COURSE: MSc Part -II

PAPER – XI

TOPIC- Molecular Biology

PREAPARED BY: Ajai Kishore Sharan

Topic-1

One gene one enzyme hypothesis

One gene–one enzyme hypothesis, idea advanced in the early 1940s that each gene controls the synthesis or activity of a single enzyme. The concept, which united the fields of genetics and biochemistry, was proposed by American geneticist George Wells Beadle and American biochemist Edward L. Tatum, who conducted their studies in the mold *Neurospora crassa*. Their experiments involved first exposing the mold to mutation-inducing X-rays and then culturing it in a minimal growth medium that contained only the basic nutrients that the wild-type, or nonmutated, strain of mold needed to survive. They found that the mutant strains of mold required the addition of specific amino acids to the minimal medium to grow. Using this information, the researchers were able to associate mutations in specific genes to the disruption of individual enzymes in the metabolic pathways that normally produced the missing amino acids. This discovery won Beadle and Tatum the 1958 Nobel Prize for Physiology or Medicine (shared with American geneticist Joshua Lederberg)

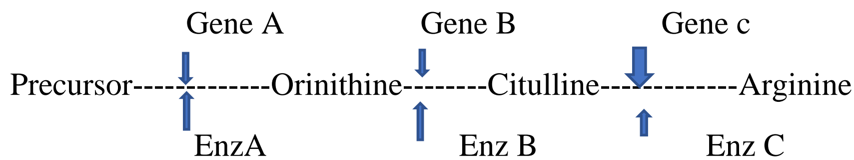
Beadle and Tatum first created Neurospora mutants by irradiating Neurospora with x-rays. They subsequently germinated sexual spores in tubes of a complete medium, or physical environment, which contained amino acids, vitamins, and other organic substances. They then transferred Neurospora to tubes of a minimal medium, which lacked some of the nutrients that the Neurospora needed to survive. Beadle and Tatum re-examined any Neurospora mutants that failed to grow in the second, minimal medium to determine whether any new growth factor requirements had been induced. In almost all cases in which a mutant was unable to survive in the minimal medium, Beadle and Tatum remedied the failure to grow by adding a chemical—either a vitamin or a specific amino acid—to the medium. The results suggested that these chemicals, which were products of genes, were necessary for the genes to encode a required enzyme in a biochemical pathway. In 1941 Beadle and Tatum published their results in "Genetic control of biochemical reactions in Neurospora," in which Beadle proposed the one gene–one enzyme hypothesis.

Beadle and latum found three types of auxotroph’s requiring amino acids ornithine, citrulline and arginine. The Prototrophs were found to have amino acid arginine in their body. Obviously, it has been synthesized from ammonia and sugar of the minimal medium. Auxotroph requiring ornithine for its growth does not contain arginine and dies due to protein deficiency. When supplied with ornithine, it is found to possess arginine. Auxotroph requiring citrulline possesses ornithine but no arginine. When citrulline is supplied, the auxotroph comes to have arginine. The nutritional mutant requiring arginine contains both ornithine and citrulline. It seems that arginine is synthesized from ammonia and sugar of the minimal medium through at least three steps each requiring its own enzyme.

Beadle and Tatum reasoned that defective enzymes are due to defective or mutant genes. Hence, genes express their effect by controlling the synthesis of enzymes. In 1948, Beadle and Tatum

The famous experiment conducted by Beadle and Tatum could draw following conclusion

1. Wild type can grow on Minimal media because they contain all the genes which is functional hence it can grow on minimal requirement of nutrient containing carbohydrate, nitrogenous substance. The wild type is also called as Prototroph.
2. Beadle and Tatum named the putative Mutant as Mutant 1, Mutant 2 and Mutant 3.
3. Mutant 1 required orinithine for Normal growth when added in Minimal media.
4. Mutant 2 required Citrulline for growth in Minimal media.
5. Mutant 3 required Arginine for growth in Minimal media.



One-Gene One-Polypeptide Hypothesis:

One gene one enzyme hypothesis has some drawbacks:

(i) All genes do not produce enzymes or their components. Some of them control other genes.

(ii) Enzymes are generally proteinaceous in nature, but all proteins are not enzymes.

(iii) Some RNAs also show enzyme activity.

(iv) A protein or enzyme molecule may consist of one or more type of polypeptides.

Yanofsky et. al. (1965) found that the enzyme tryptophan synthetase of bacterium *E. coli* consists of two separate polypeptides. A and B polypeptide A is of α-type while polypeptide B is of β-type. The synthesis of the two polypeptides is controlled by two different genes, trp-A and trp-B.

A change in any one of the two genes causes inactivation of tryptophan synthetase by stopping the synthesis of α and β-polypeptide. Inactivation of enzyme stops the synthesis of tryptophan from Indole-3-glycerol phosphate and serine.

A similar situation is seen in case of the formation of haemoglobin molecule. Haemoglobin consists of four polypeptides, 2α and 2β. Synthesis of the two types of polypeptides is controlled by two different genes located on two different chromosomes.

In view of the above-mentioned examples, "one gene one enzyme" hypothesis was changed to "one gene one polypeptide" hypothesis. According to this hypothesis "a structural gene specifies synthesis of a single polypeptide."

Topic-2

Genetic code and Wobble hypothesis

Cell is the structural and functional unit. Whereas, the structure is made by unique assembly of different and complex compounds, the function is carrying out of the metabolic processes to provide inputs for life. Both the processes are the adoptions of guided principle provided by nature. When a cell is structurally formed its next job is to carry out metabolic process. Making of a cell and its functional attribute is carrying out of complex chemical reaction in which several new chemical molecules are synthesized and broken to maintain the life. It has often been said that a cell is a chemical factory which requires various inputs to function. DNA is a genetic material which is the key molecule guiding the process of inheritance. One more function which is attributed to this DNA is that it is the carrier of genetic information. This information is being carried from cell to cell and from generation to generation. The way the genetic information exists in DNA molecule is chemical. So, it is necessary to understand the chemical nature of DNA before understanding Genetic Code.

A DNA molecule is composed of three kinds of moieties:

(i) Phosphoric acid

(ii) Deoxyribose sugar and

(iii) Nitrogen bases

The genetic information may be written in any one of the three moieties of DNA. On close examination of DNA molecule and its union during making it appears that sugar (Deoxyribose) and phosphate (Phosphoric acid) the two structural component of DNA is similar in all the nucleotides and hence it appears unlikely that these moieties of Sugar-phosphate carry the genetic information. Option falls upon nitrogen bases, the nature of which varies from one nucleotide to another. It appeared that genetic information might well depend on their sequences. The sequences of nitrogen bases (A, T, C, G) of a given segment of DNA molecule, has been found to be identical to linear sequence of amino acids in a protein molecule. The first proof of this collinearity between DNA nitrogen bases and amino acid sequence in protein molecule was first obtained during assembly of head protein of bacteriophage T4 and in *Escherichia coli*. The collinearity of amino acid molecule and DNA nucleotides has given the clue that the specific arrangement of four nitrogen bases (e.g. A, T, C and G) in DNA polynucleotide chains determines the sequence of amino acids in protein molecules. Hence, it was proposed that these four DNA bases can be considered as four alphabets of DNA molecule concerned with coded language of DNA. It was further suggested by molecular biologists that using alphabets as the genetic information may invite complex system of adoption of principles of English grammar and hence it appears that genetic information existed in DNA molecule in the form of certain language of code words. These code words utilize the services of four O nitrogen bases of DNA for its symbols.

Assignment of code to amino acid

The basic problem of a four lettered genetic code is to indicate now information written in a four-letter language (A, T, C, G) can be translated into twenty letter language (twenty amino acids of proteins). The simplest code is a singlet code (a code of single letter) in which one nucleotide codes for one amino acid. Such a code is inadequate for two amino

acids as only four amino acid will be utilized. A doubled code is also inadequate because it could specify only sixteen (4 X 4) amino acids.

	A	G	C	U
A	AA	AG	AC	AU
G	GA	GG	GC	GU
C	CA	CG	CC	CU
U	UA	UG	UC	UU

A triplet code could specify sixty-four (4X4X4) amino acids. There it is likely that there may be 64 triplet codes for 20 amino acids.

	U	C	A	G	
U	UUU	UCU	UAU	UGU	U
	UUC	UCC	UAC	UGC	C
	UUA				A
	UUG				G
C					U
C					
A					
G					
A					U
C					
A					
G					
G					U
C					
A					
G					

The first evidence to support the concept of triplet codes is provided by crick and his co-workers in 1961. When they added or deleted single or double base pairs in a bacteriophage T4 infecting *E. coli*, bacteriophages ceased to perform normal functions. However, bacteriophages with addition or deletion of three base pairs in DNA molecule, had performed normal functions. The experiment carried out by crick proposed that genetic code is in triplet form because addition of one or two nucleotides as put the reading of the code out of order, while addition of third nucleotide resulted in a return to the proper reading message. The group of nucleotides that specifies one amino acid is code word or codon.

Experimental proof that genetic code is triplet

Three hypotheses were put forward as a perspective idea to prove that genetic code is triplet in nature.

(i) The genetic code is overlapping

This hypothesis was negated as this can cause statistical regularities between adjacent acid residues. A single point mutation may lead to change in two adjacent residues. This possibility was eliminated.

(ii) Punctuated code

(iii) Unpunctuated code

Deletions of three nucleotides (or multiple of there) will restore the reading frame. This is the actual form of the code.

Most of our current knowledge of the general nature of the genetic code and the nucleotide composition has been obtained from the four main types of experimental approaches.

A. Polyuridytic acid method

Initial breakthrough in deciphering of the genetic code was achieved by Marshall Nierenberg and J. Heinrich Matthaei in 1961. They used enzyme polynucleotide phosphorylase to construct a synthetic polyribonucleotide containing only one kind of base (Uracil – found in mRNA) and added it to a cell free amino-acid-incorporating system from *E. Coli*. A polyuracil (UUU.....UUU) was prepared and to which they added different amino acids labelled with C14. Hence, this poly U system specifically stimulated the formation of a polypeptide which contained only the amino acid phenylalanine. This was concluded that code word for the amino acid phenylalanine was a sequence of three uracil nucleotides (UUU). Nirenberg influenced by the success of this work got deeply involved in this problem with Severo Ochoa to assign codon for lysine as AAA (poly A) and CCC (poly C) for proline. Attempt to assign poly G as a codon for specific amino acid failed, because GGG soon acquires a secondary structure and cannot attach to ribosomes.

B. Copolymer method

In this method, Nirenberg used mixtures of two or more ribonucleotides diphosphates and with the help of enzyme phosphorylase, prepared polyribonucleotide's. Using UDP (uracil diphosphate) and cytosine diphosphate (CDP) in the ratio 3:1, he obtained a polynucleotide which contained the triplets UUU, UUC, UCU and CUU. The triplets containing 2Cs and 1U were the least frequent. In such a poly-UC system, Nirenberg obtained polypeptide containing the amino acids phenylalanine and serine in the ratio of 3:1. This method, however, did not give the exact sequence of three bases.

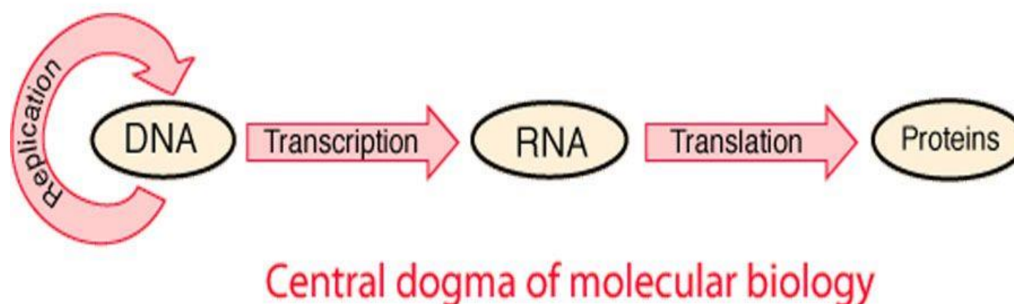
C. Binding technique Marshall W. Nirenberg and Philip Leeder in 1964 found that if a synthetic trinucleotide for a known sequence is used with ribosome and a particular aminoacyl–tRNA (t–RNA having its own specific amino acid attached), a complex will be formed provided the used codon codes for the amino acid attached to the given aminoacyl tRNA.

Codon + Ribosome + Amino acid (AA) + t RNA Ribosome-codon-AA-t RNA (complex)

Whereas, AA-tRNA can pass through nitrocellulose membrane, the ribosome - Codon - AA - t RNA complex adsorbs to the membrane. If a mixture only one of the amino acids is made radioactive, then the presence or absence of the radioactivity on the nitrocellulose membrane will show whether there is a relationship between the Codon and the amino acid which was made radioactive. Such a treatment was given by Nirenberg and his co-workers to all the 64 synthetic cordons and their respective amino acids were identified. The binding of AA-tRNA was not equally efficient in all cases; hence, sequences of bases in only 45 cordons could be worked out by this method.

4. Repetitive sequencing technique

This method of confirming the genetic code was proposed by Har Gobind Khorana and is the most authentic one. The method involves in vitro (chemical synthesis) of short segments of DNA of know base sequence with the help of enzyme DNA polymerase. From this DNA, RNA of defined base sequence is transcribed with the help of enzyme RNA polymerase. The RNA sequentially helps in the synthesis of a Polypeptide (A chain of amino acid). The principle of central Dogma is thus followed as described below:



The chain of amino acid thus formed is then determined and correlated with the base sequence of DNA and RNA. Similar type of sequence of DNA thus synthesized same sequence of amino acid. Khorana, developed methods for synthesizing polyribonucleotide's with different but repeating sequences to obtain serine, argimine and glutamic acid. Hence determined the importance of repeating units of short segments of DNA and suggested that codon AAG codes for lysine, AGA for arginine and GAA for glutamic acid. The study made by Khorana thus proved that:

1. The base sequence in DNA specifies the amino acids in proteins.
2. The information contained in DNA is conveyed through RNA.
3. The genetic code is triplet is literature and is also non-overlapping in nature.

Properties and Characteristics of the genetic code

By various experimental evidence the characteristics properties of genetic code has been established as described below:

1. The genetic code in literature is triplet

The coded language of DNA for transcription is a three lathered word. There are altogether 64 codons, for 20 amino acids. The codon assignment for different amino and has been shown below:

No. of Codons	Amino acid	Total
One Codon each	Methione, Tryptophan	02
Two Codons each	Aspargine, Asparatic acid, cysteine, Glutamine, Glutamic acid, Histidine, Lysine, Phenylalanine, Tyrosine	18
Three Codons each	Isoleucine	03
Four Codons each	Alanine, Glycine, Proline, Threonine, Valine	20
Six Codons each	Arginine, Leucine, Serine	18
Terminator Codons		03
	20 amino acids	64

The code dictionary reveals certain trend of patterns

- Amino acids with similar structural properties have related codons. The difference lies in third codon. e.g. aspartic acid has a codon GAV whereas, glutamic acid has a codon GAA.
- All codons with U in the second position specify hydrophobic amino acids (Isoleucine, Leucine, Methionine, Phenylalanine and Valine).
- All the acidic (Asparagine and Glutamic acid) and basic (Arginine and Lysine) amino acids have A or G as the second base.

3. Non overlapping

The genetic code is non overlapping, this means that adjacent codons do not overlap. In actual practice, six based code for not more than two amino acids for example UUUCCC on mRNA will code only for 2 amino acids i.e. Phenylalanine (UUU) and Proline (CCC).

4. Comma less

There is no comma (,) at the end of one codon and the begining of the next. This means that after one amino acid is coded, the second amino acid will automatically coded by the next three letters.

5. Non-ambiguity

Non-ambiguity of the codon means, a particular codon will always code for the same amino acid.

6. Universality

The genetic code is same for all the living organism. This means that same sequence of 3 bases encode for same amino acids in all life forms. Thus, UUU codes for phenylalanine and GUC for valine in all living things. The genetic code which was developed in the bacteria about 3 billion (300 crore) years ago has not undergone any change and has been preserved in its almost original form in the course of evolution.

7. Polarity

The genetic code has polarity (direction), that is, the code is always read in a fixed direction i.e. in the 5` 3` direction.

5`UUGAUC GUC UCG3`

Leu Ile Val Serine

8. Chain initiation codons

The triplet codons AUG and GUG play vital role in Escherichia coli. When these codons occur in between the two ends of a functional gene (cistron) they code for the amino acid methionine and valine. When they occur immediately after a terminaton codon (stop codon) they act as chain initiation or starter codons for the synthesis of a polypeptide chain (chain of amino acids). The codon AUG which codes for amino acid have two functions to perform:

1. When initiates protein synthesis it acts as formylated methionine but when lies in the intermediate position.

2. Hence there is a seperate species of tRNA O for initiating methionine called as tRNA met and intermediate methionine.

9. Chain termination codon

The triplets UAA, UAG and UGA do not recognize any amino acid. So, they are called as non-sense codons, as against the remaining 61 codons, which are termed as sense codons. These three codons perform the function of punctuating genetic message like a full stop at the end of the sentence. The codons UAA and UAG were discovered in bacteria and were respectively associated with the ochre and amber mutations. The codon UGA is also called as opal. The termination codons causes the release (break) of the polypeptide chain from the ribosome hence the name terminator.

Multiple recognition of codons and Wobble hypothesis

Crick (1966) proposed the 'wobble hypothesis' to explain the degeneracy of the genetic code. Except for tryptophan and methionine, more than one codon directs the synthesis of one amino acid. There are 61 codons that synthesise amino acids, therefore, there must be 61 tRNAs each having different anticodons. But the total number of tRNAs is less than 61.

This may be explained that the anticodons of some tRNA read more than one codon. In addition, identity of the third codon seems to be unimportant. For example, CGU, CGC, CGA and CGG all code for arginine. It appears that CG specifies arginine and the third letter is not important. Conventionally, the codons are written from 5′ end to 3′ end.

Therefore, the first and second bases specify amino acids in some cases. According to the Wobble hypothesis, only the first and second bases of the triple codon on 5′ → ‘3 mRNA pair with the bases of the anticodon of tRNA i.e A with U, or G with C.

The pairing of the third base varies according to the base at this position, for example G may pair with U. The conventional pairing (A = U, G = C) is known as Watson-Crick pairing (Fig. 7.1) and the second abnormal pairing is called wobble pairing.

This was observed from the discovery that the anticodon of yeast alanine-tRNA contains the nucleoside inosine (a deamination product of adenosine) in the first position (5′ → 3′) that paired with the third base of the codon (5′ → 3′). Inosine was also found at the first position in other tRNAs e.g. isoleucine and serine.

The purine, inosine, is a wobble nucleotide and is like guanine which normally pairs with A, U and C. For example, a glycine-tRNA with anticodon 5′-ICC-3′ will pair with glycine codons GGU, GGC, GGA and GGG. Similarly, a seryl-tRNA with anticodon 5′-IGA-3′ pairs with serine codons UCC, UCU and UCA (5-3′). The U at the wobble position will be able to pair with an adenine or a guanine.

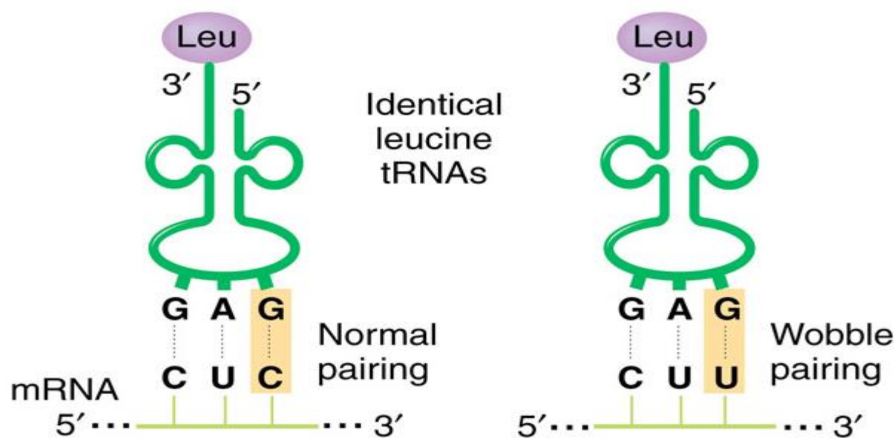
: Due to the Wobble base pairing one tRNA becomes able to recognise more than one codons for an individual amino acid. By direct sequence of several tRNA molecules, the wobble hypothesis is confirmed which explains the pattern of redundancy in genetic code in some anticodons (e.g. the anticodons containing U, I and G in the first position in 5’→ 3′ direction)

The seryl-tRNA anticodon (UCG) 5′-GCU-3′ base pairs with two serine codons, 5′-AGC-3′ and 5′-AGU-3′. Generally, Watson-Crick pairing occurs between AGC and GCU. However, in AGU and GCU pairing, hydrogen bonds are formed between G and U. Such abnormal pairing called ‘Wobble pairing’ is given in Table 7.5.

Three types of wobble pairings have been proposed:

(i) U in the wobble position of the tRNA anticodon pairs with A or G of codon,

(ii) G pairs with U or C, and(iii) 1 pairs with A, U or C.



Topic-3

Replication of DNA in Prokaryotes and Eukaryotes

Continuity of life is based upon the principle of transfer of genetic material laden with genetic information to the next generation. So, a parental cell before entering next generation must pass on its genetic information to a newly synthesized genetic material. This is an important event in the life of a cell, which is represented by S phase (synthesis phase) in the life of a cell. A cell can produce two daughter cells before undergoing cell division. During S phase the duplication of DNA takes place when a diploid cell (2N) becomes tetraploid (4N). This tetraploid cell, soon, get divided into two 2N (Diploid cells). The first question raised was that the genetic information to the newly synthesized DNA can be either semi conservative (one strand from the parent) or conservative (both the strands from the parent). Three models were also proposed to explain the replication of DNA as:

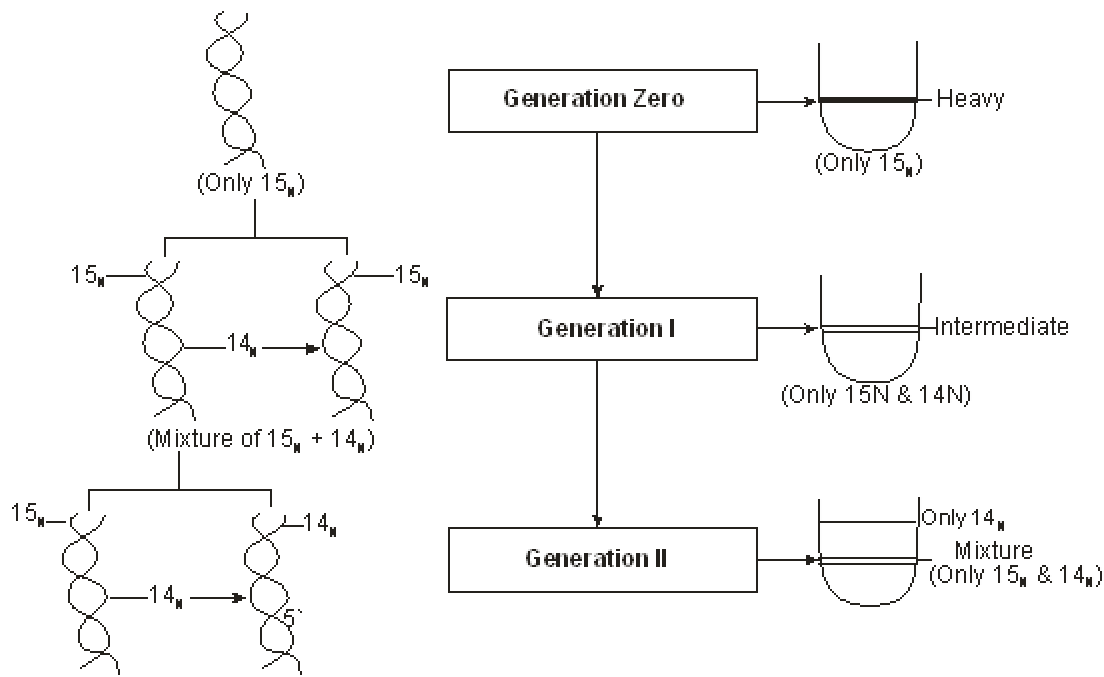
1. Semi conservative replication: One strand is new, while the other is parental.
2. Conservative replication: Both the strands are of parental origin.
3. Dispersive: Both the strands of DNA acquires fragments of parental DNA as well as newly synthesized DNA.

Evidence to support conservative mode of replication and dispersive mode of replication could not be obtained, but evidence to support semi-conservative mode of replication could be obtained by the classical experiment of Meselson and Stahl.

The Experiment

M. Meselon and F.W. Stahl (1958) differentiated parental and newly formed DNA, by incorporating normal isotope of Nitrogen (N^{14}) or heavy isotope (N^{15}). During metabolism nitrogen gets accumulated in the DNA. Culture media (for growth of bacteria) was supplemented with N^{14} and N^{15} for growth and development.

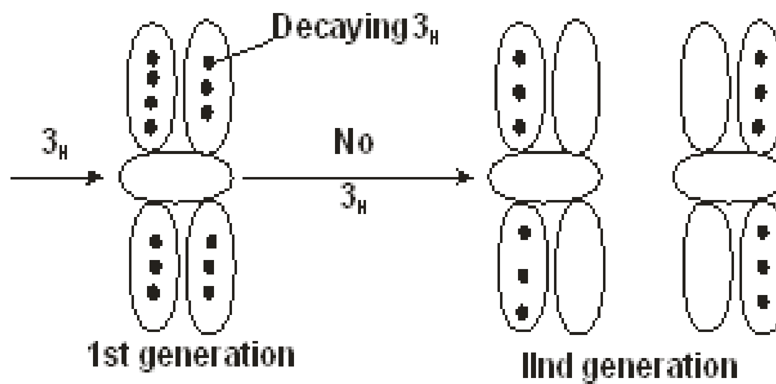
E. coli was cultivated in culture media containing N^{15} as nitrogen source for several generations. DNA was extracted and centrifuged by density gradient centrifugation (separates molecules by its density) to obtain a heavy band at the bottom of the centrifuge tube (named as generation zero). Sample was withdrawn from the N^{15} growing culture and transferred to culture media which contained N^{14} source of nitrogen. DNA was again isolated and centrifuged to obtain a band in the centrifuge tube which also shifted its position (slightly raised) and hence called as Intermediate band (named as generation I). Sample withdrawn from N^{14} culture media was sub-cultured on N^{14} media a fresh. DNA isolated from *E. coli* cells demonstrated two tubes of band (named as generation II) as light band (upper most) and intermediate band. Interpretation of Meselson and Stahl report can be shown by Figure given below



The experiment conducted by Meselson and Stahl proved that Replication of DNA is semi conservative in *E. coli*. This raised doubt regarding the mode of replication of DNA in Eukaryotes.

Replication of Chromosomes in Eukaryotes

J.H. Taylor (1958) demonstrated the semi-conservative mode of replication in DNA and hence in the chromosomes. Root tip of *Vicia faba* a member of family Fabaceae was used as experimental material. The technique of Autoradiography was used where radioactive material is used in biological system. Root tips were first fed with radioactive thymidine 3^{H} (half-life few months and then it decays). This treatment was marked as first generation. After this generation root tip (another set) was transferred to Colchicine (to induce cell division) marked as IInd generation. Slide was prepared after every generation and left as such at low temperature to all 3^{H} to decay. Based on observation made in the slide semi-conservative mode of DNA (chromosome) was established as shown in Figure given below.

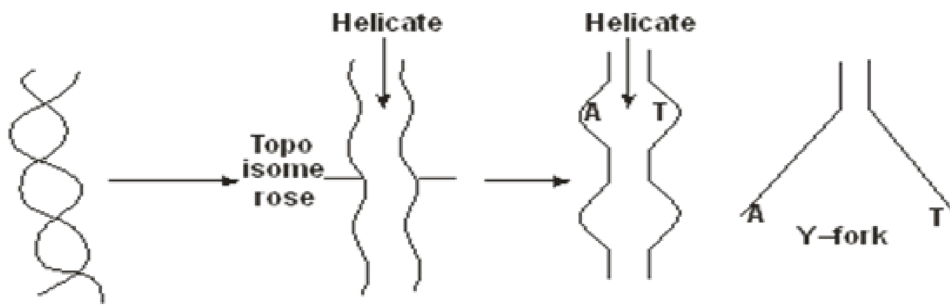


Both the chromatids in 1st generation have black dots (leaving dots of 3H during decay) but in generation II, one chromatid returned to parental type, while others have black dots. This proved that during replication of chromosome, one chromosome is old while other is fresh (new). This justifies the semi-conservative mode.

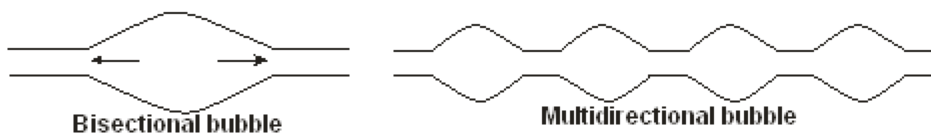
Mechanism of Replication of DNA

Replication occurs at specific times during the cell cycle. This requires soft of protein, enzymes and energy in the form of ATP. The Replication begins with initiation and ends in elongation. The initiation begins with separation of weak hydrogen bonds between the Nitrogenous bases. These two strands of DNA have nitrogenous bases complementary to each other (A to T and C to G). The separation of two strands is difficult due to their ant parallel nature and torsion generated by complementary strands. The point from where unwinding of DNA begins is called as initiation point. In viruses and also in bacteria the replicating unit in only one hence origin of replication initiates at only one point. IN Eukaryotes with large DNA molecule, there may be more than one initiation point, and each initiation point finally merge with one another.

Unwinding of the double helical ant parallel DNA is the first condition of replication. The unwinding of the DNA takes place by break down of hydrogen bonds with the help of enzyme helicases. To minimize the torsion effect of DNA helix, the enzyme Topoisomerases cut and re-joins the DNA strand to relax to torsion effect, leading to separation of DNA strand. Due to this unzipping of DNA strands replication bubbles are formed which subsequently extends as Y–shaped replication fork.



The formation of replication bubble and Y-fork in bacteria and phages is bidirectional but in eukaryote it multidirectional

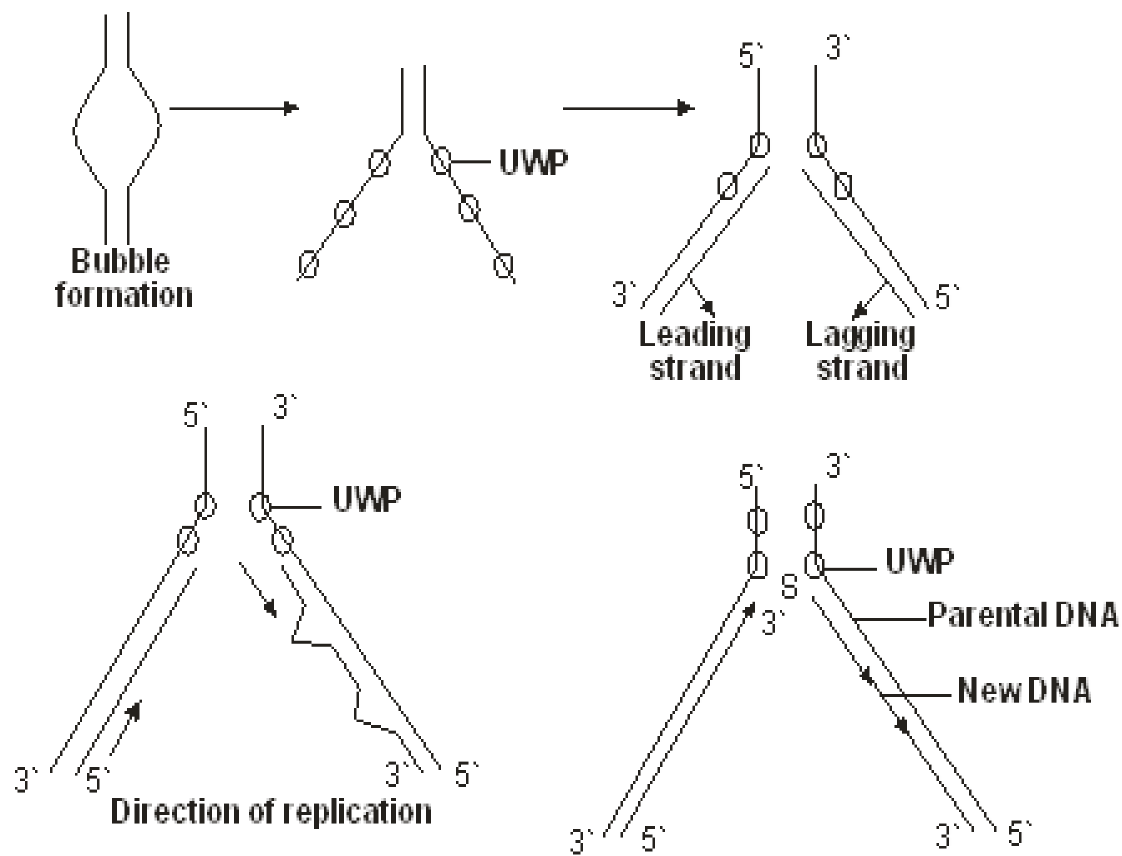


After the formation of replication fork, the unwinding protein called single strand binding protein (SSB protein) keeps the separated strand apart.

DNA directed RNA polymerase forms the RNA primer. This RNA primer forms hetero duplex structure (DNA–RNA) for sometimes. The RNA is ultimately removed enzymatically, and the gap is filled up with newly synthesized nucleotides. The enzyme DNA polymerase polymerises the nucleotides only 5`–3` direction. Because the two strands of DNA are in antiparallel direction, the two strands have to be synthesized by growing in opposite direction.

The enzyme synthesizes a new strand in a continuous pieces 5`–3` direction and is called leading strand on the other strand of DNA, the enzyme forms DNA fragments in small pieces again 5`–3` direction called as Okazaki fragments.

The Okazaki fragment soon joins with the help of enzyme DNA ligase to be called as lagging strand. The enzyme DNA primase helps in addition of new strand to the DNA template whereas enzyme exonuclease (a type of DNA polymerase) helps in removal of RNA primer. The process of DNA replication has been summarized in Figure given below



UWP= Unwinding Protein

During replication of DNA, wrong base may get incorporated, which must be removed by another set of enzymes called as DNA polymerases. This enzyme removes the wrong base and performs editing. The enzyme nuclease (a type of DNA polymerase) cuts off the wrong base and replaces it with the correct one.