into galactose and glucose. However, if there is no lactose around to be used as an energy source, the E. coli does not need to synthesize beta-galactosidase.

Prokaryotic Gene Regulation

In prokaryotes, the main site for regulation of gene expression is transcription initiation. Within a transcription unit, the activity of RNA polymerase at the promoter is regulated by 'accessory proteins'. These proteins affect the ability of RNA polymerase to recognize start sites. These proteins can act both positively (activators) or negatively (repressors).
In prokaryotic DNA, the accessibility of the promoter depends on the interaction of proteins with sequences called operators. In most operons, the operator is adjacent to the promoter elements. Moreover, in most cases, the operator has a repressor protein bound to it. Therefore, each operon has its own, specific operator and repressor. Let's understand this better using lac operon as an example.

The Lac Operon

Here, 'lac' refers to lactose. Francois Jacob and Jacque Monod were the first to elucidate the lac operon – a transcriptionally regulated system. Lac operon consists of a polycistronic structural gene regulated by a common promoter and regulatory

genes. Such arrangements are common in bacteria and are called operons. Other examples include trp operon, val operon, his operon etc.
The *lac* operon has the following parts:

- One regulatory gene – The *i* gene where 'i' is derived from 'inhibitor'. This gene codes for the repressor of the *lac* operon.
- Three structural genes –
  (i) The z gene that codes for the enzyme beta-galactosidase that hydrolyzes lactose to glucose and galactose.
  (ii) The y gene codes for the enzyme permease that increases the permeability of the cell to beta-galactosides.
  (iii) The a gene codes for transacetylase.

Lactose metabolism requires gene products of all three genes mentioned above. Lactose, the substrate for the enzyme beta-galactosidase, regulates the switching on and off of the operon. Therefore, lactose is the inducer. Let's understand how lactose switches the operon on or off.
In the absence of lactose, the i gene synthesizes the repressor which then binds to the operator region of the operon. This prevents RNA polymerase from transcribing the genes (z, y, a) on the operon. Therefore, if there is no lactose, the operon does not synthesize genes for its utilization. The action of the repressor on the lac operon is negative regulation.
In the presence of lactose, the repressor interacts with lactose and gets inactivated. Thus, RNA polymerase is free and can transcribe the genes in the operon. Therefore, if lactose is present, the operon synthesizes the genes for its utilization. Therefore, essentially, the presence of the substrate i.e. lactose regulates the synthesis of enzymes for its utilization.

Enzymes: DNA

Introduction

- At its most basic level, DNA replication is the operation of DNA polymerases producing a complementary DNA strand to the original template strand.
- To manufacture DNA, DNA helicases unwind double-stranded DNA before polymerases, generating a replication fork with two single-stranded templates.
- Replication allows a single DNA double helix to be copied into two DNA helices, which are then split into daughter cells during mitosis.
- From prokaryotes to eukaryotes, the primary enzymatic processes carried out at the replication fork are highly conserved.

- In actuality, the replication machinery is a large complex that coordinates several proteins at the replication site to create the "replisome."
- Each proliferative cell's replisome is in charge of duplicating the totality of genomic DNA.
- This mechanism is vital to all organisms because it provides for the high-fidelity transmission of hereditary/genetic information from parental cells to daughter cells.

DNA Replication Machinery

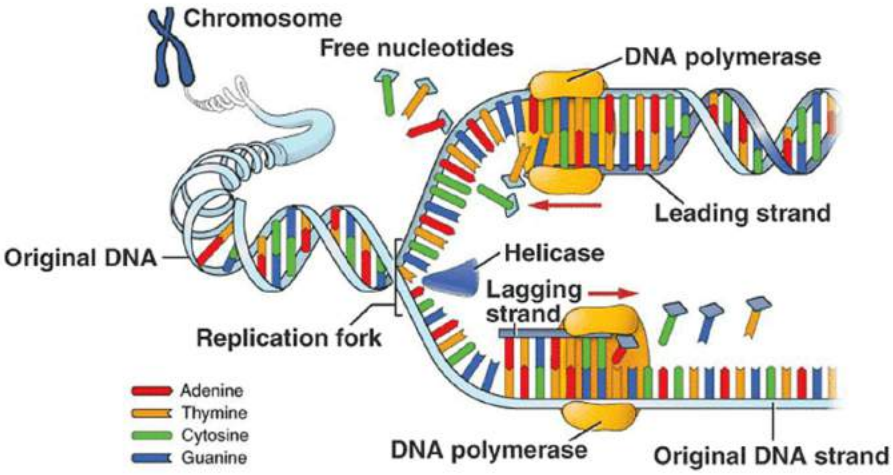
Factors involved in DNA replication and found on template ssDNAs makeup replication machineries.

- Primosotors include replication enzymes such as DNA polymerase, DNA helicases, DNA clamps, and DNA topoisomerases, as well as replication proteins such as single-stranded DNA binding proteins (SSB).
- These components work together in replication machines. All of the components involved in DNA replication are found on replication forks in most bacteria, and the complexes remain on the forks during DNA replication.
- Replisomes, or DNA replicase systems, are the replication machines. These words refer to proteins found on replication forks in general. Replisomes are not generated in eukaryotic and some bacterial cells.
- Replication factories are so named because replication machines do not move in relation to template DNAs like factories do.
- It's a process that's catalysed by enzymes. Artificial DNA primers and DNA polymerase, the principal enzyme in the replication process, can be employed to start DNA synthesis at known sequences in a template DNA molecule.
- DNA Replication Steps: Initiation, elongation, and termination are the three enzymatically catalysed and coordinated phases in DNA replication, as they are in other biological polymerisation processes.
    - **Initiation:** DNA synthesis begins at specified locations along the DNA strand known as 'origins,' which include specific coding sequences. The origin of replication (ori) with all its regulatory elements make up the replicon. The ori is where DNA replication starts, allowing plasmids to self-replicate in order to live within cells. Initiator proteins target these origins, which then attract other proteins to enhance the replication process, producing a replication complex surrounding the DNA origin. Multiple origin sites occur inside the DNA structure; these locations are referred to as replication forks when DNA replication begins.The DNA

helicase is found inside the replication complex. The double helix is unwound and each of the two strands is exposed, allowing them to be utilised as a template for reproduction. This is accomplished by hydrolysing the ATP needed to construct the nucleobase-to-nucleobase connections, therefore dissolving the link that holds the two strands together.

- **Elongation:** DNA Polymerase can begin synthesizing new strands of DNA to match the template strands once it has linked to the two unzipped strands of DNA (i.e., the template strands). Only free nucleotides can be added to the 3′ end of the primer by DNA polymerase.The new strand will be generated in a 5′ to 3′ direction since one of the template strands is read in a 3′ to 5′ direction. The leading strand refers to the freshly created strand. To launch DNA polymerase along the leading strand, DNA primase only has to synthesise an RNA primer once, at the start. This is due to the fact that DNA polymerase may lengthen the new DNA strand by reading the template 3′ to 5′ and synthesizing in the 5′ to 3′ direction, as mentioned previously.

- The lagging strand, on the other hand, is antiparallel and is read in a 5′ to 3′ direction. Continuous DNA synthesis in the 3′ to 5′ direction, as in the leading strand, would be impossible due to DNA polymerase's inability to add nucleotides to the 5′ end. Instead, RNA primers are added to the newly exposed bases on the lagging strand as the helix unwinds, and DNA synthesis occurs in fragments, but still in the 5′ to 3′ orientation as before Termination: The process of extending new DNA strands continues until either no more DNA template strands can be replicated (i.e., at the chromosome's end) or two replication forks meet and terminate. The meeting of two replication forks is uncontrolled and occurs at random throughout the chromosome's length.The freshly synthesized strands are bound and stabilized when DNA synthesis is completed. Two enzymes are required to stabilize the lagging strand: RNAse H removes the RNA primer at the start of each Okazaki fragment, and DNA ligase binds the fragments together to form a single strand.



Enzymes Machinery and Enzymes of DNA Replication

Enzymes play an important role in DNA replication. DNA replication is aided by a variety of enzymes, including DNA-dependent DNA polymerase, helicase, and ligase.

- **DNA helicase:** Helix destabilizing enzyme is another name for it. At the Replication Fork, helicase divides the two DNA strands.
- **DNA polymerase:** During DNA replication, this enzyme catalyzes the addition of nucleotide substrates to DNA in the 5′ to 3′ orientation. In addition, he proofreads and corrects errors. DNA polymerase comes in a variety of forms, each of which serves a particular purpose in various types of cells.
- **Single-stranded Binding Proteins:** Bind to ssDNA to inhibit the DNA double helix from re-annealing after DNA helicase unwinds it, preserving strand separation and promoting strand synthesis.
- **Topoisomerase:** Relaxes the DNA, which is normally super-coiled.
- **DNA gyrase:** DNA helicase relieves the strain of unwinding; this is a special sort of topoisomerase.
- **Primase:** Provides a beginning place for DNA polymerase to initiate synthesis of the new DNA strand from RNA (or DNA).
- **DNA Ligase:** DNA ligase is a commercially available enzyme that is extracted from E.coli and Bacteriophage and is used in recombinant DNA technology. DNA ligase is an enzyme that combines DNA fragments using a cloning vector.
- **Exonuclease:** Exonucleases are a group of enzymes that cleave nucleotides from the 3′ or 5′ ends of DNA and RNA strands one at a time. Endonucleases, on the other hand, hydrolyze internal phosphodiester bonds, whereas this action does not.

What are the 2 Enzymes Used in DNA Replication...?

Unwinding of the template strand and polymerisation of the daughter strands are the two fundamental steps in DNA replication. As a result, the replicative helicase and polymerase are the two primary "workhorse" enzymes in the replisome.

- Replicative DNA Polymerases: With the help of other enzymes, it aids in polymerisation, catalyzes, and regulates the entire DNA replication process. The replication process uses deoxyribonucleoside triphosphates as both a substrate and an energy source.
- Replicative DNA Helicases: The double-stranded helix must be unfolded to disclose a single-stranded template for DNA polymerases to operate. The replicative helicase is in charge of this task. The replicative helicase in eukaryotes is a hexameric complex made up of minichromosome maintenance proteins.

Conclusion

Factors involved in DNA replication and found on template ssDNAs makeup replication machineries. Promoters include replication enzymes such as DNA polymerase, DNA helicases, DNA clamps, and DNA topoisomerases, as well as replication proteins such as single-stranded DNA binding proteins (SSB). These components work together in replication machines. All of the components involved in DNA replication are found on replication forks in most bacteria, and the complexes remain on the forks during DNA replication. Replisomes, or DNA replicase systems, are the replication machines.