MICROBIAL GENETICS AND MOLECULAR BIOLOGY

M.Sc. MICROBIOLOGY

SEMESTER-II, PAPER-II

LESSON WRITERS

Prof. A. Amruthavalli
Professor
Department of Botany & Microbiology,
Acharya Nagarjuna University

Prof. K. Mallikarjuna
Professor
Department of Botany & Microbiology,
Acharya Nagarjuna University

Dr. J. Madhavi
Assistant Professor
Department of Botany & Microbiology,
Acharya Nagarjuna University

Dr. K. Nagaraju
Faculty
Department of Botany & Microbiology,
Acharya Nagarjuna University

LESSON WRITER &EDITOR

Prof. V. Umamaheswara Rao
Professor
Department of Botany & Microbiology,
Acharya Nagarjuna University

ACADEMIC ADVISOR

Prof. V. Umamaheswara Rao
Professor
Department of Botany & Microbiology,
Acharya Nagarjuna University

DIRECTOR, I/c.

Prof. V. Venkateswarlu
M.A., M.P.S., M.S.W., M.Phil., Ph.D.
Professor

CENTRE FOR DISTANCE EDUCATION
ACHARYA NAGARJUNA UNIVERSITY
Nagarjuna Nagar 522 510

Ph: 0863-2346222, 2346208
0863-2346259 (Study Material)
Website www.anucde.info
E-mail: anucdedirector@gmail.com

M.Sc. MICROBIOLOGY: MICROBIAL GENETICS AND MOLECULAR BIOLOGY

First Edition : 2025

No. of Copies :

This book is exclusively prepared for the use of students of M.Sc. Microbiology, Centre for Distance Education, Acharya Nagarjuna University and this book is meant for limited circulation only.

Published by:

Prof. V. VENKATESWARLU
Director, I/c
Centre for Distance Education,
Acharya Nagarjuna University

Printed at:

FOREWORD

Since its establishment in 1976, Acharya Nagarjuna University has been forging ahead in the path of progress and dynamism, offering a variety of courses and research contributions. I am extremely happy that by gaining 'A+' grade from the NAAC in the year 2024, Acharya Nagarjuna University is offering educational opportunities at the UG, PG levels apart from research degrees to students from over 221 affiliated colleges spread over the two districts of Guntur and Prakasam.

The University has also started the Centre for Distance Education in 2003-04 with the aim of taking higher education to the door step of all the sectors of the society. The centre will be a great help to those who cannot join in colleges, those who cannot afford the exorbitant fees as regular students, and even to housewives desirous of pursuing higher studies. Acharya Nagarjuna University has started offering B.Sc., B.A., B.B.A., and B.Com courses at the Degree level and M.A., M.Com., M.Sc., M.B.A., and L.L.M., courses at the PG level from the academic year 2003-2004 onwards.

To facilitate easier understanding by students studying through the distance mode, these self-instruction materials have been prepared by eminent and experienced teachers. The lessons have been drafted with great care and expertise in the stipulated time by these teachers. Constructive ideas and scholarly suggestions are welcome from students and teachers involved respectively. Such ideas will be incorporated for the greater efficacy of this distance mode of education. For clarification of doubts and feedback, weekly classes and contact classes will be arranged at the UG and PG levels respectively.

It is my aim that students getting higher education through the Centre for Distance Education should improve their qualification, have better employment opportunities and in turn be part of country's progress. It is my fond desire that in the years to come, the Centre for Distance Education will go from strength to strength in the form of new courses and by catering to larger number of people. My congratulations to all the Directors, Academic Coordinators, Editors and Lesson-writers of the Centre who have helped in these endeavors.

Prof. K. Gangadhara Rao
M.Tech., Ph.D.,
Vice-Chancellor I/c
Acharya Nagarjuna University.

M.Sc. MICROBIOLOGY

SEMESTER-II, PAPER-II

202MB24-MICROBIAL GENETICS AND MOLECULAR BIOLOGY

SYLLABUS

Unit-I

Experimental evidences for establishing the nucleic acids as genetic material.

Gene concept- classical concept, Modern concept (cistron, muton, recon, exon & intron); Different theories of gene concept.

Plasmids - Definition, characteristics of plasmids, types of plasmids, properties of F plasmids, R plasmids, col plasmids, Ti plasmids and other plasmids; Replication of plasmids; Isolation of plasmids; Significance of plasmids.

Mutations spontaneous mutations versus induced mutations; Types of mutations -forward, backward, suppressor, point and frame shift mutations; Radiation induced mutations - ionizing and non-ionizing radiation. Chemical mutagens Base analogues, nitrous acid, acridines, alkylating and hydroxylating agents.

UNIT-II

Genetic recombination in Bacteria - Transformation, Conjugation, Transduction (Generalized and Specialized); Gene mapping in bacteria; Gene transfer techniques Electroporation, Microinjection, Biolistics and chemical methods.

Phage Genetics- Lytic phage - Genome organization of phage T4, genetic recombination; gene expression in T4 life cycle. Lysogenic phage - λ-phage genome organization; recombination, genetics of lysogenic life cycle.

UNIT-III

Replication of DNA - Semi-conservative replication, enzymology of replication, continuous and discontinous DNA synthesis. Unidirectional replication, bi-directional replication, rolling circle replication. ygoloidoroim bris vasto to inemhsged

DNA damage and repair - Types of DNA damage- deamination, alkylation, pyrimidine dimers; Repair mechanisms – Photo reactivation, base excision repair, nucleotide excision repair, post replication and recombination repair, methyl-directed mismatch repair and SOS repair. Gene expression - Central dogma of gene action; Transcription – initiation, elongation and termination of transcription; post transcriptional processing and RNA splicing in eukaryotes; Translation initiation, elongation and termination of translation; Post translational modifications of polypeptide.

UNIT-IV

Regulation of Gene expression in bacteria - Operon concept, Inducible and repressible operons, Inducer molecules, Repressor molecules, Co-Repressor molecules.

Induction and catabolite repression of lac operon in *E. coli.*

Repression and attenuation of trp operon in *E. coli*.

Genetics of nitrogen fixation - nif genes, regulation of nif genes (local control and global control mechanisms); nod genes and their regulation.

UNIT-V

Transposable elements in bacteria - IS elements, Composite transposons, Tn3 transposons. Transposable elements in eukaryotes - Ac and Ds elements in maize, Ty elements in yeast, transposons in Drosophila, Human retrotransposons.

Mechanisms of transposition - Conservative and replicative modes.

Genome rearrangements - mating type switching in yeast; Regulation of genome activity during sporulation by special σ subunits.

REFERENCE BOOKS:

1) Gene VII - Benjamin Lewin (2000)
2) Principles of Genetics - Gardener, E.J., Simmons, M.J., Snustad, D.P. (1991)
3) Genetics - Peter J Russell (1998)
4) Microbial Genetics - David Freifelder (1990)
5) Molecular Biology - David Freifelder (2001)
6) Molecular Genetics of Bacteria - Dale JW
7) Principles of Genetics - Tamarin RH (1999)
8) An Introduction to Genetic Analysis - Suzuki,D.T et al.
9) Molecular Cell Biology - Darnell, J., Lodish, H. Baltimore, D. (1986)
10) Cell and Molecular Biology - Karp, G. (1986)
11) Molecular Genetics of Bacteria - Dale,J.W.
12) Principles of Genetics - Snustad, D.P., Simmons, M.J., Jenkins, J.B. (1997)

(202MB24)

MODEL QUESTION PAPER

M.Sc. DEGREE EXAMINATION,
MICROBIOLOGY - SECOND SEMESTER
MICROBIAL GENETICS AND MOLECULAR BIOLOGY

Time: Three hours **Maximum: 70 marks**

Answer All Questions **5 × 14 = 70M**

UNIT-I

1) a) Give an account on Classical and Modern concept of gene.

OR

b) Describe the different types of chemical mutagens.

UNIT-II

2) a) Write an account on genetic recombination in bacteria.

OR

b) Describe the lytic cycle and genome organization of T4 phage.

UNIT-III

3) a) Give an account on semi-conservative replication of DNA.

OR

b) Explain the mechanism of protein synthesis in bacteria.

UNIT-IV

4) a) Write an account on induction and catabolite repression of lac operon in *E. coli.*

OR

b) Describe the nif gene cluster and their regulation.

UNIT-V

5) a) Give an account on important transposable elements in maize, yeast and *Drosophila*

OR

b) Explain the regulation of genome activity during sporulation by special σ units.

CONTENTS

LESSON-1

EXPERIMENTAL EVIDENCES FOR ESTABLISHING THE NUCLEIC ACIDS AS GENETIC MATERIAL

1.0 OBJECTIVE:

- This lesson deals with the experimental evidences for establishing the nucleic acids as genetic material.

STRUCTURE:

1.1 INTRODUCTION

The search for genetic material started as early as the time of Aristotle who hypothesized theories of inheritance. However, more scientific evidence was given by Mendel's work on pea plants. Moreover Griffith's experiment on *S. pneumoniae* paved the way towards finding the molecule that can be referred to as the genetic material. Further establishment of concrete evidence by Avery, McLeod and Mc Carty as well as by Hershey and Chase. Genetic material must have four important properties. DNA and RNA (nucleic acids) are the genetic materials that are proven by experiments. Genetic material has the ability to carry information from one generation of an organism to the next. As we know that evolution is possible due to heritable traits and variations there are four major properties that a genetic material should possess which are

1) The genetic molecule should be able to replicate itself.
2) The genetic molecule should be able to store information.
3) The genetic molecule should be able to express the stored information.
4) The genetic molecule should be able to undergo mutation.

1.2 DNA AS GENETIC MATERIAL

1.2.1 Griffith's experiment or Transformation Experiment in Bacteria

Frederick Griffith (1928) for the first time demonstrated genetic transformation in bacterium *Diplococcuspneumoniae,* now named *Streptococcus pneumoniae* and Avery, MacLeod and McCarty (1944) later showed that DNA and not the protein, was the carrier of hereditary characters in *Diplococcuspneumoniae* or *Pneumococcus.* This bacterium causes pneumonia in mice and men. There are two strains of *Pneumococcus.*In one strain, capsule layer (slime coat) is formed of polysaccharide material and colonies are shining and "smooth" ('S' strain). In other strain, cells lack polysaccharide slime layer and colonies formed by such cells are irregular or **"rough"** ('R' strain). Smooth (S) cells are virulent and cause pneumonia but rough (R) cells are non-virulent. 'S' pneumococci are classified into types, such as, type I-S, type II-S, type III-S and so on. The cells of 'S' colony may change occasionally into 'R' type cells but the reverse change (i.e., from R to S) is almost never seen. R cells upon division always give rise to R cells. In his experiment, Griffith injected mice with living II-R pneumococci and found that the mice did not suffer. But when the mice were injected with live III-S type they suffered by pneumonia and died and when heat killed III-S bacteria were injected the mice did not suffer. However, when the mixture of living cells of non-virulent II-R and heat killed III- S cells were injected into the mice, the mice unexpectedly developed pneumonia and died. Post-mortem examination of dead mice showed the presence of both II-R and III-S type of pneumococci in the heart blood and this led Griffith to conclude that something released from the heat killed III-S cells was taken up by the avirulent R type cells which might have caused genetic transformation of living II-R bacteria into virulent cells (Fig. 1.1).

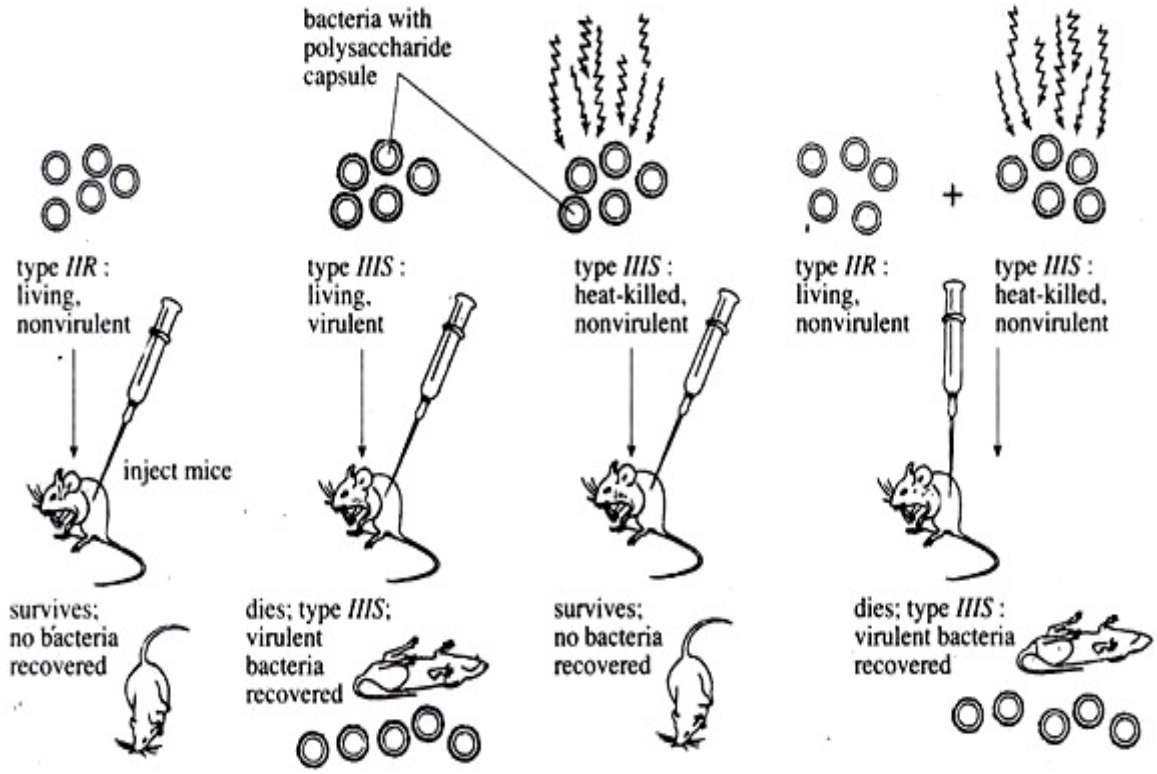


Fig-1.1: Griffth's Transformation Experiment

1.2.2 Oswald Avery, C. M. MacLeod and M. McCarty Experiment:

During World War II, in 1943, Oswald Avery, Maclyn McCarty, and Colin MacLeod working at Rockefeller University in New York, dedicated themselves to continuing the work of Griffith in order to determine the biochemical nature of Griffith's transforming principle in an in vitro system. They used the phenotype of *S. pneumoniae* cells expressed on blood agar in order to figure out whether transformation had taken place or not, rather than working with mice. The transforming principle was partially purified from the cell extract (i.e., cell-free extract of heat-killed type III S cells) to determine which macromolecule of S cell transformed the type II R-strain into type III S-strain. They demonstrated DNA to be that particular transforming principle (Fig.1.2).

- Initially, type III S cells were heat-killed, and lipids and carbohydrates were removed from the solution.
- Secondly, they treated heat-killed S cells with digestive enzymes such as RNases and proteases to degrade RNA and proteins. Subsequently, they also treated it with DNases to digest DNA, each added separately in different tubes.
- Eventually, they introduced living type IIR cells mixed with heat-killed IIIS cells onto the culture medium containing antibodies for IIR cells. Antibodies for IIR cells were used to inactivate some IIR cells such that their number doesn't exceed the count of IIIS cells that help to provide the distinct phenotypic differences in culture media that contained transformed S strain bacteria.

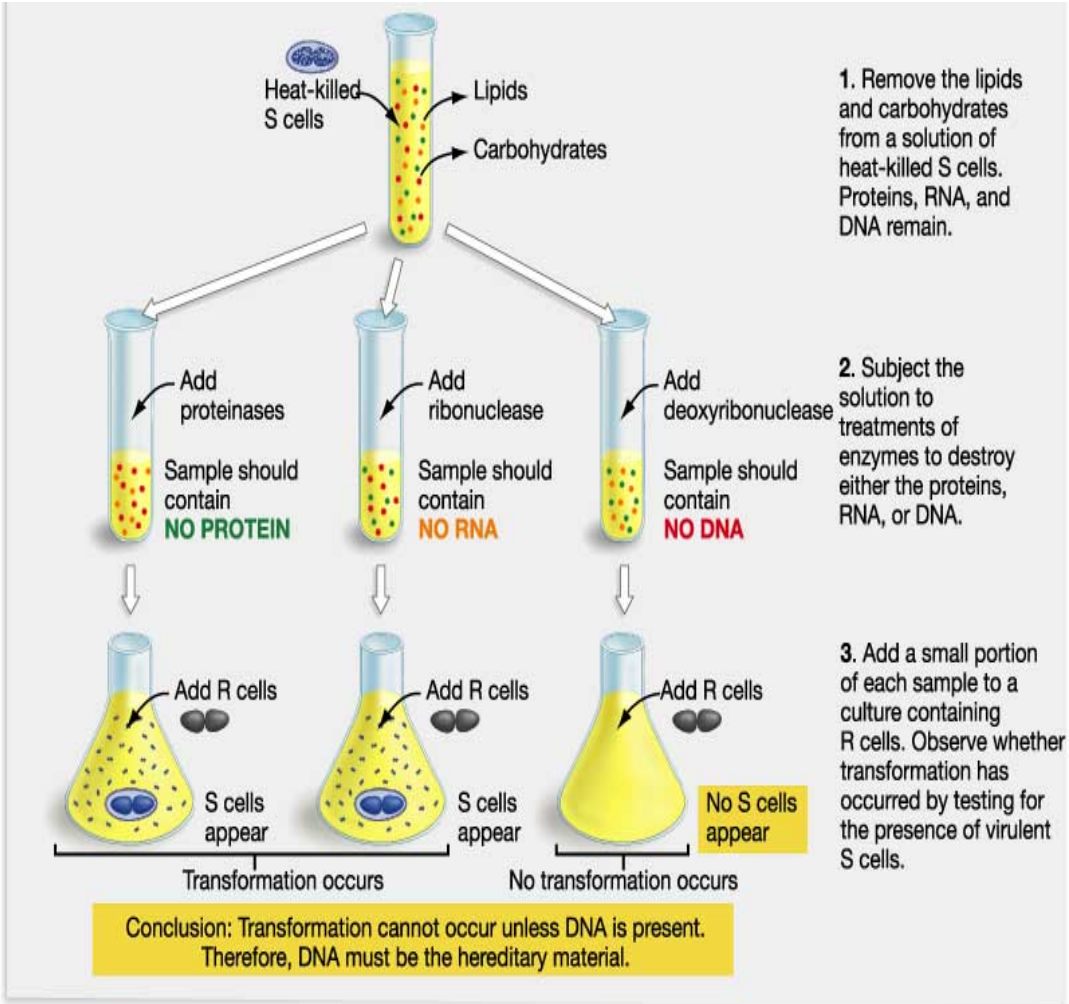


Fig-1.2: Avery, MacLeod and McCarty Experiment

1.2.3 Hershey and Chase Experiment

Although Avery and his fellows found that DNA was the hereditary material, the scientists were reluctant to accept the finding. But, not that long afterward, eight years after in 1952, Alfred Hershey and Martha Chase concluded that DNA is the genetic material. Their experimental tool was bacteriophages-viruses that attack bacteria which specifically involved the infection of *Escherichia coli* with T2 bacteriophage.T2 virus depends on the host body for its reproduction process. When they find bacteria as a host cell, they adhere to its surface and inject its genetic material into the bacteria. The injected hereditary material hijacks the host's machinery such that a large number of viral particles are released from them. T2 phage consists of only proteins (on the outer protein coat) and DNA (core) that could be potential genetic material to instruct *E. coli* to develop its progeny. They experimented to determine whether protein or DNA from the virus entered into the bacteria.

- Bacteriophage was allowed to **grow** on two of the medium: one containing a radioactive isotope of phosphorus(^{32}P) and the other containing a radioactive isotope of sulfur (^{35}S).
- Phages grown on radioactive phosphorus(^{32}P) contained radioactive P **labeled** DNA (not radioactive protein) as DNA contains phosphorus but not sulfur.
- Similarly, the viruses grown in the medium containing radioactive sulfur (^{35}S) contained radioactive ^{35}S labeled protein (but not radioactive DNA) because sulfur is found in many proteins but is absent from DNA.
- *E. coli* were introduced to be **infected** by the radioactive phages.
- After the progression of infection, the **blender** was used to remove the remains of phage and phage parts from the outside of the bacteria, followed by **centrifugation** in order to separate the bacteria from the phage debris.
- Centrifugation results in the settling down of heavier particles like bacteria in the form of pellet while those light particles such as medium, phage, and phage parts, etc., float near the top of the tube, called supernatant.

On measuring radioactivity in the pellet and supernatant in both media, ^{32}P was found in large amount in the pellet while ^{35}S in the supernatant that is pellet contained radioactively P labeled infected bacterial cells and supernatant was enriched with radioactively S labeled phage and phage parts (Fig.1.3).

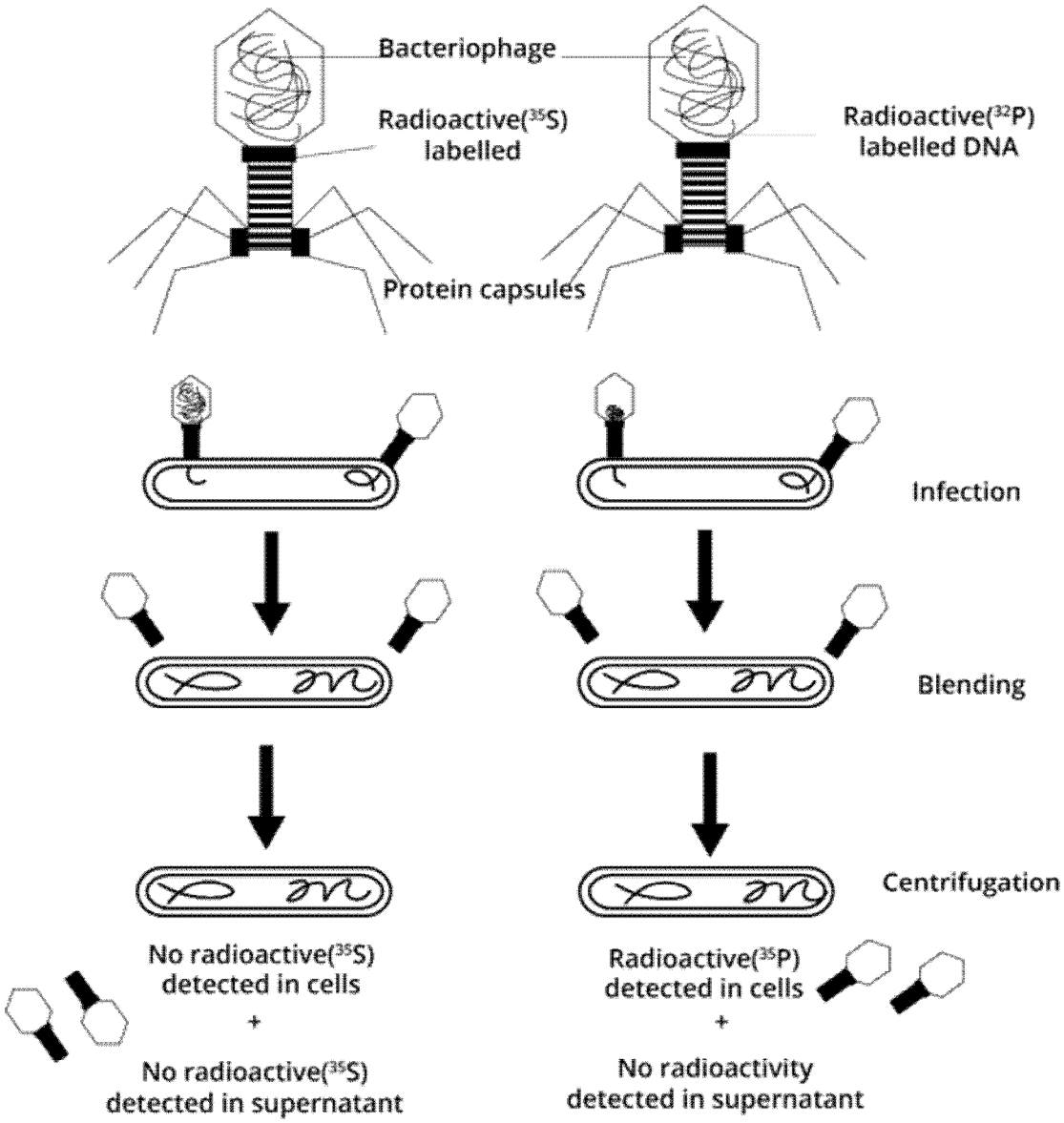


Fig-1.3: Hershey and Chase Experiment

1.3 RNA AS GENETIC MATERIAL

The genome of viruses may be DNA or RNA. Most of the plant viruses have RNA as their hereditary material. Fraenkel-Conrat (1957) conducted experiments on tobacco mosaic virus (TMV) to demonstrate that in some viruses RNA acts as genetic material (Fig.1.4). TMV is a small virus composed of a single molecule of spring-like RNA encapsulated in a cylindrical protein coat. Different strains of TMV can be identified on the basis of differences in the chemical composition of their protein coats and difference in symptoms on the tobacco leaves. By using the appropriate chemical treatments, proteins and RNA of TMV can be separated. Fraenkel-Conrat experimentally proved that in the absence of DNA, RNA acts as the genetic material. In one experiment protein and RNA components of the TMV were separated and both were used to infect the tobacco leaf separately. It was observed that in case of protein subunits, there was no symptoms on the leaf and no progeny viruses were obtained. RNA part caused the infection and symptoms appeared on the tobacco leaf. Fresh progeny of TMV was also obtained.

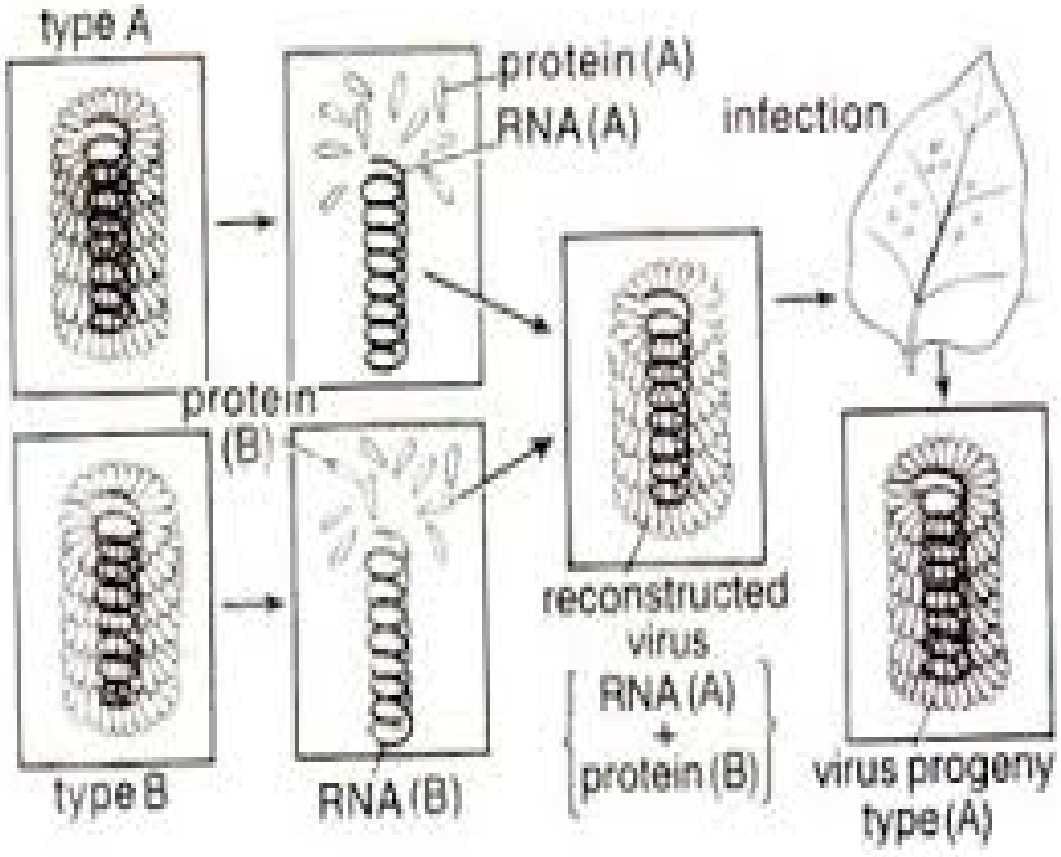


Fig-1.4: Fraenkel-Conrat Experiment

In the other experiment, two strains of TMV (type A and type B) showing different symptoms (one causing spots in random pattern and the other in a definite ring form) were taken. There protein and RNA parts were separated and chimera (hybrid) viruses were synthesized using RNA of type A and protein of type B and vice-versa. These chimera/hybrid viruses were used to infect the tobacco leaves. It was observed that symptoms on the leaf always belonged to the virus strain from which RNA was taken. Fresh progeny also belonged to the same strain. (When the hybrid or reconstituted viruses were rubbed into live tobacco leaves, the progeny viruses produced were always found to be phenotypically and genotypically identical to the parental type from where the RNA had been isolated.) These experiments proved that the genetic information of TMV is stored in the RNA and not in the protein (Fig.1.5).

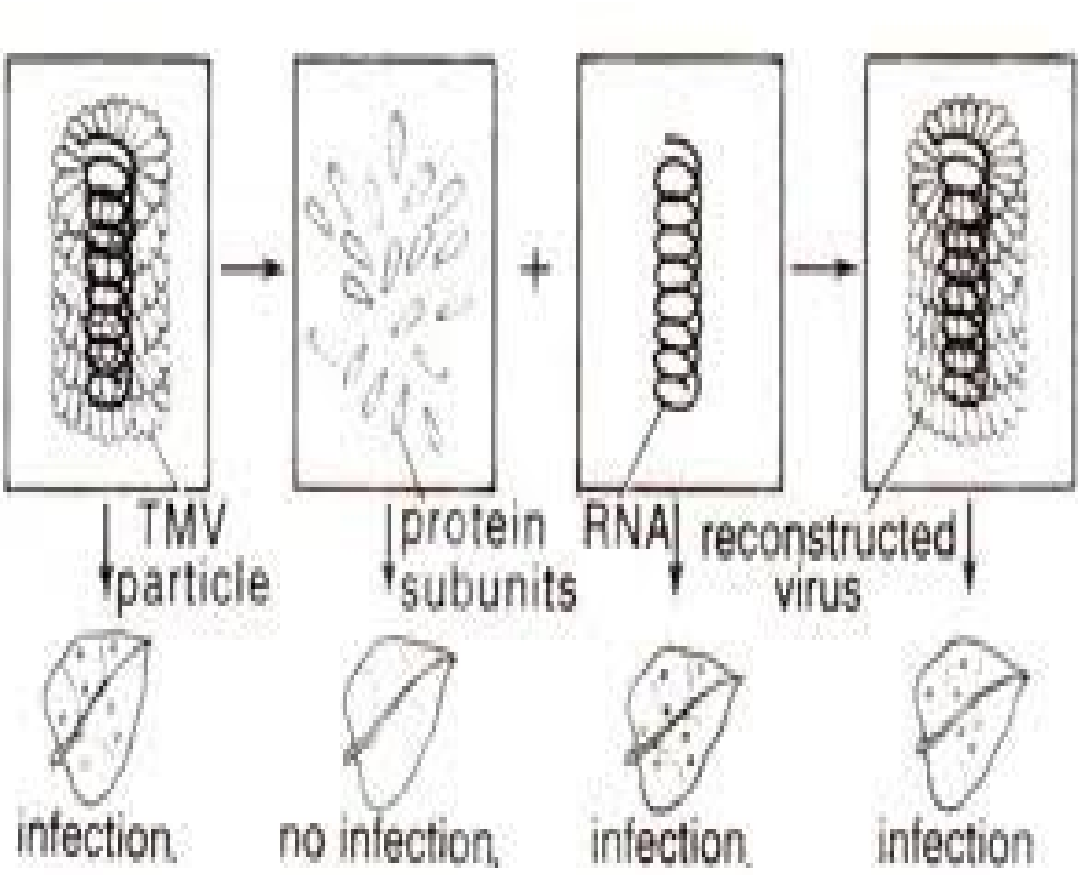


Fig-1.5: Singer Experiment

1.4 SUMMARY:

DNA, RNA and proteins are three main components play an important role in living organisms. DNA has been a widely known concept about how it stores our genetic data. However, DNA is not the only component responsible for it. RNA and proteins also play an important role determining the function of a cell. Both DNA and RNA which are collectively known as nucleic acids with almost similar chemical constituent are responsible for carrying genetic information.

1.5 TECHNICAL TERMS:

DNA, RNA, Hereditary material, Centrifugation, *Streptococcus pneumoniae*, *Escherichia coli*, T2 bacteriophage, Tobacco mosaic virus (TMV).

1.6 SELF ASSESSMENT QUESTIONS:

1) Write an account on Griffith's experiment.
2) Explain about Avery, MacLeod experiment.
3) Discuss about Hershey and Chase Experiment.
4) Explain in detail about RNA as genetic material.

1.7 SUGGESTED READINGS:

1) Campbell, N. A. and Reece J. B. (2011).Biology. IX Edition, Pearson, Benjamin, Cummings.
2) Gardner, E.J., Simmons, M.J., Snustad, D.P. (2008). Principles of Genetics. VIII Edition. Wiley India.
3) Griffiths, A.J.F., Wessler, S.R., Lewontin, R.C. and Carroll, S.B. Introduction to Genetic Analysis. IX Edition. W. H. Freeman and Co.
4) Klug, W.S., Cummings, M.R., Spencer, C.A. (2012). Concepts of Genetics. X Edition. Benjamin Cummings.
5) Lewin B. - Genes Xth edition, Wiley Eastern Limited, New Delhi
6) Powar C. B. - Genetics Vol I and II, Himalaya Publishing House, Mumbai
7) Russell, P. J. (2009). Genetics- A Molecular Approach.III Edition. Benjamin Cummings.
8) Stricberger M. W. - Genetics, 4th Edition, McMillon Public Com. Inc, New York
9) Snustad, D.P., Simmons, M.J. (2009). Principles of Genetics. V Edition. John Wiley and Sons Inc.
10) Verma P.S. and V.K. Agrawal - Cell Biology, Genetics, Molecular Biology, Evolution and Ecology.

Dr. K. Nagaraju

LESSON-2

GENE CONCEPTS – CLASSICAL CONCEPT, MODERN CONCEPT AND DIFFERENT THEORIES OF GENE CONCEPT

2.0 OBJECTIVE:

- This lesson gives a clear understanding to the students about classical and modern concepts of gene and different theories of gene concept.

STRUCTURE:

2.1 INTRODUCTION

The term "GENE" was coined in 1909 by Johannsen. Gene Theory was proposed by T.H. Morgan in 1911. In 1949, the role of genes in protein synthesis was established by L. Pauling and Ingram. The fine structure of gene was proposed in 1969 by Seymour Benzer. Mendel's experiment acted as an initial study to provide information that genes are the actual hereditary material which he called as the "Unit Factors". These factors are passed on from parents to the progeny via the germ cells and are responsible for transmission of characters in generations. Discovery of (gene) unit factor by Mendel was followed by various experiments providing evidences of them being the hereditary material. Sutton and Boveri gave "The Chromosomal Theory of Inheritance" i.e., genes are located on the chromosomes. Studies of Morgan helped in understanding the concept of crossing over and creation of genetic maps.

In addition to Sutton (1902), Morgan (1913) other scientists Bridges (1923) and Muller (1927) and many others outlined the Classical concept of gene (which included gene positioning, crossing over and their pattern of inheritance). The gene concept was modernized by Benzer in 1957, by coining the term "Cistron", "Recon" and "Muton". He emphasized about the different nature of genes and their location with respect to inheritance pattern. Founder of modern genetics, Gregor Johann Mendel was an Austrian monk He has been credited for discovering the basic principles of heredity. It took approximately 35 years for people to recognize and appreciate his work which formed the basis of modern genetics. Thus, he is recognized as the Father of Genetics. He worked with garden pea; he observed that the garden pea follows a particular pattern while passing on the traits from one generation to another. Around 1854 he himself began the experimental work by hybridizing

the pea plants and recording all the results in a well-planned manner. He chose pea plant for his plant hybridization studies as it had many varieties and the offspring took less time in completing a generation so it was easy and quickly reproducible. Until 1900, Mendel's work was not appreciated by the scientific community and directly rejected when he presented his findings. However, three botanist Hugo De Vries, Erich von Tschermak and Karl Correns are credited for Rediscovering the pioneer work of Mendel. Three of them worked independently and came to the same conclusion as Mendel. They interpreted their results in accordance to Mendel's findings and drew attention to Mendelian studies on pea.

- E R Garrod (1908) proposed One gene – one product hypothesis
- Gene theory was proposed by T H Morgan in 1911
- L Pauling and Ingram(1949) established the role of genes in protein synthesis
- Genes can be transmitted from parent to off springs.

2.2 CLASSICAL CONCEPT

Being a well learned botanist, Mendel could identify characteristic traits which were opposite in two pea plants: e.g., tall or short, smooth seed or wrinkled seed, colour of seed - yellow or green, shape of pod - wrinkled or smooth, pod colour- yellow or green, flower position- terminal or axial and flower colour- white or purple (Fig.2.1). He cross-fertilized the pea plants with opposite traits to obtain the progeny and recorded the observation. After analyzing his results he concluded three important principles referred to as the "Laws of Mendel" or the "Mendelian principles" which are –

1) **Law of Dominance:** This principle states that one factor in a pair of traits is dominant over the other in inheritance unless both factors in the pair are recessive. The progeny expresses the dominant trait.

2) **Law of Segregation:** This principle states that the offspring acquires one factor from each parent and during the production of gametes the two copies of each hereditary factor segregate. The segregation of each factor is responsible for the combination of the traits from both the parents.

3) **Law of Independent Assortment:** This principle states that the laws of chance govern which particular characteristics of the parental pairs will occur in each individual offspring. The pairing and expression of the traits in the offspring are a chance factor.

Mendel crossed the pea plants having opposite traits- i.e., tall with short, wrinkled with smooth, yellow seeds with green seeds etc. - and analyzed the results which lead to his two most important conclusions "The Law of Segregation" and "The Law of independent assortment". He gave the results using simple statistical laws and conveyed that these patterns were followed not only in pea plant but also all living beings having such traits. Unit factors in Pairs: Traits which used for the experimental analysis he called them as unit factors, which existed in pairs in individual organisms. Each unit factor came from either parent and the two forms were complimentary to each to other. For height of the plants traits were tall and small for a monohybrid cross and the output consisted of three combinations i.e., tall, small and intermediate Dominance or Recessiveness: When complimentary unit factors (inherited from either parent) are present in a single progeny, one unit factor is dominant unit factor (which is able to express itself in the F1 progeny phenotype) and other is known as recessive unit factor (which is not expressed in the phenotype of F1 progeny). We should remember that term dominant and recessive are just designatory for individual traits in the parents' genotype. For example, if we consider the height of a plant tall plants are considered dominant over the

recessive short/dwarf plants. Segregation: Formation of gametes leads to the segregation of the individual traits; in order to equalize the chances for each trait in the individual gametes. Depending upon the fusion of gametes an individual may contain a pair of similar traits (TT- all tall progeny, tt- all short progeny) and the progeny will be decided accordingly. Because of segregation and random fusion of the gametes there is equal or 50% probability of receiving either of the traits. To understand it better let us illustrate it through a monohybrid cross. Parent 1 is tall and produces only T trait whereas parent 2 is short and produces only tt trait (unit factor). Consider the gametes for Parent 1; it will be all tall "T" unit factor and for parent 2 it will be all short "t" unit factor. Following fertilization the progeny will receive one trait from each parent i.e., a tall factor and a short factor but because tall factor is a dominant to short, all the progeny are tall. On analyzing the gametes of the F1 progeny each progeny has T/tall factor which contributes to the tall height of the plant. However, selfing the F1 progeny we get four combinations for the F2 progeny which are: 1. Tall/tall (TT) 2.Tall/dwarf (Tt) 3. Dwarf/tall (tT) 4.Dwarf/dwarf (tt). A combination 1, 2, 3 yields tall plants whereas 4 yields short plant. Thus, we see that F2 progeny consists of both tall and short plants in a ratio 3:1. This ratio was well maintained in all other crosses for other traits observed by the Mendel. Mendel emphasized that it was these factors which were responsible for passing on the characteristics among the populations from one progeny to other. With time these traits/ unit factors were identified as individual alleles and we know that a person can be homozygous or heterozygous for a particular trait. He believed that single trait was controlled by a single gene. However, modern genetics has revealed the concept of multiple genes and the role of environmental factors in the expression of genes. Mendel confined his work to the pea plants and explained the inheritance of discrete hereditary material (factors) which are now known as genes. He laid the foundation of genetics and later on it was found that his observations were true for a variety of traits in many plants and animals.

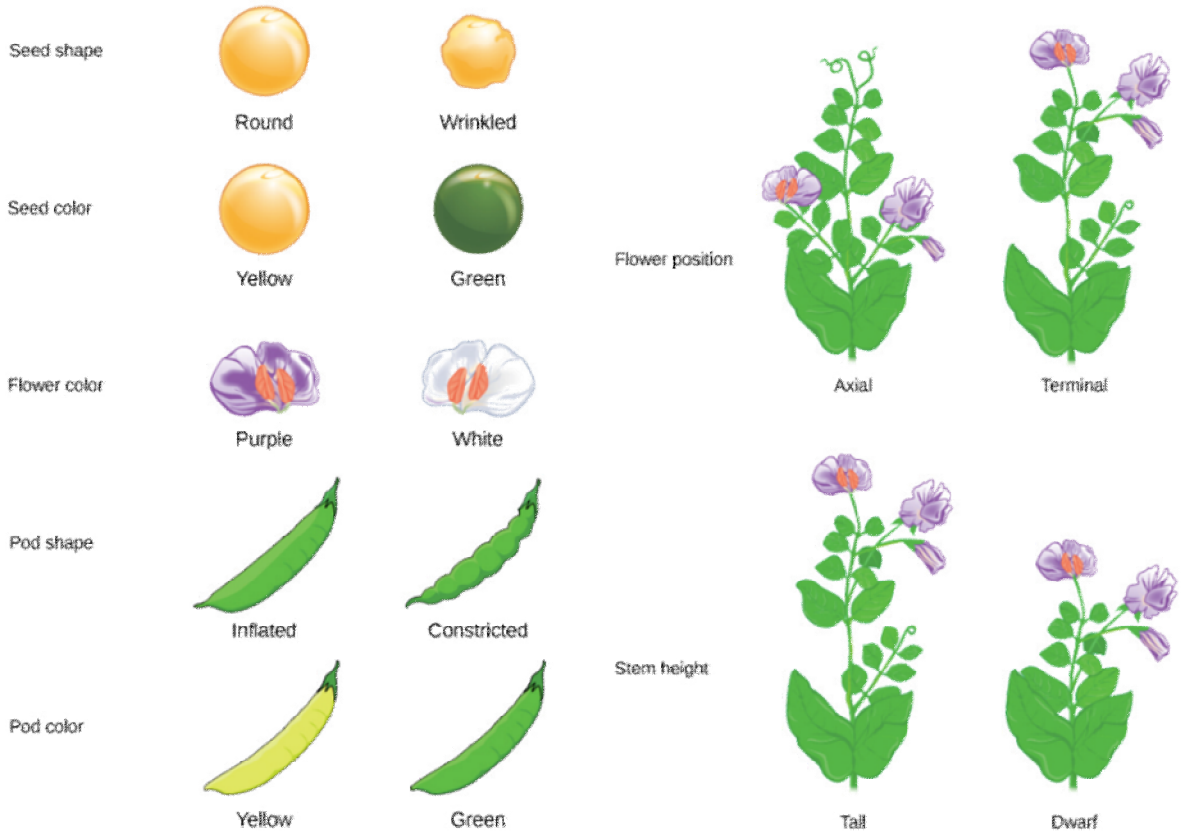


Fig.2.1: Classical Concept / Mendel Experiment

2.3 MODERN CONCEPT OF GENE

After the discovery of DNA, the gene has been defined as cistron, recon and muton. The classical gene is the smallest unit that could undergo a mutational change. A gene further divided into smaller units of function, mutation and recombination. Symour Benzer (1955-USA) coined the terms cistron, recon, and muton to explain the relationship between DNA and genetic phenomena.

a) Cistron: It is the unit of function. Cistron represents a segment of the DNA molecule and consists of a linear sequence of nucleotides, which controls some cellular function. In *E. coli* cistron may contain about 1500 base pairs. Some cistrons may contain as many as 30,000 base pairs. The cistron begin with initiation codon and ends with a terminating codon. Each cistron is responsible for coding one m-RNA molecule which in turn controls the formation of one polypeptide chain. Each cistron consists of hundreds of mutons and recons.

b) Recon: It is a unit of recombination. It is the smallest unit capable of recombining genetically. Recombination studies on microbes indicate that structurally the reconconsists of one or two pairs of nucleotides, possibly only one pair.

c) Muton: It is a unit of mutation. The shortest chromosomal unit capable of undergoing mutation has been called the muton. The muton consists of one or many pairs of nucleotides within the DNA molecule.

Types of Genes

1. **House Keeping Genes (Constitutive Genes):** They are those genes which are constantly expressing themselves in a cell because their products are required for the normal cellular activities, e.g., genes for glycolysis, ATase.
2. **Non-constitutive Genes (Luxury Genes):** The genes are not always expressing themselves in a cell. They are switched on or off according to the requirement of cellular activities, e.g., gene for nitrate reductase in plants, lactose system in *Escherichia coli*. N on- constitutive genes are of further two types, inducible and repressible.
3. **Inducible Genes**: The genes are switched on in response to the presence of a chemical substance or inducer which is required for the functioning of the product of gene activity, e.g., nitrate for nitrate reductase.
4. **Repressible Genes:** They are those genes which continue to express themselves till a chemical (often an endproduct) inhibits or represses their activity. Inhibition by an end product is known as feedback repression.
5. **Multigenes (Multiple Gene Family):** It is a group of similar or nearly similar genes for meeting requirement of time and tissue specific products, e.g., globin gene family.
6. **Repeated genes:** The genes occur in multiple copies because their products are required in larger quantity, e.g., histone genes, tRNA genes, rRNA genes, actin genes.
7. **Single copy genes:** The genes are present in single copies (occasionally 2-3 times), e.g., protein coding genes. They form 60 to 70% of the functional genes. Duplications, mutations and exon reshuffling can form new genes.

8. **Pseudogenes:** They are genes which have homology to functional genes but are unable to produce functional products due to intervening nonsense codons, insertions, deletions and inactivation of promoter regions, e.g., several of snRNA genes
9. **Processed Genes:** They are eukaryotic genes which lack introns. Processed genes have been formed probably due to reverse transcription or retroviruses. Processed genes are generally non-functional as they lack promoters.
10. **Split Genes:** They were discovered in 1977 by many workers but credit is given to Sharp and Roberts (1977). Split genes are those genes which possess extra or nonessential regions interspersed with essential or coding parts. The non-essential parts are called introns, spacer DNA or intervening sequences (IVS). Essential or coding parts are called exons. Transcribed intronic regions are removed before RNA passes out into cytoplasm. Split genes are characteristic of eukaryotes.

However, certain eukaryotic genes are completely exonic or non-split e.g., histone genes, interferon genes. Split genes have also been recorded in prokaryotes, thymidylate synthase gene and ribonucleotide reductase gene in T4. A gene that produces calcitonin in thyroid forms a neuropeptide in hypothalamus by removing an exon. Adenovirus has also a mechanism to produce 15-20 different proteins from a single transcriptional unit by differential splicing.

2.4 DIFFERENT THEORIES OF GENE CONCEPT

2.4.1 One gene-one enzyme Theory

The idea of one gene-one enzyme was introduced in early 1940's and was proved by American Geneticist George Wells Beadle and American biochemist Edward L. Tatum. Beadle and Tatum studied the growth of red bread mold *Neurospora crassa*. Their studies often referred to as the "*Neurospora* Genetics" mainly involved growing mutants of the mold on different media preparations (Fig. 2.2). The mold *Neurospora* was first exposed to x-rays to obtain various mutations. Mutations are the sudden heritable change in the genome of the organism having the capability of disrupting the mold's ability to synthesis amino acids/nutrients and grow. They subsequent germinated the mutant spores in tubes having complete medium (i.e. amino acid, vitamins and organic substances) along with proper physical environment. Later on growing these mutant strains on minimal growth media (Medium having sugars, salts and vitamins which were utilized by the enzymes of the mold to synthesis the necessary amino acids for sustaining the mold) containing only the basic nutrients in which the wild type or the non-mutated, strain of *Neurospora* could survive.

They observed that the mutants were not able to grow on the minimal media instead they required additional supplements of specific amino acids in order to grow. The effect of mutation was identified by adding the supplements one by one to test which supplement could fulfill the requirement of the missing enzyme in the mold and helped it to grow. These mutants when crossed with the normal molds were able to grow in the minimal medium revealing that these metabolic defects were outcome of recessive traits and proved the alteration in the genes. Thus, they concluded that mutants have disruption in specific gene which was responsible for the production of specific amino acid (which was being added as the supplement for the mutants to grow). Producing identifiable nutritional mutations with x-ray technique was a herculean task and they achieved success in their 299th attempt, which grew only when the minimal medium was supplemented with Vitamin B6. Thus, with patience and continued perseverance they were able to identify many more mutants revealing that genetic mutations affect metabolic pathways and gave detailed multi-step synthesis

pathway for many amino acids and vitamins. The information obtained from the experiments on *Neurospora* confirmed what Beadle had witnessed in *Drosophila*. They published their results in "Genetic control of biochemical reactions in Neurospora," in 1941, proposing the concept of one gene-one enzyme hypothesis. This hypothesis was verified but since 1940s it has undergone a slight change and today it is known as one gene one polypeptide hypothesis as stated by Yanofsky (1965). As with the advancement in modern science it became clear that an enzyme is not coded by a single gene/polypeptide.

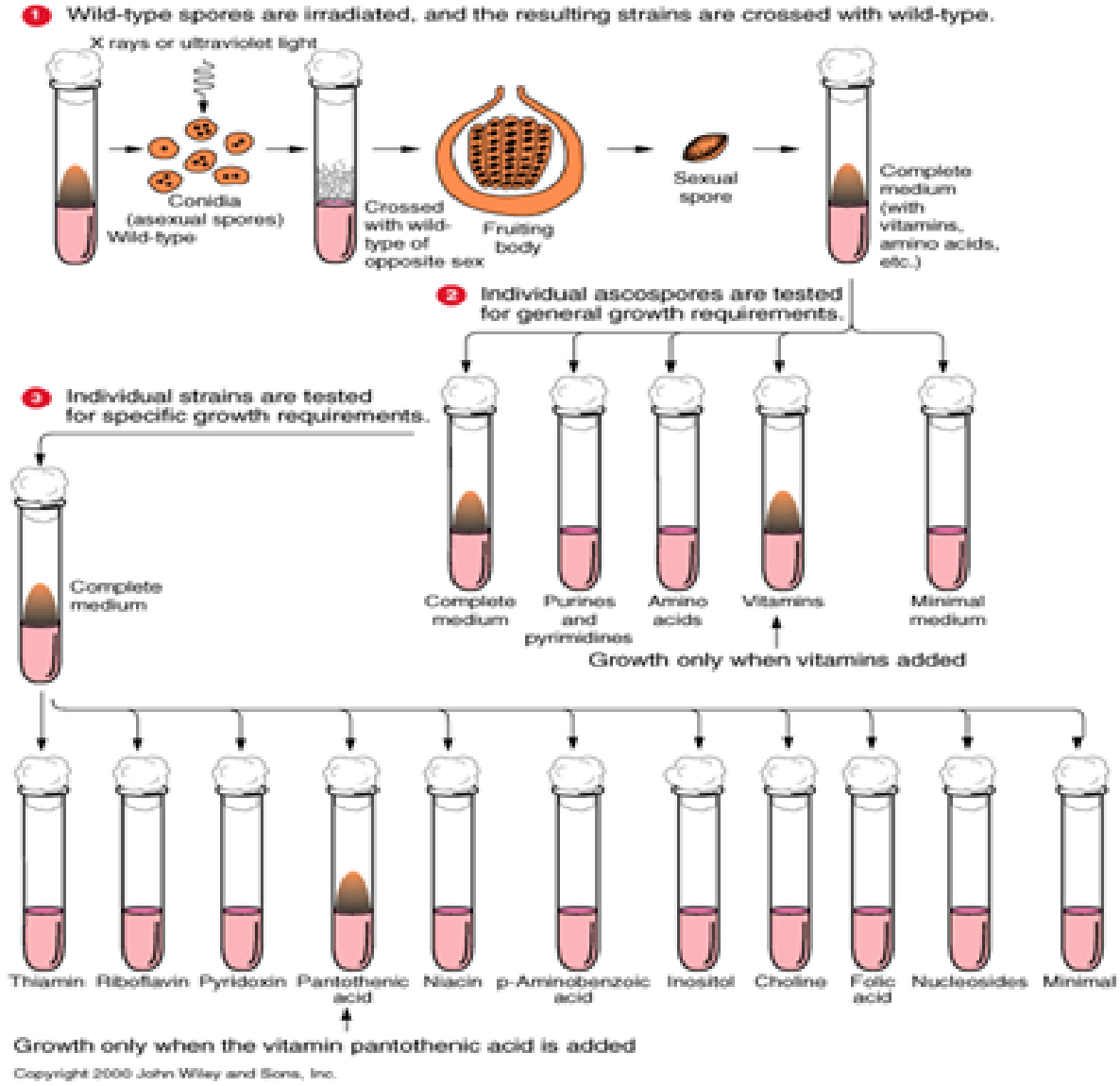


Fig-2.2: One-gene-one enzyme theory

2.4.2 One-gene-one polypeptide: Beadle and Tatum

The one-gene-one polypeptide hypothesis states that a single gene is responsible for coding and production of a single polypeptide (Fig.2.3). A gene consists of specific sequence of nucleotides. Each triplet codes for an amino acid, which form the monomeric units of polypeptides joined to each other by peptide bonds. Earlier it was known that a single gene is responsible for polypeptide (which codes the enzyme). They studied the growth of fungus *Neurosporacrassa*. In laboratory, the fungus grows on minimal medium containing agar,

sucrose, nitrate, vitamin (biotin) and inorganic minerals. This means that all supplements are present within these minimal medium for the fungal development. However, when treated with x-ray the fungus is unable to grow in minimal medium due to lack of certain growth component. The mutant spores are given individuals component of the minimal medium as a growth supplement to determine the deficient component which the mutants are unable to synthesis. The mutants give a healthy progeny when crossed with a normal/wild type and the ascospores (8) of the progeny are individually tested for their nutritional requirements. A single gene mutation is determined crossing the single arginine deficient mutant strain (a-) with wild type (a+), all the 8 ascospores of the progeny survive on a medium containing arginine, but only four ascospores survive in absence of arginine from the medium. Indicating the mutation of a single gene responsible for arginine synthesis. Arginine synthesis requires a series of steps; each of which is controlled by a single gene. The mutation of gene leads to the suppression of the specific step. This is demonstrated by the development of mutants on the minimal medium when the deficient component is added to it. Thus, on identifying the mutant we know the defect in the biosynthetic pathway. The mutation occurs in a specific chromosomal region responsible for the functionality loss of that particular enzyme (encoded within that chromosomal region). This conveys the basic enzyme relationships. The modern researchers have identified that gene is a part of DNA and responsible for synthesis of particular protein. Currently the concept of "one-gene-one enzyme" given by Beadle and Tatum has been revised to "one-gene-one polypeptide chain (protein molecule)" in view of the complexity in the structure, functions of the enzymes and the genetic variability of proteins (such as sickle cell haemoglobin). However, later on realizing the phenomenon of splicing (with which it became evident that a single gene codes for specific non-enzymatic proteins as well as individual polypeptide chains). In 1945, George Beadle modified the one gene one enzyme hypothesis to one gene one polypeptide hypothesis. Now we know that certain genes code for different types of RNA involved in protein synthesis. Especially in eukaryotes, a single gene can produce several RNA products depended on differential splicing. The direct relationship between a gene and enzyme was put forth by the *Neurospora* genetics. However later on finding that an enzyme can consists of various polypeptides it was modified to one-gene-one polypeptide hypothesis.

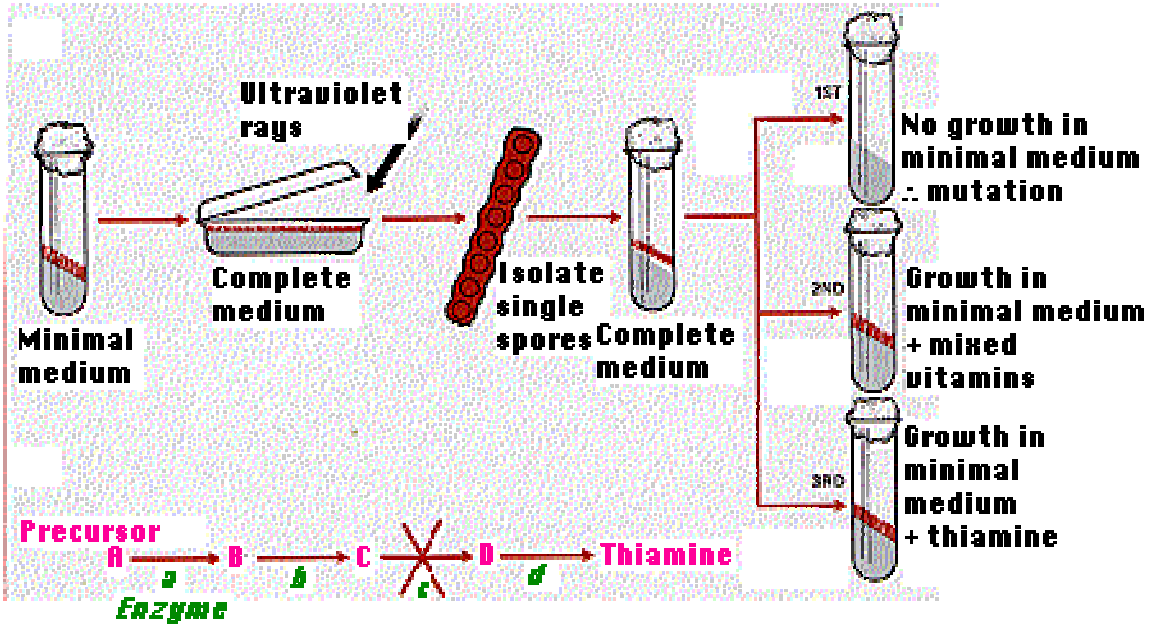


Fig.2.3: One-Gene-One Polypeptide Hypothesis

2.5 SUMMARY:

The gene concept was modernized by Benzer in 1957, by coining the term "Cistron", "Recon" and "Muton". He emphasized about the different nature of genes and their location with respect to inheritance pattern. Founder of modern genetics, Gregor Johann Mendel was an Austrian monk He has been credited for discovering the basic principles of heredity. It took approximately 35 years for people to recognize and appreciate his work which formed the basis of modern genetics. Thus, he is recognized as the Father of Genetics. He worked with garden pea; he observed that the garden pea follows a particular pattern while passing on the traits from one generation to another.

2.6 TECHNICAL TERMS:

Law of Dominance, Law of Segregation, Law of Independent assortment, Cistron, Recon,Muton, One-gene-one enzyme theory, One-gene-one polypeptide theory, *Neurospora crassa.*

2.7 SELF ASSESSMENT QUESTIONS:

1) Discuss Classical concept of gene.
2) Explain about Modern concept of Gene.
3) Write in detail about One-gene-one enzyme theory.
4) Explain One-gene-one polypeptide theory.

2.8 SUGGESTED READINGS:

1) Campbell, N. A. and Reece J. B. (2011).Biology. IX Edition, Pearson, Benjamin, Cummings.
2) Gardner, E.J., Simmons, M.J., Snustad, D.P. (2008). Principles of Genetics. VIII Edition. Wiley India.
3) Griffiths, A.J.F., Wessler, S.R., Lewontin, R.C. and Carroll, S.B. Introduction to Genetic Analysis. IX Edition. W. H. Freeman and Co.
4) Klug, W.S., Cummings, M.R., Spencer, C.A. (2012). Concepts of Genetics. X Edition. Benjamin Cummings.
5) Lewin B. - Genes Xth edition, Wiley Eastern Limited, New Delhi
6) Powar C. B. - Genetics Vol. I and II, Himalaya Publishing House, Mumbai
7) Russell, P. J. (2009). Genetics- A Molecular Approach.III Edition. Benjamin Cummings.
8) Stricberger M. W. - Genetics, 4th Edition, McMillon Public Com. Inc, New York
9) Snustad, D.P., Simmons, M.J. (2009). Principles of Genetics.V Edition. John Wiley and Sons Inc.
10) Verma P.S. and V.K. Agrawal - Cell Biology, Genetics, Molecular Biology, Evolution and Ecology.

Dr. K. Nagaraju

LESSON-3

PLASMIDS-CHARACTERISTICS, TYPES, PROPERTIES, REPLICATION, ISOLATION AND SIGNIFICANCE

3.0 OBJECTIVE:

- This lesson enriches the students with the knowledge on characters, types, properties, replication, isolation and significance of the plasmids which are extra chromosomal materials in bacteria.

STRUCTURE:

3.1 INTRODUCTION

The term plasmid was introduced in 1952 by the American molecular biologist Joshua Lederberg to refer "any extrachromosomal hereditary determinant. The term's early usage included any bacterial genetic material that exists extrachromosomally for at least part of its replication cycle, but because that description includes bacterial viruses, the notion of plasmid was refined over time to comprise genetic elements that reproduce autonomously. Later in 1968, it was decided that the term plasmid should be adopted as the term for extrachromosomal genetic element, and to distinguish it from viruses, the definition was narrowed to genetic elements that exist exclusively or predominantly outside of the chromosome and can replicate autonomously.

3.2 DEFINITION

A plasmid is a small, extra chromosomal DNA molecule within a cell that is physically separated from chromosomal DNA and can replicate independently. They are most commonly found as small circular, double stranded DNA molecules in bacteria; however,

plasmids are sometimes present in archaea and eukaryotic organisms. In nature, plasmids often carry genes that benefit the survival of the organism and confer selective advantage such as antibiotic resistance. While chromosomes are large and contain all the essential genetic information for living under normal conditions, plasmids are usually very small and contain only additional genes that may be useful in certain situations or conditions. Artificial plasmids are widely used as vectors in molecular cloning, serving to drive the replication of recombinant DNA sequences within host organisms. In the laboratory, plasmids may be introduced into a cell via transformation.

3.3 CHARACTERISTICS OF PLASMIDS

Plasmids are considered replicons, units of DNA capable of replicating autonomously within a suitable host. Plasmids are transmitted from one bacterium to another (even of another species) mostly through conjugation. This host-to-host transfer of genetic material is one mechanism of horizontal gene transfer, and plasmids are considered part of the mobilome. Unlike viruses, which encase their genetic material in a protective protein coat called a capsid, plasmids are "naked" DNA and do not encode genes necessary to encase the genetic material for transfer to a new host; however, some classes of plasmids encode the conjugative "sex" pilus necessary for their own transfer. The size of the plasmid varies from 1 to over 200 kbp, and the number of identical plasmids in a single cell can range anywhere from one to thousands under some circumstances. In order for plasmids to replicate independently within a cell, they must possess a stretch of DNA that can act as an origin of replication. The self-replicating unit, in this case, the plasmid, is called a replicon. A typical bacterial replicon may consist of a number of elements, such as the gene for plasmid-specific replication initiation protein (Rep), repeating units called iterons, DnaA boxes, and an adjacent AT-rich region. Smaller plasmids make use of the host replicative enzymes to make copies of themselves, while larger plasmids may carry genes specific for the replication of those plasmids. A few types of plasmids can also insert into the host chromosome, and these integrative plasmids are sometimes referred to as episomes in prokaryotes.

Plasmids almost always carry at least one gene. Many of the genes carried by a plasmid are beneficial for the host cells, for example: enabling the host cell to survive in an environment that would otherwise be lethal or restrictive for growth. Some of these genes encode traits for antibiotic resistance or resistance to heavy metal, while others may produce virulence factors that enable a bacterium to colonize a host and overcome its defenses or have specific metabolic functions that allow the bacterium to utilize a particular nutrient, including the ability to degrade recalcitrant or toxic organic compounds. Plasmids can also provide bacteria with the ability to fix nitrogen. Some plasmids, however, have no observable effect on the phenotype of the host cell or its benefit to the host cells cannot be determined, and these plasmids are called cryptic plasmids. Naturally occurring plasmids vary greatly in their physical properties. Their size can range from very small mini-plasmids of less than 1-kilobase pairs (Kbp) to very large mega plasmids of several mega base pairs (Mbp). At the upper end, little differs between a megaplasmid and a minichromosome. Plasmids are generally circular, but examples of linear plasmids are also known. These linear plasmids require specialized mechanisms to replicate their ends.

Plasmids may be present in an individual cell in varying number, ranging from one to several hundreds. The normal number of copies of plasmid that may be found in a single cell is called the Plasmid copy number, and is determined by how the replication initiation is regulated and the size of the molecule. Larger plasmids tend to have lower copy numbers. Low-copy number plasmids that exist only as one or a few copies in each bacterium are, upon

cell division, in danger of being lost in one of the segregating bacteria. Such single-copy plasmids have systems that attempt to actively distribute a copy to both daughter cells. These systems, which include the parABS system and parMRC system, are often referred to as the partition system or partition function of a plasmid. Plasmids may be classified in a number of ways. Plasmids can be broadly classified into conjugative plasmids and non-conjugative plasmids. Conjugative plasmids contain a set of transfer or tra genes which promote sexual conjugation between different cells. In the complex process of conjugation, plasmids may be transferred from one bacterium to another via sex pili encoded by some of the tra genes. Non-conjugative plasmids are incapable of initiating conjugation, hence they can be transferred only with the assistance of conjugative plasmids. An intermediate class of plasmids is mobilizable, and carries only a subset of the genes required for transfer. They can parasitize a conjugative plasmid, transferring at high frequency only in its presence. Plasmids can also be classified into incompatibility groups. A microbe can harbour different types of plasmids, but different plasmids can only exist in a single bacterial cell if they are compatible. If two plasmids are not compatible, one or the other will be rapidly lost from the cell.

Different plasmids may therefore be assigned to different incompatibility groups depending on whether they can coexist together. Incompatible plasmids (belonging to the same incompatibility group) normally share the same replication or partition mechanisms and can thus not be kept together in a single cell. Another way to classify plasmids is by function. There are five main classes:

- Fertility F-plasmids, which contain tra genes. They are capable of conjugation and result in the expression of sex pili.
- Resistance (R) plasmids, which contain genes that provide resistance against antibiotics or poisons. Historically known as R-factors, before the nature of plasmids was understood.
- Col plasmids, which contain genes that code for bacteriocins, proteins that can kill other bacteria.
- Degradative plasmids, which enable the digestion of unusual substances, e.g. toluene and salicylic acid.
- Virulence plasmids, which turn the bacterium into a pathogen e.g. Ti plasmid in *Agrobacterium tumefaciens*. Plasmids can belong to more than one of these functional groups.

3.4 TYPES OF PLASMIDS

Structurally every plasmid has certain essential elements (Fig.3.1). These are as follows-

- **Origin of replication (OR) –** This refers to a specific location in the strand where the replication process begins. In plasmids, this region is A=T rich region as it is easier to separate the strands during replication.
- **Selectable marker site –** This region consists of antibiotic resistance genes which are useful in the identification and selection of bacteria that contain plasmids.
- **Promoter region –** this is the region where the transcriptional machinery is loaded.
- **Primer binding site –** this is the short sequence of single-strand DNA which is useful in DNA amplification and DNA sequencing.
- **Multiple cloning sites –** This site contains various sequences where the restriction enzymes can bind and cleave the double stranded structure.
- The size of the plasmid varies from 2 kb to 200 kb.

- It is the extrachromosomal element of the cell which is not required for the growth and development of the cell.
- Most of the plasmids contain the TRA gene, which is the transferred gene and is essential in transferring the plasmid from one cell to another.

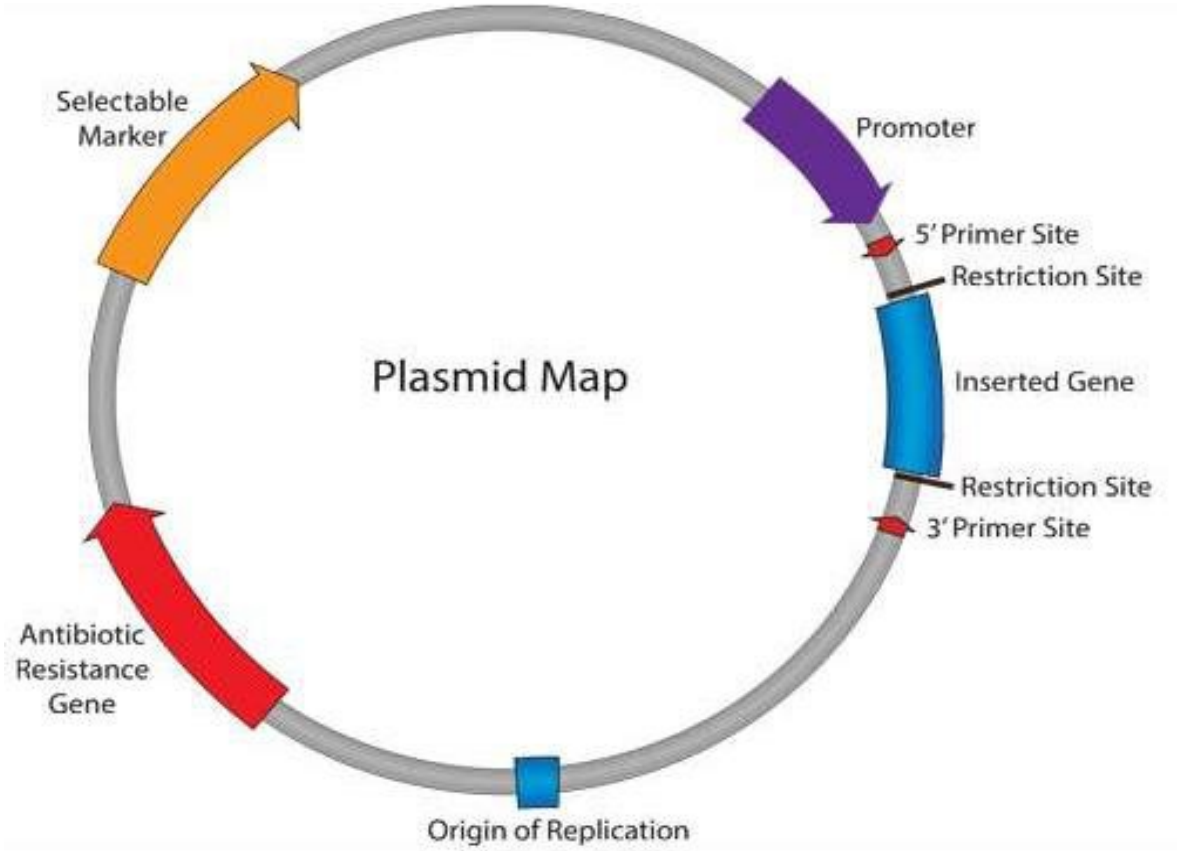


Fig-3.1: Structure of a Plasmid

Based on the presence of the TRA gene plasmids can be classified into two types:

1) **Conjugative Plasmids-**These plasmids contain TRA (transfer) gene and are commonly seen in bacteria.
2) **Non-conjugative Plasmids-**These types of plasmids lack the TRA genes.

Based on functions the plasmids can be classified into the following types:

3.4.1. F Plasmids (Fertility Plasmids)

They contain the TRA genes and hence can be transferred from one cell to another. They can replicate inside the bacterial cell. They cause the synthesis of a pilus, which is a long protein-rich structure that helps in cell-cell interaction. It also contains a sequence responsible for incompatibility (Fig.3.2).

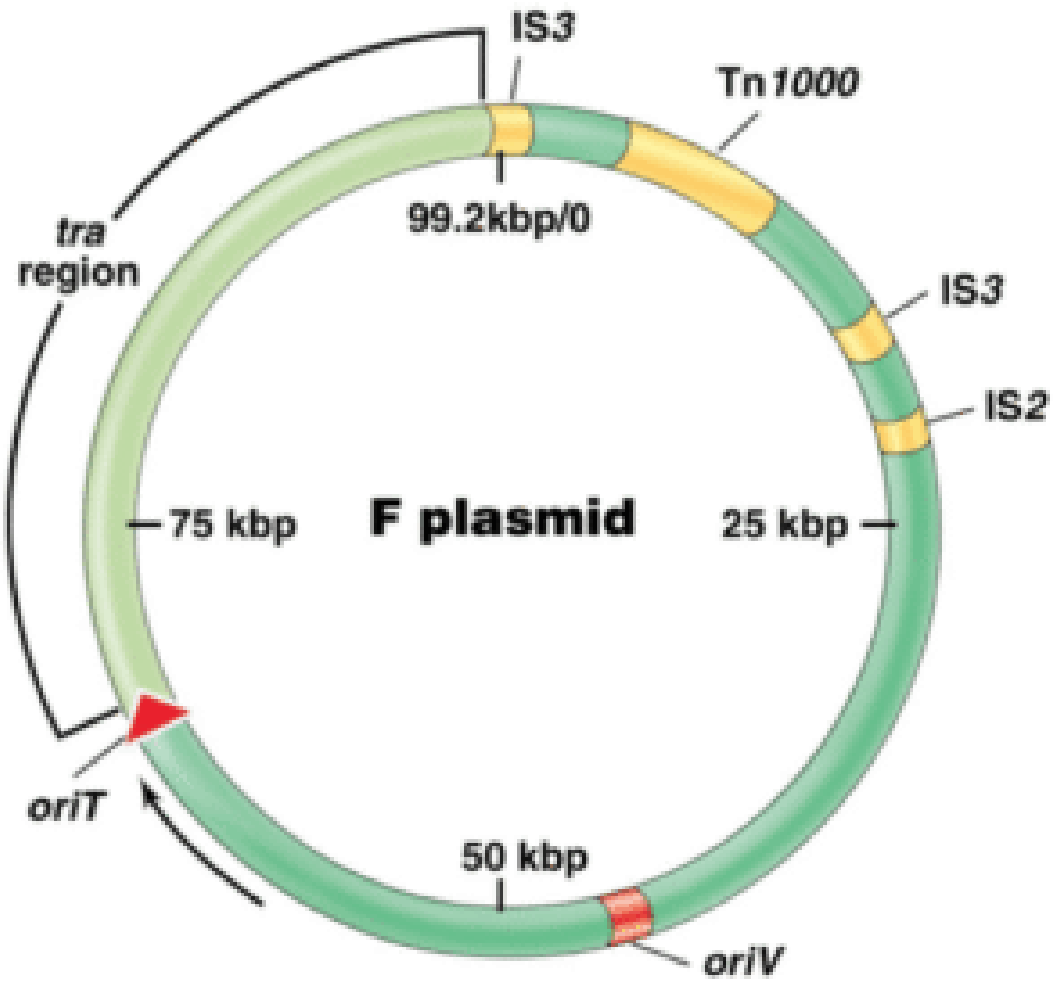


Fig-3.2: Structure of F-Plasmid

3.4.2. R plasmid (Resistance plasmids)

These plasmids contain and transmit genes for Antibiotic resistance from one cell to another. The antibiotic resistance gene protects the bacteria from antibiotics in human medicines and antibiotics naturally present in the soil. These types of plasmids are usually large in size and present in low copy numbers in the cell (Fig.3.3).

Resistance transfer factor (shortened as R-factor or RTF) is an old name for a plasmid that codes for antibiotic resistance. R-factor was first demonstrated in *Shigella* in 1959 by Japanese scientists. Often, R-factors code for more than one antibiotic resistance factor: genes that encode resistance to unrelated antibiotics may be carried on a single R-factor, sometimes up to 8 different resistances. Many R-factors can pass from one bacterium to another through bacterial conjugation and are a common means by which antibiotic resistance spreads between bacterial species, genera and even families. For example, RP1, a plasmid that encodes resistance to ampicillin, tetracycline and kanamycin originated in a species of *Pseudomonas*, from the family Pseudomonadaceae, but can also be maintained in bacteria belonging to the family Enterobacteriaceae, such as *Escherichia coli*.

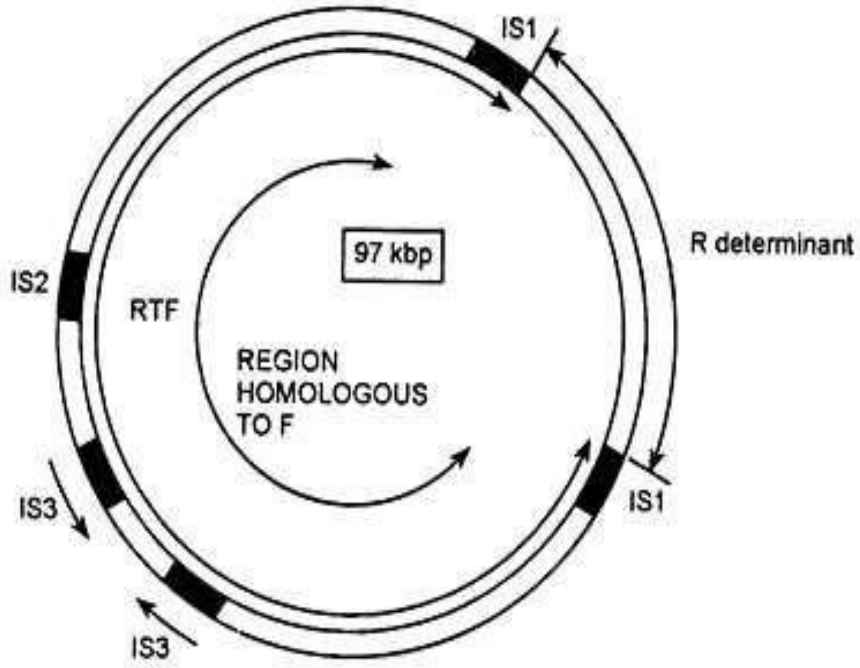


Fig-3.3: Structure of R-Plasmid

3.4.3. Col Plasmids (Colicin plasmids) or Colicinogenic plasmid:

These are known as bacteriocinogenic plasmids because they produce bacteriocins (Fig.3.4). These proteins have the ability to kill the closely related bacterial cells which lack Col plasmids. These plasmids are observed in *E. coli*.

One class of plasmids, colicinogenic (or Col) factors, determines the production of proteins called colicins, which have antibiotic activity and can kill other bacteria. Another class of plasmids, R factors, confers upon bacteria resistance to antibiotics. A colicin is a type of bacteriocin produced by and toxic to some strains of *Escherichia coli*. Colicins are released into the environment to reduce competition from other bacterial strains. Colicins bind to outer membrane receptors, using them to translocate to the cytoplasm or cytoplasmic membrane, where they exert their cytotoxic effect, including depolarization of the cytoplasmic membrane, DNase activity, RNase activity, or inhibition of murein synthesis. Virtually all colicins are carried on plasmids. The two general classes of colicinogenic plasmids are large, low-copy-number plasmids, and small, high-copy-number plasmids. The larger plasmids carry other genes, as well as the colicin operon. The colicin operons are generally organized with several major genes. These include an immunity gene, a colicin structural gene, and a bacteriocin release protein (BRP) gene. The immunity gene is often produced constitutively, while the BRP is generally produced only as a read-through of the stop codon on the colicin structural gene. The colicin itself is repressed by the SOS response and may be regulated in other ways, as well. Retaining the colicin plasmid is very important for cells that live with their relatives, because if a cell loses the immunity gene, it quickly becomes subject to destruction by circulating colicin. At the same time, colicin is only released from a producing cell by the use of the lysis protein, which results in that cell's death. This suicidal production mechanism would appear to be very costly, except for the fact that it is regulated by the SOS response, which responds to significant DNA damage. In short, colicin production may only occur in terminally ill cells. Professor Kleanthous Research Group at the University of Oxford studied colicins extensively as a model system for characterizing and investigating protein-protein interactions and recognition.

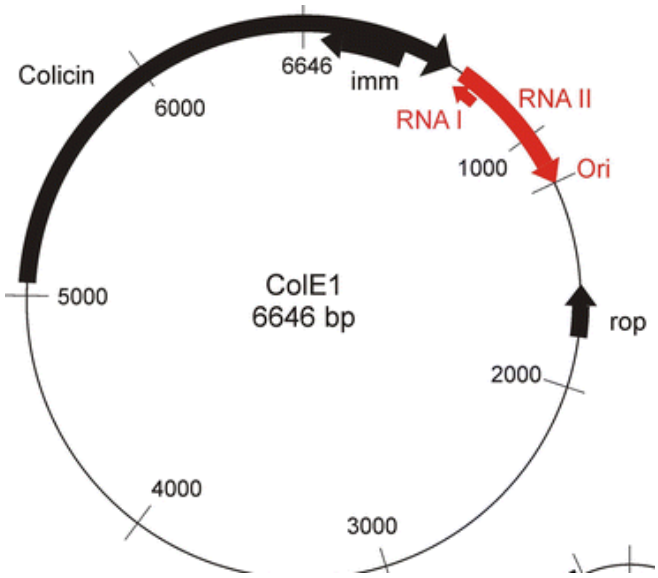


Fig-3.4: Structure of Col-Plasmid

3.4.4. Degradative Plasmids

These types of plasmids have the ability to digest unusual substances such as toluene, camphor, salicylic acid, etc. The presence of these plasmids in the organism enables the breakdown of various chemicals and substances.

3.4.5. Virulence Plasmids

These plasmids produce virulence factors that enable the bacteria to infect other cells. Bacteria containing virulence plasmids are able to infect the plant, animal, and human cells.Example – Ti plasmid is the virulence plasmid present in *Agrobacterium tumefaciens* which causes crown gall disease in plants (Fig.3.5).

A tumour inducing (Ti) plasmid is a plasmid found in pathogenic species of *Agrobacterium*, including *A. tumefaciens*, *A. rhizogenes*, *A. rubi* and *A. vitis*. Evolutionarily, the Ti plasmid is part of a family of plasmids carried by many species of Alphaproteobacteria. Members of this plasmid family are defined by the presence of a conserved DNA region known as the repABC gene cassette, which mediates the replication of the plasmid, the partitioning of the plasmid into daughter cells during cell division as well as the maintenance of the plasmid at low copy numbers in a cell. The Ti plasmids themselves are sorted into different categories based on the type of molecule, or opine, they allow the bacteria to break down as an energy source. The presence of this Ti plasmid is essential for the bacteria to cause crown gall disease in plants. This is facilitated via certain crucial regions in the Ti plasmid, including the 'vir' region, which encodes for virulence genes, and the transfer DNA (T-DNA) region, which is a section of the Ti plasmid that is transferred via conjugation into host plant cells after an injury site is sensed by the bacteria. These regions have features that allow the delivery of T-DNA into host plant cells, and can modify the host plant cell to cause the synthesis of molecules like plant hormones (e.g. auxins, cytokinins) and opines and the formation of crown gall tumours. Because the T-DNA region of the Ti plasmid can be transferred from bacteria to plant cells, it represented an exciting avenue for the transfer of DNA between kingdoms and spurred large amounts of research on the Ti plasmid and its possible uses in bioengineering. The replication of the Ti plasmid is driven by the RepC

initiator protein, which possesses two protein domains: an N-terminal domain (NTD) that binds to DNA and a C-terminal domain (CTD). Mutational analyses have shown that without a functional RepC protein, the Ti plasmid is unable to replicate. Meanwhile, the oriV sequence is around 150 nucleotides in length and is found within the repC gene. Laboratory experiments have shown that the RepC protein binds to this region, suggesting its role as the origin of replication. Therefore, while the complete process behind the replication of the Ti plasmid has not been fully described, the initial step of replication would likely depend on the expression of RepC and its binding to oriV. Of note, the RepC protein only acts in cis, where it only drives the replication of the plasmid it is encoded in and not any other plasmid also present in the bacterial cell. The Ti plasmid is maintained at low copy numbers within a bacterial cell. This is partly achieved by influencing the expression of the replication initiator RepC. When bound to ADP, RepA is activated to work with RepB, acting as a negative regulator of the repABC cassette. The levels of RepC are therefore kept low within a cell, preventing too many rounds of replication from occurring during each cell division cycle.

Furthermore, there is a small RNA known as RepE encoded between repB and repC that lowers the expression of repC. RepE is complementary to RepC and will bind with the repC mRNA to form a double-stranded molecule. This can then block the translational production of the RepC protein. Separately, the expression of the repABC cassette and hence the copy number of the Ti plasmid is also influenced via a quorum sensing system in *Agrobacterium*. Quorum sensing systems respond to bacterial population densities by sensing a molecule, known as an auto inducer, that is produced by the bacterial cells at low levels and would build up to a threshold level when there is a high density of bacteria present. In this case, the autoinducer is the N-3-oxooctanoyl-L-homoserine lactone (3-OC8-AHL) molecule, which is sensed by a regulator known as TraR. When activated, TraR will bind to regions known as tra boxes in the repABC gene cassette's promoter regions to drive expression. Therefore, a high level of population density increases the number of plasmids present within each bacterial cell, likely to support pathogenesis in the plant host.

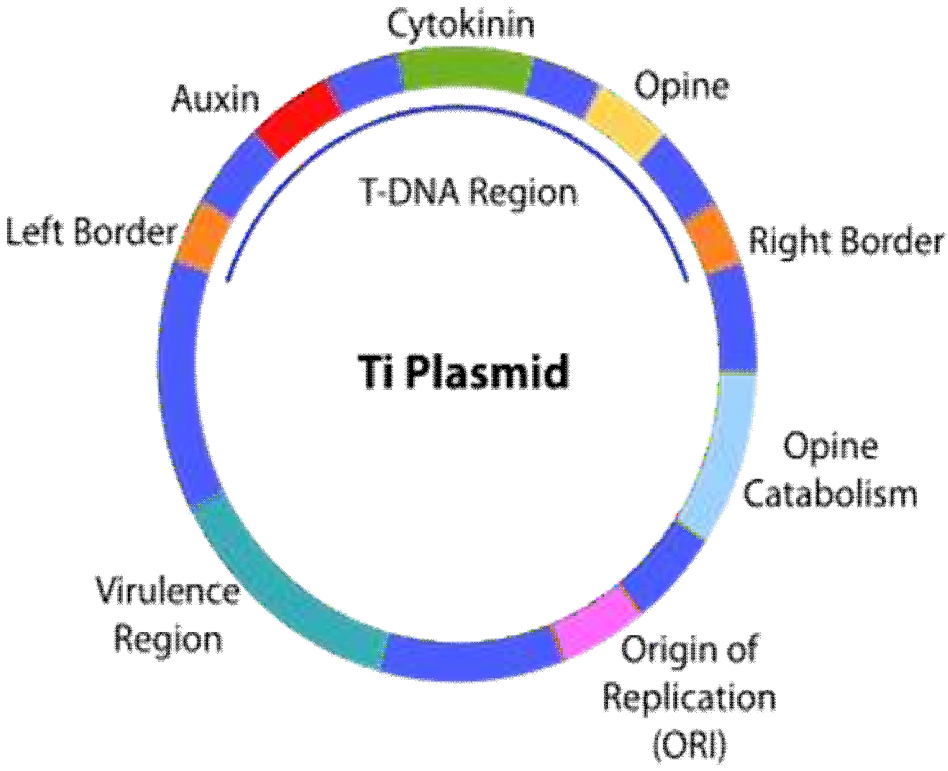


Fig-3.5: Structure of Virulence (Ti) - Plasmid

3.5 PROPERTIES OF PLASMIDS

3.5.1. F-Plasmid

The fertility factor (first named F by one of its discoverers Esther Lederberg; also called the sex factor in *E. coli* or the F sex factor; also called F-plasmid) allows genes to be transferred from one bacterium carrying the factor to another bacterium lacking the factor by conjugation. The cells containing this plasmid are designated as F^+ and those without it as F^-. F^+ bacteria are considered as male, because they can act as donor of not only the plasmid, but also chromosomal genes to the F^- cells which act as recipient and are, therefore, considered as female (Fig.3.6). The F factor is carried on the F episome, the first episome to be discovered. Unlike other plasmids, F factor is constitutive for transfer proteins due to a mutation in the gene finO. The F plasmid belongs to a class of conjugative plasmids that control sexual functions of bacteria with a fertility inhibition (Fin) system.

The most common functional segments constituting F factors are (Fig.3.2):

- OriT (Origin of Transfer): The sequence which marks the starting point of conjugative transfer.
- OriC (Origin of Replication): The sequence starting with which the plasmid-DNA will be replicated in the recipient cell.
- Tra-region (transfer genes): Genes coding the F-Pilus and DNA transfer process.
- IS (Insertion Elements) composed of one copy of IS2, two copies of IS3, and one copy of IS1000: so-called "selfish genes" (sequence fragments which can integrate copies of themselves at different locations).

When an F+ cell conjugates/mates with an F− cell, the result is two F+ cells, both capable of transmitting the plasmid to other F− cells by conjugation. The F-plasmid belongs to a class of conjugative plasmids that control sexual functions of bacteria with a fertility inhibition (Fin) system. In this system, a trans-acting factor, FinO, and antisense RNAs, FinP, combine to repress the expression of the activator gene TraJ. TraJ is a transcription factor that up regulates the tra operon. The tra operon includes genes required for conjugation and plasmid transfer. This means that an F + bacteria can always act as a donor cell. The finO gene of the original F plasmid (in *E. coli* K12) is interrupted by an IS3 insertion, resulting in constitutive tra operon expression. F + cells also have the surface exclusion proteins TraS and TraT on the bacterial surface. These proteins prevent secondary mating events involving plasmids belonging to the same incompatibility (Inc) group. Thus, each F+ bacterium can host only a single plasmid type of any given incompatibility group.

In the case of Hfr transfer, the resulting trans conjugates are rarely Hfr. The result of Hfr/F− conjugation is a F− strain with a new genotype. When Fprime plasmids are transferred to a recipient bacterial cell, they carry pieces of the donor's DNA that can become important in recombination. Bioengineers have created F plasmids that can contain inserted foreign DNA; this is called a bacterial artificial chromosome. The first DNA helicase ever described is encoded on the F-plasmid and is responsible for initiating plasmid transfer. It was originally called *E. coli* DNA Helicase I, but is now known as F-plasmid TraI. In addition to being a helicase, the 1756 amino acid (one of the largest in *E. coli*) Fplasmid TraI protein is also responsible for both specific and non-specific single-stranded DNA binding as well as catalyzing the nicking of singlestranded DNA at the origin of transfer.

The process of transfer takes place by conjugation of the F^+ cell with the F^- cell. The F-plasmid is a conjugative plasmid. When F-plasmid is integrated into the *E. coli*

chromosome, the bacterial cell changes to an Hfr-strain (high frequency of recombination). There are many sites on the *E. coli* chromosome where the F-plasmid can be integrated. Depending on the site, each integration gives rise to a different Hfr-strain. In $F^+ \times F^-$ conjugation, the plasmid alone is transmitted, but in Hfr x F^- conjugation, chromosomal genes are transmitted and rarely also the F-plasmid.The F-plasmids containing parts of the chromosomal DNA are designated as F'-plasmids. When an F-plasmid loses some of its essential genes during the excision process, the plasmid is rendered incapable of independent existence and is, ultimately, eliminated during cell division.When an F-plasmid is transmitted by conjugation to an F-recipient, it can transfer the chromosomal genes carried by it. Thereby, the recipient becomes diploid in respect of these transferred genes (because it now contains one copy of its own and another copy of the same gene transmitted by the F - plasmid). Thus, exchange of chromosomal genes may occur through F'-plasmids. This has been described as sex-duction.

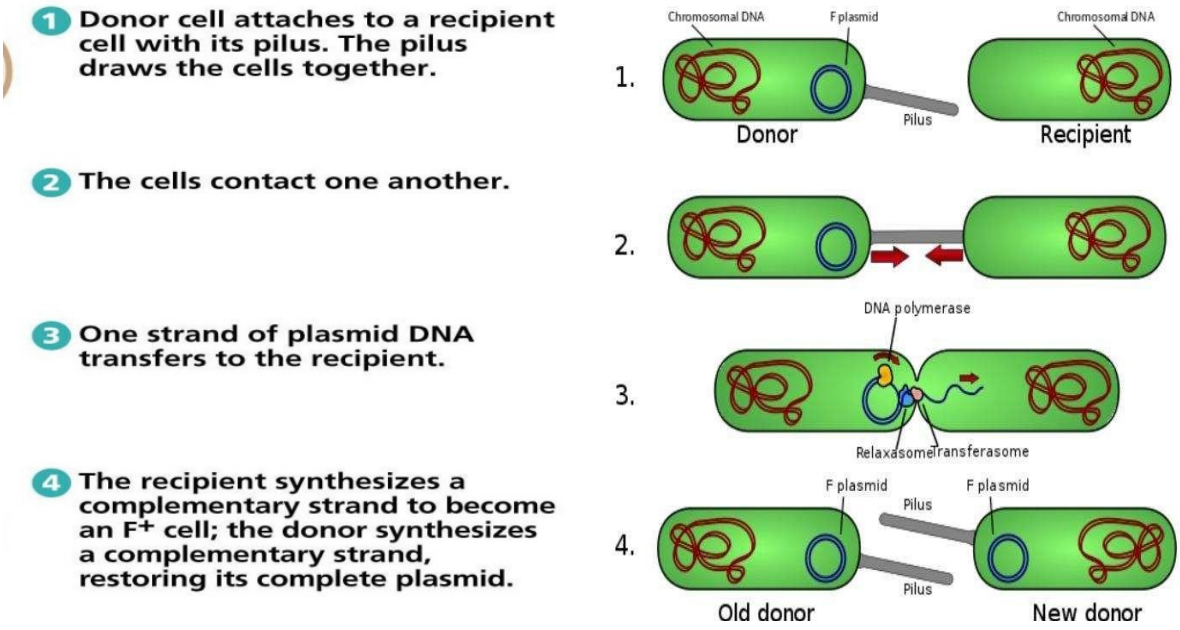


Fig-3.6: Mechanism of gene transfer in F-Plasmid

3.5.2. R-Plasmids

R-plasmids conferring resistance to various drugs individually or multiple resistance to several antibacterial agents were first discovered in Japan in the 1950s in the gastroenteritis-causing *Shigelladysenteriae*. Since then these plasmids have been found in *E. coli* and other enteric bacteria. Such plasmids have proved a great threat to the medical science. The large R-plasmids having molecular weights ranging between 30 x 10^6 Daltons are self- transmissible by conjugation with other bacteria. They are, therefore, conjugative plasmids, like the F-plasmid.

Smaller R-plasmids having molecular weights of about 5 to 6 x 10^6 Daltons are non-transmissible. Most of the self-transmissible large plasmids like R100 of *Shigella* conferring multiple drug resistance co-integrate two DNA segments joined to each other by covalent linkage to form a single double-stranded circular molecule. One DNA segment is called the resistance transfer factor (RTF), while the other segment contains the drug-resistance genes. The RTF is mainly involved in the transfer function of the R-plasmid and contains a number of genes (the transfer genes) and some others controlling replication of the plasmid in the

host cell (Fig.3.7). The resistance genes located in the other segment elaborate enzymes for destruction of the antibacterial drugs, like penicillins, streptomycin, chloramphenicol, tetracyclines, kanamycin, sulfonamides etc.

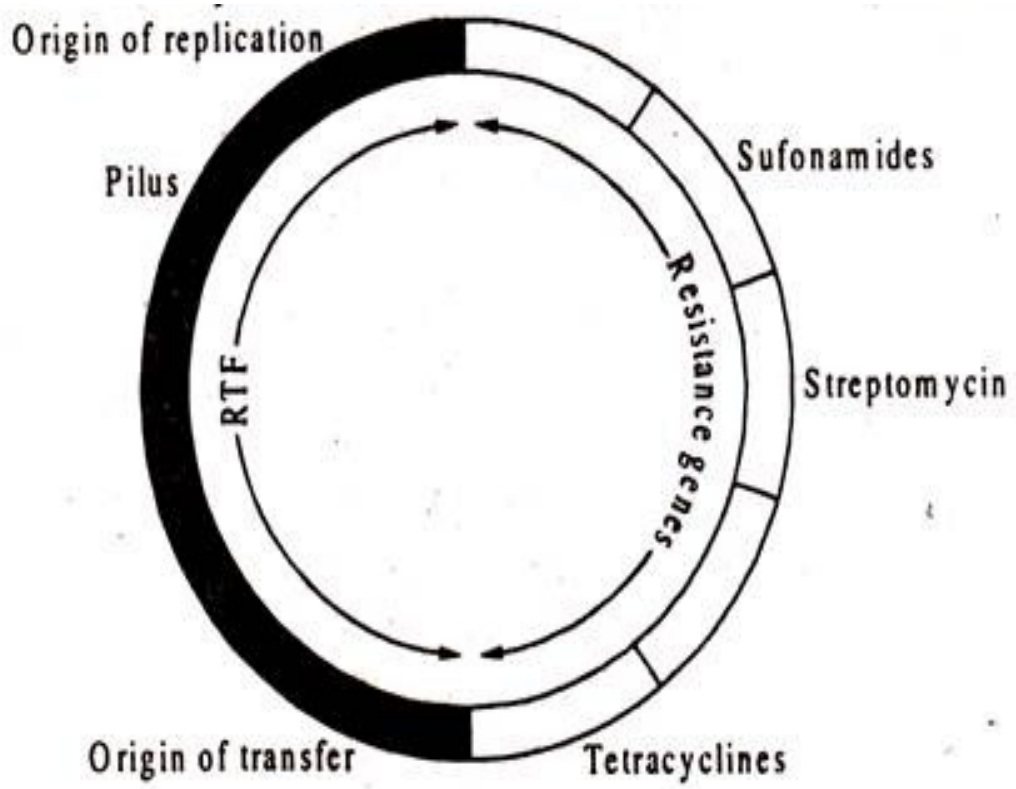


Fig-3.7: Structure of R-Plasmid

In some drug-resistant bacteria, such as *Salmonella typhimurium* strain 29, the resistance genes are located not in the same plasmid, but in separate plasmids of different size. This is sometimes known as plasmid aggregation. The transposable elements that complex transposons may carry genes for drug resistance. Such elements can be integrated into plasmids giving rise to a drug-resistance plasmid. Thus, R plasmids may be made up of a collection of transposons, each of which may carry one or more genes for antibiotic resistance. For example, Tn5 carrying a gene for kanamycin resistance may be inserted into the plasmid R100 of *Shigella* making the plasmid able to resist the antibiotic. Besides drug resistance, plasmids may also make bacterial hosts resistant to the toxic effects of heavy metals. Plasmid-coded resistance to nickel, cobalt, mercury, arsenic and cadmium has been reported in different species belonging to the genera *Pseudomonas, Escherichia, Salmonella* and *Staphylococcus.*

3.5.3. Col-Plasmids

The Col-plasmids are present in different strains of *E. coli* and they contain genes controlling synthesis of a class of proteins called colicines. Colicines are able to inhibit the growth of related bacteria which lack a Col-plasmid. Several different types of Col-plasmids have been discovered, each of which produces colicines having a different mode of inhibition of susceptible bacteria. For example, Col B induces a damage of the cytoplasmic membrane of the target bacteria and Col E2 and Col E3 cause degradation of nucleic acids.

Like R-plasmids, Col-plasmids may be self-transmissible or non-self transmissible. Large Col plasmids, like Col I and Col V-K94 having molecular weights of 60 x 10^6 Daltons or above are self- transmissible. They have a small copy number, usually 1 to 3 copies per cell. Small Col-plasmids, like Col-El, have molecular weight weighs of about 4 to 5 x 10^6 Daltons. They have a high copy number, usually 10 to 30 copies per cell. They are self-non-transmissible, but may be mobilized with the help of F-plasmid. This means that when an F^+-cell contains also a Col El plasmid and conjugates with an F- cell, the Col El plasmid can be transferred to the recipient through the mating bridge constructed by the F-plasmid. Obviously, an F-$ColEl^+$ cell is unable to mobilize the Col-plasmid to another cell, because it is unable to build a mating bridge.

In contrast, the large Col-plasmids are self-transmissible, because they have the genes for building the conjugation apparatus themselves and do not depend on the F-plasmid for transfer to other cells. Like F and large R-plasmids, the large Col-plasmids are also conjugative plasmids.Colicins belong to a general class of proteins, called bacteriocins. Many bacteria have been found to elaborate bacteriocins which are able to kill other related or even unrelated bacteria. Such proteins are coded by genes present in bacteriocinogenic plasmids.

Bacteriocins produced by different bacteria are sometimes given different names, like pyocine produced by *Pseudomonas aeruginosa*, megasine by *Bacillus megaterium*, nisin by lactobacilli, etc. In general, bacteriocins exert their antibacterial action by binding to the cell wall of the target cells and by inhibiting one of the vital metabolic processes, like replication of nucleic acids, transcription, and proteinsynthesis or energy metabolism. Bacteriocins produced by enteric bacteria help to maintain a healthy ecological balance in the human colon. Other bacteriocins produced by bacteria under natural environmental conditions probably function by eliminating competitors. Nisin produced by lactic acid bacteria has been commercially used for preservation of food and dairy products.

3.5.4. Degradative Plasmids

Degradation or dissimilation of organic compounds in course of mineralization is often controlled by plasmid-borne genes in many microorganisms. Such plasmids with genes coding for enzymes that catabolize complex organic molecules are known as degradative or dissimilation plasmids. For example, in species of *Pseudomonas*, both chromosomal and plasmid genes produce enzymes for breakdown of complex compounds. Some of the plasmid genes code for enzymes which degrade such unusual compounds like camphor, toluene, naphthalene, salicylate and complex hydrocarbons of crude petroleum. With the help of these enzymes, the bacteria can utilize these compounds as source of carbon and energy.As a result, bacteria possessing such degradative plasmids stand a much better chance of survival under conditions where only such unusual compounds are available. Normal bacteria without such plasmid-coded enzymes would perish under similar conditions.

The capability of organisms carrying degradative plasmids to metabolize unusual diverse complex compounds suggests the possibility of employing them as means of bioremediation of the polluted environment. The development of genetic engineering techniques has encouraged scientists to develop genetically improved strains of bacteria containing plasmids capable of degradation of an array of complex compounds, such as those occurring in crude petroleum. A synthetic strain of *Pseudomonas* has been developed by AnandamohanChakraborty of the University of Illinois, USA offering prospects of practical use in removing oil-spills in the oceans, caused by leakage of crude petroleum from tankers. Oil-spills prove a great danger to marine life, both plants and animals.

3.5.5 Ti-Plasmid of Agrobacterium

Ti-plasmid is a tumour-inducing large extra-chromosomal double stranded circular DNA which is present in *Agrobacterium tumefaciens,* a plant-pathogenic bacterium causing the crown-gall disease in many dicotyledonous species (Fig.3.8). Crown-gall is a tumour produced at the collar region of plants by agrobacteria which possess the Ti-plasmid. Bacteria lacking the plasmid are non-virulent.

Ti-plasmid is about 200 kilo base-pair long circular DNA. Only a small part of this large molecule, a 30 kilo base-pair long fragment is responsible for tumour formation. This fragment is called the T-DNA (T stands for transformation). When Agrobacterium infects a susceptible host plant, the Ti-plasmid is released in the host cell and a copy of the T-DNA is integrated into the genome of the host plant.The integrated T-DNA then stimulates cellular atrophy producing eventually a tumour, called a crown gall. The T-DNA insertion in plant host genome is the first instance of an inter-kingdom genetic exchange by natural means.

A notable feature of T-DNA is that once it is incorporated into the host genome, the presence of the pathogenic organism is no longer necessary for induction of tumour. Thus, a close parallelism with cancer induction in animal cell is observed. The T-DNA segment of the Ti-plasmid contains genes controlling synthesis of phytohormones, like indole acetic acid and cytokinins, as well as several other compounds, called opines. Opines, such as octopine and nopaline are used as growth substrates by agrobacteria.

The rest of the Ti-plasmid contains several genes controlling virulence (vir genes). These genes control T-DNA transfer to the host. Other genes of the plasmid control functions relating to bacterial conjugation, DNA replication and catabolism of the opines synthesised by gene products of the T-DNA segment. The T-DNA acts as a mobile unit like a transposon, but it does not have a gene, like transposaseto mediate its own mobilization. Its mobilization is effected by genes located in the Ti-plasmid, but outside T-DNA. The 30 kilo base long T-DNA is flanked on either side by 25 base pair imperfect direct repeats forming T-DNA borders.The vir genes of Ti-plasmid are involved in the generation of a transferable copy of T-DNA and its transfer to plant cell through the cell membrane and the nuclear membrane, as well as through the bacterial and plant cell walls. T-DNA is transferred as a single-stranded copy.

The copy is separated from T-DNA segment, capped at the 5′-end by a protein coded by a vir gene (vir D2) and covered by a large number of protein molecules coded by another vir gene (vir E2). This T-complex is transported to the plant cell through a membrane pore produced by another vir gene. The T-complex (ss-DNA + proteins) is about 3.6 μm long and less than 2 nm thick. The ability of *Agrobacterium tumefaciens* to transfer its Ti-plasmid to many dicotyledonous plants (but not monocotyledonous ones) opened up the possibility of introducing foreign genes into the hosts using the Ti-plasmid as a vehicle (vector).This has been practically employed to insert a gene of interest into the T-DNA segment by recombinant DNA technology. The tumour-inducing genes and other unnecessary genes of T-DNA are removed and replaced by the gene chosen for insertion. Several foreign genes have been introduced into a variety of hosts to produce transgenic plants.

Among the notable achievements is the production of transgenic plants resistant to the herbicide glyphosate and to feeding insects. Glyphosate resistance gene was isolated from Salmonella and the insect-resistance gene from *Bacillus thuringiensis* which synthesise an insecticidal protein. Another interesting achievement though not of practical significance was production of bioluminescent tomato plants by introducing the gene controlling bioluminescence in fire-fly.

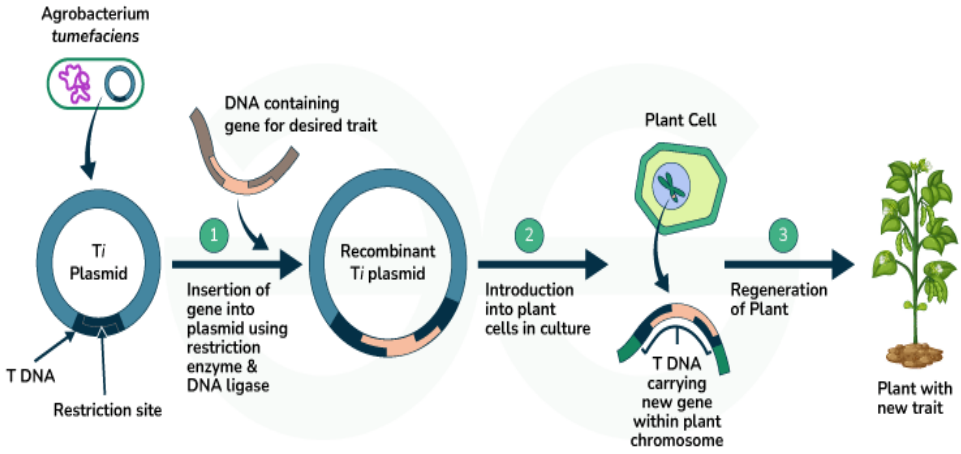


Fig-3.8: Mechanism of Ti-Plasmid transfer in *Agrobacterium*

3.5.6 Other Plasmids (pBR322)

Plasmid pBR322

In the 1980s, one of the best-studied and most often used "general-purpose" plasmid cloning vectors is pBR322. In general, plasmid cloning vectors are designated by a lowercase p, which stands for plasmid, and some abbreviation that may be descriptive or, as is the case withpBR322, anecdotal. The BR of pBR322 recognizes the work of the researchers F. Bolivar and R.Rodriguez, who created the plasmid; and 322 is a numerical designation that has relevance to these workers. Plasmid pBR322 contains 4,361 bp. As shown in Fig. 3.9, pBR322 carry two antibioticresistance genes. One confers resistance to ampicillin (Amp), and the other confers resistance totetracycline (Tet). This plasmid also has unique *BamHI*, HindIII, and *SalI* recognition sites with in the tet gene; a unique *PstI*site in the amp gene; a unique EcoRI site that is not within any coding DNA; and an origin of replication that functions only in *E. coli*. It is maintained at a high copy number in *E. coli* and cannot bereadilytransferred to otherbacteria.

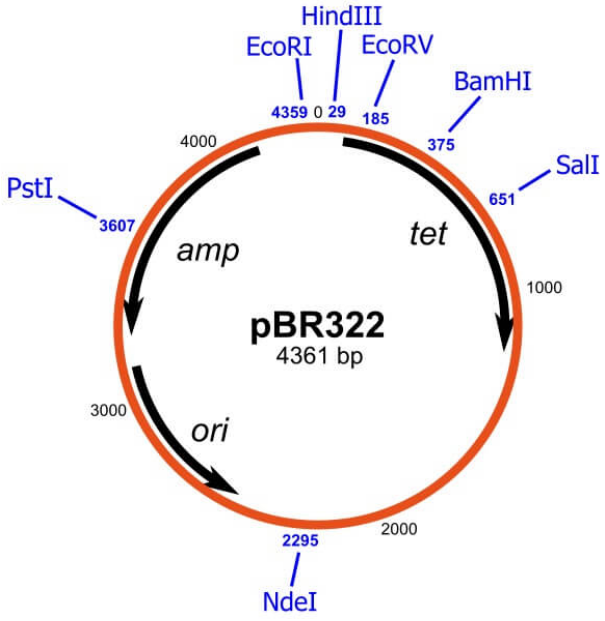


Fig-3.9: Structure of the plasmid cloning vector pBR322

Purified, closed circular pBR322 molecules are cut with are striction enzyme that lies within either of the antibiotic resistance genes and cleaves the plasmid DNA only once to create single, linear, sticky-ended DNA molecules (Fig.3.10). These linear molecules are combined with prepared target DNA from a source organism. This DNA has been cut with the same restriction enzyme, which generates the same sticky ends as those on the plasmid DNA. The DNA mixture is then treated with T_4DNA ligase in the presence of ATP. Under these conditions, a number of different ligated combinations are produced, including the original closed circular plasmid DNA. To reduce the amount of this particular unwanted ligation product, the cleaved plasmid DNA preparation is treated with the enzyme alkaline phosphatase to remove the 5'-phosphate groups from the linearized plasmid DNA. As a consequence, T_4 DNA ligase cannot join the ends of the dephosphorylated linear plasmid DNA. However, the two phosphor diester bonds that are formed by T_4DNA ligase after the ligation and circularization of alkaline phosphatase-treated plasmid DNA with restriction endo nuclease-digested source DNA, which provides the phosphate groups, are sufficient to hold both molecules together, despite the presence of two nicks. After transformation, these nicks are sealed by the host cell DNA ligase system. In addition, although unwanted, fragments from the source DNA are also joined to each other by T_4DNAligase.

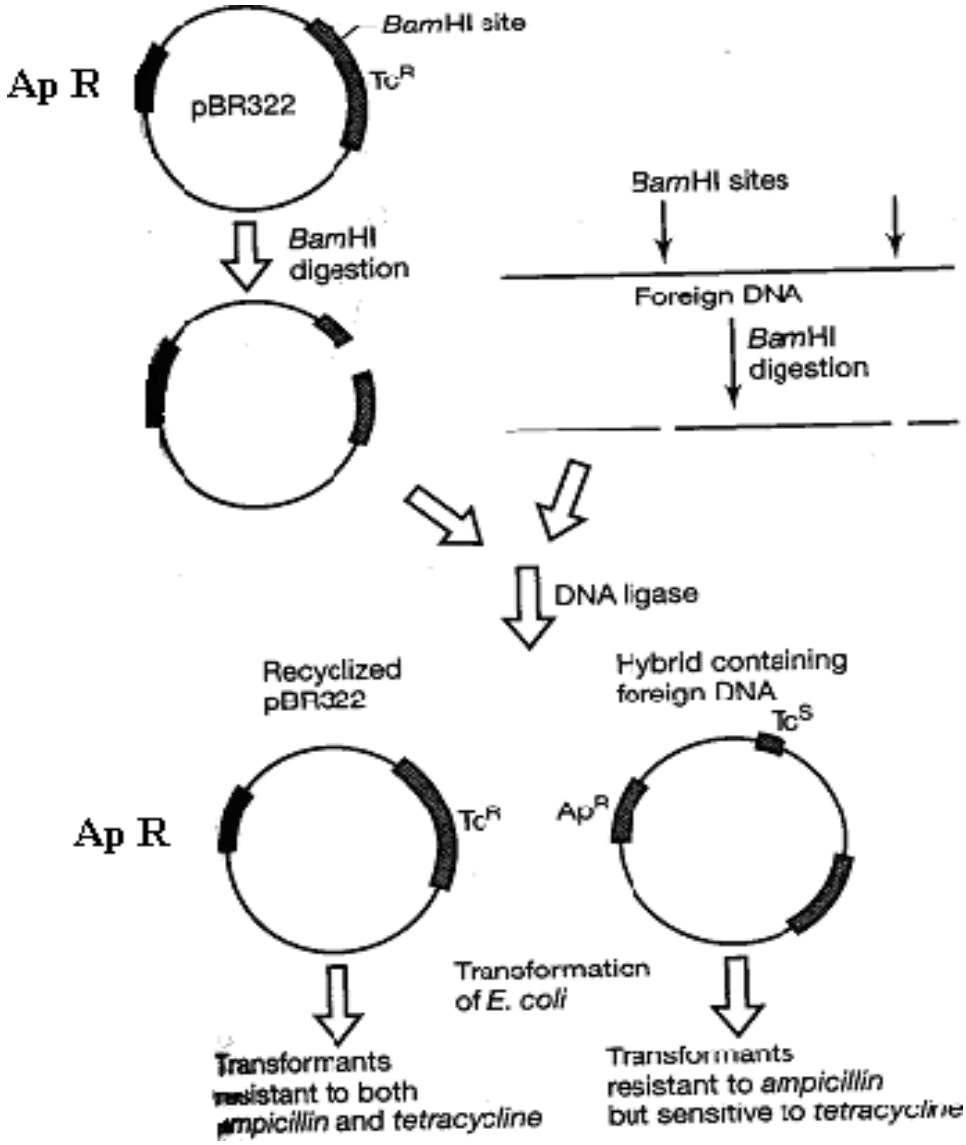


Fig-3.10: The use of plasmid pBR322 as acloning vector

Plasmidp UC19

The plasmid pBR322 is a well-conceived cloning vector. It has only a few unique cloning sites, and its use involves a time-consuming selection procedure. Thus, it was inevitable that other systems would be developed. For example, the plasmid pUC19 is 2,686 bp long and contains an ampicill in resistance gene, B-galactosidasegene *(IaeZ')* of the lactose operon of *E. coli,* a *lacI* gene that produces a repressor protein that regulates the expression of the *IaeZ'* gene, a short sequence with multiple unique cloning sites (i.e., EcoRI, *SacI, KpnI, XmaI, SmaI,* Bam HI, *XbaI, SalI, HindI, Aeel, BspMI, PstI, SphI,* and *HindIII),* and the origin of replication from pBR322 (Fig.3.11).

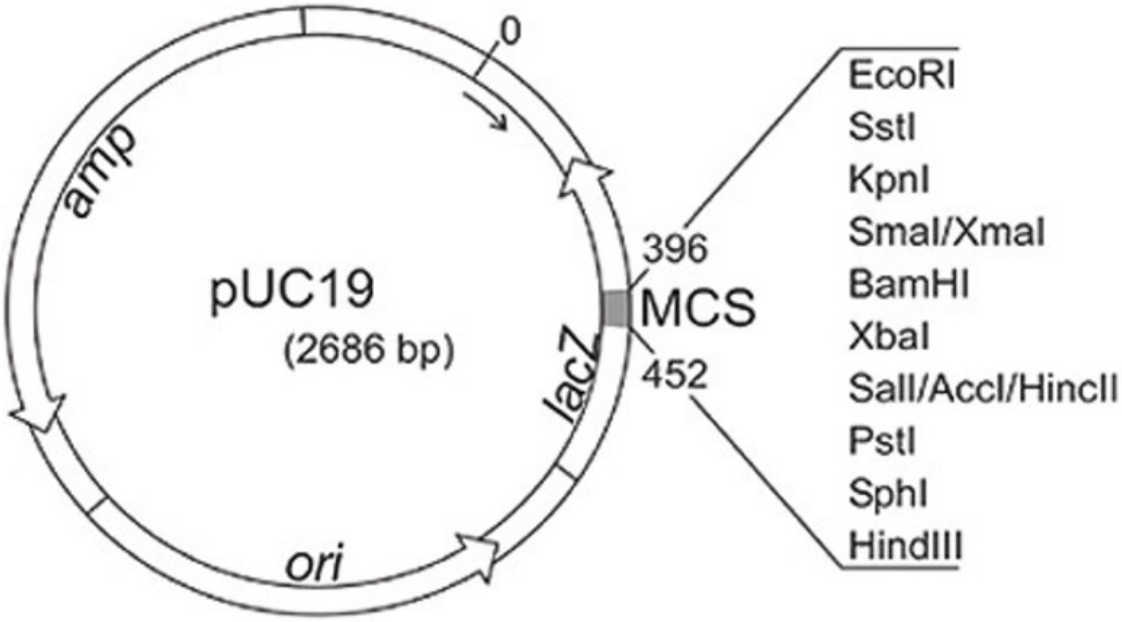


Fig-3.11: Genetic map of the plasmid cloning vector pUC19

For a pUC19 cloning experiment, DNA from a source organism is cut with one of the restriction endo nucleases for which there is a recognition site in the multiple cloning sequence. This source DNA is mixed with pUC19 plasmids that have been treated initially with the same restriction endo nuclease and subsequently with alkaline phosphatase. After ligation with T_4 DNA ligase, the reaction mixture is introduced by transfore1ation into a host cell which can synthesize that part of β -galactosidase (*Lac* Z α) that combines with the product of the *lacZ'* gene to form a functional enzyme. The treated host cells are plated onto medium that contains ampicillin, IPTG, and X-Gal. Non transformed cells cannot grow in the presence of ampicillin. Cells with circularized plasmids can grow with ampicillin in the medium; because they can form functionalβ-galactosidase, they produce blue colonies. In contrast, host cells that carry a plasmid-cloned DNA construct produce white colonies on the same medium.

In addition to pBR322 and the pUC-series plasmids, many other cloning vectors have been devised. Some vector shave inventive selection systems to identify clones with insert-vector constructs. For example, a vector that is derived from the pUC series carries a gene that, when expressed, encodes a protein that kills the cell. This cell-killing gene (suicide gene) is fused in the correct reading frame to the *lacZ'* gene. A cell with an intact plasmid and no IPTG in the medium does not synthesize the suicide protein. With an insert and IPTG, a

non functional suicide protein isproduced because the insert, in all likelihood, disrupts the reading frame of the suicide gene. Bycontrast, cells with a plasmid and no insert in the presence of IPTG synthesize the suicide protein and are killed. Non transformed cells are sensitive to an antibiotic, whereas recombinant clone shave as part of the vector a gene that confers resistance to the antibiotic. In other words, the only surviving cells in the presence of IPTG and antibiotic are those that carry insert DNA. One suicide gene that has been used in this way encodes an enzyme that prevents the rejoining of double-stranded breaks in the chromosomal DNA. These DNA breaks accumulate and are responsible for the death of the cell.

Although a number of vectors have ingenious designs, in principle they all retain the two basic requirements of recombinant DNA technology: a choice of cloning sites and an easy way to identify cells that have plasmid cloned DNA constructs. It should be noted that unique restriction endo nuclease sites have a dual function in recombinant DNA research: they are essential for inserting DNA into a cloning vector, and they allow an inserted DNA sequence to be recovered from the vector. In other words, after a piece of DNA has been cloned into a site, it can be retrieved by cutting a purified plasmid-cloned DNA construct with the original restriction endo nuclease because the insertion event creates two recognition sites at the ends of the cloned DNA sequence. Occasionally, the initial recognition site is destroyed. Then the cloned piece of DNA is difficult to retrieve. A recovered DNA fragment can be cloned into cloning vectors for DNA sequencing or vectors that have been specially designed to achieve high levels of expression (transcription and translation) of the cloned gene.

3.6 PLASMID REPLICATION

Bacterial plasmid replication is not dependant on its nuclear genome replication with long intermissions between replication proceedings occurring during the course of cell division. Definite plasmid copy number depends on plasmid type, host organism and the growth conditions. Unintended aberrations from normal copy number are attuned. However dominant and recessive copy mutants to the wild type do exist.

Plasmid replication mechanisms

There are three types of plasmid replications namely rolling circle, Col E1 type and Iteron contain replication.

Rolling circle: Rolling circle replication mechanism is specific to bacteriophage family M13 and the fertility F factor which encodes for sex pili formation during recombination by means of conjugation. Fragments smaller than 10 kilo bases are usually replicate by this replication mechanism and reported in some Gram positive bacteria. It allows the transfer of single stranded replication product at a faster rate to the recipient cell through pilus as in case of fertility factor or to the membrane in case of phage.Rolling circle occurs to a covalently closed circular piece of double-stranded DNA. A nick is produced in one of the strands by enzyme nickases creating a 5’ phosphate and a 3’ hydroxyl. Free 3’ hydroxyl will be used by DNA polymerase to make new DNA pushing the old nicked strand off of the template DNA (Fig.3.12)

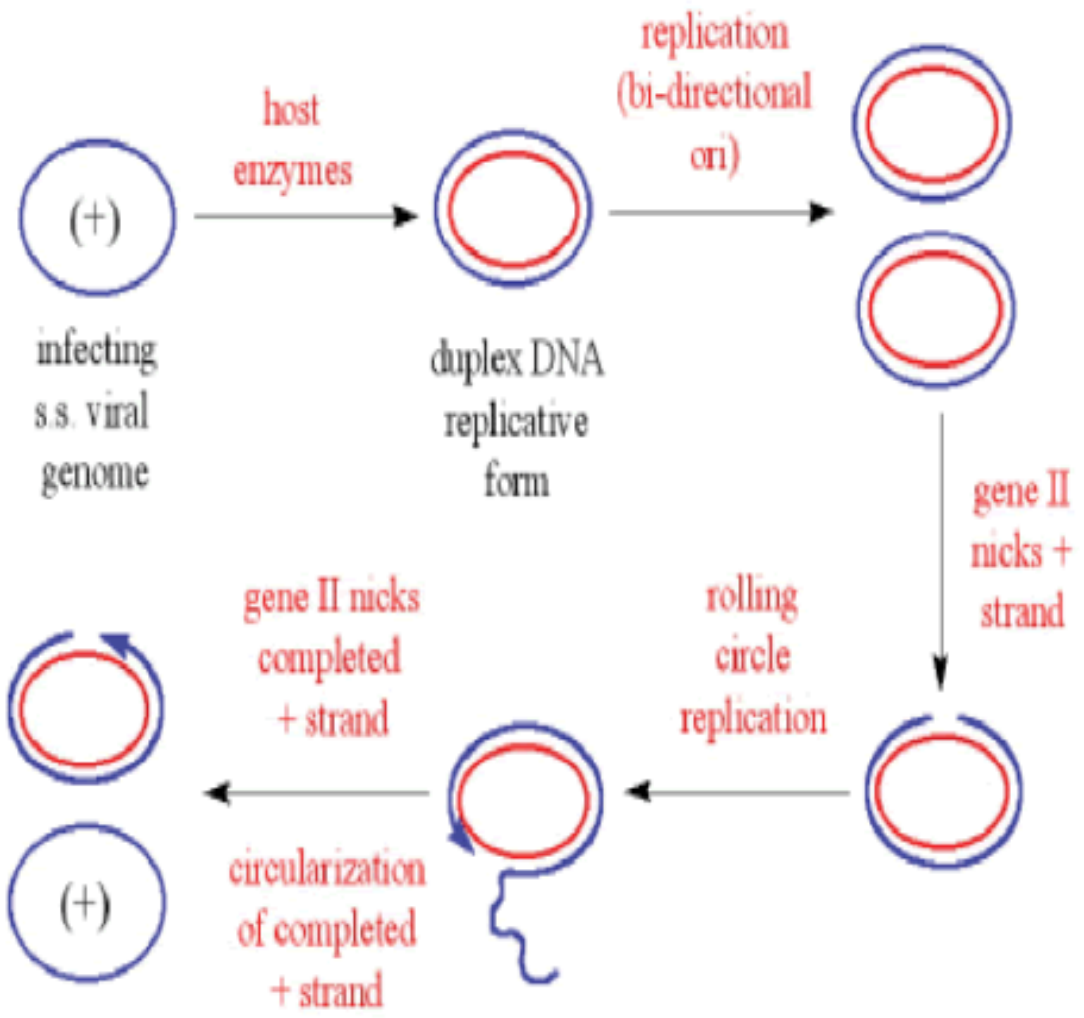


Fig-3.12: Mechanism of Rolling Circle Mechanism in Plasmid DNA

Col E1 type Replication: Col E1 replication is a negative regulation mechanism which enables the plasmid to control its own copy numbers by involving RNA type I, RNA type II, Rom protein, and the plasmid itself (Fig.3.13). Col E1 replication is initiated by means of RNA-RNA interactions and does not rely on replication initiation protein encoded by the plasmid to regulate its copy number. RNA type II that originates 555 base pairs upstream from the replication origin of Col E1 plasmid is transcribed which marks the start of Col E1 replication. A determined hybrid with the DNA strand is formed by a loop enriched in G nucleotide positioned 290 of RNAII and a C-rich region on the template strand positioned 20 nucleotides upstream from the origin. Several stems and loops are exhibited by the newly formed secondary structure. A DNA/RNA hybrid is recognized by enzyme RNase and dissociates the RNA hybrid to the 3' end of RNAII. The resultant RNA primer is linked to the plasmid with a free 3' hydroxyl group. This RNA enables replication of DNA to begin by providing DNA polymerase a specific site to initiate nucleotides synthesis. Consequently DNA synthesis is commenced with the leading strand is happening.

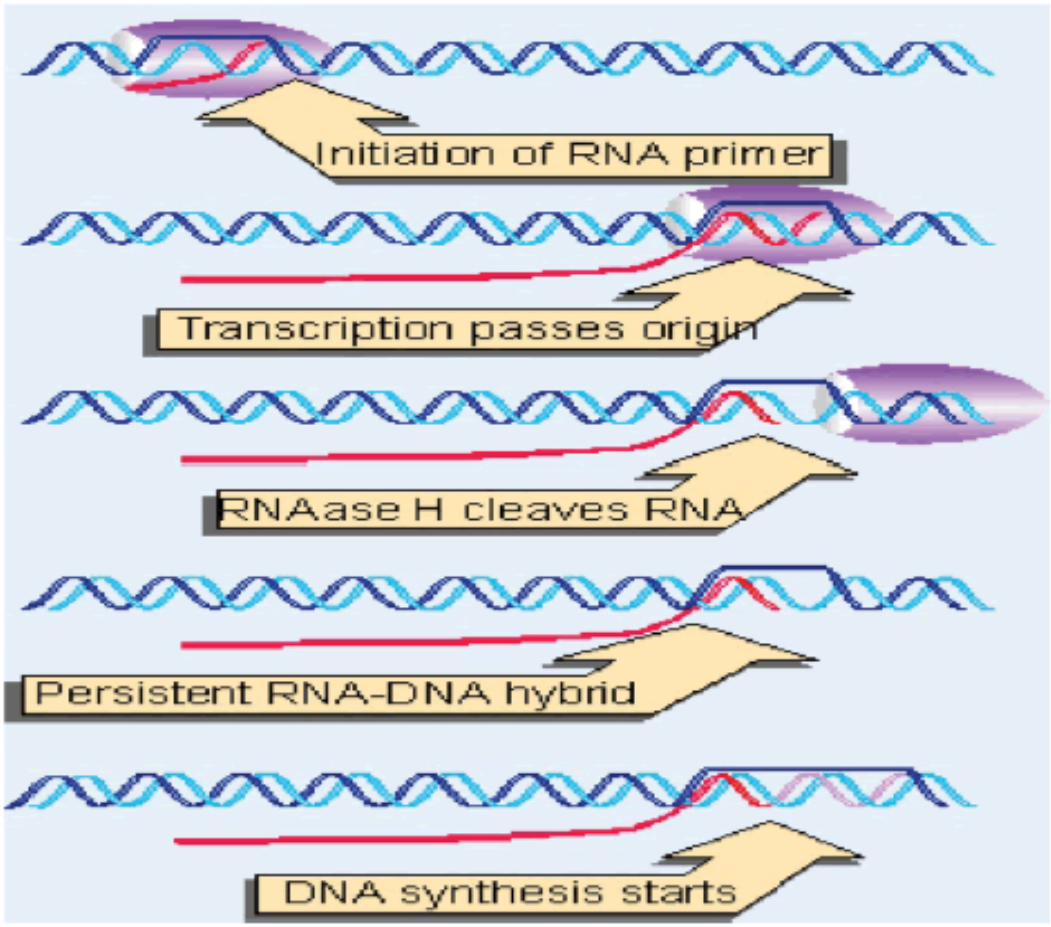


Fig-3.13: Mechanism of Col E1 type replication in Plasmid DNA

Plasmid amplification is provided in *Escherichia coli* bacteria cells. Plasmid linearization by restriction cleavage can be ordered as a follow-up service. Such an operation is recommended especially when the plasmid is used as a PCR standard. Amplified plasmids are delivered either in midiprep or maxiprep quantities. Plasmid curing Plasmid curing occurs naturally through cell division or by treating the cells with any chemical or physical agents. The inhibition of conjugational transfer of antibiotic resistance plasmid can be used to decrease the spread of antibiotic resistance plasmid in the environment. Regulation of copy number Plasmids must regulate their copy number (average number of plasmid copies per cell) to ensure that they do not excessively burden the host or become lost during cell division. Plasmids may be either high copy number plasmids or low copy number plasmids; the regulation mechanisms between these two types are often significantly different. Biotechnology applications may involve engineering plasmids to allow a very high copy number. For example, pBR322 is a low copy number plasmid (~20 copies/cell) from which several very high copy number cloning vectors (~1000 copies/cell) have been derived. High copy number plasmids, also called relaxed plasmids, require a system to ensure that replication is inhibited once the number of plasmids in the cell reaches a certain threshold. Relaxed plasmids are generally regulated through one of two mechanisms: antisense RNA or iteron binding groups. Low copy number plasmids, also called stringent plasmids, require tighter control of replication.

3.7 ISOLATION OF PLASMID

Both genomic DNA isolation and plasmid DNA isolation are extraction procedures vital for the isolation of DNA to be used in the different principles and techniques in the field of biotechnology. Particularly, the isolation of DNA finds its applications in a polymerase chain reaction, gel electrophoresis and some sequencing techniques of DNA (Fig.3.14). The important difference between genomic DNA isolation and plasmid DNA isolation is that genomic DNA isolation makes use of strong lysis, involving the mechanical or enzymatic disintegration of the cell membranes to give out the genomic DNA into the solution. This DNA can be purified from the proteins and lipid membrane. On the other hand, the plasmid DNA isolation makes use of temperate alkaline lysis to obtain plasmid DNA into the solution with the genomic DNA. The separation of plasmid DNA from genomic DNA requires neutralization with potassium acetate.

Genomic DNA Isolation

- It is a simpler process as strong lysis is the only essential step to release the genomic DNA in the solution. Lysis of bacteria, plants and yeast requires the enzymatic breakdown of the cell wall (that is rigid and strong), before the plasma membrane is mechanically disrupted
- Once lysis is performed, the sample is purified using spin filter membrane technology along with guanidine salts that facilitate the association to silica

Plasmid DNA Isolation

- Plasmid DNA is extracted from cells in this process
- Here recognizing and using the apt lysis method for successful separation of plasmid DNA from the genomic DNA, the process can be a bit complicated
- The process includes – Cell Cultivation, Re-suspension, Cell Lysis, Neutralization, Cleaning and Concentration

Harvest and Lysis of Bacteria:

1) Transfer 1.5 ml of broth containing cells into 1.5 ml centrifuge tubes and centrifuge the cells at 13000 rpm for 1 min. Discard supernatant.
2) Re-suspend the cells in 200 l of Re-suspension solution. Pipette up and down or vortex for mixing.
3) Add 200 l of Lysis Solution. Invert gently to mix. Do not vortex. Allow to clear for about 5 min.
4) Preparation of clear lysate: Add 350 l of Neutralization solution (S3). Invert 4-6 times to mix.
5) Centrifuge at 13,000 rpm for 10 minutes.
6) Preparation of binding column: Add 500 l Column Preparation Solution to binding column in a collection tube.
7) Spin at 13,000rpm for 1 min. Discard flow through.
8) Binding plasmid DNA to column: Transfer the cleared lysate to the binding column
9) Centrifuge at 13,000 rpm for 1 min. Discard the flow through.

10) Wash to remove contaminants: Add 750 µl Wash solution to column. Spin for 1 min. Discard flow through.

11) Spin for 1 min to dry the column (dry by speed vacuum). Allow at room temperature for 2-5 min.

12) Elution of purified Plasmid DNA: Transfer the column to a fresh collection tube. Add 70 µl of Elution Solution (or sterile sigma water). Spin for 1 min at 13,000 rpm. The eluate contains plasmid DNA. Store the plasmid DNA at - 20°C or -80°C.

A. Preparation of Agarose Gel for Electrophoresis of the isolated plasmid DNA:

1) Prepare 0.8% Agarose solution in 1X TBE.
2) Pour the agarose into Gel casting trays pre-set along with combs. Allow the agarose gel to polymerize and then remove combs without breaking the wells. Submerge the gel into the horizontal electrophoresis tank containing 1X TBE buffer.

B. Electrophoresis of the isolated plasmid DNA

1) Take about 5 µl of the (eluate) plasmid DNA and add 2 µl of Sample buffer.
2) Load about 7 µl into each well along with a marker DNA or Ladder.
3) Run electrophoresis at constant voltage and allow the DNA to run in 1x TBE running buffer, keeping track of the dye front.
4) Remove the gel and place in a solution of ethidium bromide for staining DNA for 30min. Remove the gel with gloves, rinse with water and place in the Gel Documentation system / Trans illuminator and visualize the DNA bands in the UV light.
5) Quantify the DNA concentration from the standard (ladder) used.

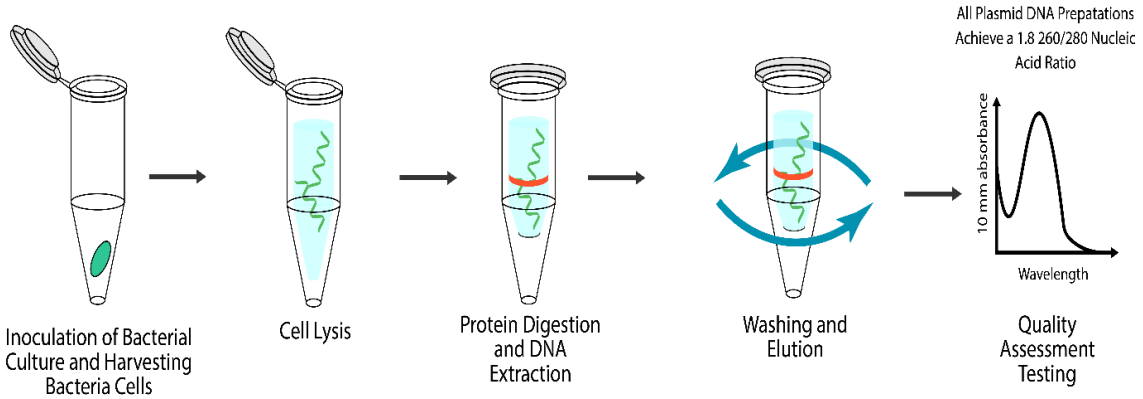


Fig-3.14: Isolation of plasmid DNA

3.8 SIGNIFICANCE OF PLASMIDS:

The presence of plasmids in a cell can also have other biological significance such as

- Nodulation and symbiotic nitrogen fixation: *Rhizobium*

- Transfer genetic information for a biochemical pathway for the degradation of organic compounds such as octane, camphor, naphthalene, salicylate etc. *Pseudomonas*.
- Pigment production: *Erwinia, Staphylococcus*
- Lactose, sucrose, urea utilization, nitrogen fixation: Enteric bacteria
- Plasmids can be constructed artificially (artificial plasmids are called vectors) and are used to introduce foreign DNA into another cell of interest. Plasmids play crucial roles in genetic engineering, molecular cloning and various areas of Biotechnology.
- Plasmids are also used For gene transfer as potential treatment in gene therapy ,so that it can express the protein lacking cells
- Plasmids encoding can help to deliver a therapeutic gene to a specific site ,so that cell damage & cancer causing Mutations are avoided
- It is the less expensive and mass production of protein can be obtained. Example: Insulin hormone production.

3.9 SUMMARY

A plasmid is a small, extrachromosomal DNA molecule within a cell that is physically separated from chromosomal DNA and can replicate independently. They are most commonly found as small circular, double stranded DNA molecules in bacteria; however, plasmids are sometimes present in archaea and eukaryotic organisms. In nature, plasmids often carry genes that benefit the survival of the organism and confer selective advantage such as antibiotic resistance. While chromosomes are large and contain all the essential genetic information for living under normal conditions, plasmids are usually very small and contain only additional genes that may be useful in certain situations or conditions. Artificial plasmids are widely used as vectors in molecular cloning, serving to drive the replication of recombinant DNA sequences within host organisms.

3.10 TECHNICAL TERMS

F-plasmids, Col plasmids, R-plasmids, Conjugative and Non-conjugative plasmids, Degradative plasmids, Col-plasmids, *Agrobacterium tumefaciens.*

3.11 SELF ASSESSMENT QUESTIONS:

1) Write about Characteristics of plasmids.
2) Explain about types of plasmids.
3) Write in detail about Properties of F-plasmid and Col plasmids.
4) Explain about Replication of plasmid.
5) Add a note on isolation of plasmid DNA.
6) Discuss about significance of plasmids.

3.12 SUGGESTED READINGS:

1) Campbell, N. A. and Reece J. B. (2011). Biology. IX Edition, Pearson, Benjamin, Cummings.
2) Gardner, E.J., Simmons, M.J., Snustad, D.P. (2008). Principles of Genetics. VIII Edition. Wiley India.
3) Griffiths, A.J.F., Wessler, S.R., Lewontin, R.C. and Carroll, S.B. Introduction to Genetic Analysis. IX Edition. W. H. Freeman and Co.
4) Klug, W.S., Cummings, M.R., Spencer, C.A. (2012). Concepts of Genetics. X Edition. Benjamin Cummings.
5) Lewin B. - Genes Xth edition, Wiley Eastern Limited, New Delhi
6) Powar C. B. - Genetics Vol. I and II, Himalaya Publishing House, Mumbai
7) Russell, P. J. (2009). Genetics- A Molecular Approach.III Edition. Benjamin Cummings.
8) Stricberger M. W. - Genetics, 4th Edition, McMillon Public Com. Inc, New York
9) Snustad, D.P., Simmons, M.J. (2009). Principles of Genetics.V Edition. John Wiley and Sons Inc.
10) Verma P.S. and V.K. Agrawal - Cell Biology, Genetics, Molecular Biology, Evolution and Ecology.

Dr. K. Nagaraju

LESSON-4

MUTATIONS-SPONTANEOUS AND INDUCED MUTATIONS, TYPES OF MUTATIONS AND CHEMICAL MUTAGENS

4.0 OBJECTIVE:

- To acquaint the students about the different types of mutations occur in living forms and their effects.

STRUCTURE:

4.1 INTRODUCTION

Mutation is the process by which a DNA base pair change or a chromosome change is produced. Thus, a mutation may be the result of any detectable change that affects DNA's chemical or physical constitution, its replication, its phenotypic function, or the sequence of one or more DNA base pairs. A mutation can be transmitted to daughter cells and even to succeeding generations, thereby giving rise to mutant cells or mutant individuals. If a mutant cell gives rise only to somatic cells in multi cellular organisms, a mutant spot or area is produced, but the mutant characteristic is to passed on to the succeeding generation. This type of mutation is called a somatic mutation. However, mutations in the germ-line of sexually reproducing organisms maybe transmitted by the gametes to the next generation, producing an individual with the mutation in both its somatic and germ line cells. Such mutations are called germ-line mutations. A somatic mutation affects the individual in which it happens, while a germ-line mutation affects individuals of the subsequent generations.

Mutations can occur spontaneously, but they can also be induced experimentally by the application of amutagen, any physical or chemical agent that significantly increases the frequency of mutational events above the spontaneous mutation rate. Mutations that result from treatment with mutagens are called induced mutations and naturally occurring mutations are spontaneous mutations. There are no qualitative differences between spontaneous and induced mutations. The manifestation of a mutant phenotype is typically the result of a change in DNA that results in the altered function or production of a protein.

4.2 SPONTANEOUS MUTATIONS VERSES INDUCED MUTATIONS

4.2.1 Spontaneous Mutations: The spontaneous mutations occur suddenly in the nature and their origin is unknown. They are also called as "Background Mutations" and have been reported in many organisms such as, *Oenothera*, maize, bread molds, bacteria, viruses, *Drosophila*, mice, man, etc. Spontaneous mutations occur naturally in the genome. They generally occur due to error during replication, mitosis, meiosis, etc. Mutations may also occur due to mobile genetic elements or transposons. Spontaneous mutations occur with non-zero probability even given a healthy, uncontaminated cell. They can be characterized by the specific change: Tautomerism - A base is changed by the repositioning of a hydrogen atom, altering the hydrogen bonding pattern of that base, resulting in incorrect base pairing during replication. Spontaneous mutations occur naturally without any apparent or known cause, i.e., they are not induced by an exposure to a mutagenic agent. There are two possible sources of origin of spontaneous mutations - (1) errors during DNA replication, and (2) mutagenic effects of the natural environment of the organisms. The DNA polymerase incorporates wrong bases during DNA replication with a frequency of 10^{-5}. But DNA polymerase also has $3' \rightarrow 5'$ exonuclease activity which enables it to proofread the newly synthesized DNA strands and, thereby, correct the errors made during DNA replication. The DNA polymerase proofreads every nucleotide newly added to the growing end of the polynucleotide chains. It detects a mis-matched nucleotide, incorporates the correct nucleotide in its place and then proceeds with the further synthesis of the chain.

The contribution of environment to spontaneous mutations is not clearly understood. Some of the solar radiations are surely mutagenic and are likely contributors to spontaneous mutations. For example, the UV rays of sunlight are highly mutagenic but they are unable to penetrate more than skin-deep in animals and higher plants. The mutagenic action of UV rays becomes dramatically apparent in human beings suffering from a genetic disease called xeroderma pigmentosum. These individuals lack an effective mechanism for the repair of damage to DNA caused by UV rays. As a result, they develop mild to severes kin cancer in the areas of body exposed to sunlight. On the other hand, both eukaryotic and prokaryotic genomes contain some mobile DNA elements, e.g., insertion sequences, which integrate at specific sites in their genomes and produce gene mutations. The rate of spontaneous mutations is very low and generally ranges between 10^{-8} and 10^{-10} per nucleotide per generation for forward mutations in different genes of bacteria and viruses. For eukaryotes the estimates range from 10^{-7} to 10^{-9} per nucleotide per generation. The rates of reverse mutations have been estimated in prokaryotes and are generally much lower than those for forward mutations.

Tautomerism is caused by certain chemical mutagens. In the next replication purines pair with pyrimidines and the base, pair is altered at a particular locus. The uncommon forms are unstable and at the next replication, the cycle reverts to their normal forms.

Depurination - Loss of a purine base (A or G) to form an apurinic site (AP site). Depurination is known to play a major role in cancer initiation.

Deamination - Hydrolysis changes a normal base to an atypical base containing a keto group in place of the original amine group. Examples include C → U and A → HX (hypoxanthine), which can be corrected by DNA repair mechanisms; and 5MeC (5-methylcytosine) → T, which is less likely to be detected as a mutation because thymine is a normal DNA base.

The main causes of spontaneous mutations are, Replication errors, Slipped strand mispairing, Wobble base pairing, Depurination, Deamination, Tautomerism and unequal crossing over.

4.2.2 Induced Mutations

Induced mutations do not occur spontaneously. They are induced through various chemical and physical agents known as mutagens. Mutagens greatly enhance the frequency of mutation. Besides naturally occurring spontaneous mutations, the mutations can be induced artificially in the living organisms by exposing them to abnormal environment such as radiation, certain physical conditions (i.e., temperature) and chemicals. Mutations produced due to the treatment with either a chemical or physical agents are called induced mutations. The agents capable of inducing mutations are known as mutagens and their capacity for inducing mutations is termed as mutagenic property. Induced mutations are useful in two different ways – (1) in genetic and biochemical studies, and (2) in crop improvement. The process of inducing mutations through treatment with a mutagen is known as mutagenesis. Since the rate of spontaneous mutation is so low, geneticists use mutagens to increase mutation frequency so that a significant number of organisms have mutations in the gene being studied. Treatments with various mutagens increase the mutation rate by several orders of magnitude. Some potent chemical mutagens may produce mutations at the rate of more than 1% per gene/generation. Generally two classes of mutagens, radiation and chemical, are used and both involve specific mechanisms of action.

Differences between Spontaneous Verses Induced Mutations

Spontaneous Mutation	Induced Mutation
Spontaneous mutations occur naturally and mainly due to error in replication	**Induced mutations occur due to physical or chemical agents**
Occurs due to slippage in natural processes	**Induced by mutagens**
Caused due to replication error, tautomeric shift, transposable genetic elements, unequal cross overs, etc.	**Caused due to base modification, base analogues, intercalating agents, base mispairing, radiations, etc.**
E.g. sickle cell anaemia	**E.g. skin cancer due to prolonged exposure to radiations**

4.3 TYPES OF MUTATIONS

4.3.1 Forward mutations and Backward Mutations

Point mutations generally fall into two classes in terms of their effects on the phenotype in comparison to the wild type. **Forward Mutations** are mutations that cause the genotype to change from wild type to mutant, and **Reverse Mutations** or reversions or back mutations are the mutations that cause the geno type to change from mutant to wild type. A genere version is a mutational event that causes a change from a mutant phenotype to wild-type or partially wild-type function. Reversion of an on sense mutation, for instance, occurs when a base pair change results in a change of them RNA non sense codon to acodon for an amino acid, the mutation is a true reversion. If the reversion is to some other amino acid, the mutation is a partial reversion, and complete function may be restored.Reverse mutations could occur in different ways. In a true reverse mutation, the original base pair sequence of the wild type may be restored. Thus, if a GC pair of the wild type sequence is replaced by an

AT pair to produce a forward mutation, a true reverse mutation could again substitute a GC pair in that position. Sometimes a different base pair may be inserted at the site of the altered pair which had produced the forward mutation. Thus when GC is replaced by AT, the reversion may be due to substitution by CG instead of GC. This produces a reverse phenotype even though its sequence differs from the wild type in a single base pair.

4.3.2 Suppress or Mutations

The effects of a mutation may be diminished or suppressed by another mutation that occurs at a different site from the original mutation and this is referred as suppressor mutation. This suppressor mutation is also called second-site mutation. A suppressor mutation does not result in a reversal of the original mutation; instead, it makes or compensates for the effects of the initial mutation. There are two major classes of suppressor mutations: those which occur with in the same gene as the original mutations but at a different site called intragenic suppressors, and those occurring in a different gene called intergenic suppressors. Both intragenic and intergenic suppressors operate to allow the production of functional or partially functional copies of the protein which were initially rendered inactive by the original deleterious mutation. Thus, function can be restored only when both the original mutation and the suppressor mutation are present together in the same cell.

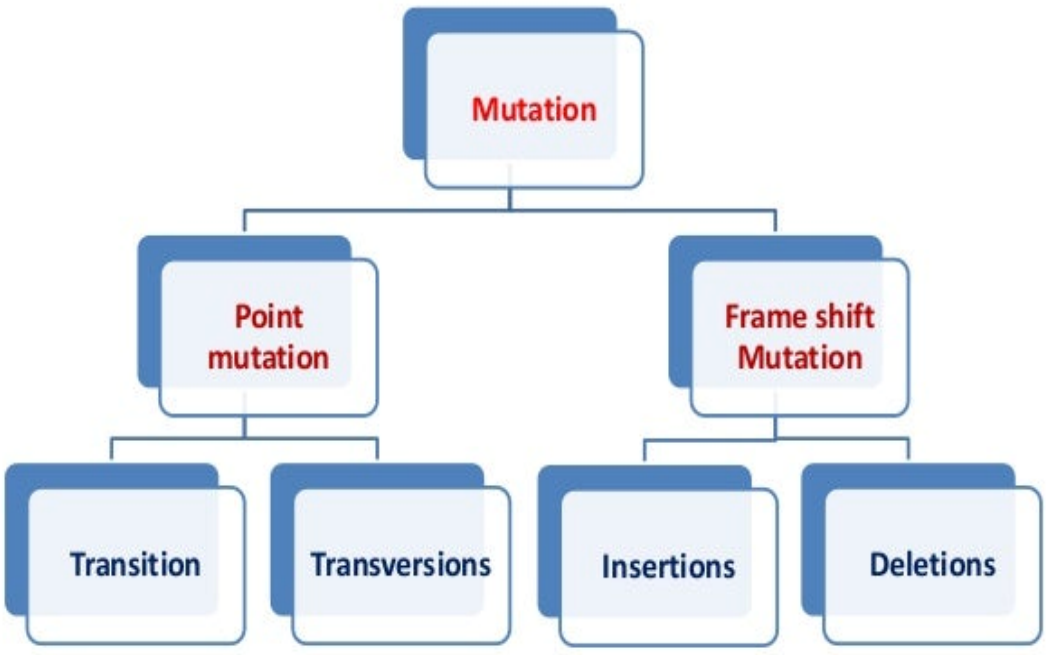
Intragenic suppressors act in one of two ways: by altering a different nucleotide in thesame codon in which the original mutation occurred, or by altering a nucleotide in a different codon. As an example, a DNA sequence of three base pairs in the wild type specifies the mRNA codon 5′-CGU-3′, which is read as arginine. The original or first mutation is a GC to AT transversion at the first base pair, resulting in the mRNAcodon5′-AGU-3′, which specifies serine. The suppressor or second mutation is a TA to AT transversion at the third position, giving the mRNA codon 5′-AGA-3′, which is an arginine codon. Thus, with both mutations, the protein will be completely functional in cells.Intergenic suppression is the suppression of a mutational defect by a second mutation in another gene. Genes that cause suppression of mutations in other genes are called suppressor genes. Many intergenic suppressors work by changing the way the mRNA encoded by the mutant gene is read. Each suppressor gene can suppress the effects of only one type of nonsense, missense, or frame shift mutation; hence, suppressor genes can suppress only a small proportion of the point mutations that theoretically can occur within a gene. On the other hand, a given suppressor gene will suppress all mutations for which it is specific, whatever gene the mutation is in.

Suppressors of nonsense mutations have been well characterized, particularly in *E. coli* and yeast. The suppressor genes in this case often are mutant tRNA genes. That is, particular tRNA genes can mutate so that their anticodons recognize a chain terminating codon and put an amino acid into the chain. Thus, instead of polypeptide chain synthesis being stopped prematurely as a result of a nonsense mutation, the altered or suppressor tRNA inserts an amino acid at that position, and full or partial function of the polypeptide may be restored. There are three classes of nonsense suppressors, one for each of the nonsense codons UAG, UAA, and UGA. For example, a gene for a tyrosine tRNA which has the anticodon 3′-AUG-5′is mutated so that the tRNA has the anticodon 3′-AUC-5′, the mutated suppressor tRNA which will still carry tyrosine will read the nonsense codon 5′-UAG-3′. So, instead of chain termination occurring, tyrosine is inserted at that point in the polypeptide. How functional the complete protein will depend on the effects of the inserted tyrosine in the protein. If it is an important part of the protein, then the incorrect amino acid may not restore

function to a significant degree. If it is in a less crucial area, the protein may have some or complete function.

4.3.3 Point mutations

The gene mutation in which there is only a single changed base pair of DNA is called point mutation. A point mutation may be a base substitution, a base insertion or a base deletion, but the term most frequently refers to a base substitution. Point mutation is a type of genetic mutation where one of the base pairs in the DNA sequence is altered either by insertion or deletion. Point mutation, as the name indicates, occurs at a particular point of the DNA sequence as a result of changes in one particular base pair. Usually, point mutations arise as a result of mistakes during DNA replication; however, in some cases, modification of DNA can be induced by X-rays or ultraviolet radiation. The effects of point mutations are often based on the specifics of the mutation as these can range from no effect to deleterious effects affecting protein production, composition, and function. Point mutation is a type of mutation that affects the structure of the chromosome, induced by changes in the nucleotide sequences. Point mutations can have important effects on the production of proteins as nucleotides are read in triplets, and the changes in the sequences might cause changes in the sequence of amino acids in the protein. Besides, the changes in the sequences at the beginning or the end of the sequence can also result in premature termination or initiation. Thus, the severity of mutation depends on the position of the mutation.



Categories of Mutations

Point mutations can be differentiated into two types viz., Transitions and Transversions depending on the nature of the substitution of the base pairs (Fig.4.1).

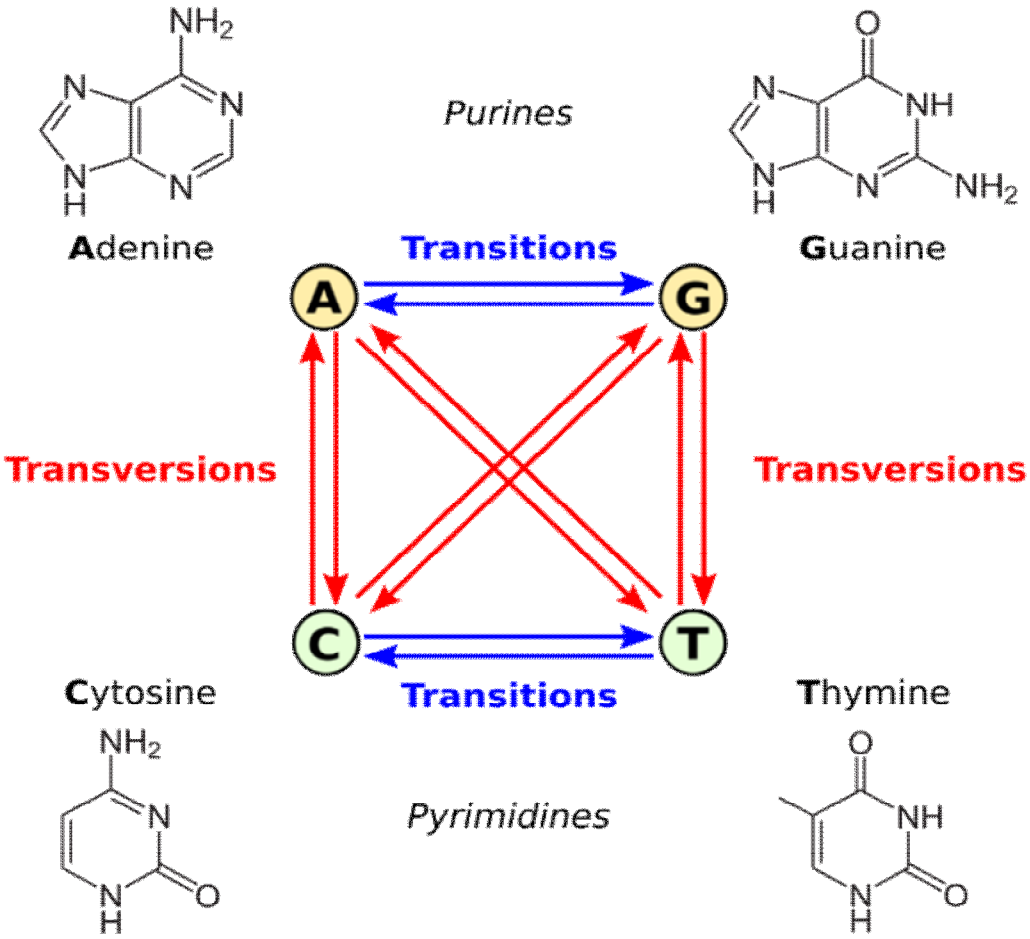


Fig-4.1: Types of mutations

4.3.4. Transition Mutations

Transition mutations are mutations that occur due to the substitution of a pyrimidine base by another pyrimidine base or the substitution of one purine base for the other. Since the base pairs in a double-stranded DNA bind to specific partners on the corresponding strands, the substitution of one nitrogen base for another results in a different set of base pairs. Transition mutations are caused by oxidative deamination and tautomerization of the base pairs. These are more common in genomes due to the molecular mechanism involved to generate them, even though there is just a single possible transition. However, transition mutations are less likely to result in amino acid substitutions and therefore exist as silent substitutions in genomic populations like single nucleotide polymorphisms.

4.3.5 Transversion Mutations

This is also a specific type of base pair substitution mutation which involves a change from a purine-pyrimidine base pair to a pyrimidine-purine base pair. The four types of transversion mutations are AT to TA, GC to CG, AT to CG, and GC to TA. Transversion mutations are the mutations where the purine base substitutes for a pyrimidine base and vice versa. Transversion mutations are usually spontaneous, but these can be induced by radiation or alkylating agents. A reversible transversion is possible in the case of spontaneous mutation.There are two possible transversions, but these are less likely to occur as it is difficult to substitute a double ring structure for a single ring structure. A transversion,

however, has a more pronounced effect as the third nucleotide codon of the DNA is less tolerant to transversion. A transversion seldom results in codons that encode for the amino acids, which brings about changes in the protein.

4.3.6 Frame shift Mutations

A frame shift mutation is a type of genetic mutation resulting from the alteration of a number of nucleotides in DNA sequences (Fig.4.2).Frame shift mutation occurs as a result of a change in the reading frame of the sequence. Since the codons for gene expression occur in the form of triplets, the difference in nucleotides results in the change in the frame. Frame shift mutations are caused by the addition or deletion of nucleotides from a DNA sequence, resulting in the shift of frame. Substitution cannot result in frame shift mutation as it requires the production of a completely different amino acid product. In some cases, the mutation might be spontaneous, resulting from some mistake during the processing of the DNA.The change in DNA sequence can occur either at a particular point or over a region of the sequence. The change causes a shift in the reading frame to produce a different product. Besides, frame shift mutation can also occur via induced mutation in the presence of physical agents like X-ray and UV radiation. Frame shift mutation might also be induced by the production of molecules like reactive oxygen species that influence DNA composition. Frame shift mutations can occur in both the single-stranded and double-stranded DNA as a result of different causes.

It occurs as a result of the addition or deletion of nucleotides in the sequence of DNA. The addition or deletion of nucleotide causes a shift of the reading frame of mRNA. In an mRNA, each codon is represented by three bases without punctuation, and insertion or deletion of nucleotide changes the entire frame. So frame shift mutation brings greater phenotypic change than point mutation. Insertion or deletion of one or two base pairs of nucleotide causes a shift in frame. However, insertion or deletion of three base pair adds or remove a whole codon, this result in the addition or removal of single amino acid from polypeptide chain. Frameshift mutations are apparent in severe genetic diseases such as Tay–Sachs disease; they increase susceptibility to certain cancers and classes of familial hypercholesterolemia; in 1997, a frame shift mutation was linked to resistance to infection by the HIV retrovirus.

4.3.7 Deletion:

- Deletion of the DNA segment occurs in anaphase movement or during reorganizing of the nuclei or digestion by nucleases.
- Deletion can either be terminal or intercalary, where the terminal deletion results from the removal of the terminal section of a chromosome.
- The intercalary deletion occurs as a result of the loss of an intermediate section or portion of the chromosome.
- Deletion is the more common mechanism of frame shift mutation that produces a change in the reading frame.

4.3.8 Insertion

- Insertion is the addition of nucleotides into a DNA sequence in a way that changes the reading frame of the sequence.

- It occurs within the microsatellite regions of the DNA in the presence of DNA polymerase. The number of nucleotides and the position of insertion can be different and also determines the severity of the mutation.
- Usually, insertion and subsequent frame shift mutation cause the active translation of a gene while producing a premature stop codon.

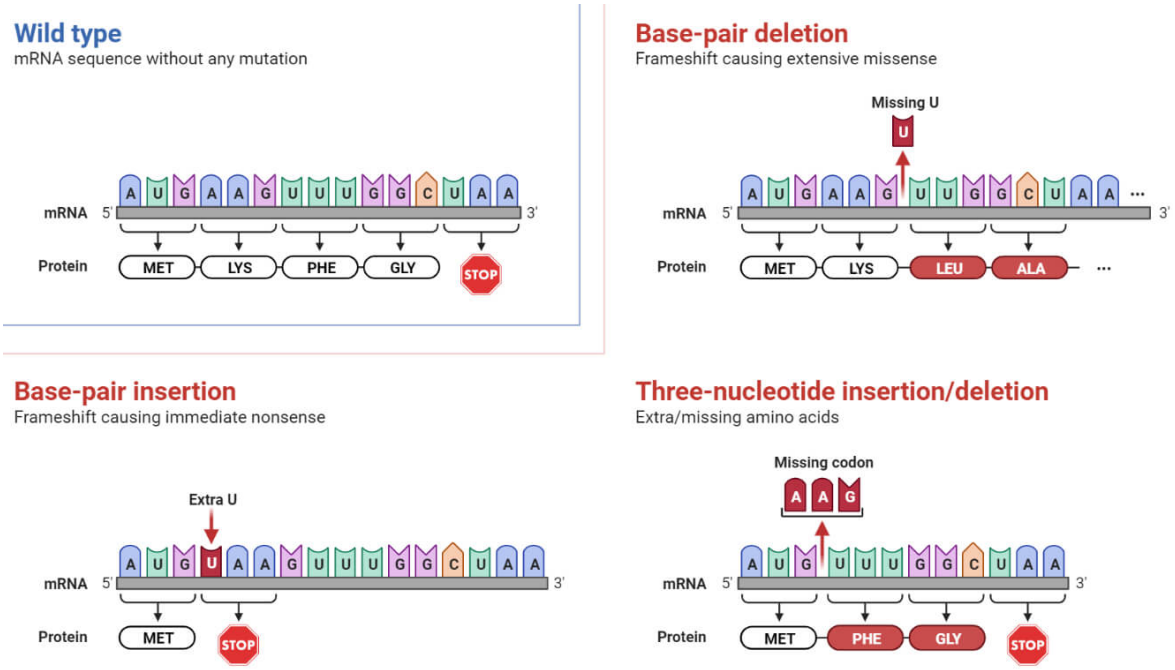


Fig-4.2: Frame shift mutations

4.4 RADIATION INDUCED MUTATIONS

Both X-rays and UV rays are used to induce mutations. X-rays are an example of ionizing radiation. Ionizing radiation can penetrate tissues, hence the use of X-rays as a diagnostic tool. Collision of ionizing radiation with atoms in its path gives rise to ions and reactive chemical radicals that can break chemical bonds, including those in DNA. That is, the products of ionizing radiation can induce chromosome breakages, chromosome rearrangements, and damage to DNA, e.g., point mutations. In fact, ionizing radiation is the leading cause of gross chromosomal mutations in humans. At certain low levels of ionizing radiation, point mutations are commonly produced and at these levels, there is a linear relationship between the rate of point mutations and radiation dosage.

UV rays are non-ionizing and they have insufficient energy to induce ionizations. However, UVlight is a useful mutagen, and at high enough doses it can kill cells. UV light is used as a sterilizing agent in some applications. Ultraviolet light causes mutations because the purine and pyrimidine bases in DNA absorb light very strongly in the UV range (254 to 260 nm). At this wavelength UV light induces gene mutations primarily by causing photochemical changes in the DNA. One of the effects of UV radiation on DNA is the formation of abnormal chemical bonds between adjacent pyrimidine molecules in the same strand. This bonding is induced mostly between adjacent thymines, forming the structures called thymine dimers, usually designated T ^T,C^C,C^T and T^C pairs are also produced by UV radiation, but in much lower amounts.This unusual pairing produces a bulge in the DNA strand and disrupts the normal pairing of the Tswith the corresponding As on the opposite strand. Many of the thymine dimers are repaired leaving no trace of the original damage.

4.5 CHEMICAL MUTAGENS

There is a long list of chemicals which are used as mutagens. Detailed treatment of such chemicals is beyond the scope of this discussion. The chemical mutagens can be divided into four groups, viz., (a) Alkylating agents, (b) Base analogues, (c) Acridine dyes, and (d) Others. A brief description of some commonly used chemicals of these groups is presented below.

4.5.1. Base Analogues

Base analogues refer to chemical compounds which are very similar to DNA bases. Such chemicals sometimes are incorporated in DNA in place of normal base during replication (Fig. 4 and 5). Thus, they can cause mutation by wrong base pairing. An incorrect base pairing results in transitions or transversions after DNA replication. The most commonly used base analogues are 5-bromouracil (5BU) and 2-aminopurine (2AP). 5-bromouracil is similar to thymine, but it has bromine at the C5 position, whereas thymine has CH3 group at C5 position. The presence of bromine in 5BU enhances its tautomeric shift from keto form to the enol form. The keto form is a usual and more stable form, while enol form is a rare and less stable or short lived form. Tautomeric change takes place in all the four DNA bases, but at a very low frequency. The change or shift of hydrogen atoms from one position to another either in a purine or in a pyrimidine base is known as tautomeric shift and such process is known as tautomerization. The base which is produced as a result of tautomerization is known as tautomeric form or tautomer. As a result of tautomerization, the amino group (-NH2) of cytosine and adenine is converted into imino group (-NH). Similarly keto group (C = 0) of thymine and guanine is changed to enol group (-OH). 5BU is similar to thymine, therefore, it pairs with adenine (in place of thymine). A tautomer of 5BU will pair with guanine rather than with adenine. Since the tautomeric form is short-lived, it will change to keto form at the time of DNA replication which will pair with adenine in place of guanine. In this way it results in AT GC and GC —> AT transitions.

Normally, 5-bromouracil pairs with adenine justas thymine does, but it occasionally mispairs with guanine, leading to a transition (T · A→5BU· A→5BU · G→C · G). Through mispairing, 5-bromouracil can also be incorporated into a newly synthesized DNA strand opposite guanine. In the next round of replication 5-bromouracil pairs with adenine, leading to another transition (G · C→G · 5BU→A · 5BU→A · T) (Figs.4.3 and 4.4)

Another mutagenic chemical is 2-aminopurine (2AP), which is a base analog of adenine. Normally, 2-aminopurine base pairs with thymine, but it may mispair with cytosine, causing a transition mutation (T · A→T · 2AP→C · 2AP→C · G). Alternatively, 2-aminopurine may be incorporated through mispairing into the newly synthesized DNA opposite cytosine and then later pair with thymine, leading to a C · G→C · 2AP→T · 2AP→T · A transition.

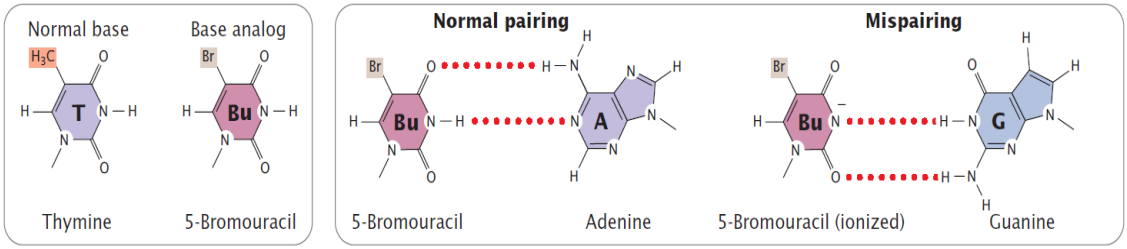


Fig-4.3: Function of 5 bromo uracil (5BU)

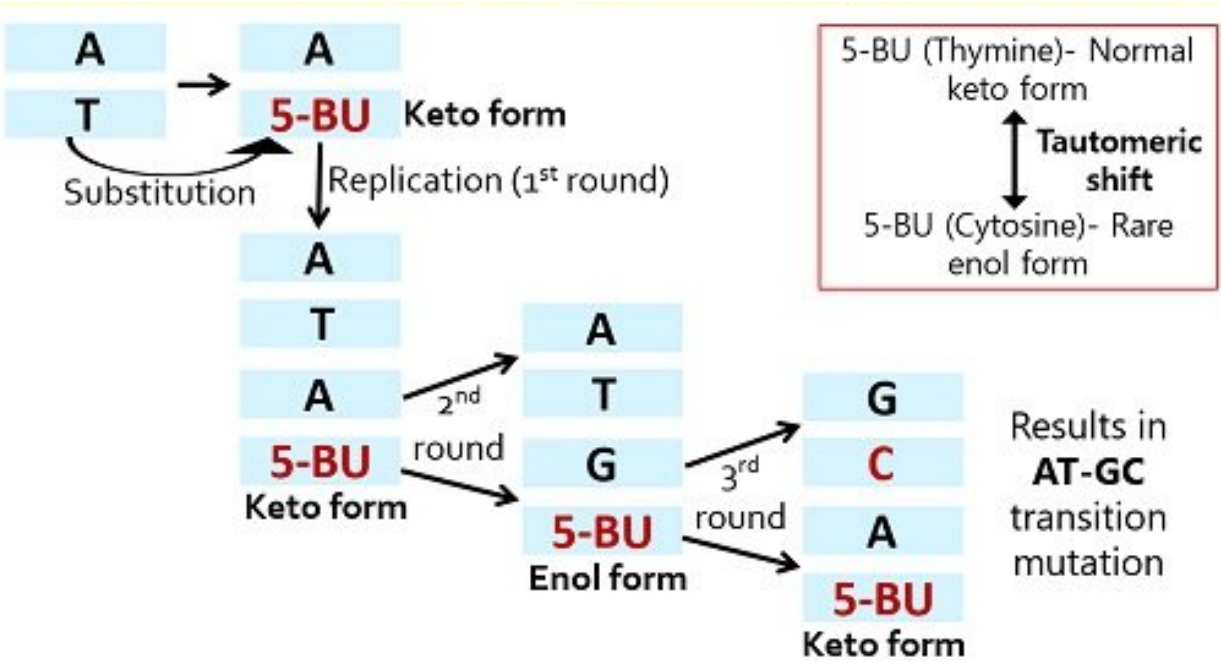


Fig-4.4: Mechanism of base analog by 5-bromouracil

4.5.2. Deaminating Agents (Nitrous acid)

Other important chemical mutagens are nitrous acid and hydroxyamine (Fig.4.5). Their role in induction of mutation is briefly described here. Nitrous acid is a powerful mutagen which reacts with C6 amino groups of cytosine and adenine. It replaces the amino group with oxygen (+ to – H bond). As a result, cytosine acts like thymine and adenine like guanine. Thus, transversions from GC —> AT and AT —> GC are induced. Hydroxylamine is a very useful mutagen because it appears to be very specific and produces only one kind of change, namely, the GC —> AT transition. All the chemical mutagens except base analogues are known as DNA modifiers.

Nitrous acid (HNO_2) is a potent mutagen which acts both on replicating and non-replicating DNA. Nitrous acid is a deaminating agent that removes amino groups from the bases. It leads to oxidative deamination of cytosine, adenine and guanine in which the amino groups are replaced by keto (=O) groups yielding uracil, hypoxanthine and xanthine, respectively. Treatment of guanine with nitrous acid produces xanthine, but since this purine base has the same pairing properties as guanine, no mutation results. However, when cytosine is treated with nitrous acid, uracil is produced which pairs with adenine. The deamination of cytosine by nitrous acid, then, produces a CG-to-TA transition mutation during replication. Likewise, nitrous acid modifies adenine to produce hypoxanthine, a base that pairs with cytosine rather than thymine, thus resulting in an AT-to-GC transition mutation. Thus nitrous acid produces both AT to GC as well as GC to AT transitions and these transitions are produced only when the DNA treated with HNO_2 undergoes replication. A nitrousacid-induced mutation can be reverted by a second treatment with nitrous acid.

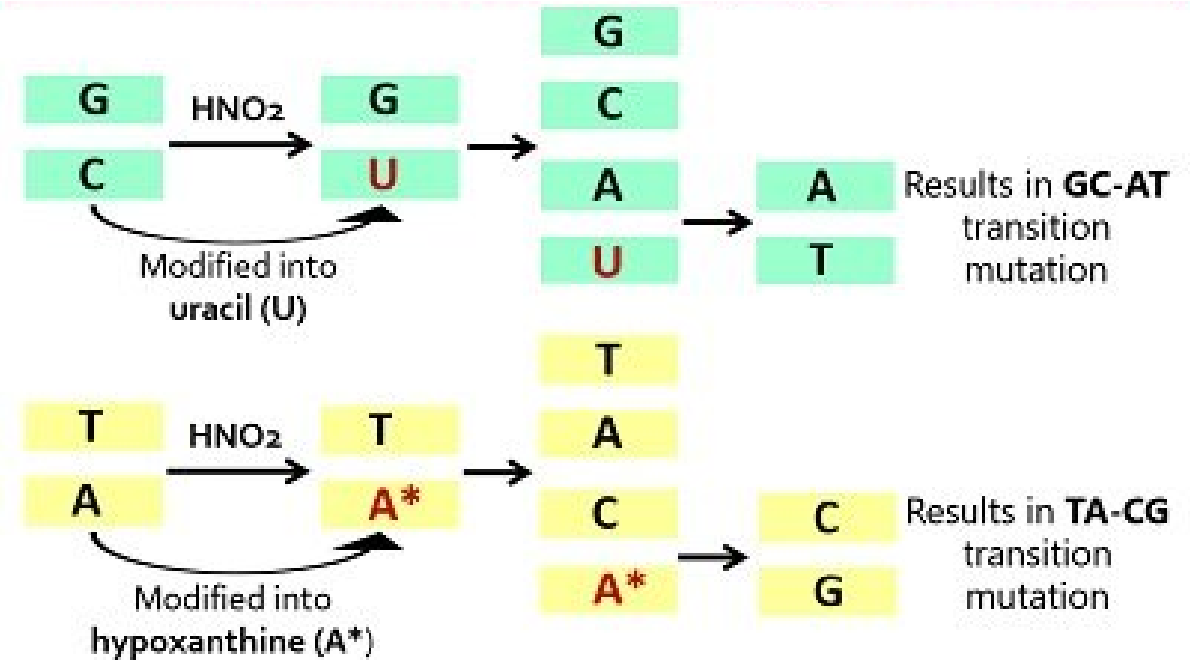


Fig-4.5: Mechanism of Deaminating agent (HNO2)

4.5.3 Acridine Dyes

Acridine dyes are very effective mutagens. Acridine dyes include, pro-flavin, acridine orange, acridine yellow, acriflavin and ethidium bromide (Fig.4.6). Out of these, pro-flavin and acriflavin are in common use for induction of mutation. Acridine dyes get inserted between two base pairs of DNA and lead to addition or deletion of single or few base pairs when DNA replicates. Thus, they cause frame shift mutations and for this reason acridine dyes are also known as frame shift mutagens. Proflavin is generally used for induction of mutation in bacteriophages and acriflavin in bacteria and higher organisms.

The intercalating agents produce mutations by sandwiching themselves (intercalating) between adjacent bases in DNA, distorting the three-dimensional structure of the helix and causing single-nucleotide insertions and deletions in replication. These insertions and deletions frequently produce frameshift mutations, and so the mutagenic effects of intercalating agents are often severe. Because intercalating agents generate both additions and deletions, they can reverse the effects of their own mutations.

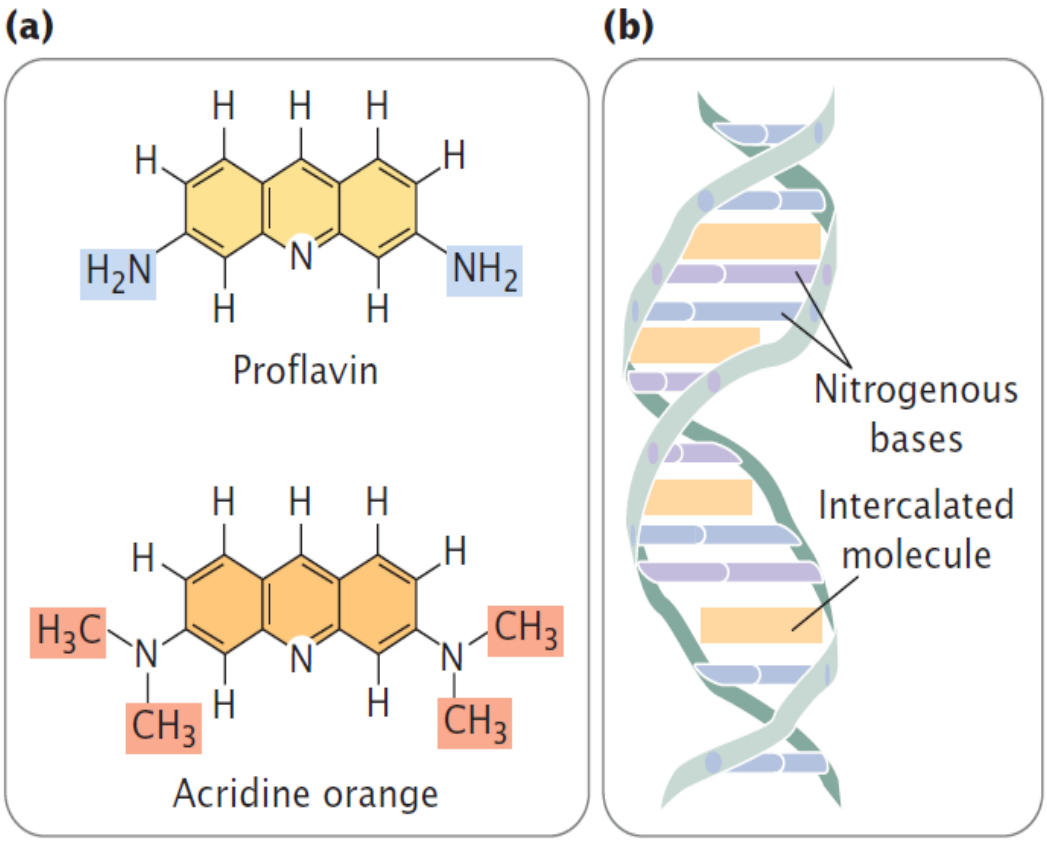


Fig-4.6: Intercalating agents: a) Proflavin and Acridine orange. b) Insert themselves between adjacent bases in DNA

4.5.4 Alkylating Agents

This is the most powerful group of mutagens. They induce mutations especially transitions and transversions by adding an alkyl group (either ethyl or methyl) at various positions in DNA. Alkylation produces mutation by changing hydrogen bonding in various ways. The alkylating agents include ethyl methane sulphonate (EMS), methyl methane sulphonate (MMS), ethylene imines (EI), sulphur mustard, nitrogen mustard, etc. (Fig.4.7). Out of these, the first three are in common use. Since the effect of alkylating agents resembles those of ionizing radiations, they are also known as radiomimetic chemicals. Alkylating agents can cause various large and small deformations of base structure resulting in base pair transitions and transversions. Transversions can occur either because a purine has been so reduced in size that it can accept another purine for its complement, or because a pyrimidine has been so increased in size that it can accept another pyrimidine for its complement. In both cases, diameter of the mutant base pair is close to that of a normal base pair.

For example, ethylmethylsulfonate (EMS) adds an ethyl group to guanine, producing *O*6-ethylguanine, which pairs with thymine. Thus, EMS produces C · G→T · A transitions. EMS is also capable of adding an ethyl group to thymine, producing 4-ethylthymine, which then pairs with guanine, leading to a T · A→C · G transition. Because EMS produces both C · G→T · A and T · A→C · G transitions, mutations produced by EMS can be reversed by additional treatment with EMS.

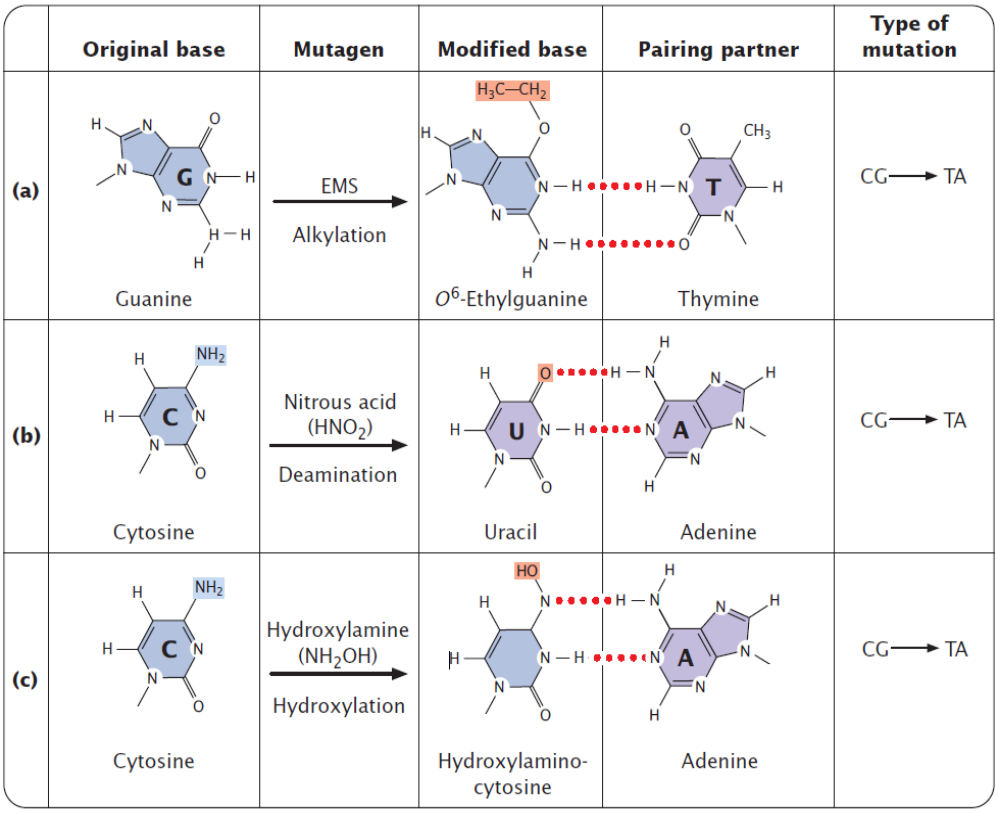


Fig-4.7: Different Types of Alkylating Agents

4.5.5. Hydroxylating Agents

Hydroxylamine is a very specific base modifying mutagen that adds a hydroxyl group to cytosine, converting it into hydroxyl amino cytosine. This conversion increases the frequency of a rare tautomer that pairs with adenine instead of guanine and leads to C · G→T· A transitions. Because hydroxylamine acts only on cytosine, it will not generate T · A→C · G transitions; thus, hydroxylamine will not reverse the mutations that it produces.

Reactive forms of oxygen (including superoxide radicals, hydrogen peroxide, and hydroxyl radicals) are produced in the course of normal aerobic metabolism, as well as by radiation, ozone, peroxides, and certain drugs. These reactive forms of oxygen damage DNA and induce mutations by bringing about chemical changes in DNA. For example, oxidation converts guanine into 8-oxy-7,8-dihydrodeoxyguanine, which frequently mispairs with adenine instead of cytosine, causing a G · C→T · A transversion mutation.

4.6 SUMMARY

Mutation is the process by which a change in DNA base pairs or a change in the chromosomes is produced. Mutations at the level of chromosomes are called chromosomal mutations. Mutations in the sequences of genes at the level of the base pair are called gene mutations. Gene mutations may occur by a substitution of one base pair for another or by the addition or deletion of one or more base pairs. The consequences of a gene mutation to an organism depend upon a number of factors, especially the extent to which the amino acid coding information is changed. A missense mutation causes the substitution of one amino acid for another, and nonsense mutation cause premature termination of synthesis of the polypeptide.

The effects of a gene mutation can be reversed either by reversion of the gene sequence to its original state, or by a mutation at a site distinct from that of the original mutation. The latter is called a suppressor mutation. Suppressor mutations that occur within the same gene as the original mutation are intragenic suppressors. They act either by altering a different nucleotide in the same codon affected by the original mutation, or by altering a nucleotide in a different codon. Suppressor mutations that occur in a different gene from the original mutation are called intergenic suppressors. Often, suppression by this class of suppressors involves a tRNA with an altered anticodon.

Mutations occur spontaneously at a low rate. The mutation rate can be increased through the use of mutagens like irradiation and certain chemicals. Typically, mutagens are used by researchers so that a mutant of interest is more likely to be found in a population of cells. Chemical mutagens work in a number of different ways, such as by acting as base analogues, by modifying bases, or byintercalating into the DNA. The later one results in frameshift mutations, while the others result in base pair substitution mutations.

4.7 TECHNICAL TERMS:

Spontaneous mutations, Induced mutations, Transitions, Transversions, 5-bromouracil, Acridine dyes, Insertion, Deletion, Pro-flavin, Acridine orange, Acridine yellow, Acriflavin, Ethidium bromide, Ethyl methane sulphonate (EMS), Methyl methane sulphonate (MMS), Ethylene imines (EI).

4.8 SELF ASSESSMENT QUESTIONS

1) Define mutation? Explain the different types of mutations.
2) What is a mutagen? Explain different types of chemical mutagens and their mode of actions.
3) Discuss spontaneous and induced type of mutations.
4) Explain Radiation induced mutations.
5) Add a note on Point and frame shift mutations.

4.9 SUGGESTED READINGS:

1) Freifelder, D. Molecular Biology (1990)-Narosa Publishing House, New Delhi.
2) Gardner, E.J., Simmons, M.J. and Snustad, D.P. Principles of Genetics(2001)-John Wiley & Sons, Inc., New York.
3) Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics(1997)-John Wiley & Sons, Inc., New York.
4) Russel, P.J. Genetics (1998)-TheBenjamin / Cummings Publishing Company, Inc., California.
5) Singh, B.D. Fundamentals of Genetics(2001)-Kalyani Publishers, New Delhi.

Dr. K. Nagaraju

LESSON-5

GENETIC RECOMBINATION IN BACTERIA

5.0 OBJECTIVE:

- This lesson enlightens the students with knowledge on mechanisms of recombination in bacterial cells.

STRUCTURE:

5.1 INTRODUCTION

In order to study the genetics of an organism it is necessary that some kind of sexuality occurs in this organism. For long time, bacteria and viruses were believed to reproduce through purely asexual methods. J.Lederberg and E.L.Tatum demonstrated sexuality in bacteria for the first time and this open new area of research. Genetic recombination in bacteria is the process by which genetic material is exchanged between different bacterial cells or between a bacterium and foreign DNA. This process contributes to **genetic diversity**, allowing bacteria to adapt to changing environments, develop antibiotic resistance, and evolve new traits. Unlike eukaryotes, bacteria reproduce **asexually** by binary fission, producing identical daughter cells. However, they can exchange genetic information through several **horizontal gene transfer (HGT)** mechanisms-**Transformation, Conjugation and Transduction** which result in recombination.

5.2 TRANSFORMATION:

Bacterial transformation is defined as the transfer of free DNA released from a donor bacterium into the extracellular environment that result in assimilation and usually an expression of the newly acquired trait in a recipient bacterium. Bacterial transformation is

based on the natural ability of bacteria to release DNA which is then taken up by another competent bacterium. The success of transformation depends on the competence of the host cell. Competence is the ability of a cell to incorporate naked DNA in the process of transformation. There are two types of transformations –

1. Natural Transformation

- Occurs in some species naturally (e.g., *Streptococcus pneumoniae, Bacillus subtilis, Neisseria gonorrhoeae*).
- These bacteria have specific proteins that make their cell membranes **competent** (able to take up DNA).

2. Artificial Transformation

- Done in the lab to introduce recombinant DNA into bacteria (e.g., *E. coli* in genetic engineering).
- Competence is **induced** using:
  - **Chemical methods** (e.g., $CaCl_2$ treatment followed by heat shock)
  - **Physical methods** (e.g., electroporation-using an electric field to create pores in the membrane).

Steps in Natural Transformation:

Development of Competence: The bacterial cell becomes capable of taking up DNA.

Binding of DNA: Free DNA from the environment binds to the cell surface.

DNA Uptake: One strand of the DNA is degraded, and the other is transported into the cytoplasm.

1) **Integration:**
   The DNA may:
   - Recombine with the bacterial chromosome (if homologous)
   - Exist as a plasmid (if it has an origin of replication)
2) **Expression:**
   New genes are expressed, potentially giving the bacterium new traits (e.g., antibiotic resistance, virulence factors).

Griffith Experiment:

British bacteriologist Frederick Griffith reported the first demonstration of bacterial **transformation**-a process in which external DNA is taken up by a cell, thereby changing its morphology and physiology. Griffith conducted his experiments with *Streptococcus pneumoniae*, a bacterium that causes pneumonia. Griffith worked with two strains of this bacterium called rough (R) and smooth (S). The two cell types were called “rough” (R) and “smooth” (S) after the appearance of their colonies grown on a nutrient agar plate.

The R strain is non-pathogenic (does not cause disease). The S strain is pathogenic (disease-causing), and has a capsule outside its cell wall. The capsule allows the cell to escape the immune responses of the host mouse.

When Griffith injected the living S strain into mice, they died from pneumonia. In contrast, when Griffith injected the live R strain into mice, they survived. In another experiment, when he injected mice with the heat-killed S strain, they also survived. This experiment showed that the capsule alone was not the cause of death. In a third set of experiments, a mixture of live R strain and heat-killed S strain were injected into mice, and-to his surprise-the mice died. Upon isolating the live bacteria from the dead mouse, only the S strain of bacteria was recovered. When this isolated S strain was injected into fresh mice, the mice died. Griffith concluded that something had passed from the heat-killed S strain into the live R strain and transformed it into the pathogenic S strain. He called this the *transforming principle*. These experiments are now known as Griffith's transformation experiments (Fig.5.1).

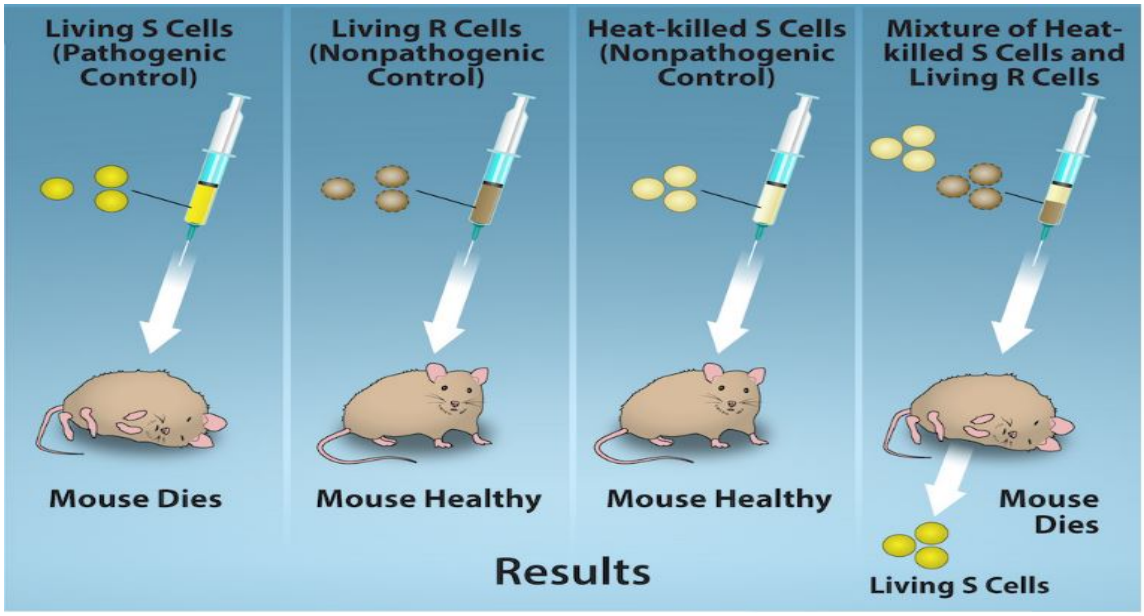


Fig-5.1: Griffith's Transformation Experiment

What occurred in Griffith's experiments in 1928 could not be fully explained at that time. Scientists then were unaware of the role of DNA blueprints for a living thing. Further, with such limited knowledge about DNA and what it does, scientists then were also unaware that bacteria are capable of "picking up" DNA from other bacterial cells or picking up DNA from their environment.

In Griffith's experiment, DNA from the dead S strain were transferred into the live R strain. This DNA provided the R strain with instructions for building the S strain's capsule. As a result, the live R strain cells began to produce the S strain's capsule and thereby developed the ability to evade the host's immune system. The result is the R strain acted in the same way as the S strain, became pathogenic, and caused death in the mice. In essence, what happened in these circumstances is **bacterial transformation**. Bacterial transformation is a process where bacteria absorb DNA from their environment resulting in new characteristics in the bacteria. The new characteristics that transformed bacteria (also called **transgenic** bacteria) exhibit is dependent on what types of genes are present in the DNA that they took up from their environment.

Applications:

- **Genetic engineering** (inserting recombinant plasmids into *E. coli*)

- **Production of proteins** (e.g., insulin, enzymes)
- **Gene mapping and cloning**
- **Studying gene function and regulation**

5.3 CONJUGATION

In biology, **conjugation** is a form of horizontal gene transfer (HGT), along with transformation and transduction, which allows the exchange of genetic material between two organisms or cells. It is observed as a sexual process in eukaryotes, for example, in certain protozoan ciliates. However, the process is far more common and has been most extensively studied in bacteria. Conjugation was discovered in the Gram negative bacterium *Escherichia coli* in 1946 by Joshua Lederberg and Edward Tatum, for which Lederberg was awarded Nobel Prize in physiology and medicine. It is one of the critical mechanisms in the evolution of bacteria and facilitates the spread of traits such as antibiotic resistance, virulence factors and metabolite utilization pathways across populations. Broad-host-range conjugative plasmids have been used in molecular biology to introduce recombinant genes into bacterial species that are refractory to routine transformation or transduction methods

Conjugation involves the transfer of genetic material between two bacterial cells - Donor F^+ cell containing fertility factor **(F plasmid**), and **Recipient cell (F^- cell)** without the F plasmid. But it is mediated **by the F (fertility) plasmid;** the F plasmid carries genes that code for the **sexpilus** and proteins needed for DNA transfer. It **requires physical contact** between cells - unlike the transformation or transduction (Fig.5.2).

Steps in Conjugation

- The F plasmid contains *tra* locus, which includes the *pilin*. This gene, along with some regulatory proteins results in the formation of pilli on the F^+ cell surface.
- The proteins present in the pilli attach themselves on the F^- cell surface. The pilli are responsible for making contact between the cells, but the transfer of plasmid doesn't occur through the pilli.
- The *traD* enzyme, located at the base of the pilus, initiates membrane fusion.
- Once the conjugation is initiated, enzyme relaxase creates a nick in the conjugative plasmid at the *oriT*.
- The nicked strand (called the T strand) then unwinds and is transferred to the recipient cell in the 5'-3' direction.
- The complementary strand is synthesized in both cells; thus, both the donor and recipient are F^+.
- In certain F^+ bacterial cells, the F element infrequently (about once in every 10,000 F^+ cells) becomes associated with the main bacterial chromosome in such a way that a copy of the chromosome instead is transferred through the conjugation tube from donor to recipient cell.
- In the insertion process, the circular F element breaks at a particular point and becomes a linear segment of the bacterial chromosome.
- An F^+ cell that carries such an integrated F element is known as an **Hfr cell** (Hfr stands for the *high frequency of recombination*).
- The integrated F element of Hfr cells is ordinarily replicated passively along with the bacterial chromosome and in this way it is transmitted from one Hfr generation to the next.

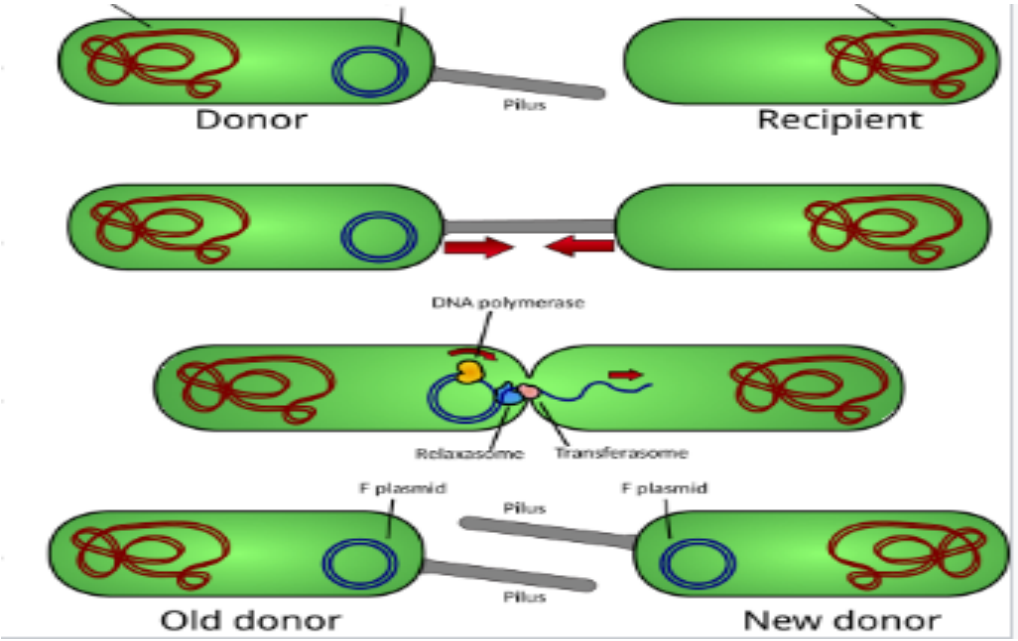


Fig-5.2: Conjugation in Bacteria

Types of Conjugation Systems

1) **$F^+ \times F^-$ Conjugation:**
   - Transfer of the F plasmid from donor (F^+) to recipient (F^-).
   - Result: Both cells become F^+.
2) **Hfr × F^- Conjugation (High-Frequency Recombination):**
   - The F plasmid integrates into the donor's chromosome.
   - During conjugation, **chromosomal genes** can be transferred to the recipient.
   - Usually, the **entire chromosome is not transferred**, so the recipient remains **F^-** but may acquire new genes.
3) **F′ (F prime) Conjugation:**
   - The F plasmid excises from the chromosome, taking along some **host chromosomal DNA**.
   - The F′ plasmid is then transferred to the recipient, resulting in **partial diploidy** for some genes.

Examples of Bacterial Conjugation

- *Agrobacterium tumefaciens* causes crown gall tumor in plants by transferring the T DNA element, a part of the *Ti* (tumor-inducing) plasmid present in this bacterium, into a plant cell where the T element becomes incorporated into the plant cell's genome.
- Conjugative plasmids encoding antimicrobial resistance genes are called R plasmids which are transferred through *Shigella* spp. that might result in a widespread outbreak of antibiotic-resistant *Shigella*-mediated dysentery

Significance of Conjugation

- **Genetic diversity:** Introduces new genetic material into bacterial populations.
- **Antibiotic resistance:** Resistance genes (on **R plasmids**) can spread rapidly between bacteria.
- **Biotechnology:** Used to transfer genes between bacterial strains for research or industrial purposes.

5.4 TRANSDUCTION:

The transfer of bacterial DNA from a donor cell to a recipient cell with the assistance of a bacteriophage is referred to as bacteriophage-mediated genetic material transfer. This process, is known as transduction. the process described by Zinder and Lederberg in 1952 pertaining to certain bacteria, specifically those responsible for causing mouse typhoid, such as Salmonella typhimurium. Also has been observed in various bacterial species, including E. coli, Proteus, Shigella, and Staphylococcus. In this process, bacteriophages, which infect bacteria, use host cells to multiply and while assembling they sometimes pack the bacterial DNA with them. Later, when these viruses infect new bacterial cells, the bacterial genome that they carry may get inserted into the host genome. Transduction is commonly used in genetic engineering for inserting foreign DNA into the host cell.

Steps of Transduction (Fig.5.3)

1) Infection of the bacterial cell by bacteriophage.
2) The virus uses the host machinery to make multiple copies either directly by the lytic cycle or first gets incorporated into the bacterial genome by the lysogenic cycle and followed by the lytic stage.
3) During the assembly of bacteriophages, the bacterial genome also gets packed by mistake in the viral head alongside the viral genome. In the lysogenic cycle, during excision of prophage, some parts of the bacterial genome that flank the prophage are also excised and go inside the assembled viral head together with the viral genome.
4) When these viruses infect another bacterial cell, they inject the viral DNA as well as donor DNA into the host cell.
5) The bacterial DNA either forms plasmids or gets inserted into the recipient DNA if it is homologous to the recipient genome. Most of the time it remains as an extrachromosomal DNA. It can also get inserted with the prophage if it is a temperate phage. So, the fate depends on the portion of bacterial DNA and also on the nature of bacteriophages.

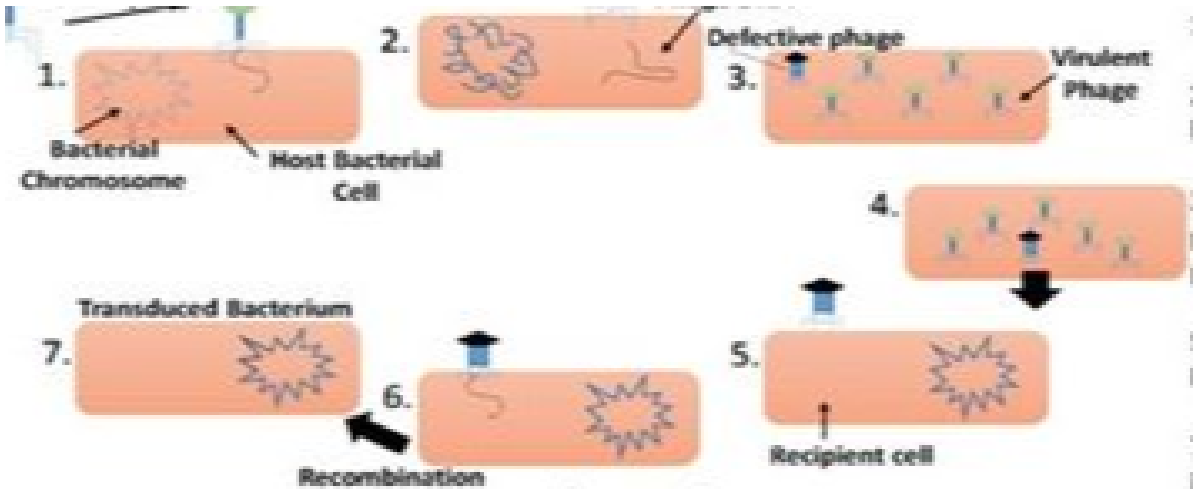


Fig-5.3: Transduction

Types of Transduction:

Transduction is common in both virulent and temperate phages, i.e. by lytic or lysogenic cycle. Transduction is of two types:

- Generalized Transduction – In this, the phage can carry any part of DNA.
- Specialized Transduction – In this, the phage carries only the specific part of DNA.

5.4.1 Generalized Transduction

Bacteriophage infection initiates a process facilitated by specific DNA segments called prophage particles in the bacterial cell's cytoplasm. In the course of a lysogenic bacterial cell's infection, bacterial DNA fragments while the bacteriophage's nucleic acid utilizes bacterial enzymes for new phage component synthesis. Simultaneously, these developing phage particles integrate the bacterial DNA fragments into their genetic material. Later, these progeny phage particles transfer the bacterial cell's genetic material to newly infected bacterial cells, illustrating the process of generalized transduction where bacteriophages actively aid in the transfer of bacterial DNA fragments.

Example: The best known generalized transducing phages are P22 in S. typhimurium and P1 in E.coli.

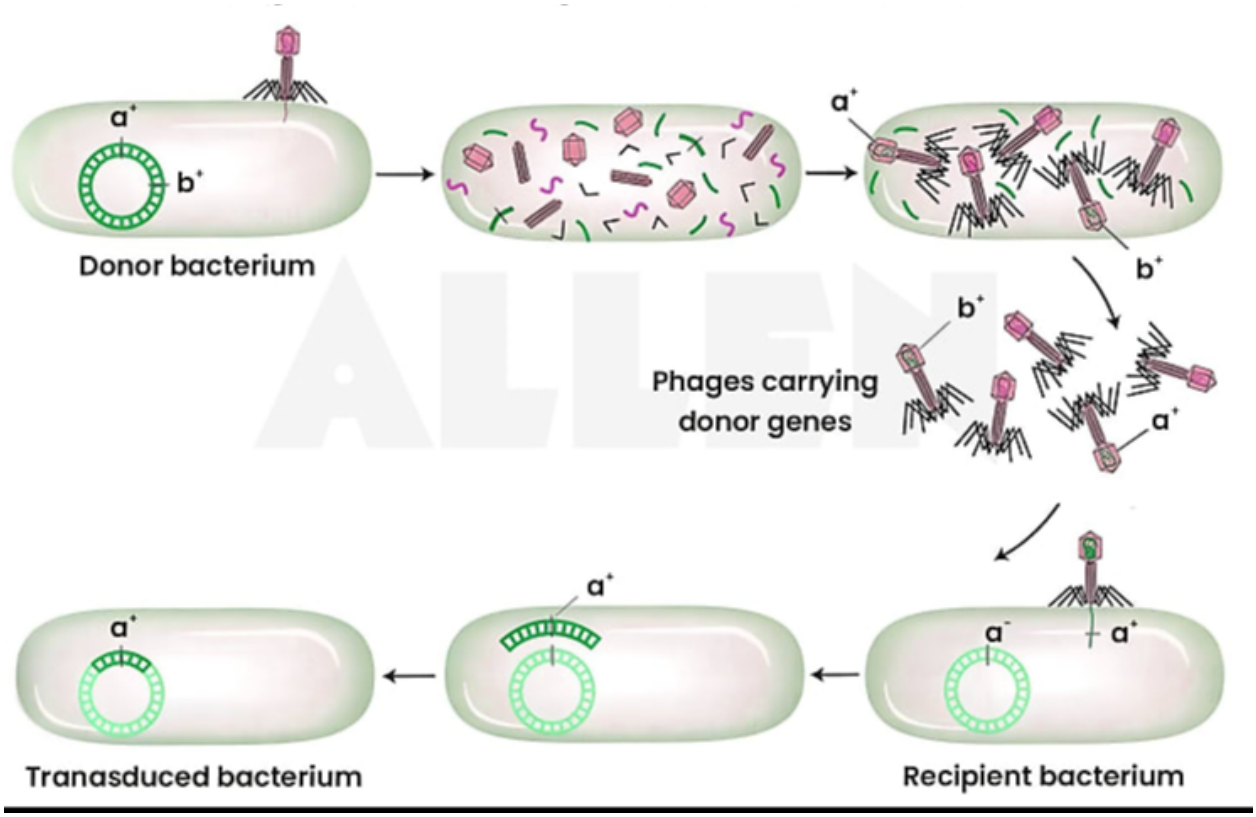


Fig-5.4: Generalized transduction

5.4.2 Specialized transduction:

As discovered by Andre Lwoff et al. in 1953, is a phenomenon where certain bacterial strains can endure extended periods of bacteriophage infection without bacterial cell lysis. In these lysogenic bacteria, the bacterial DNA joins with the phage DNA and replicates together. This state is maintained by a repressor protein that inhibits phage particle synthesis. When this protein production ceases, the bacterial cell begins to produce phage components. Before phage particle synthesis initiates, both phage and bacterial DNA degrade, and some bacterial genes are incorporated into the phage DNA, replicating alongside it. The progeny phage particles differ from their parent phage. When they infect new bacterial cells, they transmit closely linked bacterial genes. This specialized transduction transfers specific genes closely associated with the phage DNA. Example: Bacteriophage lambda (**λ**) is the best-known specialized transducing phage; ***λ*** carries only the gal (required for the utilization of galactose as an energy source) and bio (essential for the synthesis of biotin) genes from one E. coli cell to another.

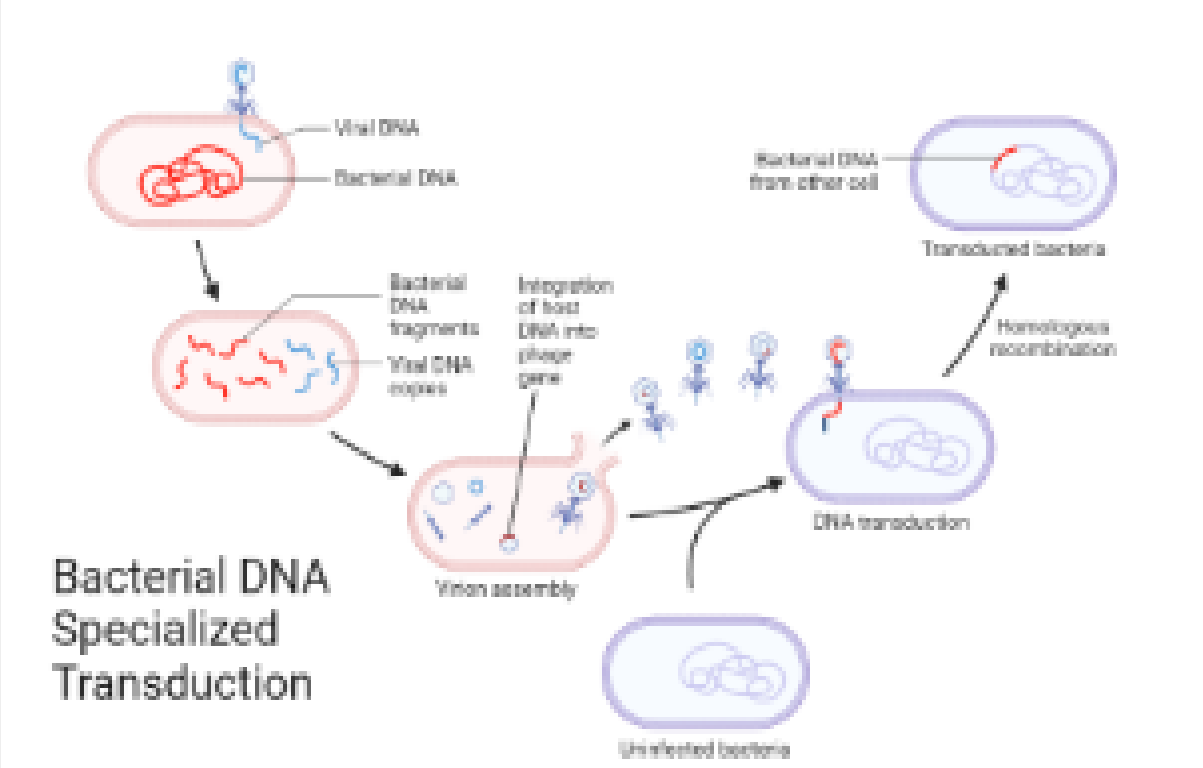


Fig-5.5: Specialized Transduction

Transduction in bacterial genetics is crucial for various reasons Gene Exchange: It helps bacteria diversify their genetic makeup, potentially gaining new abilities like antibiotic resistance by swapping genes through bacteriophages. Gene Therapy and Engineering: Bacteriophage-derived vectors are used to deliver therapeutic genes, aiding in treating genetic disorders and advancing genetic engineering. Research Tool: It is vital for understanding how genes function, are expressed, and transferred within bacteria. Evolutionary Influence: Transduction shapes bacterial evolution by introducing new genetic material, impacting their ability to adapt and their ecological roles. Genetic engineering Use: It's instrumental in creating GMOs and producing specific proteins, such as insulin or vaccines, with applications across fields like medicine and agriculture.

5.5 SUMMARY:

Gene transfer in bacteria is the non-reproductive exchange of genetic material, primarily occurring through three mechanisms: **conjugation**, **transformation** and **transduction**. Conjugation involves direct cell-to-cell contact, transformation is the uptake of free DNA from the environment, and transduction is the transfer of DNA via a bacteriophage virus. These processes, which are forms of horizontal gene transfer (HGT), allow bacteria to rapidly acquire new traits like antibiotic resistance, facilitating adaptation.

5.6 TECHNICAL TERMS:

Griffith, Transformation, Conjugation, Transduction, F-plasmid, Donor, Recipient, Hfr strain.

5.7 SELF ASSESSMENT QUESTIONS:

1) Give a detailed account on transformation.
2) Give a detailed account on conjugation.
3) Explain the process of transduction.

5.8 SUGGESTED READINGS:

1) *The Genetics of Bacteria and Their Viruses* by W. Hayes
2) Richard J. Epistein: Human Molecular Biology.
3) Microbial Genetics" by David Freifelder.
4) "Principles of Gene Manipulation and Technology" by Old & Primrose**.**
5) "Modern Microbial Genetics" edited by Streips and Yasbin**.**
6) "Microbial Genetics" by Sylwia Okoń, Beata Zimowska, and Mahendra Rai.

Dr. J. Madhavi

LESSON-6

GENE TRANSFER TECHNIQUES

6.0 OBJECTIVE:

- To make the students familiar with the different gene transfer techniques that are useful to achieve the recombination in bacteria.

STRUCTURE:

6.1 INTRODUCTION

Gene transfer is a natural mechanism that occurs in different organisms and scientists also use it to make genetically modified organisms, by introduction of exogenous DNA into targeted cells like viruses, bacteria, plant and animal cells and their stable expression gives rise to transgenic plants and animals. The transferred gene is transgene and the whole process is referred to as transgenesis. The objective of artificial gene transfer is to change the phenotype of an organism by changing its genotype or gene expression. For achieving genetic transformation in plants, the basic pre-requisite is the construction of a vector which carries the gene of interest flanked by the necessary controlling sequences, i.e., the promoter and terminator, and deliver the genes into the host plant. There are two kinds of gene transfer methods in plants. It includes –

Direct methods are those methods which don't use bacteria as mediators for integration of DNA into host genome. These methods include electroporation, microinjection, biolistic and chemical methods.

6.2 GENE TRANSFER METHODS

6.2.1 Electroporation

Electroporation, also called electro-permeabilization (EP), is a physical method of introducing genetic material, proteins, drugs, or other desired molecules into cells for various

purposes with the application of electric pulses to create temporary pores in the cell membrane, delivering the molecules to the cells.Neumann et al. first introduced the concept of electroporation in 1982. Over the past few decades, electroporation has made significant advancements. Initially, it was limited to simple in vitro experiments. It can now be used in gene transfer, gene therapy, drug delivery, and various in vivo therapeutic approaches.

Principle of electroporation involves applying electric pulses to cells, which creates temporary pores in the cell membrane. These pores allow the entry of molecules, such as DNA, proteins, drugs, and other substances into the cells. The pulses change the electrical potential across the membrane, leading to the formation of temporary pores in the phospholipid bilayer that forms the cell membrane. These pores allow molecules to pass through the membrane. Once the molecules enter the cells through the temporary pores, they can interact with the cellular mechanism, influencing cellular processes and functions (Fig.6.1).

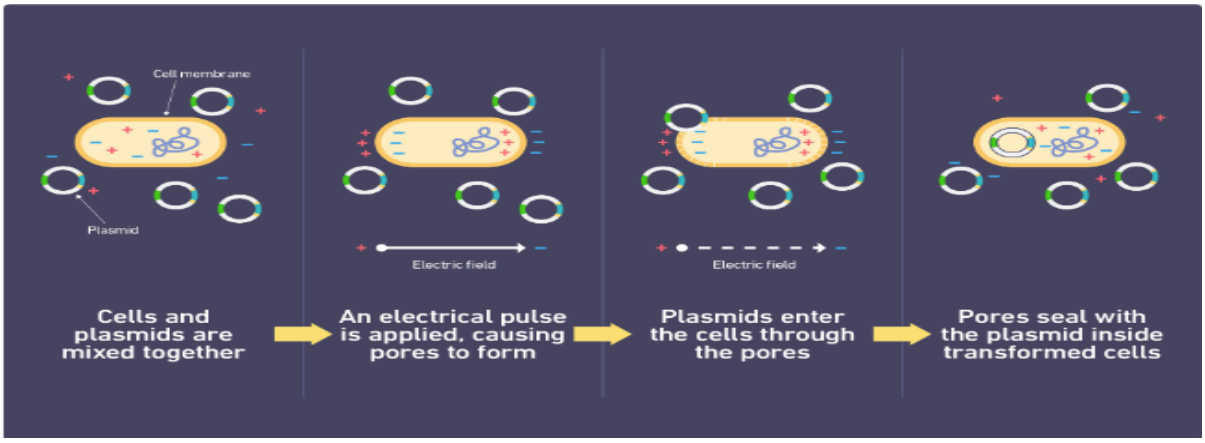


Fig-6.1: Electroporation of Cells

Process

The specific steps and materials involved in electroporation may vary based on the cell type, and electroporation equipment. The steps below provide a general overview of the electroporation process.

1) Preparation of Cells for Electroporation

- First, the cells of interest are cultured and grown to the desired growth phase.
- Then the cells are harvested by centrifugation and resuspended in an appropriate electroporation buffer.
- The DNA or other molecules of interest are also prepared. This may involve isolating and purifying the DNA or preparing a solution of the desired molecules.

2) Electroporate the Cells

- The cell-DNA mixture is transferred into an electroporation cuvette, which is a specialized chamber designed for electroporation.
- The cuvette is placed in the electroporation device called an electroporator.
- An electric pulse is applied to the cells by activating the electroporation device.

3) Recovery of Cells

- After the electroporation pulse, the cuvette is immediately removed from the electroporation device.
- Then the cells are allowed to recover briefly at room temperature or on ice before being transferred to a normal cell growth medium. During this time, the cell membranes reseal, and the introduced molecules are incorporated into the cells.

4) Culturing and Harvesting the Transfected Cells

- Finally, the transfected cells are cultured and harvested.
- After the recovery period, the cells are transferred from the cuvette to a suitable growth medium that promotes their survival and growth.
- The cells are incubated under appropriate culture conditions based on the experimental requirements.

Applications

- Electroporation is widely used in the transformation of bacterial cells to make them competent and capable of taking up foreign DNA for various purposes like recombinant protein production, cloning, and other biotechnological applications.
- Electroporation is also used in transfection. Similar to bacterial transformation, electroporation allows the manipulation of eukaryotic cells by engineering their genetic material.
- Electroporation is widely used in genetic engineering to introduce foreign DNA into cells. For example, it can be used for the transformation of plant cells, allowing the introduction of foreign DNA into plants for crop improvement.

Advantages

- Electroporation is a versatile method that can be used in various cell types, including bacteria, yeast, plant cells, and mammalian cells.
- Electroporation can be used to deliver genetic material into cells that are difficult to transfect.
- Electroporation provides consistent and reproducible results.
- Electroporation does not require the use of vectors for gene transfer. It is a safe and non-toxic method.

Limitations

- High-voltage electric pulses used in electroporation can cause cell damage or cell death if the parameters are not optimized.
- Electroporation requires specialized equipment, including an electroporator, cuvettes, and appropriate voltage settings.

6.2.2 Microinjection

Microinjection is one of the physical methods of gene transfer used to introduce DNA or other genetic materials directly into a cell using a small glass needle or micropipette (Fig.6.2). This method allows efficient transfer and integration of desired genes into the host cell's genome. Dr.Marshall A. Barber introduced the concept of DNA microinjection during

the early 19th century. Since then, it has continually evolved to keep up with the advancements in biomedical fields. The principle of microinjection is based on the direct delivery of genetic material into individual cells using a fine glass needle called a micropipette, a positioning device known as a micromanipulator, and a microinjector. The process is performed under a powerful microscope. The genetic material is delivered into the cell by applying hydrostatic pressure to release a fluid containing the DNA through the micropipette. The small tip diameter of the micropipette and the precise movements enabled by the micromanipulator allows the precise delivery of desired materials.

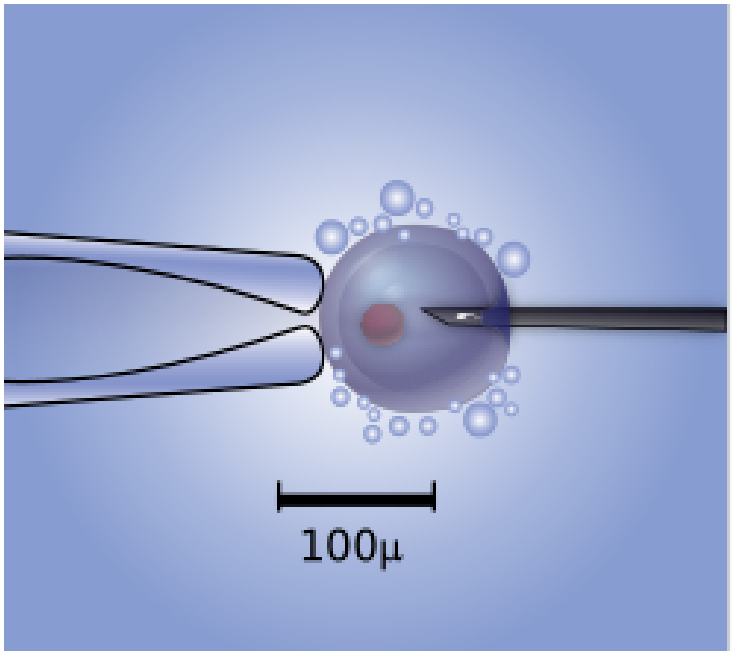


Fig-6.2: Microinjection Method

Method:

- First, the glass micropipette or needle is prepared by heating and stretching the glass to form a fine tip at the heated end. The resulting tip size is usually around 0.5 mm in diameter, resembling an injection needle.
- The entire process of delivering foreign materials using microinjection is performed under a powerful microscope for precise manipulation and observation.
- The cells to be microinjected are placed in a container.
- A holding pipette is placed near the target cell. It uses gentle suction to hold the cell during the injection process.
- The micropipette, containing the desired contents, is mounted on a micromanipulator, which allows precise positioning and movement of the pipette. The micropipette is lowered near the cell.
- The micropipette is carefully inserted into the cell membrane, either into the cytoplasm or the nucleus, depending on the target location for the genetic material.
- After delivering the contents into the cell, the empty micropipette needle is slowly and carefully withdrawn from the cell, minimizing damage to the cell and its membrane. Proper technique and high-quality micropipettes help minimize cell death during this step.

Applications of Microinjection

- Microinjection is commonly used in the production of transgenic animals, introducing foreign DNA into fertilized eggs to study gene function and create disease models.
- Microinjection has been widely used in various studies, including human embryo research and injection of mitotic cells.
- In neuroscience, microinjection is useful for working with primary cultured human neurons, delivering proteins, peptides, and cDNA constructs into the cytosol of human neurons which is difficult to perform using other gene transfer techniques.

Advantages of Microinjection

- Unlike other gene transfer methods, microinjection does not require the use of selection markers such as antibiotic-resistance genes. This makes the process easier and removes the need for additional steps to identify and isolate transformed cells.
- Microinjection allows precise delivery of materials in terms of volume and timing, which can be challenging with other methods like electroporation or transfection.
- Injected cells can be easily identified by co-injecting a marker dye or fluorescently labelled proteins.
- Microinjection requires less protein preparation compared to electroporation, making it advantageous for experiments involving less abundant or expensive proteins and peptides.
- Microinjection is less stressful to cells, reducing cell death that is commonly observed with chemical transfection or viral infection methods.

Disadvantages of Microinjection:

- Microinjection requires vehicle controls to assess potential effects on cell viability and to ensure accurate results.
- The technical expertise required to master the microinjection method and maintain cell viability is another limitation of this method.
- The use of alternative techniques like transfection, infection, and electroporation has become more popular compared to microinjection for delivering non-permeable foreign materials into single cells.

6.2.3 Biolistic Method

Biolistic method is one of the physical methods of gene transfer used to introduce foreign genetic materials directly into cells or tissues using high-velocity particles (Fig.6.3). The method is known by alternative names, including Particle Bombardment, Particle Gun, Ballistics, Microprojectile Bombardment and Particle Acceleration methods. John Sanford and Ed Wolf developed the concept of micro projectile bombardment. This technique was initially developed for plant cell transformation and later adapted for mammalian cells. The main principle of the microprojectile bombardment method is to deliver genetic materials directly into target cells by bombarding them with micro carrier particles at high velocity. The microcarrier particles, usually gold or tungsten, are coated with the genetic

material of interest. The high-velocity stream shoots the microcarrier particles toward the target cells, allowing them to penetrate the cell membrane and enter the cytoplasm. Once inside the cells, the genetic material separates from the microcarrier particles and can be used by the cell's machinery for gene expression.

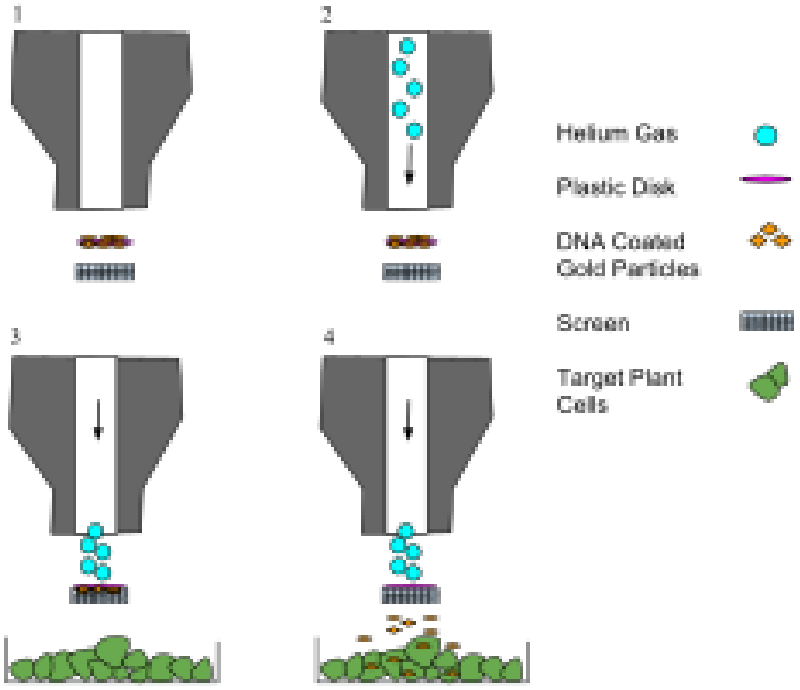


Fig-6.3: Biolistic Method

Steps:

- The process begins by preparing the microcarrier particles for coating with the genetic material of interest. The microcarrier particles are usually made of gold or tungsten.
- Then, the desired genetic materials are coated onto the surface of the microcarrier particles.
- Once the microcarrier particles are coated, they are loaded into the gene gun device. The gene gun shoots out particles at a high velocity using a strong pulse of pressurized helium.
- When the pressure inside the gene gun reaches a critical level, the rupture disk bursts, creating a strong wave of gas. This wave pushes the microcarrier that contains microcarrier particles toward the target cell. The macrocarrier hits the stopping screen. The stopping screen prevents the passage of the microcarrier but allows the coated microcarrier to pass through it.
- The target cells are placed under a vacuum in the main chamber of the gene gun. The cells are placed on a petri dish or culture plate.
- When the high-velocity particles reach the target cells, they penetrate the cell membrane and enter the cytoplasm.
- After the microcarrier particles reach the target cells, the genetic material of interest separates from the microcarrier particles. Once separated, the genetic materials can be used by the cell's machinery, leading to the expression of the genes.

Applications

- Microprojectile bombardment is widely used for introducing foreign genes into plant cells and producing genetically modified plants with improved traits such as disease resistance and higher yield.
- Microprojectile bombardment has also been used to generate transgenic animals with specific desired traits.
- This method of gene delivery allows the study of gene function and expression patterns in different tissues.
- This method also has applications in gene therapy for delivering therapeutic genes directly into target tissues to treat genetic disorders, cancer, and other diseases.

Advantages:

1) It is a fast and relatively simple method for delivering desired genetic materials into cells.
2) This method is not dependent on host specificity or species limitations.
3) Unlike other methods that need the cell wall to be removed, gene gun delivery can penetrate the intact cell wall. This simplifies the process and can be used to transform a wider range of cells

Disadvantages

1) This method requires specialized equipment. The initial investment for the required equipment and materials can be costly.
2) The target cells may experience physical damage due to the high-velocity particles leading to reduced cell viability.
3) When using the gene gun method on a larger scale, its effectiveness tends to decrease.
4) Another limitation is that the introduced DNA can randomly integrate into the host genome. This can result in unpredictable patterns of gene expression.

6.2.4 Chemical Methods:

- Cell membrane is a sheet like assembly of amphipathic molecules that separate cells from their environment. These physical structures allow only the controlled exchange of materials among the different parts of a cell and with its immediate surroundings. DNA is an anionic polymer, larger molecular weight, hydrophilic and sensitive to nuclease degradation in biological matrices. They cannot easily cross the physical barrier of membrane and enter the cells unless assisted.
- Chemical methods are transfection techniques that make use of carrier molecules to overcome the cell-membrane barrier. Different from what the name implicates, there are no chemical reactions taking place between the carrier molecule and the nucleic acid or any cellular component.
- The principle consists of the interaction of negatively charged nucleic acids with positively charged carrier molecules, like polymers or lipids, enabling the nucleic acid to come into contact with the negatively charged membrane components and incorporating the gene into the cell by endocytosis and later releasing it into the cytoplasm.

- These synthetic compounds are introduced near the vicinity of recipient cells thereby disturbing the cell membranes, widening the pore size and allowing the passage of the DNA into the cell.

An ideal chemical used for DNA transfer should have the ability to-

- Protect DNA against nuclease degradation.
- Transport DNA to the target cells.
- Promote the import of DNA into the nucleus.

The commonly used methods of chemical transfection are

1) Lipofection
2) Calcium phosphate
3) DEAE dextran

Liposome Mediated Gene Transfer (Lipofection):

Lipofection, also known as "lipid transfection" or "liposome-based transfection," uses a lipid complex to deliver DNA to cells. Liposomes are artificial phospholipid vesicles used for the delivery of a variety of molecules into the cells. They may be multi-lamellar or unilamellar vesicles with a size range of0.1 to 10 micrometer or 20-25 nanometers, respectively. They can be preloaded with DNA by two common methods- membrane-membrane fusion and endocytosis thus forming DNA- liposome complex. This complex fuses with the protoplasts to release the contents into the cell. Animal cells, plant cells, bacteria, yeast protoplasts are susceptible to lipofection method.

Liposomes can be classified as either Cationic Liposome or pH-sensitive

1) Cationic Liposomes

- Cationic liposomes are positively charged liposomes which associate with the negatively charged DNA molecules by electrostatic interactions forming a stable complex.
- Neutral liposomes are generally used as DNA carriers and helpers of cationic liposomes due to their non-toxic nature and high stability in serum. A positively charged lipid is often mixed with a neutral co-lipid, also called helper lipid to enhance the efficiency of gene transfer by stabilizing the liposome complex (lipoplex). Dioleoylphosphatidyl ethanolamine (DOPE) or dioleoylphosphatidyl choline (DOPC) are some commonly used neutral co-lipids.
- The negatively charged DNA molecule interacts with the positively charged groups of the DOPE or DOPC. DOPE is more efficient and useful than DOPC due to the ability of its inverted hexagonal phase to disrupt the membrane integrity.
- The overall net positive charge allows the close association of the lipoplex with the negatively charged cell membrane followed by uptake into the cell and then into nucleus.

- The lipid: DNA ratio and overall lipid concentration used in the formation of these complexes is particularly required for efficient gene transfer which varies with application.

2) Negatively Charged Liposomes

- Generally, pH-sensitive or negatively-charged liposomes are not efficient for gene transfer. They do not form a complex with it due to repulsive electrostatic interactions between the phosphate backbone of DNA and negatively charged groups of the lipids. Some of the DNA molecules get entrapped within the aqueous interior of these liposomes.
- However, formation of lipoplex, a complex between DNA and anionic lipids can occur by using divalent cations (e.g. Ca2+, Mg2+, Mn2+, and Ba2+) which can neutralize the mutual electrostatic repulsion. These anionic lipoplexes comprise anionic lipids, divalent cations, and plasmid DNA which are physiologically safe components.
- They are termed as **pH sensitive** due to destabilization at low pH.

The efficiency of both *in vivo* and *in vitro* gene delivery using cationic liposomes is higher than that of pH sensitive liposomes. But the cationic liposomes get inactivated and unstable in the presence of serum and exhibit cytotoxicity. Due to reduced toxicity and interference from serum proteins, pH-sensitive liposomes are considered as potential gene delivery vehicles than the cationic liposomes. (Fig.6.4).

Liposome Action:

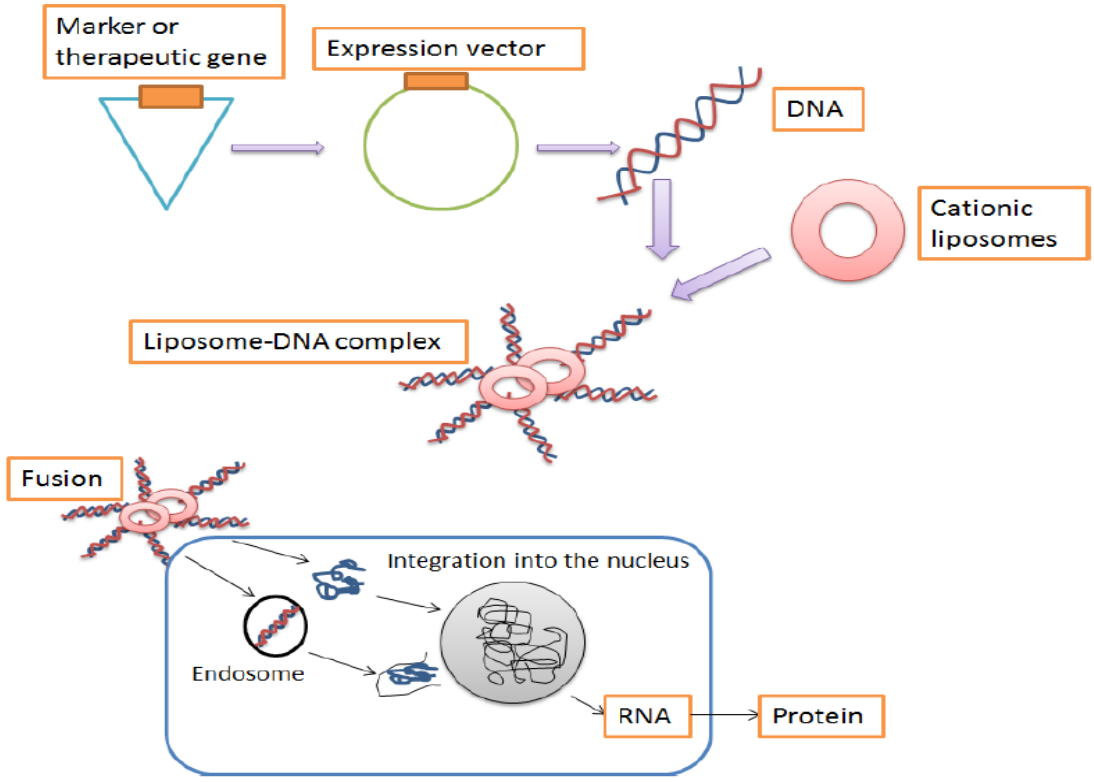


Fig-6.4: Schematic representation of liposome action in gene transfer

In addition, liposomes can be directed to cells using monoclonal antibodies which recognize and bind to the specific surface antigens of cells along with the liposomes. Liposomes can be prevented from destruction by the cell's lysosomes by pre- treating the cells with chemicals such as chloroquine, cytochalasin B, colchicine etc. Liposome mediated transfer into the nucleusis still not completely understood.

Advantages:

- Efficient delivery of nucleic acids to cells in a culture dish.
- Delivery of the nucleic acids with minimal toxicity.
- Protection of nucleic acids from degradation.
- Easy to use, requirement of minimal steps and adaptable to high-throughput systems.

Disadvantages:

- It is not applicable to all cell types.

Calcium Phosphate Transfection:

This method is based on the precipitation of plasmid DNA and calcium ions by their interaction. In this method, the precipitates of calcium phosphate and DNA being small and insoluble can be easily adsorbed on the surface of cell. This precipitate is engulfed by cells through endocytosis and the DNA gets integrated into the cell genome resulting in stable or permanent transfection (Fig.6.5).

Uses:

- This method is mainly used in the production of recombinant viral vectors.
- It remains a choice for plasmid DNA transfer in many cell cultures and packaging cell lines. As the precipitate so formed must coat the cells, this method is suitable only for cells growing in monolayer and not for suspension cultures.

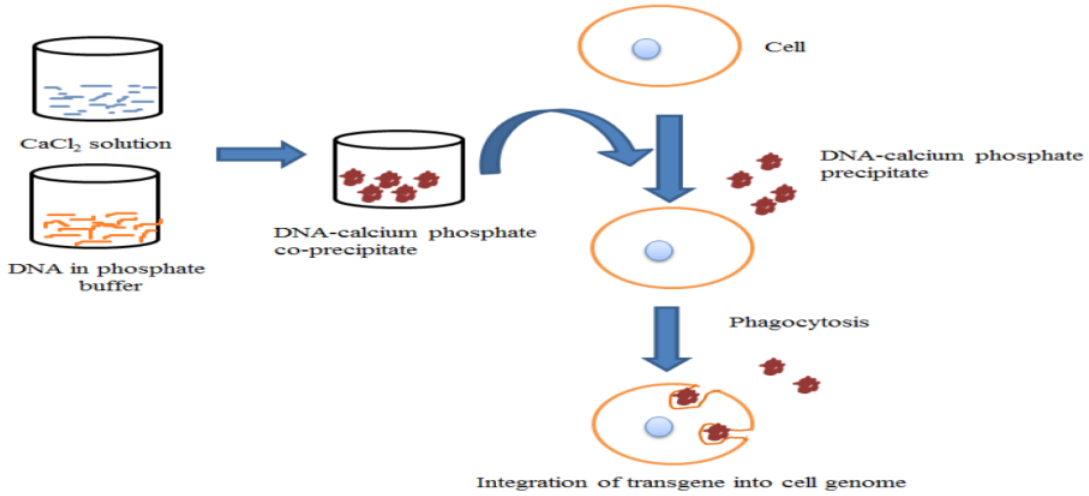


Fig-6.5: A schematic representation of transfection by Calcium phosphate precipitation

Advantages:

- Simple and inexpensive.
- Highly efficient (cell type dependent) and can be applied to a wide range of cell types.
- Can be used for stable or transient transfection.

Disadvantages:

- Toxic especially to primary cells.
- Slight change in pH, buffer salt concentration and temperature can compromise the efficacy.
- Relatively poor transfection efficiency compared to other chemical transfection methods like lipofection.
- Limited by the composition and size of the precipitate.

DEAE-Dextran (Diethylaminoethyl Dextran) mediated DNA transfer

- Diethylaminomethyl dextran (DEAE-dextran) is a soluble polycationic carbohydrate that promotes interactions between DNA and endocytic machinery of the cell.
- In this method, the negatively charged DNA and positively charged DEAE – dextran form aggregates through electrostatic interaction and form a polyplex. A slight excess of DEAE – dextran in mixture results in net positive charge in the DEAE – dextran/ DNA complex formed. These complexes, when added to the cells, bind to the negatively charged plasma membrane and get internalized through endocytosis. Complexed DNA delivery with DEAE-dextran can be improved by osmotic shock using DMSO (Dimethyl sulfoxide) or glycerol.
- Several parameters such as number of cells, polymer concentration, transfected DNA concentration and duration of transfection should be optimized for a given cell line.

Advantages:

- Simple and inexpensive.
- More sensitive.
- Can be applied to a wide range of cell types.

Disadvantages:

- Toxic to cells at high concentrations.
- Transfection efficiency varies with cell type.
- Can only be used for transient transfection but not for stable transfection.

6.3 SUMMARY:

Gene transfer involves indirect and direct methods. Various methods have been using in genetransfer techniques like Microinjenction, elcroporation and chemical methods are widely using for transfer of genetic material efficiently to produce novel characters of organisms.

6.4 TECHNICAL TERMS:

Gene transfer, Microinjection, Electroporation, Biolistic.

6.5 SELF ASSESSMENT QUESTIONS:

1) Explain in detail about the direct gene transfer methods.
2) Discuss various chemical methods of gene transfer techniques**.**

6.6 SUGGESTED READINGS:

1) Microbial Genetics" by David Freifelder.
2) "Principles of Gene Manipulation and technology" by Old & Primrose.
3) "Modern Microbial Genetics" edited by Streips and Yasbin.
4) "Microbial Genetics" by SylwiaOkoń, BeataZimowska, and MahendraRai.

Dr. J. Madhavi

LESSON-7

PHAGE GENETICS

7.0 OBJECTIVE:

- Students will understand the structure, genome organization, life cycles, and recombination of T4 and Lambda bacteriophages

STRUCTURE:

7.1 INTRODUCTION

Bacteriophages are obligate intracellular parasite on bacteria that uses bacterial machinery system for its own multiplication and development. These are commonly referred as “phage”. Bacteriophages were jointly discovered by Frederick Twort (1915) in England and by Felix d'Herelle (1917) at the Pasteur Institute in France. “Bacteriophage” term was coined by Felix d'Herelle. Some of the examples of bacteriophages are, spherical phages such as φX174 (ssDNA), filamentous phages such as M13 (ssDNA), T-even phages such as T2, T4 and T6 that infect *E.coli*, temperate phages such as λ and μ.

The pathogenicity of an organism is determined by its virulence factors. Virus virulence factors determine whether an infection will occur and how severe the resulting viral disease symptoms are. Viruses often require receptor proteins on host cells to which they specifically bind. Typically, these host cell proteins are endocytosed and the bound virus then

enters the host cell. Model organisms of virulent viruses that have been extensively studied include virus T4 and other T-even bacteriophages which infect *Escherichia coli* and a number of related bacteria.Typical phages have hollow heads (where the phage DNA or RNA is stored) and tunnel tails, the tips of which have the ability to bind to specific molecules on the surface of their target bacteria. The viral DNA is then injected through the tail into the host cell, where it directs the production of progeny phages often over a hundred in half an hour. These "young" phages burst from the host cell (killing it) and infect more bacteria.

7.2 T4 PHAGE STRUCTURE

The T4 bacteriophage has a complex structure consisting of an icosahedral head that contains a double-stranded DNA genome. This head is connected to a contractile tail with a hollow core, a contractile sheath, a baseplate, and six long tail fibers that attach to the host cell. The entire virion is approximately 200 nm long and 90 nm wide, and the tail fibers are crucial for host recognition and attachment (Fig.7.1).

Fig-7.1: T4 Bacteriophage Structure

7.2.1 T4 Genome Organization:

More than 200 T4 genes have been identified. These genes account for about 90% of the DNA; thus, perhaps 20 genes remain to be found. T4 genes of known function can be divided into two classes: 82 metabolic genes and 53 particle-assembly-genes. Of the 82 metabolic genes, only the 22 genes involved in DNA synthesis, transcription, and lysis are essential. The remaining 60 metabolic genes duplicate bacterial genes; particles in which these genes are mutated will grow, although occasionally they will have a smaller burst size. Of the 53 assembly genes, 34 codes for structural proteins, and 19 codes for the synthesis of enzymes and protein factors that are required catalytically for assembly. Thus, 17% of the DNA of phage T4 encodes essential metabolic functions, 39% is necessary for phage assembly, and 44% serves nonessential metabolic functions.

A genetic map of some of these genes is shown in Figure 7.2. A notable feature of the map is that genes having related functions are often adjacent and transcribed as part of poly cistronic mRNA molecules. This is an efficient arrangement, allowing the synthesis of functionally related proteins to occur at nearly the same time and minimizing the number of regulatory elements required. Not all functionally related genes, however, are part of single transcription units, and some transcription units contain functionally discrete genes. The tendency to cluster related genes is common in many phage systems.

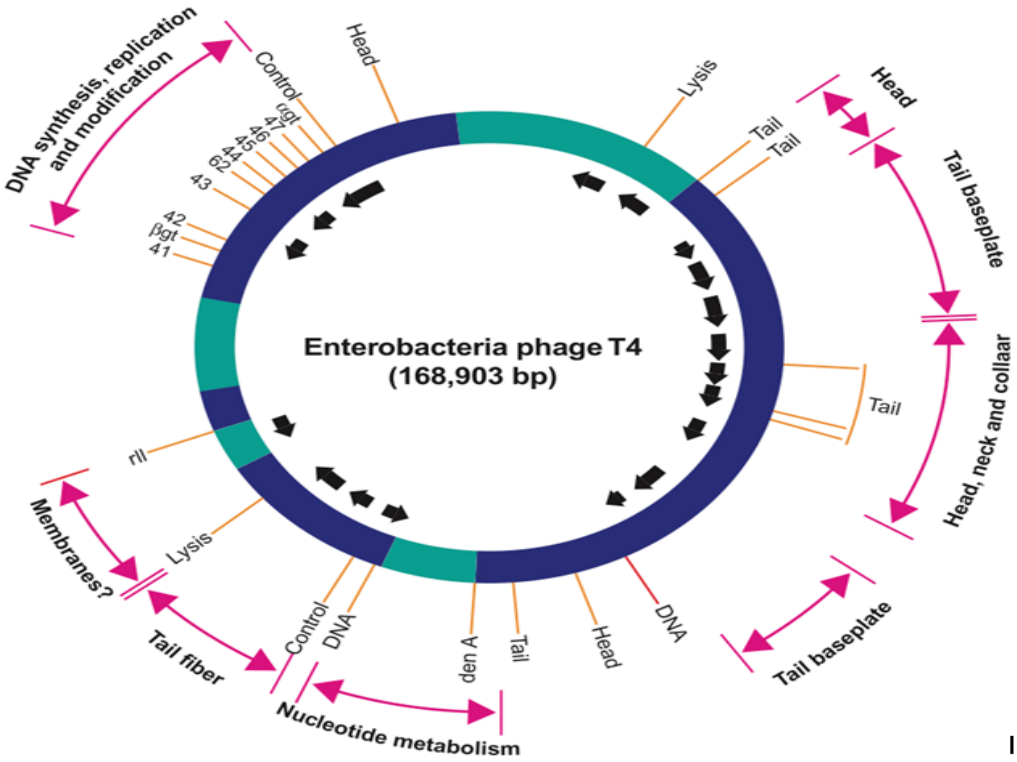


Fig-7.2: Genetic Map of T4 Bacteriophage

7.2.2 T4 Life Cycle

This life cycle is also known as virulent cycle because phages multiply inside the host and lyse the cell at the end of its life cycle (Fig.7.3). After attachment of tail fiber to host, genetic materials are injected inside the host. The time period between the entry of genetic material inside the host and release of mature phage after end of life cycle is termed as eclipse period. Synthesis of phage components and its packaging into mature phages takes place in this period. After infection, the genetic material of phages uses host biosynthetic machinery for replication, transcription and translation. Structural proteins of phages (capsid, tail etc.) are also synthesized inside the host using host biosynthetic machinery. After synthesis, genetic materials are packed inside the capsid and tail is attached on it. This process is called as maturation of phages. In lytic phage, phages also synthesized lysis protein. Bacterial cells are lysed due to accumulation of phage lysis protein and mature phages are released into the medium. Around 10-1000 phages are released from the bacterial cell. The average yield of phages per infected bacterial cell is known as burst size.

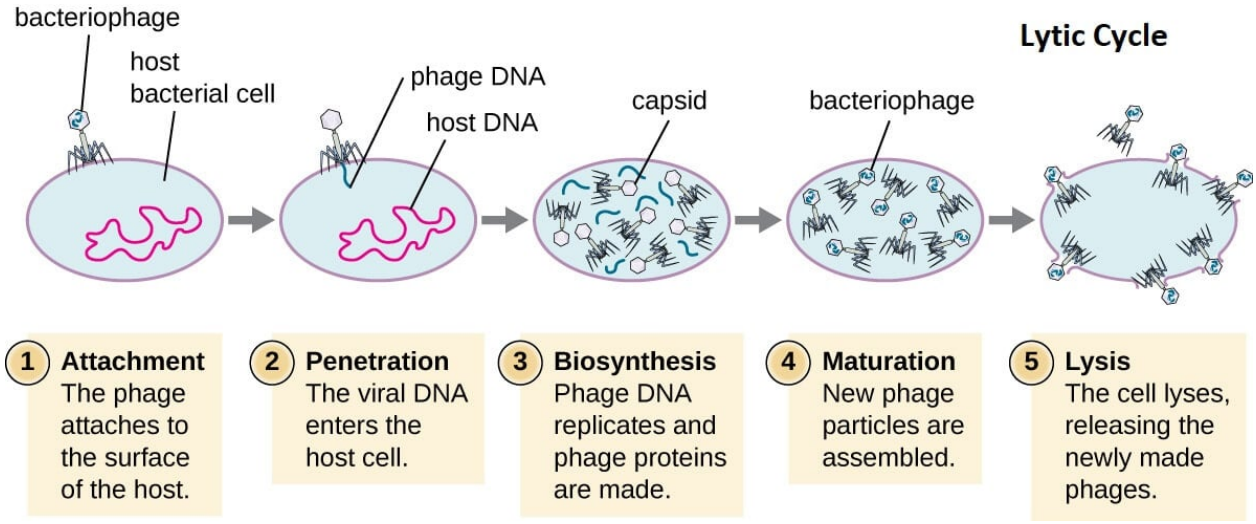


Fig-7.3: Lytic Cycle of T4 bacteriophage

7.2.3 Genetic Recombination

Genetic recombination is crucial for T4 propagation, particularly for DNA replication and packaging.

- **Recombination-Dependent Replication (RDR):** The T4 genome is linear and terminally redundant (has repeated sequences at both ends). While initial replication starts at specific origins, the primary mechanism for continuous DNA synthesis later in the infection cycle is RDR.
- **Mechanism:** RDR is initiated when the 3' end of a single-stranded DNA segment (often generated at the ends of linear DNA molecules or stalled replication forks) invades a homologous sequence in another DNA molecule, forming a D-loop structure. This invading strand serves as a primer for DNA polymerase, allowing replication to continue.
- **Concatemer Formation:** Repeated rounds of RDR result in the formation of large DNA concatemers (long, multi-genome-length molecules).
- **Packaging:** These concatemers are the substrates for packaging. Phage-encoded terminase enzymes cut the concatemer into unit-length genomes, which are then inserted into empty pro-heads. The "headful" packaging mechanism ensures each phage receives a full, circularly permuted, terminally redundant genome.

The T4 recombination system involves several key proteins, including the UvsX recombinase (analogous to *E. coli*RecA), the Gp32 single-strand DNA-binding protein, and the UvsY mediator protein. This system serves as a key model for understanding homologous recombination, DNA repair, and the intricate coupling of recombination and replication in all organisms.

7.2.4 Gene Expression in T4 Lytic Life Cycle

In the T4 lytic life cycle, gene expression is tightly controlled in a temporal cascade to produce proteins in a specific order (early, middle, late), while genetic recombination is an essential process used to repair DNA, initiate the secondary phase of replication, and produce the long DNA concatemers needed for packaging into new phage heads.

The T4 phage orchestrates gene expression primarily by modifying the host *E. coli* RNA polymerase (RNAP).

- **Early Genes (0–5 minutes post-infection):** Immediately after the phage DNA is injected, the host RNAP transcribes the T4 early genes using its own sigma factor (σ70sigma to the 70th power σ70). Early gene products include enzymes that:
    - **Shut off host gene expression:** T4-encoded proteins modify the host RNAP's alpha subunits and degrade the host's cytosine-containing DNA to divert the cellular machinery to viral functions.
    - **Encode replication and recombination proteins:** These initial proteins prepare the cell for DNA replication.
- **Middle Genes (2–10 minutes post-infection):** The expression of middle genes is activated by early proteins (like AsiA and MotA) that bind to the host RNAP and alter its promoter specificity, redirecting it to recognize middle promoters. These genes encode additional proteins for DNA metabolism, replication, and transcription factors for the next phase.
- **Late Genes (After 9 minutes post-infection):** Transcription of late genes, which primarily encode structural proteins for the phage head, tail, and tail fibers, requires further modification of the RNAP with new phage-encoded factors (gp33 and gp55) that replace the host's σ70sigma to the 70th power σ70 subunit. This stage is also coupled to active DNA replication, ensuring that structural components are only made when sufficient DNA is available for packaging.

7.3 LYSOGENIC PHAGE

Viruses are often very specific as to which hosts and which cells within the host they will infect. This feature of a virus makes it specific to one or a few species of life on earth. So many different types of viruses exist that nearly every living organism has its own set of viruses that try to infect its cells. Even the smallest and simplest of cells, prokaryotic bacteria, may be attacked by specific types of viruses. Among these, the lysogenic cycle is a mechanism where viruses and hosts form an intriguing, silent partnership. The bacteriophage lambda, with its interactions with the bacterium *Escherichia coli*, serves as a representative model to delve deeper into this cycle. In contrast to the aggressive takeover seen in the lytic cycle, the lysogenic cycle offers a more patient approach. Here, the viral genome becomes a silent tenant within the host's DNA, lying in wait for the right conditions to act. This strategy allows the virus to survive and spread without causing direct harm to the host.

- Key difference between the lytic pathway and the lysogenic pathway is that here new virus particles are not immediately released and will not immediately cause disease once they infect a host cell.

7.3.1 Structure of Lambda phage

Morphological structure of phage has anicosahedron head and a tail (Fig.7.4). The head is made of protein of several types and contains a 46,500 bp long genomic DNA. The phage λ contains double stranded circular DNA of about 17 µm in length packed in protein head of capsid. The head is 55 nm in diameter consisting of 300-600 capsomers (subunits) of 37,500 Daltons.The capsomers are arranged in clusters of 5 and 6 subunits i.e. pentamers and hexamers. The head is joined to a non-contractile 180 µm long tail by a connector. The tail

consists of 35 stacked discs. It ends in a fiber.There is a hole in capsid through which passes this narrow neck portion expanding into a knob like structure inside. The tail possesses a thin tail fiber (25 nm long) at its end which recognizes the hosts. Also the tail consists of about 35 stacked discs or annuli. Unlike T-even phage, it is a simple structure devoid of the tail sheath.

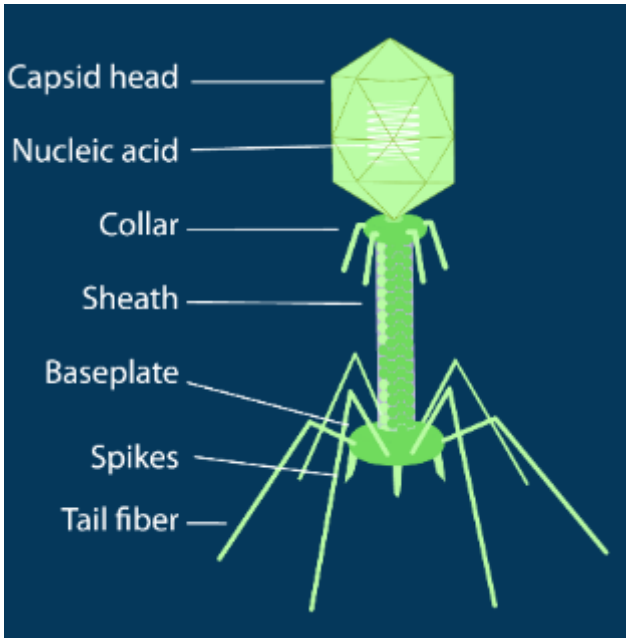


Fig-7.4: Structure of Lambda Phage

7.3.2 Phases of the Lysogenic Cycle

Viral Genome Integration:

- **Prophage Formation:** The viral DNA that integrates into the host cell's genome is termed a prophage. This integrated form allows the virus to essentially hide within the host, avoiding detection and immediate destruction.

Prophage: The latent form of a bacteriophage in which the viral genome is incorporated into the bacterial host's chromosome and replicated along with it without causing cell lysis.

- **Site-Specific Integration:** This integration is not random. Viruses like the lambda phage have specific sites where they prefer to integrate, ensuring stability and minimal disruption to the host genome.

Dormancy and Propagation

- **Silent Operation:** Once the prophage is in place, the viral genes remain largely inactive. However, the host cell operates normally, going about its regular functions.
- **Cell Division and Viral Spread:** As the host cell divides, it also replicates the viral DNA, passing the prophage to each of its progeny. This method allows the virus to spread silently within a population.

Induction and Lytic Entry

- **Stress Activation:** Certain environmental or cellular stressors can awaken the dormant prophage. Once activated, it can initiate the lytic cycle, producing more viruses and ultimately leading to the host cell's destruction.

7.3.3 Lambda phage life cycle

- **Attachment and Entry:** The lambda phage recognizes and binds to specific receptors on the E. coli cell surface, and then it injects its DNA into the bacterium.
- **Decision-making:** Post-entry, the phage must choose between the lytic and lysogenic paths. Environmental factors and the health of the host cell influence this decision.
- **Integration in the Lysogenic Pathway:** Opting for lysogeny, the lambda DNA integrates into a specific site in the *E. coli* chromosome. With this, the bacterium becomes a lysogen, bearing the viral DNA silently within.
- **Propagation:** As the lysogen divides, the viral DNA propagates, spreading to new generations.

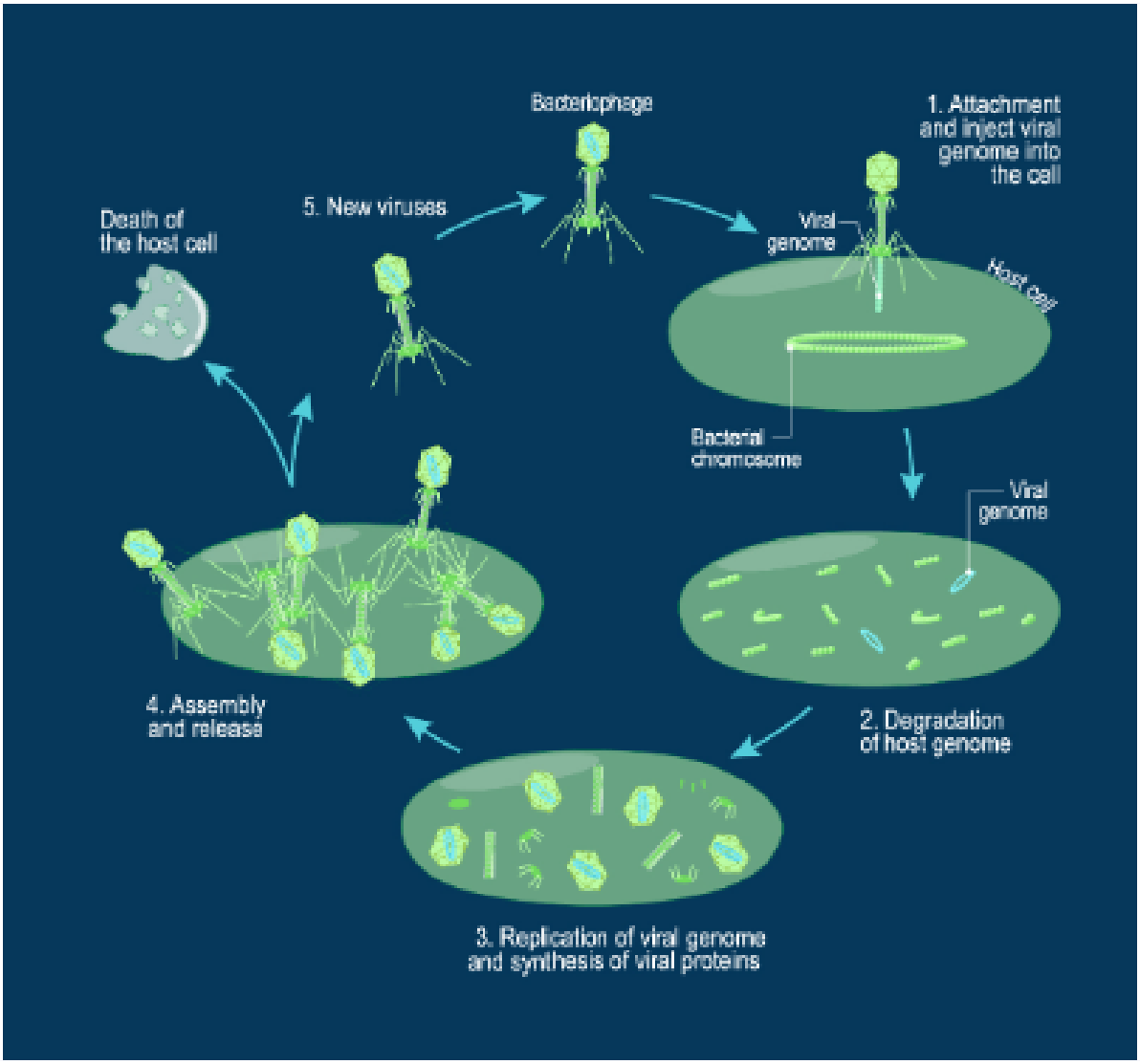


Fig-7.5: Lambda Phage Life Cycle

Genome Organization of Phage Lambda

Lambda DNA is a linear and double stranded duplex of about 17 µm in length. It consists of 48, 514 base pairs of known sequence. Both the ends of 5′ terminus consist of 12 bases which extend beyond the 3′ terminus nucleotide. This result in single stranded complementary

region commonly called cohesive ends. The cohesive ends form base-pairs and can easily circularize. Consequently a circular DNA with two single strand breaks is formed. The double stranded region formed after base pairing of complementary nucleotides is designated as COS. The events of circularization occur after injection of phage DNA into *E.coli* cell where the bacterial enzyme, i.e., *E.coli* DNA ligase, converts the molecule to a covalently sealed circle. The genetic map of lambda phage is given in figure 7.6.

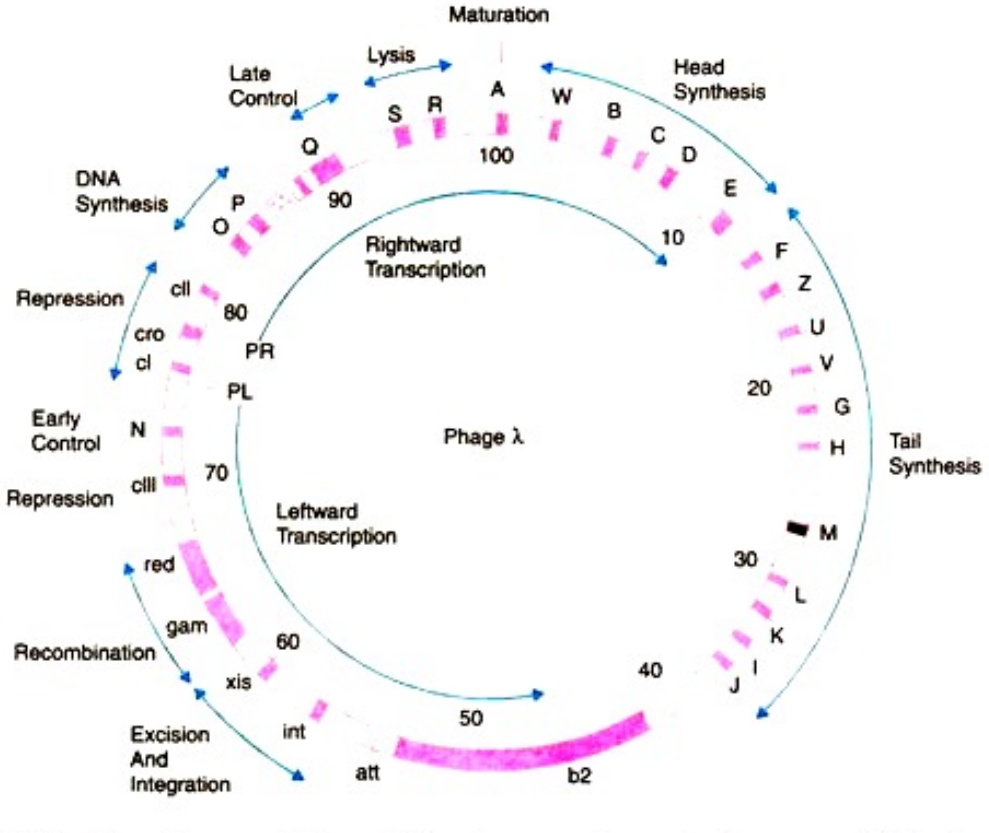


Fig-7.6: Genetic Map of Lambda Phage

7.3.4 Recombination

The primary form of recombination in the lysogenic cycle is site-specific recombination, where the phage genome is physically integrated into the host bacterium's chromosome to form a prophage. This process involves specific DNA sequences and a phage-encoded enzyme.

Site-Specific Recombination in the Lysogenic Cycle

In a temperate phage, such as the well-studied lambda phage that infects *E. coli*, the genetic material enters the host cell and circularizes. The choice between the lytic or lysogenic cycle depends on various factors, including host cell conditions and the multiplicity of infection.

If conditions favor lysogeny, site-specific recombination occurs in several key steps:

- **Attachment Sites:** The phage genome has a specific attachment site called *attP* (phage attachment site), and the bacterial chromosome has a corresponding site called *attB* (bacterial attachment site).
- **Enzymes:** The recombination is facilitated by a phage-encoded enzyme called integrase (Int) and a host-encoded protein, IHF (Integration Host Factor).

- **Integration:** The Int protein, often with the help of IHF, forms a higher-order DNA-protein complex (intasome) at the *attP* site. This complex mediates a precise crossover event between the *attP* and *attB* sites, resulting in the linear insertion of the phage DNA into the bacterial chromosome.

The integrated phage genome, or prophage, is then replicated passively along with the host's DNA during normal bacterial cell division and passed to all daughter cells. This state can be maintained indefinitely until environmental stressors (like UV light or certain chemicals) trigger its excision and a switch to the lytic cycle.

Recombination Types

While site-specific recombination is essential for *establishing* lysogeny, other types of recombination are also relevant to lysogenic phages:

- **Specialized Transduction:** During the *excision* phase, the prophage usually leaves the host chromosome precisely. However, errors can occur, leading to the rare excision of a portion of adjacent host genes along with the phage DNA. This composite genome can then be packaged into a new phage particle and transferred to a new host, a process called specialized transduction.
- **Generalized Transduction:** Some phages (e.g., P1 phage), though often referred to in the context of lysogeny, use generalized transduction, where any random segment of the host DNA can be packaged into a phage head during the lytic cycle and transferred to another bacterium.
- **Homologous Recombination:** This general recombination pathway, involving host cell machinery (like the RecA protein), can be involved in the repair of damaged phage DNA during mixed infections or in genetic engineering techniques used in laboratories.

In the lysogenic cycle, the phage's genetic material (DNA) integrates into the host bacterium's chromosome, becoming a dormant prophage. The host cell continues to replicate normally, passing the prophage to daughter cells, which is a form of genetic inheritance for the virus. Under stress, the prophage can excise itself, initiating the lytic cycle to produce new phages.

7.3.5 Genetics of Lysogenic Life Cycle

The genetic map of phage λ is remarkable and a characteristic of the map is the clustering of genes according to their functions. For example, the head and tail synthesis, replication and recombination genes are arranged in four distinct clusters. These genes can also be grouped into three major operons viz. right operon, left operon and immunity operon. The right operon is involved in the vegetative function of the phage e.g. head synthesis, tail synthesis and DNA replication leading lytic cycle. The left operon is associated with integration and recombination events of lysogenic cycle. The immunity operon products interact with DNA and decide whether the phage will initiate lytic cycle or lysogenic cycle. Singer et al (1977) have given the nucleotide sequence of ϕX174. Genetic map of bacteriophage has been given by Echols and Murialdo (1978).

i) Head Synthesis Genes

At the left end of phage genome the head genes viz. A, W, B, C, D, E are located which are associated with phage DNA maturation and head proteins.

ii) Tail Synthesis Genes

The genes F, Z, U, V, G, H, M, L, K, I, J are clustered just right to head genes and code for tail proteins.

iii) Excision and Integration Genes

The gene xis codes protein that excises the phage DNA from the bacterial chromosomes, and int coded protein is involved in integration of phage DNA into the bacterial chromosome.

iv) Recombination

The two genes int and xis codes at att P for site-specific recombination. The three red genes code for three proteins at normal frequency for general recombination. The redL codes for exonuclease, red B for beta-protein and red V for gamma protein. The gamma protein inhibits exonuclease V.

v) Positive Regulation Gene

The genes N and R are the positive regulation genes. The proteins code4by these genes increase the rate of transcription of other genes. Protein coded by N gene induces the transcription of cll, Q, P, A, red, gam, xis and int, whereas the protein coded by Q gene stimulates the transcription of head, tail and lysis genes. The N and Q genes are also required in plaque formation, in the absence of which the number of phage particles would be less but not zero.

vi) Negative Regulation Genes

The cl gene acts as a repressor and its product maintains the prophage in the lysogenic form in bacterial host. Moreover, the cll and cIII assist the d gene in lysogeny. The proteins encoded by cro binds to PL and PR and reduce the expression of cl, N, red and xis genes.The interactions between Q proteins encoded by cro and phage repressor occur in host cell and the result decides the operation of lytic or lysogenic cycle.

vii) DNA Synthesis Genes

The two genes O and P are involved in synthesis of phage DNA. The origin of DNA replication lies within the coding sequence for gene Q which encodes a protein for initiation of DNA replication, and the gene that generates the cohesive ends is located adjacent to one of the ends. The function of gene N is required in transcriptional process of these genes.

viii) Lysis Genes

The S and R genes control the lysis of bacterial cell envelope which occurs at the end of lytic cycle.

7.4 SUMMARY:

The lytic cycle is a rapid, destructive viral replication process where a virus hijacks a host cell's machinery to produce new viruses, which are then released by bursting the cell. In contrast, the lysogenic cycle is a more covert cycle where the viral DNA integrates into the host's chromosome and is copied passively during cell division without immediate harm, only to later enter the lytic cycle if triggered.

7.5 TECHNICAL TERMS:

Lambda phage, Lysogenic cycle, Recombination, Genome Organization, Genetic Map.

7.6 SELF ASSESSMENT QUESTIONS:

1) Explain in detail about the gene expression in T4 life cycle.
2) Explain the structures of T4 and Lambda bacteriophages.
3) Explain the primary difference between the lytic and lysogenic cycles, using the bacteriophage lambda as an example.

7.7 SUGGESTED READINGS:

1) "Microbial Genetics" by David Freifelder**.**
2) "Principles of Gene Manipulation and Technology" by Old and Primrose**.**
3) "Modern Microbial Genetics" edited by Streips and Yasbin**.**
4) "Microbial Genetics" by Sylwia Okoń, Beata Zimowska, and Mahendra Rai.

Dr. J. Madhavi

LESSON-8

DNA REPLICATION

8.0 OBJECTIVE:

- This lesson plan is aimed to make the students to understand the mechanism of replication of DNA and the enzymology of replication process.

STRUCTURE:

8.1 INTRODUCTION

Genetic information is transferred from parent to progeny organisms by a faithful replication of the parental DNA molecules. In prokaryotes, the replication of DNA occurs during Log phase of growth and conjugation process. In eukaryotes, replication of DNA occurs during the 'S' sub-phase of interphase in both mitosis and meiosis. Replication of double-stranded DNA is a complicated process and involves various enzymes. The complexity is from the facts that: (1) a supply of energy is required to unwind the helix, (2) the single strands resulting from the unwinding tend to form intrastrand base pairs, (3) a single enzyme can catalyze only a limited number of physical and chemical reactions and many reactions are needed in replication, (4) several safeguards have evolved that are designed both to prevent replication errors and to eliminate the rare errors that do occur and (5) both circularity and the enormous size of DNA molecules impose geometric constraints on the replicative system.

There are three hypothetically possible modes of DNA replication viz., (1) Dispersive, (2) Conservative and (3) Semi-conservative. But the most practical mode of replication that normally occurs in vivo is the semi-conservative replication. In dispersive replication, the old DNA molecule would break into several pieces, each fragment would replicate, and the old

and new segments would recombine randomly to yield progeny of DNA molecules with a combination of old and new segments along its length. According to the conservative replication, the two newly synthesized strands would associate to form one double helix, while the two old strands would remain together as one double helix. In contrast, in the semi-conservative mode of DNA replication, each newly synthesized strand of DNA would remain associated with the old strand against which it was synthesized. Thus each progeny DNA molecule would consist of one 'old' and one 'newly synthesized' strand.

8.2 SEMI-CONSERVATIVE REPLICATION OF DNA

The semi-conservative mode of DNA replication was postulated by Watson and Crick along with the double-helix model of DNA. The main features of this mode of DNA replication are:

1) A progressive separation of the two strands of a DNA molecule
2) Complementary base-pairing of the bases located in the single-stranded regions thus produced with the appropriate free deoxyribonucleotides
3) Formation of phosphodiester linkages between the neighbouring deoxyribo nucleotides that have base-paired with the single-stranded regions, thereby producing the regions of new strand
4) This ensures that the base sequences of the new strands are strictly complementary to those of the old strands
5) As a result, each DNA molecule produced by replication has one 'old' and one 'new' strand
6) The base sequence of a newly synthesized strand is dictated by the base sequence of the old strand, since the old strand serves as a template for the synthesis of the new strand

Evidence for Semi-conservative Replication

The evidence for semi-conservative replication of DNA was first presented by Meselson and Stahl in 1958. They grew *E. coli* on 15N (a heavy isotope of ^{14}N) for 14 cell generations so that the nitrogen present in DNA bases of these cells was ^{15}N. DNA having ^{15}N has a detectably higher density than that having ^{14}N, therefore they are called heavy and light DNA, respectively. The heavy and light DNAs can be readily separated through equilibrium density gradient centrifugation as they form distinct bands in the centrifuge tube. Indensity gradient centrifugation, a heavy salt solution, $CsCl_2$ is centrifuged at 30,000-50,000 rpm for 48-72 hrs which leads to the formation of a linear gradient of increasing density from the top of centrifuge tube to the bottom of tube. When DNA is centrifuged in such a solution, it will move to a position where the density of salt solution is the same as that of DNA.

Meselson and Stahl transferred the *E. coli* cells grown on ^{15}N medium to a medium containing normal ^{14}N. They withdrew samples from these *E. coli* cells after approximately one, two and three cell generations. DNA from these cell samples was separated and subjected to density gradient centrifugation. After one cell generation, the DNA formed a single band intermediate between the heavy and light DNAs. The DNA obtained after two cell generations formed two bands of comparable intensity, one of the bands was intermediate and the other was light in density. The same two bands were recovered in the DNA isolated after three cell generations, although the intermediate band was relatively lower in intensity than the light band.

These findings can be readily explained on the basis of semi-conservative replication of DNA. The DNA from *E. coli* cells grown on ^{15}N had ^{15}N in both the strands; therefore it was heavier than the normal DNA having ^{14}N. When these *E. coli* cells were allowed one semi- conservative replication of their DNA on the ^{14}N medium, each of the resultant DNA molecules would have one heavy (^{15}N) and one light (^{14}N) strand. Therefore, these DNA molecules would have intermediate density. One more semi-conservative replication of these DNA molecules in the ^{14}N medium would generate two types of DNA molecules – (i) half of the molecules would have one heavy and one light strand (intermediate density) and (ii) the remaining half would have both light strands (light density). These molecules would obviously form one intermediate and one light band. On the third round of DNA replication on ^{14}N, the intermediate density DNA molecules would yield half intermediate and half light molecules, while all the molecules obtained from light DNA molecules would be light. This is the reason for the lower intensity of the intermediate band after three cell generations. All the expectations based on the semi-conservative mode of DNA replication are satisfied by these findings to the full extent.

8.3 ENZYMOLOGY OF DNA REPLICATION

DNA replication involves several proteins and enzymes. In *E. coli*, at least two dozen gene products are involved in DNA replication. Many of these gene products have been purified and their roles in DNA replication have been studied in vitro. Many of these proteins were first identified through the studies of mutants of *E. coli*, while some enzymes eg., ligase, DNA polymerase I etc., were first discovered biochemically.

8.3.1 DNA Polymerase

DNA polymerase is the chief enzyme of DNA replication. The DNA polymerase activity was first demonstrated by Kornberg in 1956. It catalyzes the covalent addition of deoxyribo nucleotides to the 3′-OH of a preexisting polynucleotide called as a primer. Clearly this enzyme cannot initiate the synthesis of a poly nucleotide, it can only add nucleotides to a primer polynucleotide.

All known polymerases have an absolute requirement of a free 3′-OH on the primer and, therefore, the direction of synthesis of the new chain is always 5′ to 3′. They also require a template DNA strand which specifies the base sequence of the new DNA chain. These enzymes utilize the 5′ triphosphates of the four deoxyribonucleotides viz., dATP, dGTP, dCTP and dTTP.

The complementary base pairing during DNA replication, purely on the basis of chemical and physical forces involved in the process, would give rise to error at a frequency of approximately 10^3 per base pair replicated. The actual rate of errors in bacteria seems to be about 10^{-8} to 10^{-10}. This phenomenal decline in the error rate may be due to the following two processes: (1) DNA polymerase may scrutinize the bases entering into pairing with the bases of DNA strand being replicated and allow only the correct pairing, and (2) it does scrutinize the base pair after the new base has been added to the chain and deletes the wrong ones (the process is called 'proof reading'). DNA polymerase definitely proof-reads new chains.

DNA polymerases isolated from prokaryotes and eukaryotes differ from each other in several respects.

Prokaryotic DNA Polymerases

Three different types of DNA polymerases, designated as DNA polymerase I, II, and III, have been isolated from prokaryotes.

DNA Polymerase I (Kornberg Enzyme): DNA polymerase I or Kornberg enzyme was the first to be isolated from *E. coli* by Arthur Kornberg and his associates and was used for in vitro DNA synthesis in 1957. For quite sometime, it was believed that this enzyme catalyses DNA replication in vivo. But it is now known that this enzyme is chiefly a DNA repair enzyme.

DNA polymerase I possesses the following activities: (1) 5′→ 3′ polymerase, (2) 3′ → 5′ exonuclease and (3) 5′ → 3′ exonuclease. The 5′ → 3′ and 3′ → 5′ exonuclease activities appear to be located in different sites of the enzyme.

An exonuclease is an enzyme that degrades nucleic acids from one end and is unable to make internal cuts in a polynucleotide. The 3′ → 5′ exonuclease activity enables the enzyme to remove deoxyribonucleotides, one by one, from the 3' end of a DNA strand. This activity is crucial for 'proof reading' or 'editing' of the polynucleotide chain being newly synthesized.

Whenever the DNA chain being synthesized has a wrong base insertion, the 3′ → 5′ exonuclease activity of the enzyme removes the wrong base, and the 5′ → 3′ polymerase activity reinitiates the synthesis of the chain. This activity reduces the chances of errors in DNA replication.

On the other hand, the 5′ → 3′ exonuclease activity functions in the removal of the DNA segments damaged by UV light, irradiation and other agents. This activity excises a small groups of nucleotides, up to 10. An endonuclease must cleave the DNA strand close to the site of damage before the 5′ → 3′ exonuclease action of the DNA polymerase I may take place.

This activity of DNA polymerase is also responsible for the excision of RNA primers during DNA replication, and the ribonucleotides are immediately replaced by appropriate deoxyribonucleotides due to the 5′ → 3′ polymerase activity of the enzyme.

The DNA polymerase I enzyme is coded by the gene polA. It is a single chain polypeptide of 109,000 daltons and can be cleaved by proteolytic treatments into a large and a small segments. The larger segment is of 68,000 daltons and is known as 'Klenow fragment' which exhibits the 5′→ 3′ polymerase and 3′ → 5′ exonuclease activities. The small fragment is of 65,000 daltons and possesses the 5′ → 3′ exonuclease activity.

DNA Polymerase II and III: The other two DNA polymerases, DNA polymerase II and III, were discovered in polA mutants of *E. coli* which are deficient in DNA polymerase I. The pol A mutants replicate their DNA at normal rates but are deficient in DNA repair activity.

The DNA polymerase II has 5′→ 3′ polymerase activity and 3′→ 5′ exonuclease activity. The exact in vivo function of this enzyme is not known, although it most likely functions in DNA repair, particularly, in the absence of DNA polymerases I and III.

DNA polymerase III possesses 5′ → 3′ polymerase and 3′ → 5′ exonuclease activities and is responsible for DNA replication in vivo. This enzyme has a higher affinity for nucleotide triphosphates than DNA polymerases I and II, and catalyzes the synthesis of DNA chains at very high rates. This enzyme has several active subunits (Fig. 8.1) the α subunit of 130,000 daltons which possesses polymerase activity, while another subunit, ε is of 25,000 daltons exhibits the proof reading ability. Generally, the α and ε subunits function in conjunction with each other. Three other subunits, called δ, γ and τ, confer onto the enzyme the ability to remain on a sing template.

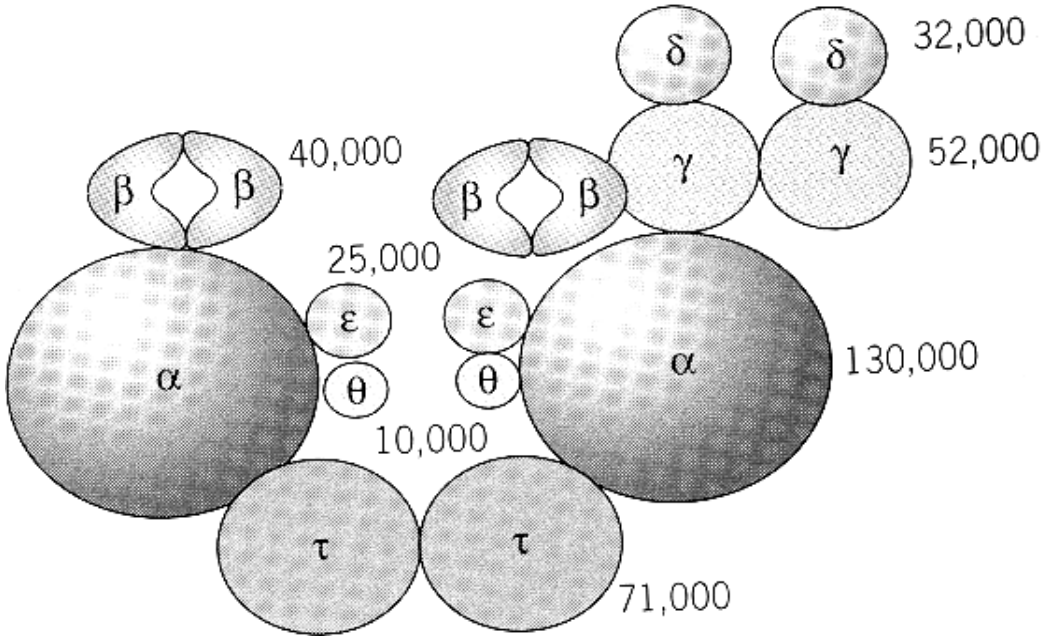


Fig-8.1: Structure of Prokaryotic DNA Polymerase

The major properties of the three DNA polymerases are summarized in the table 8.1 below.

Table 8.1 – Properties of DNA polymerases I, II and III of *E.coli*

Characteristic	DNA Pol I	DNA Pol II	DNA Pol III
Molecular weight (Daltons)	109,000	120,000	>250,000
Constitution	Monomer	Not known	Heteromultimer
Molecules per cell	400	Not known	10-20
Nucleotides polymerized at 37°C/min/molecule	Up to 1,000	Up to 50	Up to 15,000
Affinity for 5′ triphosphates of deoxyribonucleosides	Low	Low	High
5′ → 3′ polymerase	Present	Present	Present
3′→5′ exonuclease	Present	Present	Present
5′→3′ exonuclease	Present	Absent	Absent
Functions in	DNA repair, excision of RNA primers	DNA repair	DNA replication, the real replicase

Eukaryotic DNA Polymerases

There are four different types of eukaryotic DNA polymerases called α, β, γ and δ. DNA polymerases α and β are confined to the nuclei of cells. DNA polymerase α most likely catalyzes chromosome replication in eukaryotes. DNA polymerase γ is found in mitochondria and chloroplasts, and is believed to be responsible for the replication of chromosomes of these organelles. DNA polymerase δ has been isolated from calf thymus and rabbit bone marrow, but the functions of this enzyme are not yet known. Of the four polymerases only DNA polymerase δ possesses the 3′ → 5′ exonuclease activity. And none of the eukaryotic polymerases has the 5′ → 3′ exonuclease activity required for the removal of RNA primers during DNA replication. This function is carried out by some other enzyme in eukaryotes.

8.3.2 Unwinding Proteins

The unwinding of DNA double helix to generate single-stranded regions of the DNA duplex for DNA replication begins at the origin site, and is the first event in replication initiation. In *E. coli*, it requires six proteins, viz., DnaA, DnaB, DnaC, HU, gyrase and SSB, for the formation of a prepriming complex in vitro. The OriC of *E. coli* has four 9 bp consensus repeats in its right side, while three 13 bp consensus repeats occur in its left side. The first step in the formation of prepriming complex is the binding of DnaA protein to the 9 bp repeats, one molecule to each repeat. More molecules of Dna A go on binding cooperatively till 20-40 molecules form a core and the OriC DNA folds around this core. The Dna A molecules act on the three 13 bp repeats in the left side of the OriC. As a result, each of these sites 'melts' to become opened up and get separated to yield single-stranded regions.

Six molecules of DnaB form a hexamer to which six monomers of DnaC become associated and this large aggregate binds to the single-stranded structure generated in the 13 bp repeats of OriC. DnaB provides the helicase activity which actually unwinds DNA. It actually replaces the DnaA from the 13 bp repeats and begins the unwinding of DNA. Gyrase provides the swivel that allows one strand to unwind around the other. Without this action unwinding would generate torsion in the DNA molecule. The SSB protein bind to the single-stranded segment of DNA so generated and stabilizes it and prevents duplex formation. The HU protein is a histone-like protein of *E. coli* is not absolutely necessary for replication initiation in vitro but it enhances the reaction. The initiation of replication generally requires about 60 bp to be unwound or single-stranded. ATP is required for activities of helicase, gyrase, primase and DNA polymerase III.

8.3.3 Primase

These enzymes catalyze the synthesis of RNA primers which are a prerequisite for the initiation of DNA replication in the vast majority of organisms. Primase is quite distinct from the RNA polymerase which is normally used for the transcription of DNA. The primase of *E. coli* and certain viruses are also called DnaG protein since it is produced by the Dna G gene.

8.3.4 Polynucleotide Ligase

Ligase enzyme is an important enzyme both in DNA replication and in DNA repair. DNA ligase catalyzes the formation of phosphodiester linkage between the 5′-phosphoryl group of one nucleotide and the 3′-OH group of the immediate neighbour nucleotide at the side of a nick in a DNA strand. The property of formation of phosphodiester linkage by DNA

ligase is similar to that of DNA polymerase. But the DNA ligase cannot fill in the gaps in a DNA strand, while DNA polymerases do this precisely. In addition ligase seals the nicks left behind by DNA polymerase I during DNA repair and among the Okazaki fragments generated during discontinuous DNA replication. DNA ligases isolated from different organisms differ in their properties from each other.

8.4 MECHANISM OF REPLICATION

DNA replication is a complex process and involves various enzymes (Fig. 8.2). The replication of DNA is catalyzed by the enzyme DNA polymerase. Replication of DNA begins at certain unique and fixed point called 'origin site'. Two enzymes, DNA gyrase and DNA helicases, bind to the origin points and induce the unwinding of complementary strands of DNA double helix. Certain proteins, called single-stranded binding proteins, bind to the single-stranded regions thus produced to keep them in single-stranded condition. An enzyme called primase initiates transcription of the strand whose 3′-end is single-stranded (3′→5′ strand) which generates a 10-60 nucleotide-long primer RNA transcribed in 5′→3′ direction. The free 3′-OH of this primer RNA provides the initiation point for DNA polymerase for the sequential addition of deoxyribonucleotides. DNA polymerase has an absolute requirement for a free 3′-OH of a pre-existing polynucleotide for the initiation of DNA replication. DNA polymerase III progressively adds deoxyribonucleotides to the free 3′-OH of this growing polynucleotide chain so that the replication of the 3'→5′ strand of the DNA molecule is continuous i.e., growth of the new strand in 5′→3′ direction.

The replication of the second strand (5′→3′ strand) of the DNA molecule is discontinuous. The replication of this 5′→3′ strand begins somewhat later than that of the 3′→5′ strand. Consequently, a segment of the 5′→3′ strand of a DNA molecule always replicates later than the homologous segment of the 3′→5′ strand. Therefore, the 3′→5′ strand of a DNA molecule is known as the 'leading strand', while the 5′→3′ strand is termed as the 'lagging strand'. When the replication of the 3′→5′ strand has progressed for some time, primase initiates the synthesis of RNA primer on the 5′→3′ strand close to the replication fork and away from the origin. The RNA synthesis progresses towards the origin in the 3′→5′ direction of the lagging strand. The 3′-OH of this primer RNA provides the initiation point for DNA polymerase to catalyze the replication of the 'lagging strand'. The replication of lagging strand proceeds in the direction which is opposite to that of the leading strand. However, on both leading and lagging strands, the new strand is synthesized from the 5′-end to the 3′-end.

The replication of lagging strand generates small polynucleotide fragments called 'Okazaki fragments' named after the scientist R. Okazaki, who first identified them. The replication of 5′→3′ strand is discontinuous in that it has to be initiated several times, and every time one Okazaki fragment is produced. These Okazaki fragments are about 1000-2000 nucleotides long in *E. coli*, while they are only 100-200 nucleotides long in eukaryotes.

The RNA primer associated with the newly synthesized DNA strands / Okazaki fragments are most likely digested by the DNA polymerase I in prokaryotes. This enzyme also catalyzes the filling of the gaps so generated in the new strands through semi-conservative replication of the old strands. The Okazaki fragments, after the gap-filling by DNA polymerase I, are joined together by the enzyme polynucleotide ligase which catalyzes the formation of phosphodiester bonds between the immediate neighbour nucleotides of the adjacent fragments. The replication fork generated at the origin may produce an 'eye' in a linear DNA molecule, while in a circular DNA it generates a 'θ-structure'. The eye and θ-structures shall be produced irrespective of whether the DNA replication is unidirectional or bi-directional.

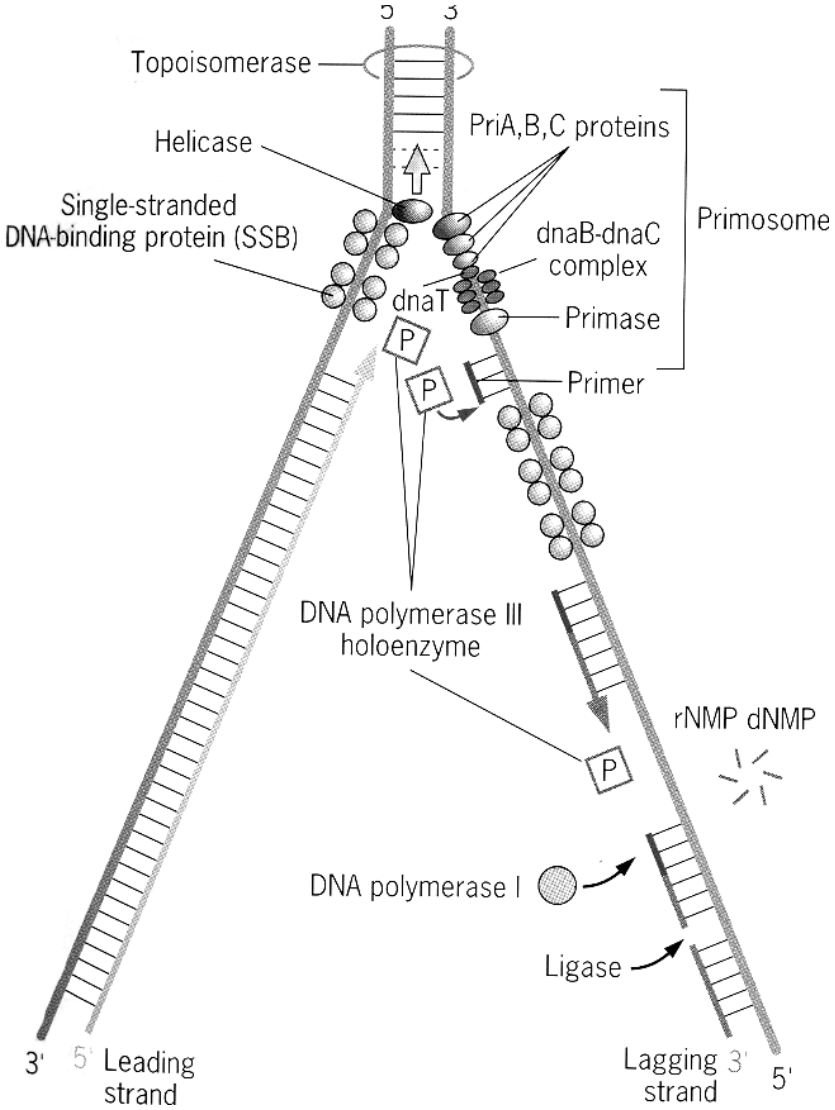


Fig-8.2: Mechanism of DNA Replication

8.4.1 Unidirectional and Bidirectional Replication of DNA

If the replication of the DNA proceeds only in one direction from the origin point, it is referred as unidirectional replication. In bi-directional replication, the process proceeds in both the directions. In unidirectional replication, one branch point remains at a fixed position with respect to the replication bubbles and this position defines the replication origin. In bi-directional replication both branch points move with respect to the bubbles, so each branch point is a replication fork. Bi-directional replication has been widely observed with phage, bacterial, and plasmid DNA. A small number of phages and plasmids use the unidirectional mode exclusively. Bi-directional replication from a fixed origin has also been demonstrated for several organisms with chromosomes that replicate as linear structures.

Replication of the chromosome of phage T7, begins at a unique site near one end to form a so-called 'eye' structure and then proceeds bidirectionally until one fork reaches the nearest end. Replication of DNA molecules in the chromosomes of eukaryotes is also bi-directional. However, bidirectional replication is not universal. The chromosome of coliphage P2, which like the lambda chromosome is circular during replication, replicates unidirectionally from a unique origin.

8.4.2 Rolling Circle Replication of DNA

The rolling circle model of DNA replication applies to the replication of several viral DNAs such as ΦX174 and λ and to the replication of the *E. coli* F factor during conjugation and transfer of donor DNA to a recipient.

The unique aspect of rolling-circle replication is that one parental circular DNA strand remains intact and rolls while serving as a template for the synthesis of a new complementary strand (fig. 8.3). The first step is the generation of a specific cut or nick in one of the two strands at the origin of replication by sequence-specific endonuclease producing 3′-OH and 5′-phosphate termini.

The 5′ end of the cut strand is then displaced from the circular molecule. This creates a replication fork structure and leaves a single stranded stretch of DNA that serves as a template for the addition of deoxyribonucleotides to the free 3′ end by DNA polymerase III using the intact circular DNA as a template.

This new DNA synthesis occurs continuously as the 5′ cut end continues to be displaced from the circular molecule; thus, the intact circular DNA is acting as the leading strand template. The 5′ end of the cut DNA strand is rolled out as a free "tongue" of increasing length as replication proceeds. This single-stranded DNA tongue becomes covered by single stranded binding proteins.

New DNA synthesis on the displaced DNA occurs in the 5′-to-3′ direction i.e., from the circle out toward the end of the displaced DNA. With further displacement, new DNA synthesis again must begin at the circle and move outward along the displaced DNA strand. Thus, synthesis on this strand occurs discontinuously as the displaced strand acts as lagging strand template.

In this case, primase synthesizes short RNA primers that are extended as DNA in the form of Okazaki fragments by DNA polymerase III. The RNA primers are ultimately removed and adjacent Okazaki fragments are joined through the action of DNA ligase. The so formed linear molecule will become circularized to form into a circular molecule.

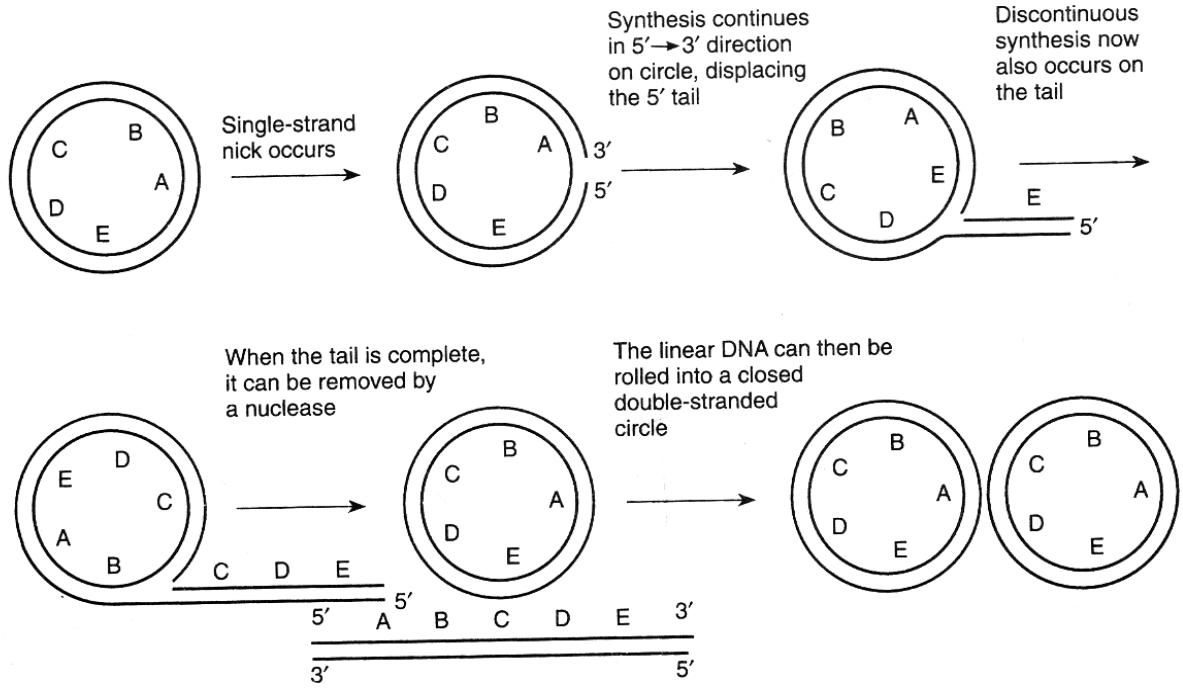


Fig-8.3: Rolling circle mechanism of Replication

8.5 SUMMARY:

DNA replication is semi-conservative, which indicates that one half of the parent duplex is transmitted to each of the daughter cells during cell division. The mechanism of replication was first suggested by Watson and Crick as part of DNA model structure. Replication begins at a single origin on the circular bacterial chromosome and proceeds outward in both directions as a pair of replication forks. The initiation of strand separation at the origin site involves a number of proteins. DNA synthesis and editing or proof-reading are catalyzed by a family of DNA polymerases. To catalyze the polymerization reaction, the enzyme requires all four deoxyribonucleoside triphosphates, a template strand to copy, and a primer containing a free 3′-OH to which nucleotides can be added.

The synthesis of new strand is always carried out in 5′→3′ direction. One of the newly synthesized strands (the leading strand) grows toward the replication fork and is synthesized continuously. The other newly synthesized strand (the lagging strand) grows away from the fork and is synthesized discontinuously.

Events at the replication fork require a variety of different types of proteins having specialized functions. These proteins include DNA gyrase, DNA helicase, SSBs, Primase etc. The DNA polymerase III is the prime replicase responsible for replication and the DNA polymerase I is for editing or proof-reading function and for the removal of RNA primers after the DNA strand synthesis. And the DNA ligase seals the fragments of the lagging strand into a continuous polynucleotide.

8.6 SELF ASSESSMENT QUESTIONS:

1) Describe the mechanism of DNA replication.
2) Write an essay on the enzymology of DNA replication.
3) Explain the semi-conservative model of DNA replication.

8.7 SUGGESTED READINGS:

1) Freifelder, D. *Molecular Biology* (1990) - Narosa Publishing House, New Delhi.
2) Gardner, E. J., Simmons, M. J. and Snustad, D. P. *Principles of Genetics* (2001) - John Wiley & Sons, Inc., New York.
3) Snustad, D. P., Simmons, M. J. and Jenkins, J. B. *Principles of Genetics* (1997) - John Wiley & Sons, Inc., New York.
4) Russel, P. J. *Genetics* (1998) - The Benjamin/Cummings Publishing Company, Inc., California.
5) Singh, B. D. *Fundamentals of Genetics* (2001) - Kalyani Publishers, New Delhi.

Prof. A. Amruthavalli

LESSON-9

DNA DAMAGE AND REPAIR

9.0 OBJECTIVE:

- To make the learner to understand about the types of DNA damages and the mechanisms for the repair of the damage.

STRUCTURE:

9.1 INTRODUCTION

Spontaneous and induced mutations constitute damage to the DNA of a cell or an organism. Especially with high doses of mutagens, the mutational damage can be considerable. Both prokaryotic and eukaryotic cells have a number of repair systems to deal with damage to DNA. All of these systems use enzymes to make the correction. Some of the systems directly correct the lesion while others first excise the lesion creating a single-stranded gap and then synthesize new DNA for the resulting gap. If the repair systems are unable to correct all of the lesions, the result is a mutant cell or, if too many mutations remain, death of the cell. Clearly, DNA repair systems are very important for the survival of the cell. The fact that the repair systems are not 100 percent efficient makes it possible to isolate mutants for study. At high doses of mutagens, repair systems are unable to correct all of the damage, and cell death can result.

9.2 TYPES OF DNA DAMAGES

9.2.1 Alkylation:

The transfer of an alkyl group, either a methyl or ethyl group, to reactive sites on the bases

or phosphates is referred to as alkylation. This reaction is catalyzed by some alkylating agents namely nitrogen mustard, sulphur mustard, methyl methanesulphonate, ethyl ethanesulphonate etc. This addition results in the base modification and thereby base pair potential. A particular reactive site is the oxygen of carbon 6 in the guanine. The alkylated guanine pairs with thymine and produces a GC-to-AT transition when the DNA is replicated. Alkylation damage like this can be removed by specific DNA repair enzymes. In this repair system the modified base is not removed from the DNA. In this case, an enzyme encoded by the ada gene, called O6-methylguanine methyltransferase, recognizes the O6-methylguanine in the DNA and removes the methyl group, thereby changing it back to its original form.

9.2.2 Deamination

Deamination is the removal of an amino group from a base. This is one of the most common chemical event that occurs to produce a spontaneous mutation. This deamination of cytosine to uracil (Fig. 9.1) can also be catalyzed by Nitrous acid (HNO_2), which is a very potent mutagen. For example, the deamination of cytosine produces uracil. Uracil is not a normal base in DNA, although it is a normal base in RNA. The deamination is a disastrous change because the deamination product, uracil, pairs with adenine rather than with guanine. This has two effects: (1) an incorrect base will appear in mRNA and (2) an adenine instead of a guanine will occur in newly replicated DNA strands. An intracellular repair system exists to remove most of the uracils produced by deamination of cytosine, thereby minimizing the mutational consequences of this event. However, if the uracil is not repaired, an adenine will be incorporated in the new DNA strand opposite it during replication. Ultimately this will result in the conversion of a CG base pair to a TA base pair i.e., a transition mutation.

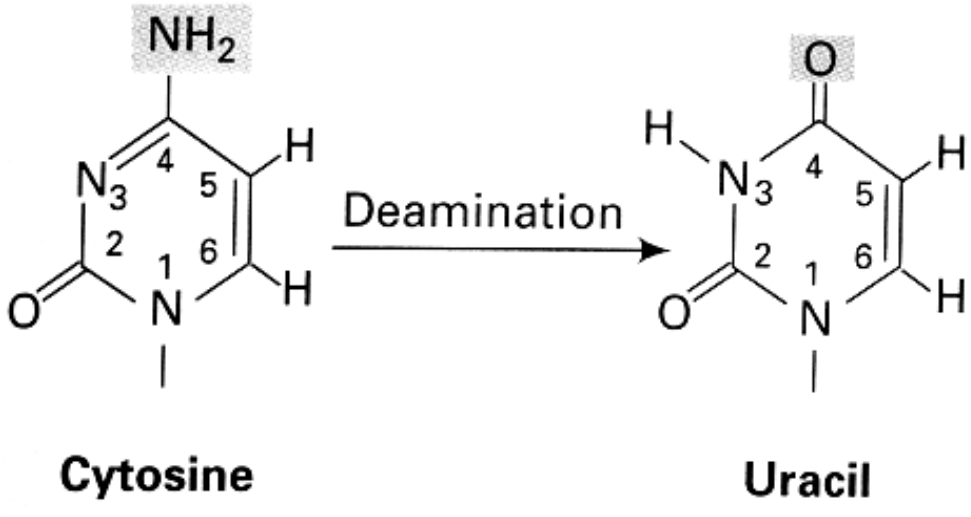


Fig-9.1: Deamination of Cytosine

In this case, an uracil-dependent DNA glycosylase enzyme detects the individual unnatural base and catalyze its removal from the deoxyribose sugar to which it is attached. This catalytic activity leaves a gap in the DNA where the base was removed. This hole is called an AP site (apurinic site or apyrimidinic site). The enzyme AP endonuclease recognizes the hole and cuts the DNA backbone on the 3′ side beside the missing base. This leaves a primer end from which DNA polymerase I initiates repair synthesis and subsequently the gap is filled followed by sealing of nick by DNA ligase.

9.2.3 Pyrimidine Dimers

Ultraviolet light (UV) rays are non-ionizing and they have insufficient energy to induce ionizations. However, UV is a useful mutagen and at high enough doses it can kill cells. In some applications, UV light is used as a sterilizing agent. Ultraviolet light causes mutations because the purine and pyrimidines in DNA absorb light very strongly in the UV range (254 to 260 nm). At this wavelength UV light induces gene mutations primarily by causing photochemical changes in the DNA. One of the effects of UV radiation is the formation of abnormal chemical bonds between adjacent pyrimidine molecules in the same strand. This bonding is induced mostly between adjacent thymines, forming the thymine dimers (Fig. 9.2), usually designated as T^T.

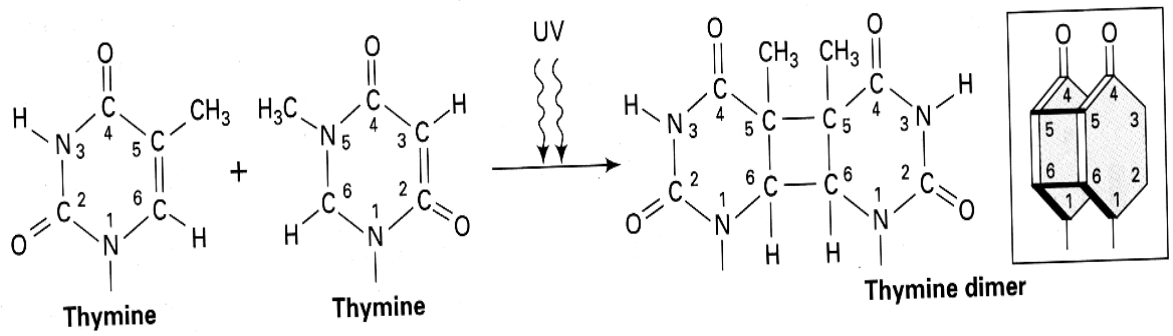


Fig-9.2: Formation of Thymine Dimer

Formation of a thymine dimer involves the disruption of double bonds between 4 and 5 carbon residues of the two thymine molecules as well as of the hydrogen bonds between these thymines and the corresponding adenine residues of the complementary DNA strand, and the formation of covalent bonds between the two 4 and the two 5 carbons of the two adjacent thymines. This unusual pairing produces a bulge in the DNA strand and disrupts the normal pairing of the Ts with the corresponding As on the opposite strand. So, the significant effects of the presence of thymine dimers are (1) the DNA helix becomes distorted as the thymines, which are in the same strand, are pulled toward one another and (2) as a result of distortion, hydrogen bonding to adenines in the opposing strand, and though possible is significantly weakened. This structural distortion causes inhibition of advance of the replication fork as the DNA polymerase is unable to catalyze replication of a DNA molecule in the region distorted by a pyrimidine dimer formation. This UV damage of DNA can be repaired by different mechanisms viz., photoreactivation, excision repair and post-replication recombination repair.

9.3 DNA DAMAGE REPAIR MECHANISMS

9.3.1 Light-Dependent Repair:

This light-dependent repair mechanism for DNA repair is also be called as 'Light repair' or 'Photoreactivation' (Fig. 9.3). This involves in the direct correction and repair of UV-light induced thymine dimmers. It is an enzymatic cleavage of thymine dimers. The dimers are reverted directly to the original form by exposure to visible light in the wavelength range of 320 to 370 nm. In E. coli, it is catalyzed by a light-activated enzyme called 'DNA photolyase' or 'Photolyase' or 'PR enzyme' which is encoded by the phr gene. The photolyase enzyme searches for and binds to the thymine dimers on the DNA, and uses light energy to

cleave the covalent cross-links between the thymine residues of the dimer. Photolyase will bind to thymine dimers in DNA in the dark, but it cannot catalyze cleavage of the bonds joining the thymine moieties without energy derived from visible light, specifically light within the blue region of the spectrum. Photolyase also splits cytosine dimers and cytosine-thymine dimers. Thus, when UV light is used to induce mutations in bacteria, the irradiated cells are grown in the dark for a few generations to maximize the mutation frequency.

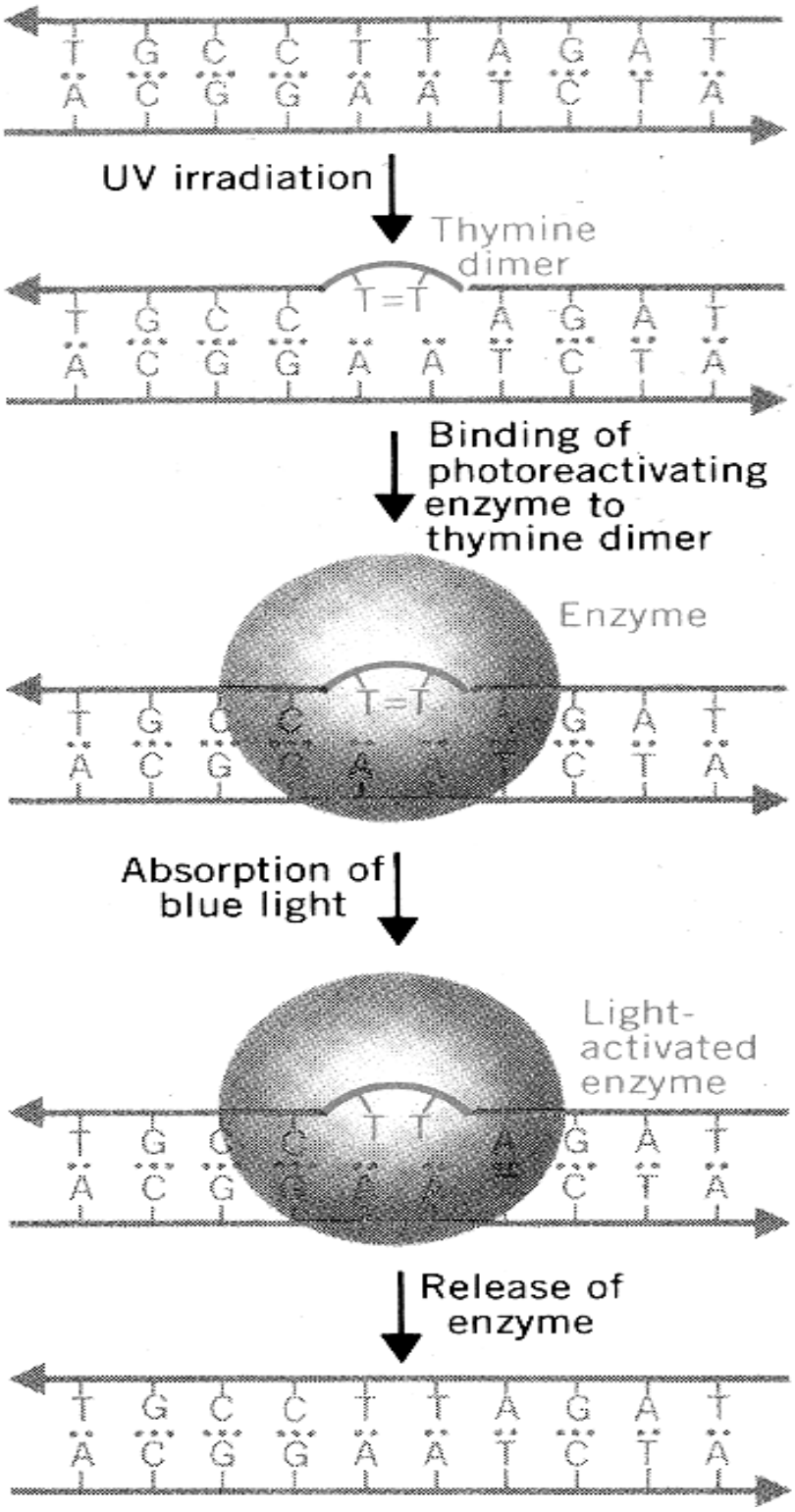


Fig-9.3: Photoreactivation Mechanism

9.3.2 Methyl directed Mismatch Repair

The DNA polymerases proofread the DNA strands during their synthesis, removing any mismatched nucleotides at the 3′ termini of growing strands. The mismatch repair pathway provides a backup to this replicative proofreading by correcting mismatched nucleotides remaining in DNA after replication. Many mismatched base pairs left after DNA replication may be corrected by this mechanism. Mismatches often involve the normal four bases in DNA. For example a T may be mispaired with a G. Because both T and G are normal components of DNA, mismatch repair systems need some way to determine whether the T or the G is the correct base at a given site. The repair system makes this distinction by identifying the template strand, which contains the original nucleotide sequence, and the newly synthesized strand, which contains the misincorporated base. This distinction can be made based on the pattern of methylation in newly replicated DNA. In E. coli, the A in palindromic GATC is usually methylated by the action of dam-methylase, an enzyme encoded by the dam gene. Thus, both A nucleotides in the DNA segment are methylated. However, after replication, the parental DNA strand has a methylated A nucleotide in the GATC sequence, while the A nucleotide in the GATC of the newly replicated DNA strand is not methylated until a short time after its synthesis. Therefore, for a short while after replication, the parental strand has methylated GATC sequence, while the new strand has an unmethylated GATC sequence, called hemimethylation. The mismatch repair system uses this difference in methylation state to excise the mismatched nucleotide in the nascent strand and replace it with the correct nucleotide by using the methylated parental strand of DNA as template (Fig.9.4).

In E. coli, the products of three genes namely mutS, mutL, and mutH are involved in the initial stages of mismatch repair. The first step in the repair process is the binding of the mutS-encoded protein, MutS, to the mismatch, which may be a single base pair mismatch. MutH and MutL proteins then join the complex of MutS and DNA. This results in the DNA looping which brings the unmethylated GATC sequence located about 1 to 2 kb away from the replication site close the mismatch. MutH contains a GATC-specific endonuclease activity that cleaves the unmethylated strand at hemimethylated GATC sites either 5′ or 3′ to the mismatch. The incision sites may be 1000 nucleotide pairs or more from the mismatch. The subsequent excision process requires MutS, MutL, DNA helicase II (Mut U or Uvr D), and an appropriate exonuclease. If the incision occurs at a GATC sequence 5′ to the mismatch, a 5′ → 3′ exonuclease like E. coli exonuclease VII is required. If the incision occurs 3′ to the mismatch, a 3′ → 5′ nuclease activity like that of E. coli exonuclease I is needed. After the excision process has removed the mismatched nucleotide from the unmethylated strand, DNA polymerase fills in the gap, and DNA ligase seals the nick. The homologs of the E. coli MutS and MutL proteins have been identified in both Saccharomyces cerevisiae and humans which indicate the presence of similar mismatch repair pathways occur in eukaryotes. In fact, mismatch excision has been demonstrated in vitro with nuclear extracts prepared from human cells. Thus mismatch repair is probably a universal or nearly universal mechanism for safeguarding the integrity of genetic information stored in double stranded DNA.

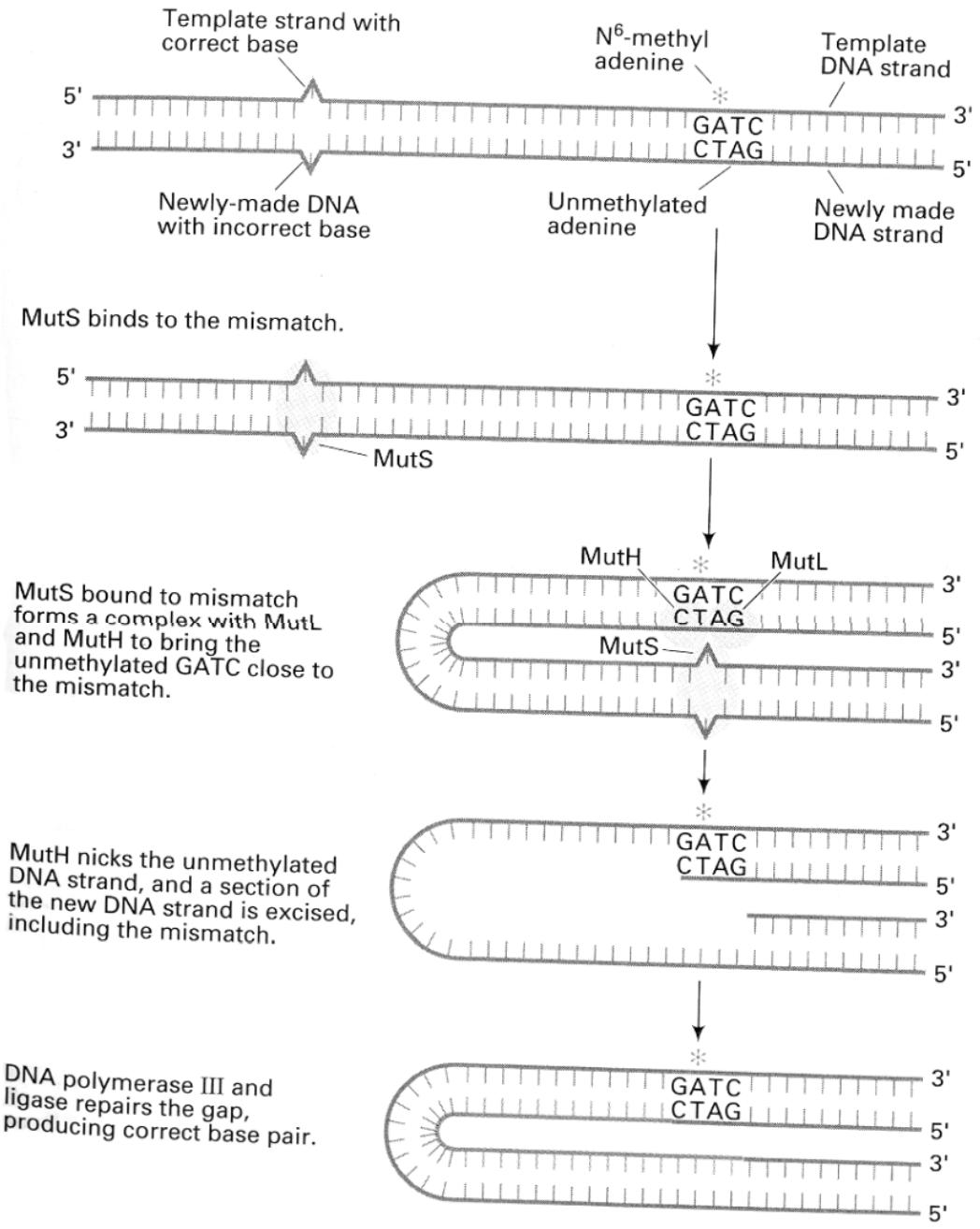


Fig-9.4: Mismatch Repair Mechanism

9.3.3 Nucleotide Excision Repair

The nucleotide excision repair pathways remove larger defects like thymine dimmers and bases with bulky side-groups from DNA. This pathway is operative in the dark, and occurs by very similar mechanisms in E. coli and humans. In nucleotide excision repair, a unique excision nuclease activity produces cuts on either side of the damaged nucleotides and excises an oligonucleotide containing the damaged bases. This nuclease is called as 'excinuclease' to distinguish it from the endonucleases and exonucleases that play other roles in DNA metabolism.

In E. coli, excinuclease activity requires the products of three genes viz., uvrA, uvrB, and uvrC. A trimeric protein containing two UvrA polypeptides and one UvrB polypeptide recognizes the defect in DNA, binds to it, and uses energy from ATP to bend the DNA at the damaged site. The UvrA dimer is then released, and UvrC protein binds to the UvrB-DNA complex. The UvrB protein cleaves the fifth phosphodiester bond from the damaged nucleotides on the 3′ side, and the UvrC protein hydrolyzes the eight phosphodiester linkage from the damage on the 5′ side. The uvrD gene product, DNA helicase II, releases the excised oligomer. In the last steps of the pathway, DNA polymerase I replaces UvrB protein and fill in the gap using the complementary strand as template followed by the sealing of the nick left by polymerase by DNA ligase.

Nucleotide excision repair mechanism in humans occurs through a pathway similar to the one in E. coli, but it involves about four times as many proteins. In humans, the excinuclease activity requires at least 17 polypeptides. Protein XPA (for xerodermapigmentosum protein A) recognizes and binds to the damaged nucleotides in DNA. It then recruits the other proteins required for excinuclease activity. In humans, the excised oligomer is 29 nucleotides long rather than the 12-mer removed in E. coli. The gap is filled by either DNA polymerase δ or ε in humans, and DNA ligase completes the job.

9.3.4 Post-Replication Recombination Repair

In E. coli, light-dependent repair, excision repair, and mismatch repair can be eliminated by mutations in the phr gene, uvr gene and mut gene, respectively. In multiple mutants deficient in these repair mechanisms, still another DNA repair system called Post-replication repair is operative. When a DNA polymerase III encounters a thymine dimer in a template strand, its progress is blocked. DNA polymerase restarts DNA synthesis at some position past the dimer, leaving a gap in the nascent strand opposite the dimer in the template strand. At this point, the original nucleotide sequence has been lost from both strands of this progeny double helix. The damaged DNA molecule is repaired by a recombination-dependent repair process mediated by the E. coli recA gene product. The RecA protein, which is required for homologous recombination, stimulates the exchange of single strands between homologous double helices. During post-replication repair, the RecA protein binds to the single strand of DNA at the gap and mediates pairing with the homologous segment of the sister double helix. The gap opposite the dimer is filled with the homologous DNA strand from the sister DNA molecule. The resulting gap in the sister double helix is filled in by DNA polymerase, and the nick is sealed by DNA ligase. The thymine dimer remains in the template strand of the original progeny DNA molecule, but the complementary strand is now intact. If the thymine dimer is not removed by the nucleotide excision repair system, this post-replication repair must be repeated after each round of DNA replication.

The mechanism of this repair is given in figure 9.5 below.

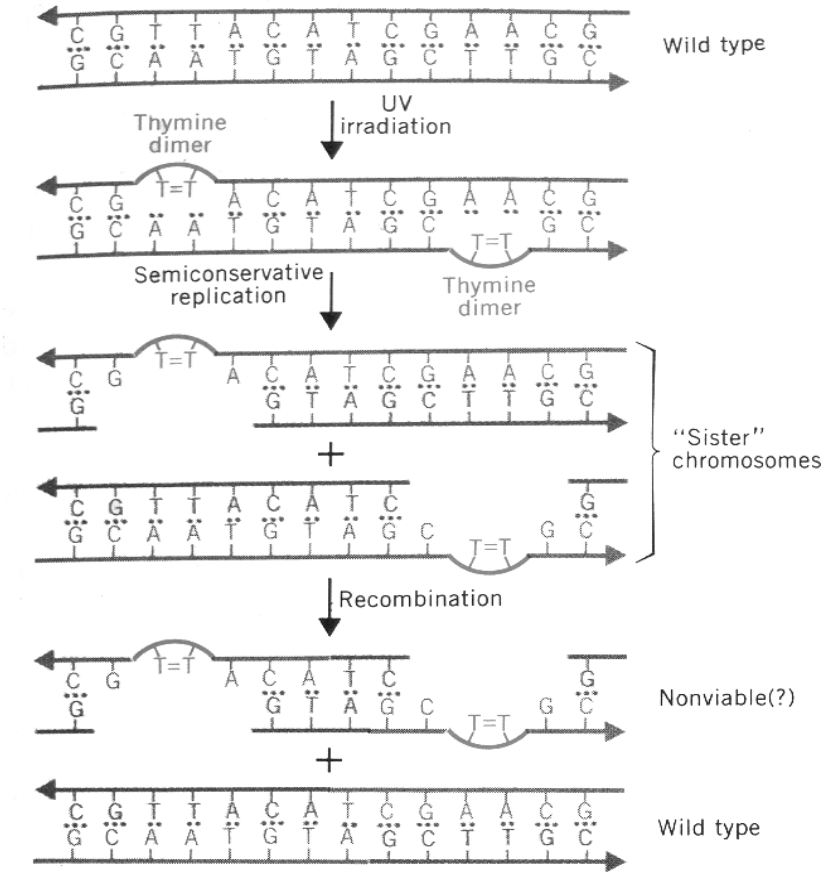


Fig-9.5: Post-Replication Recombination Repair

9.3.5 SOS Repair:

When the DNA of *E. coli* cells is heavily damaged by mutagenic agents such as UV light, the cells take some drastic steps in their attempt to survive. They go through a so-called SOS response, during which a whole battery of DNA repair, recombination, and replication proteins are synthesized. SOS repair includes a bypass system that allows DNA chain growth across damaged segments at the cost of fidelity of replication. It is an error-prone process; even though intact DNA strands are formed, the strands contain incorrect bases. The principle involved is that survival with mutations is better than no survival at all. The SOS response appears to be a somewhat desperate and risky attempt to escape the lethal effects of heavily damaged DNA. When the error-prone repair system is operative, mutation rates increase sharply. SOS repair is thought to invoke a relaxation of the editing system in order to allow polymerization to proceed across a dimer despite the distortion of the helix.

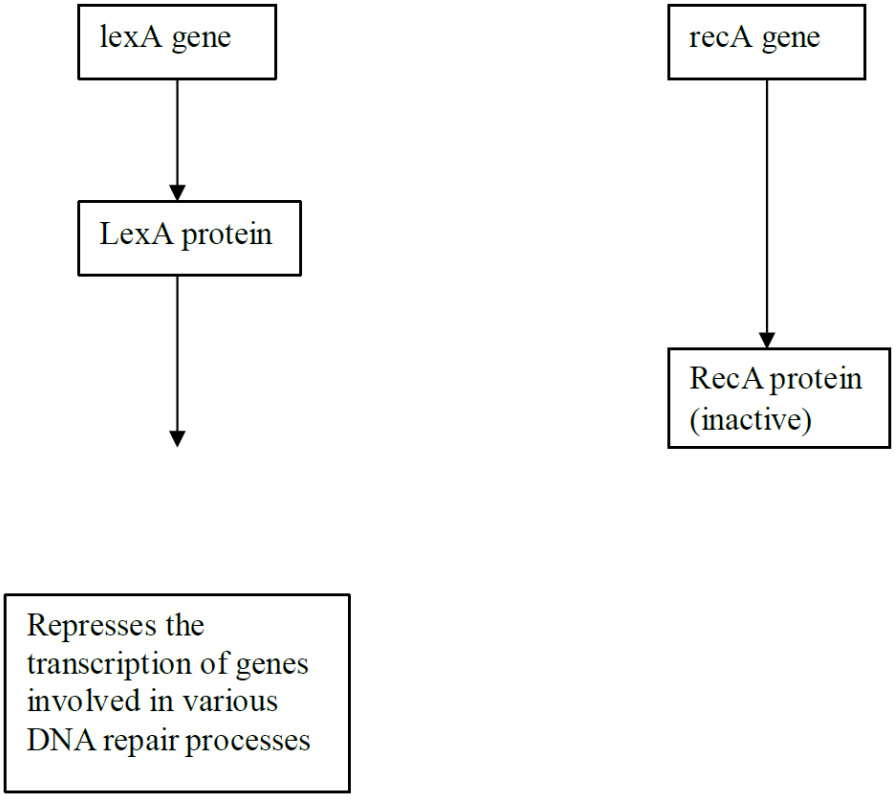
The SOS response has been studied best in E. coli, where the synthesis of many of the enzymes involved in the repair of DNA damage is regulated by the SOS system. Two genes are key to controlling the SOS system: lex A and recA. The E. coli cells with mutant recA and lexA genes have their SOS response permanently turned on. In the uninduced state, when there is no DNA damage, the lexA encoded protein, LexA functions as a repressor to prevent transcription of about 17 genes whose protein products are involved in the repair of various kinds of DNA damages, including excision repair of gaps. All of the genes involved have in common a 20-nucleotide regulatory sequence called the SOS box.

The recA encoded protein, RecA, is a regulator of the induction of the SOS response. When there is sufficient damage of DNA, somehow the RecA protein becomes activated, perhaps by binding to single-stranded DNA. The activated RecA stimulates the LexA protein to cleave itself and this relieves the repression of the DNA repair genes. As a result, the DNA repair genes are transcribed, and DNA repair proceeds. After the DNA damage is dealt with, RecA again becomes inactivated, and newly synthesized LexA protein again acts to repress the DNA repair genes.

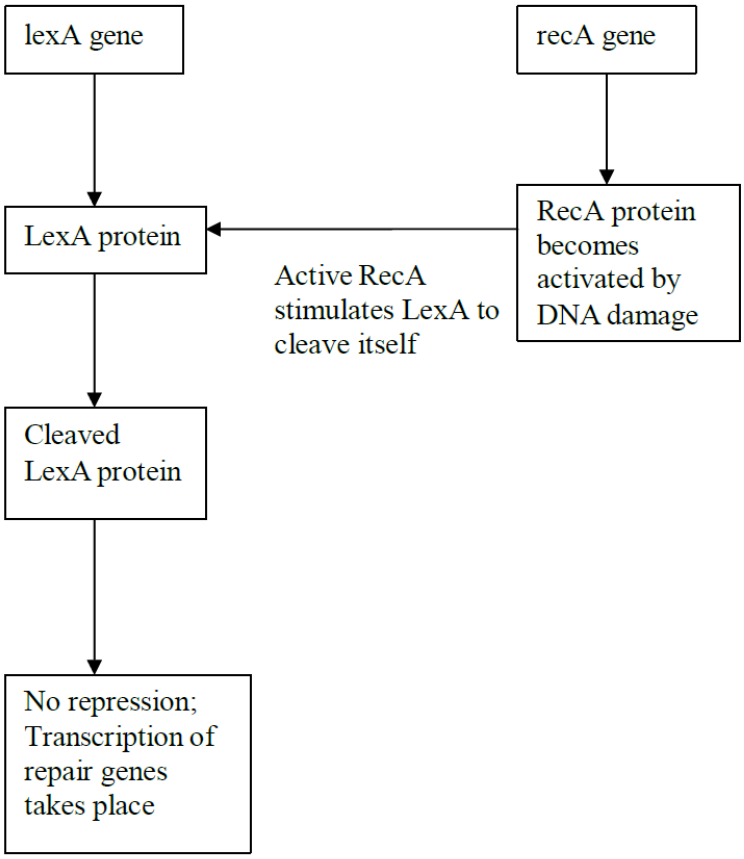
The role of the RecA product is two fold. One of these is its regulatory function and the other is a direct effect on editing. The RecA protein binds tightly to single-stranded DNA but only very weakly to double-stranded DNA. The distortion resulting from a pyrimidine dimer produces a short stable single stranded region to which RecA binds. When DNA polymerase III encounters a dimer site to which RecA is bound, RecA interacts with the ε subunit of the polymerase, which is responsible for editing, and inhibits the editing function. As a result, the replication fork advances. The presence of RecA at the dimer site inhibits editing and causes the mispaired base to remain in the daughter strand as a mutation.

SOS repair also requires two other genes, umuC and umuD. The role of the gene products is not clearly known but three hypotheses have been suggested: (1) they facilitate tight binding of RecA at the small distortion, (2) they facilitate binding of pol III to be distorted region, or (3) they enable the pol III to leave the damaged site either by releasing pol III from RecA or RecA from the DNA.

Un-Induced State



Induced State



9.4 SUMMARY:

Genetic damage can occur to the DNA spontaneously, through replication errors, or through treatment with radiation or chemical mutagens. If the genetic damage is not repaired mutations will result, and if there has been too much damage, cell death may result. Damage through alkylation, deamination and UV irradiation are quite common. Cells possess a number of repair mechanisms that function to correct at least some damage to DNA. These repair mechanisms include: (1) repair by DNA polymerase proofreading, in which a base pair mismatch in DNA being synthesized is immediately repaired by 3′ → 5′ excision; (2) photo reactivation of pyrimidine dimers induced by UV light; (3) excision repair, in which pyrimidine dimers and other DNA damage that distorts the DNA helix are excised and replaced with new DNA; (4) repair of damaged bases by glycolysaes and AP endonuclease; and (5) repair by mismatch correction in which the methylation state of a DNA sequence signals which DNA strand is newly synthesized so that the mismatched base on that strand is corrected. Any DNA damage that is not repaired may result in a mutation and may have the potential to be lethal to the cell. The collective array of repair enzymes, then, serves to reduce mutation rates for spontaneous errors by several orders of magnitude.

9.5 TECHNICAL TERMS:

Alkylation, Deamination, Photo reactivation, Mismatch repair, Recombination repair, SOS repair.

9.6 SELF ASSESSMENT QUESTIONS:

1) What are the thymine dimers. Explain the mechanisms for the removal of thymine dimers.
2) Describe the various mechanisms of DNA damage repair systems.
3) Explain the methyl mediated mismatch repair of DNA damage.
4) Write in detail about the SOS response.

9.7 SUGGESTED READINGS:

1) Freifelder, D. Molecular Biology (1990) - Narosa Publishing House, New Delhi.
2) Gardner, E.J., Simmons, M.J. and Snustad, D.P. Principles of Genetics (2001) - John Wiley & Sons, Inc., New York
3) Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4) Russel, P.J. Genetics (1998) - The Benjamin/Cummings Publishing Company, Inc., California
5) Singh, B.D. Fundamentals of Genetics (2001) - Kalyani Publishers, New Delhi.

Prof. A. Amruthavalli

LESSON-10

GENE EXPRESSION

10.0 OBJECTIVE:

- The purpose of this lesson plan is to make the reader very clear about the nature of gene expression in prokaryotes.

STRUCTURE:

10.1 INTRODUCTION

Crick, in 1958, proposed that the information present in DNA in the form of base sequence is transferred to RNA and then from RNA it is transferred to protein in the form of amino acid sequence, and that this information does not flow in the reverse direction, that is, from protein to RNA to DNA. The DNA molecules provide the information for their own replication. This relationship between DNA, RNA and protein molecules is known as the **Central Dogma**. The structure, function, development and reproduction of an organism depend on the properties of the proteins present in each cell and tissue. A protein consists of one or more chains of amino acids. Each chain of amino acids is called a polypeptide. The sequence of amino acids in a polypeptide chain is coded for by a gene that is a specific base pair sequence in DNA. When a protein is needed in the cell, the genetic code for that protein's amino acid sequence must be read from the DNA and processed into the finished protein. Two major steps occur in the process of protein synthesis – **transcription** and **translation**.

Transcription is the transfer of information from a double-stranded, template DNA molecule to a single-stranded RNA molecule. Translation or protein synthesis is the

conversion, in the cell, of the messenger RNA (mRNA) base sequence information into the amino acid sequence of a polypeptide. Unlike the DNA replication, transcription and translation generally occur throughout the cell cycle.

10.2 BACTERIAL RNA POLYMERASE

The RNA polymerase is similar in most bacteria but has a different structure in cyanobacteria and archa ebacteria. In most bacteria, a singletype of enzymeis responsible for the transcription of all the various types of RNA, e.g., mRNA, rRNA and tRNA. The *E. coli* complete RNA polymerase enzyme is capable of recognizing the appropriate promoter site for initiating the transcription and to continue the synthesis of RNA using DNA as template. The complete enzyme is known as the 'holo enzyme'. The holoenzyme consists of two components namely the 'core enzyme' and the 'sigma factor'. The holoenzyme is composed of different subunits plus sigma factor and is symbolized as $\alpha_2\beta\beta'\sigma$. The core enzyme cannot initiate transcription at the proper sites, but it can synthesize RNA using DNA as a template. The core RNA polymerase consists of four polypeptides which are of three types –

- The α subunit- this polypeptide is coded by rpoA gene. It is 40,000 daltons in molecular weight and present in two copies / core enzyme molecule. It functions in binding with promoter of DNA.
- The β subunit - this subunit is coded by rpoB gene. It is 1,55,000 daltons in molecular weight and present in a single copy/enzyme molecule. It is involved in binding with the incoming nucleotides for RNA synthesis.
- The β′ subunit - the β′ polypeptide is coded by rpoC and 1,60,000 daltons in molecular weight. It is involved in binding with the template DNA i.e., the single stranded stretch of DNA generated to allow the transcription to proceed.

Each subunit contributes to the function of the core enzyme as a whole, which should have at least the following four distinct functional sites viz., a DNA unwinding site, the site binding to the antisense strand, a site binds to the sense strand and a DNA rewinding site.

The sigma factor is involved in binding of the RNA polymerase to the promoter DNA and possibly with the transcription initiation. It is not involved in the RNA synthesis as it is released when the RNA chain reaches to 8-9 bases, the core enzyme then continues transcription. The *E. coli* sigma factor (σ^{70}) is a single polypeptide of 85,000 daltons and is coded by the rpoD gene. The chief function of the sigma factor is to ensure the stable binding of RNA polymerase to only at the promoters. The binding of RNA polymerase to any DNA segment is called loose binding. The association of σ with the core enzyme drastically reduces its ability for loose binding with DNA.

The association of core enzyme with σ factor is a critical factor in making the holoenzyme capable of recognizing the promoter binding sites for transcription initiation. When the holoenzyme forms a tight binding complex with DNA, σ factor contacts the DNA upstream of the start point. In the case of tight binding, a short stretch of the DNA associated with the holoenzyme becomes single-stranded and such complexes are very stable. The loose binding of the enzyme yields a closed binary complex in which the DNA is present as normal duplex in association with the enzyme. In contrast, the tight binding of the enzyme gives rise to the open binary complex in which a part of the DNA associated with the enzyme is single stranded. When RNA synthesis begins, the open binary complex is converted into a ternary complex since the enzyme is now complexed with a DNA-RNA hybrid segment as well.

10.3 TRANSCRIPTION

Transcription begins with the attachment of RNA polymerase holoenzyme with the promoter region of transcription unit and comes to an end when the core enzyme reaches the terminator site and dissociates from the DNA. The entire process may be divided into three steps viz., Initiation, Elongation and Termination. Mechanism of transcription is given in the figure 10.1 below.

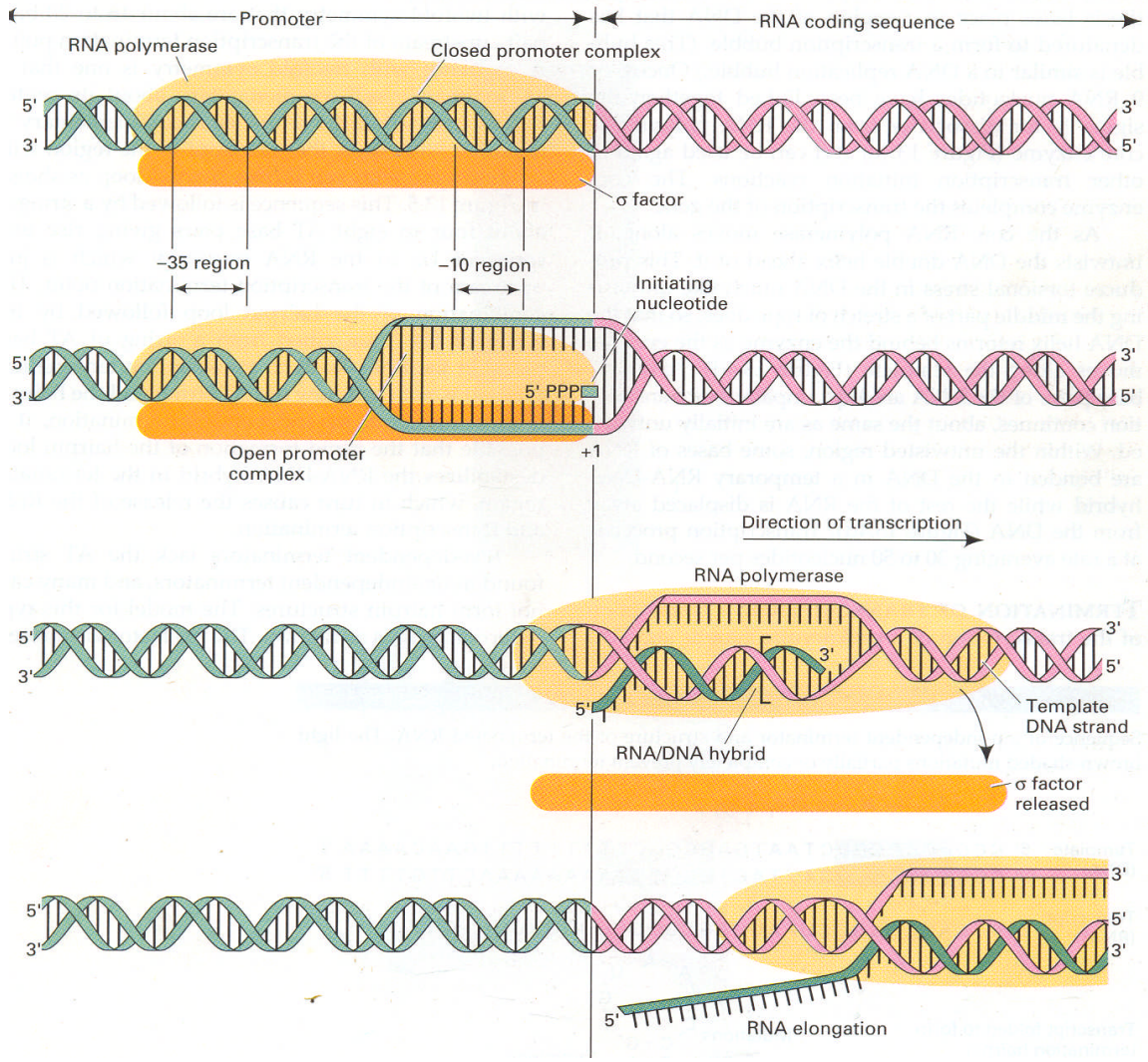


Fig-10.1: Mechanism of Transcription

10.3.1 Initiation of RNA Chain

Initiation of RNA chain involves three steps - (1) binding of the RNA polymerase holoenzyme to a promoter region in DNA, (2) the localized unwinding of the two strands of DNA by RNA polymerase, providing a template strand free to base pair with incoming ribonucleotides, and (3) the formation of phosphodiester bonds between the first few ribonucleotides in the nascent RNA chain. Initially the enzyme binds to the duplex DNA of the promoter region forming the closed binary complex. The sigma sub unit of RNA polymerase mediates its binding to promoter in DNA. Two DNA sequences in most promoters of *E. coli* genes are considered to be critical in specifying the initiation of transcription. These consensus sequences are generally called as–35 sequence and –10

sequence as they are centered at 35 and 10 base pairs upstream from the base pairat which transcription starts. The –10 consensus sequence in the non-template strand is **TATAAT** and the –35 consensus sequence is **TTGACA**. The sigma subunit initially recognizes and binds to the –35 sequence and this sequence is sometimes called the recognition sequence. The AT-rich –10 sequence facilitates the localized unwinding of DNA, which is an essential prerequisite to the synthesis of a new RNA chain. The formation of closed binary complex induces strand separation in the stretch of DNA duplex bound to the enzyme. The melted region is <17 bp long and begins from the midpoint of the –10 hexamer consensus sequence of the promoter, and extends just beyond the start point so that the latter becomes available for transcription initiation. The complex of the melted DNA with the enzyme is known as the open binary complex. Once the template strand becomes available, the enzyme begins to incorporate the RNA nucleotides beginning at the start point. As in the case of DNA replication, the RNA synthesis also proceeds in the 5′ → 3′ direction and the template DNA strand must be oriented as3′ → 5′.

10.3.2 Elongation of the RNA Chain

RNA synthesis takes place in a region of the DNA that has denatured to form a transcription bubble. Once 8 or 9 RNA nucleotides have been linked together, the sigma factor dissociates from the RNA polymerase core enzyme and can be used again in other transcription initiation reactions. The core enzyme completes the transcription of the gene. The covalent extension of RNA chains takes place within the transcription bubble, a locally unwound segment of DNA. The RNA polymerase molecule contains both DNA unwinding and DNA rewinding activities. RNA polymerase continuously unwinds the DNA double helix ahead of the polymerization site and rewinds the complementary DNA strands behind the polymerization siteas it moves along the double helix. In *E. coli*, the average length of a transcription bubble is 18 nucleotide pairs, and about 40 ribonucleotides moves along the DNA molecule. Within the untwisted region, some bases of RNA are bound to the DNA in a temporary RNA-DNA hybrid while the rest of the RNA is displaced away from the DNA.

10.3.3 Termination of RNA Chain

Termination of the transcription of a prokaryotic gene is signaled by controlling elements called terminators. When the core enzyme reaches the terminator site, there is no further addition of ribonucleotides to the RNA chain. The RNA molecule dissociates from the template DNA strand, and the core enzyme frees itself from the template DNA molecule. The DNA strands in the melted region re-associate to form normal double helix and the process of transcription catalyzed by these polymerase molecules comes to an end. The released core enzyme will soon bind to a segment of DNA irrespective of the base sequence, and would begin the search for a promoter once the sigma factor binds to it. Basing on the requirement of a polypeptide or protein namely ‘rho-factor’, the terminators are divided into two groups – (1) simple or rho-independent terminators, and (2) rho-dependent terminators or type II terminators.

Rho-independent terminators consist of sequences with twofold symmetry that are about 16 to 20 base pairs upstream of the transcription termination point. A sequence with twofold symmetry is one that is approximately self-complementary about its center, that is, one half of the sequence is complementary to the other half. Thus, the transcript of the region with twofold symmetry can form a hairpin loop. This sequence is followed by a string of about four to eight AT base pairs giving rise to a series of Us in the RNA transcript which is

just upstream of the transcription termination point. The combination of the hairpin loop followed by the string of Us in the RNA leads to transcription termination. the rapid formation of the hairpin loop destabilizes the RNA-DNA hybrid in the terminator region, which in turn causes the release of the RNA and transcription termination.

Rho-dependent terminators lack the AT string found in rho-independent terminators, and many cannot form hairpin structures. The rho-factor is a protein with two domains – one domain binds to RNA, and the other domain binds to ATP. For transcription termination, rho-factor first binds to ATP, and is activated by it. The activated rho then binds to the RNA transcript at some recognition site located upstream of the termination site. It then moves along the RNA till it catches the RNA polymerase. When the rho-factor catches the RNA polymerase at the terminator site, it leads to unwinding of the RNA-DNA hybrid and interacts with the RNA polymerase itself which leads to the release of the RNA transcript, the RNA polymerase and the rho-factor from the DNA template bringing the termination of transcription.

10.4 TRANSLATION

The translation process or the protein synthesis can be divided into three stages – (1) polypeptide chain initiation, (2) chain elongation, and (3) chain termination. The main features of the initiation step are binding of mRNA to the ribosome, selection of the initiation codon, and binding of the charged tRNA bearing the first amino acid (fMet or Met). In the elongation stage there are two processes – joining together two amino acids by peptide formation and moving the mRNA and the ribosome with respect to one another in order that each codon can be translated successively. This movement is called translocation. In the termination stage the completed protein is dissociated from the synthetic machinery and the ribosomes are released to begin another cycle of synthesis.

10.4.1 Initiation of polypeptide chain

Initiation comprises all the events that precede the formation of the first peptide bond. In prokaryotes, the initiation complex is formed when the initiation codon, AUG, of them RNA binds to the P site of the 30S subunit of a ribosome, and at RNA^{met} molecule carrying formyl methionine base pairs with the codon AUG. I^fn *E. coli*, the formation of initiation complex requires a number of initiation factors, viz., GTP, and the factors IF1, IF2, and IF3. In addition to the AUG initiation codon, the initiation of protein synthesis in prokaryotes requires other information coded in the base sequence of an mRNA molecule upstream to the initiation codon. These sequences serve to align the ribosome on the message in the proper reading frame so that polypeptide synthesis can proceed correctly. Most of these binding sequences are purine-rich and are about 8 to 12 nucleotides upstream from the initiation codon. Evidence from the work of John Shine and Lynn Dalgarno indicates that this purine-rich sequence, and other nucleotides in this region, are complementary to a pyrimidine rich region at the 3′ end of 16S rRNA. The mRNA region that binds in this way is known as *Shine-Dalgarno* sequence. The formation of complementary base pairs between the mRNA and 16S rRNA allows the ribosome to locate the true sequence in the mRNA for the initiation of protein synthesis.

In *E. coli*, translation commences with the formation of a 30S initiation complex which involves the recognition of the AUG initiation codon by the small ribosome subunit. At the beginning of this process, three protein initiation factors, IF1, IF2, and IF3, are bound to the 30S ribosomal sub unit along with a molecule of GTP. The fMet-tRNA.fMet and them

RNA then attach to the 30S-IF-GTP complex to form the 30S initiation complex. IF3 is released as a result of this process. Next, the 50S subunit binds, leading to GTP hydrolysis and the release of IF1 and IF2. The final complex is called the 70S initiation complex (Fig. 10.2). All the three IF molecules are recycled for use in other initiation reactions. the 70S ribosome has two binding sites for aminoacyl-tRNA, the peptidyl (P), and aminoacyl (A) sites and the fMet-tRNA.fMet is bound to the mRNA in the P site.

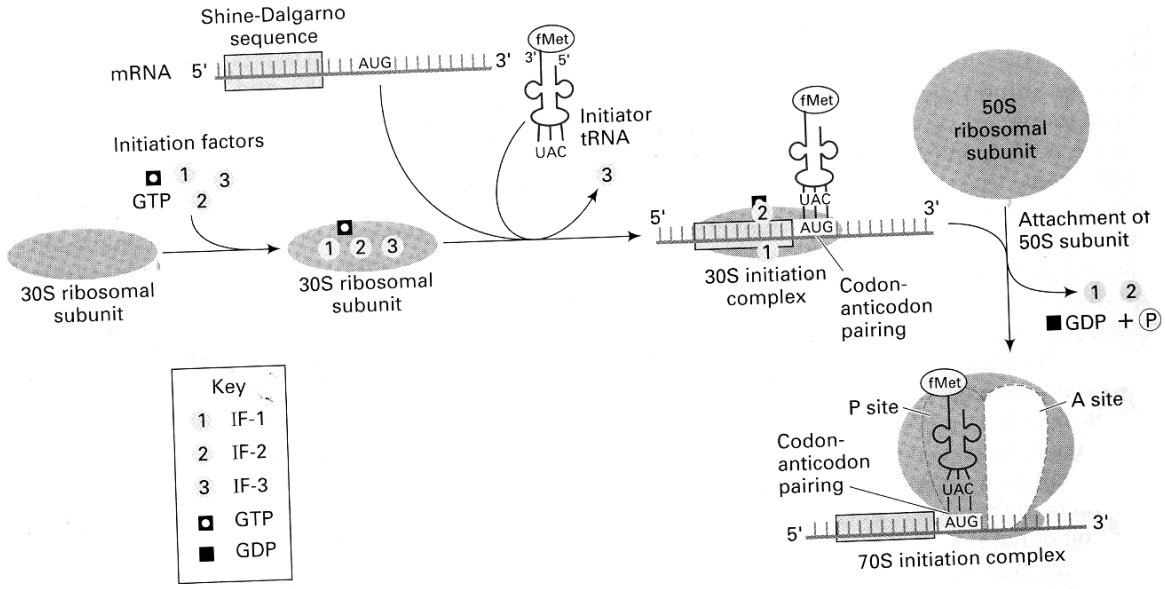


Fig-10.2: Formation of Final 70S Initiation Complex

10.4.2 Elongation of polypeptide chain

After initiation, the elongation phase of translation (Fig.10.3) begins. This phase has three steps–

(1) The binding of aminoacyl-tRNA (charged tRNA) to the ribosome

(2) The formation of a peptide bond

(3) The movement (translocation) of the ribosome along them RNA, one codon at a time

At the start of the elongation phase in prokaryotes, the fMet-tRNA.fMet is hydrogen-bonded to the AUG initiation codon in the peptidyl (P) site of the ribosome in the 70S initiation complex. FMet-tRNA.fMet is the only tRNA to bind before the complete ribosome is formed. The orientation of this tRNA-codon complex exposes the next codon in them RNA in the aminoacyl (A) site. Next, the appropriate aminoacyl-tRNA binds to the exposed mRNA codon in the A site. This aminoacyl-tRNA is brought to the ribosome complexed with the protein elongation factor EF-Tu and a molecule of GTP. When the amino acyl-tRNA binds to the codon in the A site, the GTP is hydrolyzed and EF-Tu is released bound to the GDP produced. This EF-Tu is recycled. First, a second elongation factor, EF-Ts, binds to EF-Tu and displaces the GDP. Next, GTP binds to the EF-Tu—EF-Ts complex to produce an Ef-Tu-GTP complex simultaneously with the release of EF-Ts. The aminoacyl-tRNA binds to the EF-Tu-GTP, and that complex can then bind to the A site in the ribosome when the appropriate codon is exposed.

When ribosome maintains the two amino acyl-tRNAs in the correct positions (one in P site and another in A site), a peptide bond formation occurs between the two amino acids. The peptide bond formation occurs in two steps. The first step is the breakage of the bond between the carboxyl group of the amino acid and the tRNA in the P site. The second step is the formation of the peptide bond between the now-freed amino acid and the amino acid attached to the tRNA in the A site. This reaction is catalyzed by peptidyl transferase. Once the peptide bond has formed, a tRNA without an attached amino acid (an uncharged tRNA) is left in the P site. The tRNA in the A site, now called peptidyl-tRNA, has the first two amino acids of the polypeptide chain attached to it.

The last step in the elongation cycle is 'translocation'. Once the peptide bond is formed and the growing polypeptide chain is on the tRNA in the A site, the ribosome moves one codon along the mRNA toward the 3′ end. In prokaryotes, the translocation requires the activity of another protein elongation factor, EF-G. An EF-G-GTP complex binds to the ribosome, and translocation then takes place along with movement of the uncharged tRNA away from the P site. The uncharged tRNA moves from the P site and then binds to the Exit site or E site. After translocation, the EF-G is then released in a reaction requiring GTP hydrolysis; EF-G can then be reused. During the translocation step the peptidyl-tRNA remains attached to its codon on the mRNA. And since the ribosome has moved, the peptidyl-tRNA is now located in the P site. The exact mechanism for the physical translocation of the ribosome is not known. After the translocation is completed, the A site is vacant. An amino acyl-tRNA with the correct anti codon binds to the newly exposed codon on the mRNA in the A site. The whole process is repeated until translation terminates at a stop codon.

In both prokaryotes and eukaryotes, once the ribosome moves away from the initiation site on the mRNA, the initiation site is open for another initiation event to occur. Thus many ribosomes may simultaneously be translating each mRNA. The complex between an mRNA molecule and all the ribosomes that are translating it simultaneously is called a polyribosome or polysome. An average mRNA may have eight to ten ribosomes synthesizing protein from it. Simultaneous translation enables a large amount of protein to be produced from each mRNA molecule.

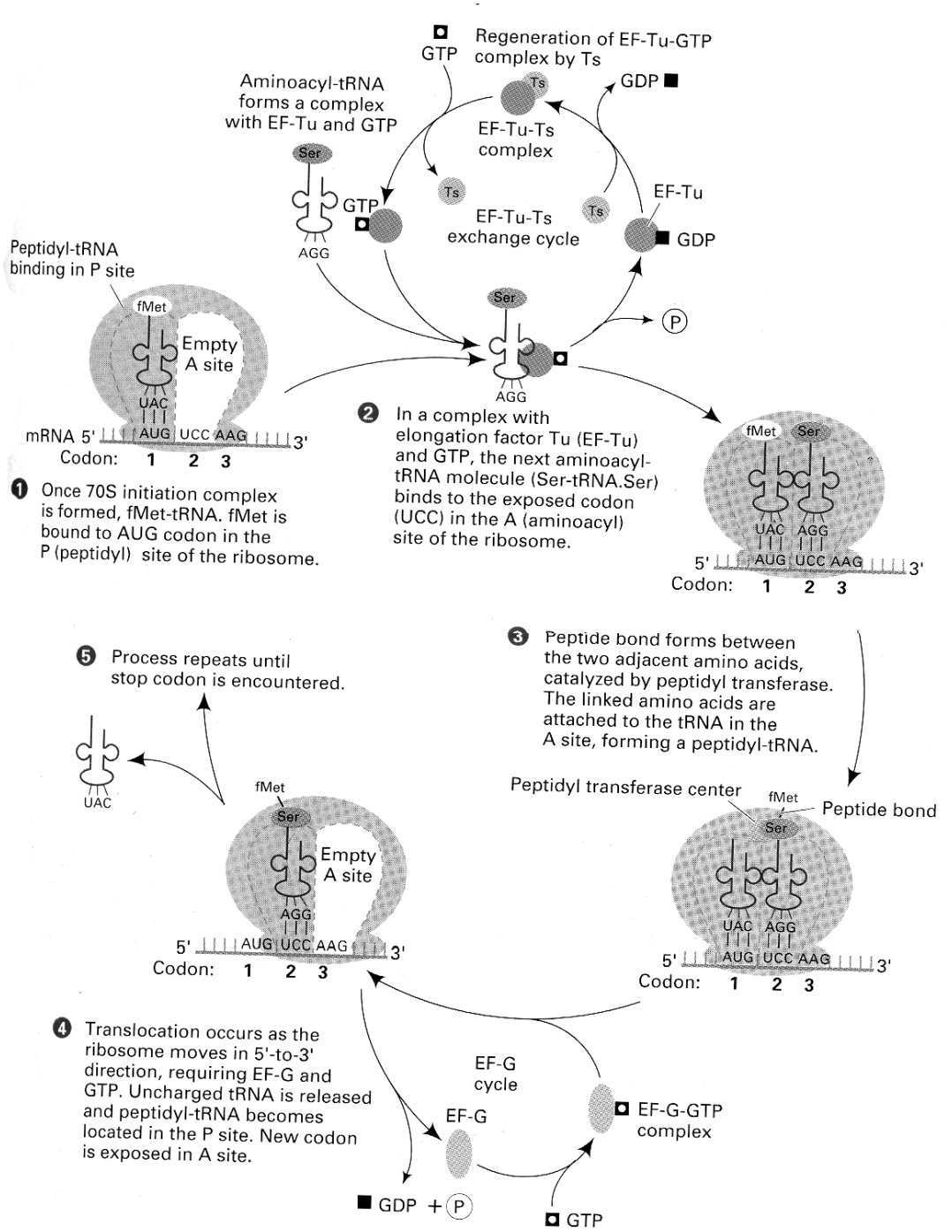


Fig-10.3: Elongation and Translocation steps during Translation

10.4.3 Termination of Translation

Elongation continues until the polypeptide coded for in the mRNA is completed. The end of a polypeptide chain is signaled by one of the three stop codons, UAG, UAA, and UGA, which are the same in both prokaryotes and eukaryotes. The stop codons do not code for any amino acid, and so no tRNAs in the cell have anti codons for them. The ribosome recognizes a chain termination codon only with the help of proteins called termination factors or release factors (RF), which read the chain termination codons, and then initiate a series of specific termination events.

The *E. coli* has three RFs – RF1, RF2, and RF3 – and each is a single polypeptide. Factor RF1 recognizes UAA and UAG, while RF2 recognizes UAA and UGA. Thus these two release factors have overlapping specificity in codon recognition. Factor RF3, which does not recognize any of the stop codons, stimulates the termination events. The specific termination events triggered by the release factors are: (1) release of the polypeptide from the tRNA in the P site of the ribosome in are action catalyzed by peptidyl transferase, and then (2) release of the tRNA from the ribosome and dissociation of the two ribosomal sub units from them RNA. The initiating amino acid, formyl methionine is usually cleaved from the completed polypeptide.

10.5 COMPARISON OF PROTEIN SYNTHESIS BETWEEN PROKARYOTES AND EUKARYOTES

Characteristic	Prokaryotes	Eukaryotes
RNA Polymerase		
Types	One in each species	Three; RNA polymerase I, II and III
Molecular weight	4,80,000 daltons	5,00,000 daltons
Subunits	$\alpha_2\beta\beta'\sigma$	2 large sub units +< 10 smaller subunits
Transcription factors	Not known	Many different types of transcription factors
Promoter	A simpler and relatively Smaller sequence	A relatively larger sequence and Has variable modules
Enhancer	A part of the promoter may act in a manner similar to enhancer	Present at variable distances from the promoter
Transcription initiation	The holo enzyme binds to initiates promoter and transcription	Transcription factors first bind to promoter, then RNA polymerase associates with them and initiates
Transcription Complex		
Composition	Core RNA polymerase +σ factor	RNA polymerase + transcription factors
Separation of components	σ factor dissociates from the core enzyme after initiation	Transcription factors dissociate only when transcription is complete

mRNA	Polycistronic RNA transcripts In general, there is no processing after transcription.	Monocistronic RNA transcripts Mature mRNA molecules are processed from longer precursor hnRNA molecues.
	Very short life of 2-5 minutes; degradation may begin even before transcription is complete.	Much longer life of ~ 6hr or more; mRNA in seeds may survive more than 100 years
	Poly–(A) tail not found.	Most mRNA's have 100-200 adenine nucleotides at the 3′-end; producing a poly-(A) tail
	5′cap is not present.	Partiallymethylated5′cap present.
	Shine-Dalgarno consensus Present at the 5′ end.	Shine-Dalgarno consensus is absent.
rRNA and ribosomes	Ribosomes of 70S; dissociate into 30S and 50S subunits.	Ribosomes of 80S; dissociates into 40S and 60S subunits.
	Ribosomes free in the cytoplasm.	Ribosomes attached to endoplasmic reticulum as well as free in the cytoplasm.
	rRNA of 16S (30S subunit), 23S and 5S (50S subunit)	rRNA is 18S (40S subunit),28S, 7.8S and 5S (60S subunit)
Translation	Translation often simultaneous with transcription.	Translation is not simultaneous with transcription.
	Translation occurs while mRNA is attached to the chromosome.	mRNA has to be processed and transported out of the nucleus as translation occurs mostly in the cytoplasm.
	Translation initiation involves base-pairing between 16S rRNA and mRNA in the Shine-Dalgarno consensus sequence.	Translation initiation is based on the recognition of the 5′ cap of mRNA by the 40S subunit of ribosomes and by some proteins.
	Formyl methionine is incorporated by the initiation codon AUG.	Methionine is incorporated at the initiation point by the codon AUG.
	Initiation site of mRNA first binds with the 30S subunit of ribosome which then accepts formyl methionly-$tRNA_f^{met}$ complex.	Methionyl-tRNAfmet first binds to the 40S subunit, to which an mRNA molecule is then attached to form the initiation complex.

10.6 SUMMARY:

The gene expression refers to the transcription and translation of mRNA which leads to the synthesis of protein. When a gene is expressed, the DNA base pair sequence is transcribed into the base sequence of an RNA molecule. When a gene is transcribed, only one of the two DNA strands is copied. The direction of RNA synthesis is 5′ to 3′ and the reaction is catalyzed by RNA polymerase. The promoter sequences of the gene specify where the transcription is to begin and terminator sequence specify where transcription is to stop. In bacteria, there is only one type of RNA polymerase. The promoter is recognized by a complex between the RNA polymerase core enzyme and a protein factor called sigma. Once transcription is initiated correctly, the sigma factor dissociates from the enzyme and is reused in other transcription initiation events. Termination occurs in one of the two ways, rho-dependent and rho-independent termination processes.

Protein synthesis occurs on ribosomes, where the genetic message encoded in mRNA is translated. Amino acids are brought to the ribosome on charged tRNA molecules. Each tRNA has ananticodon, which binds specifically to a codon in them RNA. As a result, the correct amino acid sequence is achieved by (1) the specific binding of each amino acid to its own specific RNA, and (2) the specific binding between the codon of the mRNA and the complementary anti codon in the tRNA. The initiator codon for the start of the protein is AUG. In prokaryotes, the sequence upstream to the AUG codon required for the small ribosomal subunit binding is the Shine- Dalgarno sequence. The initiation of protein synthesis requires the initiation factors. During elongation phase, the polypeptide chain is elongated one amino acid at a time. This occurs simultaneously with the movement of the ribosome towards the 3′ end of the mRNA one codon at a time. Protein factors called elongation factors play important catalytic roles. The signal for polypeptide chain growth to stop is the presence of a chain terminating codon (UAG, UAA, or UGA) in the mRNA. Specific protein factors called release factors read the stop codon and initiate the events characteristic of protein synthesis termination.

10.7 SELF ASSESSMENT QUESTIONS:

1) Write an essay on the protein synthesis in prokaryotes.
2) Explain the transcription process in prokaryotes.
3) Give an account on the translation process in prokaryotes.

10.8 SUGGESTED READINGS:

1) Freifelder, D. Molecular Biology (1990) - Narosa Publishing House, New Delhi.
2) Gardner, E.J., Simmons, M.J. and Snustad, D.P. Principles of Genetics (2001) - John Wiley & Sons, Inc., New York
3) Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4) Russel, P.J. Genetics (1998) - The Benjamin/Cummings Publishing Company, Inc., California
5) Singh, B.D. Fundamentals of Genetics (2001) - Kalyani Publishers, New Delhi

Prof. A. Amruthavalli

LESSON-11

REGULATION OF GENE EXPRESSION-LAC OPERON

11.0 OBJECTIVE:

- The main intention of this lesson plan is to enlighten the students regarding the regulation of lac genes at the transcription level that encode the enzymes required for lactose utilization in *E.coli*.

STRUCTURE:

11.1 INTRODUCTION

Microorganisms exhibit remarkable capacities to adapt to diverse environmental conditions. This adaptability depends in part on their ability to turn on and turn off the expression of specific sets of genes in response to changes in the environment. Most microorganisms, exhibit a striking ability to regulate the expression of specific genes in response to environmental signals. The expression of particular genes is turned on when the products of these genes are needed for growth. Their expression is turned off when the gene products are no longer needed. Clearly, the ability of an organism to regulate gene expression in this way increases its ability to reproduce under a variety of environmental conditions. The synthesis of gene transcripts and translation products requires the expenditure of considerable energy. By turning off the expression of genes when their products are not needed, an organism can avoid wasting energy and can utilize the conserved energy to synthesize products that maximize growth rate. Gene expression in prokaryotes is regulated at several different levels: transcription, mRNA processing, mRNA turnover, translation, and enzyme function. However, the regulatory mechanisms with the largest effects on phenotype act at the level of transcription.

Certain gene products, such as tRNA molecules, rRNA molecules, ribosomal proteins, RNA polymerase subunits, and enzymes catalyzing metabolic processes that are frequently referred to as cellular ‘housekeeping’ functions are essential components of almost all living

cells. Genes that specify products of this type are continually being expressed in most cells. Such genes are said to be expressed constitutively and are referred to as 'constitutive genes' or 'housekeeping genes'. Other gene products are needed for cell growth only under certain environmental conditions. Constitutive synthesis of such gene products would be wasteful, using energy that could otherwise be utilized for more rapid growth and reproduction.A change in the set of gene products synthesized in a prokaryotic cell involves regulatory mechanisms that control gene expression. Genes whose activity is controlled in response to the needs of a cell or organism are called 'regulated genes'.In bacterial systems in which several enzymes act in sequence in a single metabolic pathway, it is often the case that either all of these enzymes are present or all are absent. This phenomenon is called coordinateregulation. This results from the control ofsynthesis of a single polycistronic mRNA that encodes all of the gene products.

11.2 OPERON CONCEPT

Induction and repression of gene expression can be accomplished bythe same mechanism, with one relatively minor modification. Based on the results of their studies of genetic mutations affecting the regulation of the synthesis of the lactose utilization enzymes, Francois Jacob and Jacques Monod proposed a classical model to explain the regulation of genes encoding the enzymes required for lactose utilization in *E. coli*. This model of mechanism is called 'operon model'. For their proposal, Jacob and Monod received the Nobel Prize in 1965. By definition, an operon is a cluster of genes, the expressions of which are regulated together by operator-regulator protein interactions, plus the operator region itself and the promoter. According to the proposal, the transcription of one or a set of contiguous structural genes is regulated by two controlling elements. One of these elements, the 'regulator or repressor gene', encodes a protein called the 'repressor'. Under the appropriate conditions, the repressor binds to the second element, the operator.

When the repressor binds to operator, transcription of the structural genes cannot occur. Transcription is initiated at RNA polymerase-binding sites called promoters located just upstream from the structural genes. Operator regions are usually located between the promoters and the structural genes that they regulate. Whether the repressor will bind to the operator and turn off the transcription of the structural genes in an operon is determined by the presence or absence of effector molecules in the environment. These effector molecules are usually small molecules such as amino acids, sugars, and similar metabolites. In the case of inducible operons, the effector molecules are inducers. Those active on repressible operons are called co-repressors. These effector molecules, inducers and co-repressors, are bound by the repressors and cause changes in the three- dimensional structures of the repressors. Conformations changes in protein structure resulting from the binding of small molecules are called all osteric transitions. In the case of repressors, the allosteric changes caused by the binding of effector molecules usually alter their ability to bind to operator regions. Inducible operons and repressible operons can be distinguished from one another by determining whether the naked repressor or the repressor-effector molecule complex is active in binding to the operator.

In the case of inducible operon, the free repressor binds to the operator, turning off transcription. When the effector molecule (the inducer) is present, it is bound by the repressor, causing the repressor to be released from the operator, so the repressor-inducer complex cannot bind to the operator. Thus the addition of inducer turns on or induces the transcription of the structural genes in the operon. In the case of repressible operon, the situation is just reversed. The free repressor cannot bind to the operator and repressor-effector

molecule (co-repressor) complex is active in binding to the operator. Thus transcription of the structural genes in a repressible operon is turned on in the absence of and turned off in the presence of the effector molecule. Except for this difference in the operator-binding behavior of the free repressor and the repressor-effector molecule complex, inducible and repressible operons are identical. They operate by exactly the same mechanism.

11.3 TYPES OF REGULATION OF TRANSCRIPTION

There are several common patterns of regulation of transcription. These depend on the type of metabolic activity of the system being regulated. In a catabolic or degradative system, the concentration of the substrate of an enzyme early in the pathway often determines whether the enzymes in the pathway are synthesized. Enzymes that are involved in the catabolic or degradative pathways, such as in lactose, galactose, or arabinose utilization, are characteristically inducible. This induction occurs at the level of transcription. Induction alters the rate of synthesis of enzymes, not the activity of existing enzyme molecules. Induction is different from enzyme activation, which occurs when the binding of a small molecule to an enzyme increases the activity of the enzyme, but does not affect its rate of synthesis.

In contrast, in an anabolic or biosynthetic pathway the final product is the regulatory substance. Bacteria can synthesize most of the organic molecules, such as amino acids, purines, pyrimidines, and vitamins, required for growth. For example, the *E. coli* genome contains five genes encoding enzymes that catalyze steps in the biosynthesis of tryptophan. These five genes must be expressed in *E. coli* cells growing in an environment devoid of tryptophan in order to provide adequate amounts of this amino acid for ongoing protein synthesis. And a regulatory mechanism has evolved in E. coli that turns off the synthesis of the tryptophan biosynthetic enzymes when external tryptophanis available. This process of turning off the expression of genes in response to an environmental signal is called repression. A gene whose expression has been turned off in this way is said to repressed; when its expression is turned on, the gene is said to be derepressed and the response is called derepression. Enzymes that are components of anabolic or biosynthetic pathways often are repressible. Like induction, repression also occurs at the level of transcription. Repression should not be confused with feed-back inhibition, which occurs when the product of a biosynthetic pathway binds to and inhibits the activity of the first enzyme in the pathway, but does not affect the synthesis of the enzyme.

The molecular mechanisms for each of the regulatory patterns vary quite widely but usually fall in one of the two major categories – negative regulation and positive regulation. In negative regulation, an inhibitor is present in the cell that keeps the transcription turned off. In this mechanism, the association of a specific protein called repressor with the operator DNA prevents the transcription of the operon. The transcription is believed to be prevented due to a change in the configuration of the operator DNA so that RNA polymerase is unable to perform its function. Since the regulation of gene action in such a system is achieved by the prevention of transcription by a repressor, it is known as negative control. In positive regulation, an effector molecule which may be a protein, a small molecule, or a molecular complex activates a promoter. In positive control of transcription, association of as pecific protein, termed as activator, to a segment of DNA in the promoter gene or to RNA polymerase enables the transcription of the operon. The promontory effect of the activator is believed to be due to its effect on DNA configuration in the transcription initiator region, which then becomes more favourable for the action of RNA polymerase. Negative and positive regulations are not mutually exclusive, and some systems are both positively and negatively regulated and need two regulators.

11.4 LACTOSE UTILIZATION AS A CARBON SOURCE IN *E.COLI*

E. coli is able to grow in a simple medium containing salts including a nitrogen source and a carbon source such as glucose. These chemicals provide molecules that can be manipulated by the enzymatic machinery of the cell to produce everything the cell needs to grow and reproduce, such as nucleic acids, proteins, and lipids. The energy for these biochemical reactions in the cell comes from the metabolism of glucose, a process that is of central importance to the functioning of a bacterial cell and of cells of all organisms. The enzymes required for glucose metabolism are coded for by constitutive genes.

If lactose, or one of several other sugars, is provided to *E. coli* as a carbon source instead of glucose, a number of enzymes are rapidly synthesized; these enzymes are needed for the metabolism of this particular sugar. The enzymes are synthesized because the genes that code for them become actively transcribed in the presence of the sugar; and the genes are inactive if the sugar is absent. In other words, the genes are regulated genes whose products are needed only at certain times. Biochemical analysis had shown that when the lactose, a disaccharide, is the sole carbon source in growth medium, three proteins are synthesized –

1) **β-galactosidase**: This enzyme has two functions(a) It catalyzes the breakdown of lactose into its two component mono saccharides, glucose and galactose; and (b) It catalyzes the isomerization of lactose to 'allolactose' a compound important in the regulation of expression of the lactose utilization genes.
2) **β-galactosidepermease**: This is also called as M protein. This M protein is found in the *E. coli* cytoplasmic membrane and is needed for the active transport of lactose from the growth medium into the cell.
3) **β-galactosidetransacetylase**: The function of this enzyme is poorly understood.

In a wild-type *E.coli* that is growing in a medium containing glucose but no lactose, only a few molecules of each of the above three enzymes are produced, indicating a low level of expression of the three genes that code for the proteins. For example, only an average of three molecules of β-galactosidase is present in the cell under these conditions. If lactose but no glucose is present in the growth medium, the number of molecules of each of the three enzymes increases coordinately about a thousand fold. This occurs because the three essentially inactive genes are now being actively transcribed and the resulting mRNA translated. This process is called coordinate induction. The inducer molecule is directly responsible for the increased production of the three enzymes is actually allolactose, not lactose. Allolactose is produced from lactose by β- galactosidase. Further, the mRNAs for the enzymes have a relatively short half-life, so the transcripts must be made continually in order for the enzymes to be produced. So, when lactose is no longer present, transcription of the three genes is stopped and any mRNAs already present are broken down thereby resulting in a drastic reduction in the amounts of the three proteins in the cell.

11.5 THE LACTOSE OPERON IN *E.COLI*: INDUCTION AND CATABOLITE REPRESSION

Jacob and Monod proposed the operon model largely as a result of their studies of the lactose (lac) operon in *E. coli*. More is known about this operon than any other operon. The lac operon (Fig. 11.1) contains a promoter (P), an operator (O), and three structural gens,

*lac*Z, *lac*Y, and *lac*A, encoding the enzymes β-galactosidase, β-galactoside permease, and β-galactoside transacetylase, respectively. The β-galactoside permease pumps lactose into the cell, where β-galactosidase cleaves it into glucose and galactose. The exact biological role of the transacetylase is not clearly understood.

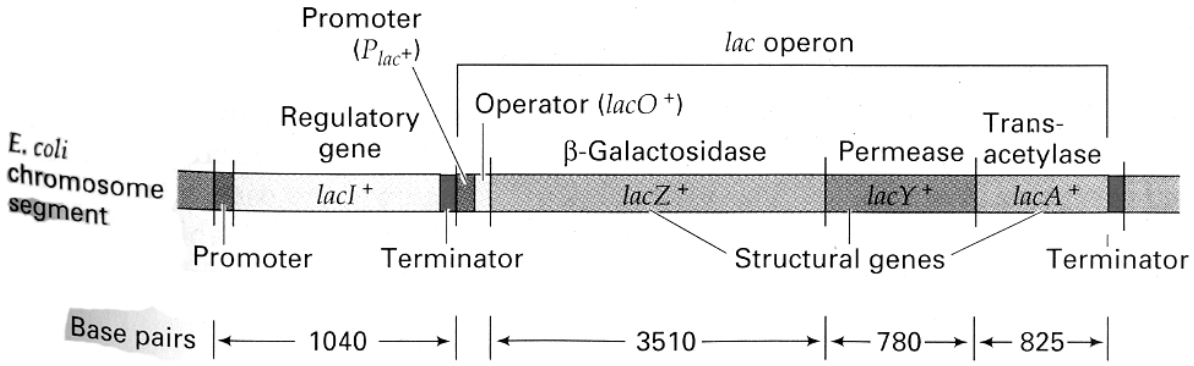


Fig.11.1 Structure of Lac Operon

11.5.1 Induction

The lac operon is an inducible operon; the *lac*Z, *lac*Y, and *lac*A genes aree xpressed only in the presence of lactose. The *lac* regulator gene, designated as *I* gene, encodes a repressor that is 360 amino acids long. The active form of the *lac* repressor is a tetramer containing four copies of the *I* gene product. In the absence of inducer, the repressor binds to the *lac* operator sequence, which, in turn prevents RNA polymerase from catalyzing the transcription of the three structural genes (Fig.11.2). A few molecules of the *lac* Z, *lac* Y, and *lac* A gene products are synthesized in the un induced state, providing a low background level of enzyme activity. This background activity is essential for induction of the *lac* operon, because the inducer of the operon, 'allolactose', is derived from lactose in a reaction catalyzed by β-galactosidase. Once formed, allolactose is bound by the repressor, causing the release of the repressor from the operator (Fig. 11.3). In this way, allolactose induces the transcription of the *lac*Z, *lac*Y, and *lac*A structural genes. The *lac I* gene, the *lac* operator, and the *lac* promoter were all initially identified genetically by the isolation of mutant strains that exhibited altered expression of the *lac* operon genes. Mutations in *I* gene and the operator frequently result in the constitutive synthesis of the *lac* gene products. These mutations are designated I^- and O^c, respectively.

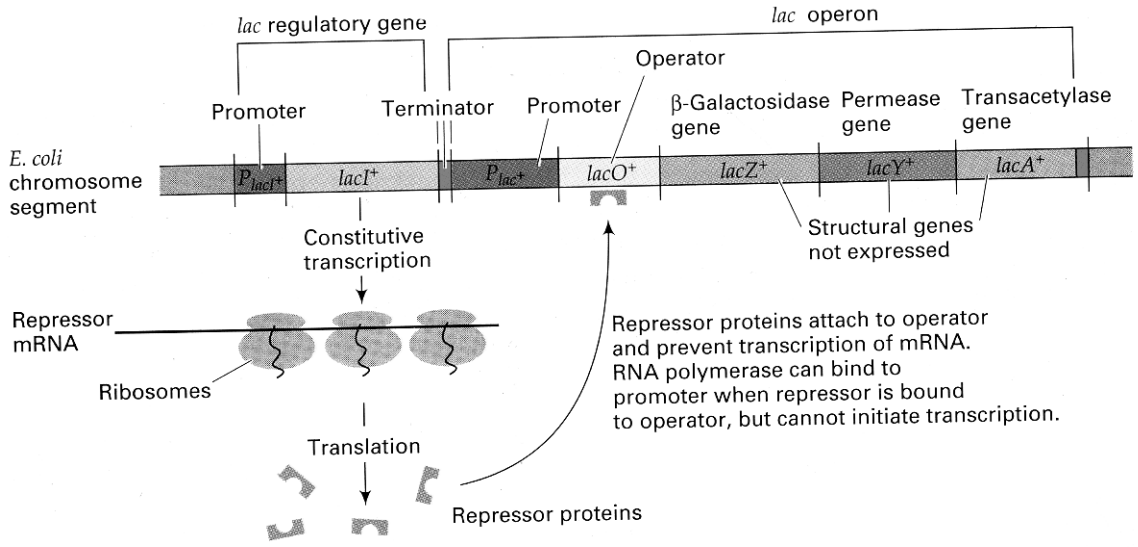


Fig.11.2 Regulation of Lacoperon in the Absence of Lactose

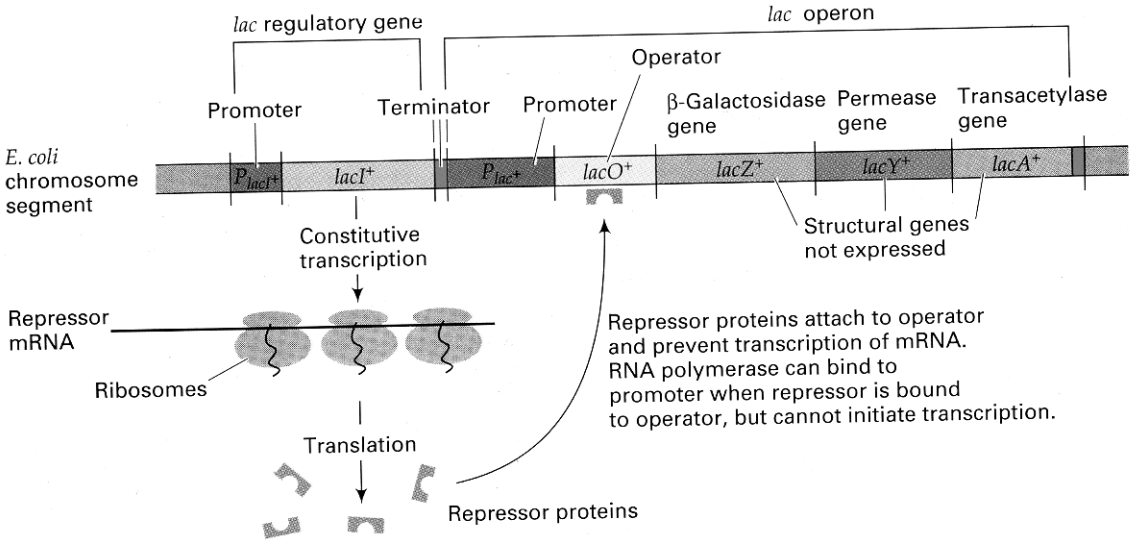


Fig.11.3 Regulation of Lacoperon in the Presence of Inducer

11.5.2 Catabolite Repression

Jacob and Monod proposed the operon model to explain the induction of the biosynthesis of the enzymes involved in lactose utilization when this sugar is added to the medium in which *E. coli* cells are growing. However, the presence of glucose has long been known to prevent the induction of the *lac* operon, as well as other operons control ling enzymes involved in carbohydrate catabolism. This phenomenon is called as 'catabolite repression' or the 'glucose effect'. This assures that glucose is metabolized when present, in preference to other, less efficient, energy sources.

The catabolite repression of the *lac* operon and several other operons is mediated by a regulatory protein called CAP (Catabolite Activator Protein) and a small effector molecule

called cyclic AMP (adenosine-3′,5′ -phosphate; abbreviated cAMP). Because CAP binds cAMP when this mononucleotide is present at sufficient concentrations, it is sometimes called the cyclic AMP receptor protein. Infact, the *lac* promoter contains two separate binding sites – one for RNA polymerase and one for the CAP-cAMP complex. The CAP-cAMP complex must be present in its binding site in the lac promoter in order for the operon to be induced. The CAP-cAMP complex thus exerts positive control over the transcription of the *lac* operon. It has an effect exactly opposite to that of repressor binding to an operator. Although the precise mechanism by which CAP-cAMP stimulates RNA polymerase binding to the promoter is still uncertain, its positive control of the *lac* operon transcription is firmly established by the results of both in vivo and in vitro experiments. CAP is known to function as a dimer, thus, like the *lac* repressor, it is multimeric in its functional state.

Only the CAP-cAMP complex binds to the *lac* promoter; in the absence of cAMP, CAP does not bind. Thus cAMP acts as the effector molecule, determining the effect of CAP on *lac* operon transcription. The intracellular cAMP concentration is sensitive to the presence or absence of glucose. High concentrations of glucose cause sharp decrease in the intracellular concentration of cAMP. How glucose controls the cAMP concentration is not clear. Perhaps glucose, or some metabolite that forms in the presence of sufficient concentrations of glucose, inhibits the activity of adenylcyclase, the enzyme that catalyzes the formation of cAMP from ATP. The presence of glucose results in a decrease in the intracellular concentrations of cAMP. In the presence of a low concentration of cAMP, CAP cannot bind to the *lac* operon promoter. In turn, RNA ploymerase cannot bind efficiently to the *lac* promoter in the absence of bound CAP-cAMP. The overall result of the positive control of transcription of the *lac* operon by the CAP-cAMP complex is that in the presence of glucose *lac* operon transcription never exceeds 2 percent of the induced rate observed in the absence of glucose. By similar mechanisms, CAP and cAMP keep the arabinose and galactose operons of *E. coli* from being induced in the presence of glucose.

11.6 SUMMARY:

From the studies of the regulation of expression of the three lactose - utilizing genes in *E. coli* bacteria, Jacob and Monod developed a model that is the basis for the regulation of gene expression in a large number of bacterial and bacteriophage systems. The genes for the enzymes are contiguous in the chromosome and are adjacent to a controlling site, an operator, and a single promoter. This complex constitutes a transcriptional regulatory unit called an operon. A regulator gene, which may or may not be nearby, is associated with the operon. The addition of an appropriate substrate to the cell results in the coordinate induction of the operon's structural genes. With respect to the *lac* operon, induction occurs as lactose binds with a repressor protein that is encoded by the regulator gene, inactivating it and preventing it from binding to the operator. As a result, RNA polymerase can transcribe the three genes onto a single polygenic mRNA. As long as lactoseis present, mRNA continues to be produced and the enzymes are made. When lactoseis no longer present, the repressor protein is no longer inactivated, and it binds to the operator, thereby preventing RNA polymerase from transcribing the *lac* genes.

If both glucose and lactose are present in the medium, the lactose operon is not induced because glucose is the preferred energy source. This phenomenon is called catabolite repression and involves cellular levels of cyclic AMP. In the presence of lactose and absence of glucose, cAMP complexes with CAP to form a positive regulator needed for RNA polymerase to bind to the promoter. Addition of glucose results in a lowering of cAMP level so no CAP-cAMP complex is produced and, therefore, RNA polymerase cannot bind to the promoter and transcribe the *lac* genes.

11.7 SELF ASSESSMENT QUESTIONS:

1) Describe the mechanism of regulation of *lac* operon.
2) What is Operon Concept. Explain the Negative control and Positive control mechanisms of gene regulation.

11.8 SUGGESTED READINGS:

1) Freifelder, D. Molecular Biology (1990) - Narosa Publishing House, New Delhi.
2) Gardner, E.J., Simmons, M.J. and Snustad, D.P. Principles of Genetics (2001) - John Wiley & Sons, Inc., New York.
3) Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York.
4) Russel, P.J. Genetics (1998) – The Benjamin / Cummings Publishing Company, Inc., California.
5) Singh, B.D. Fundamentals of Genetics (2001) - Kalyani Publishers, New Delhi.

Prof. V. Uma Maheswara Rao

LESSON-12

REGULATION OF GENE EXPRESSION - TRP OPERON

12.0 OBJECTIVE:

- This lesson plan is aimed to give a clear idea to the student about the repressive mechanism of gene regulation in the case of tryptophan operon.

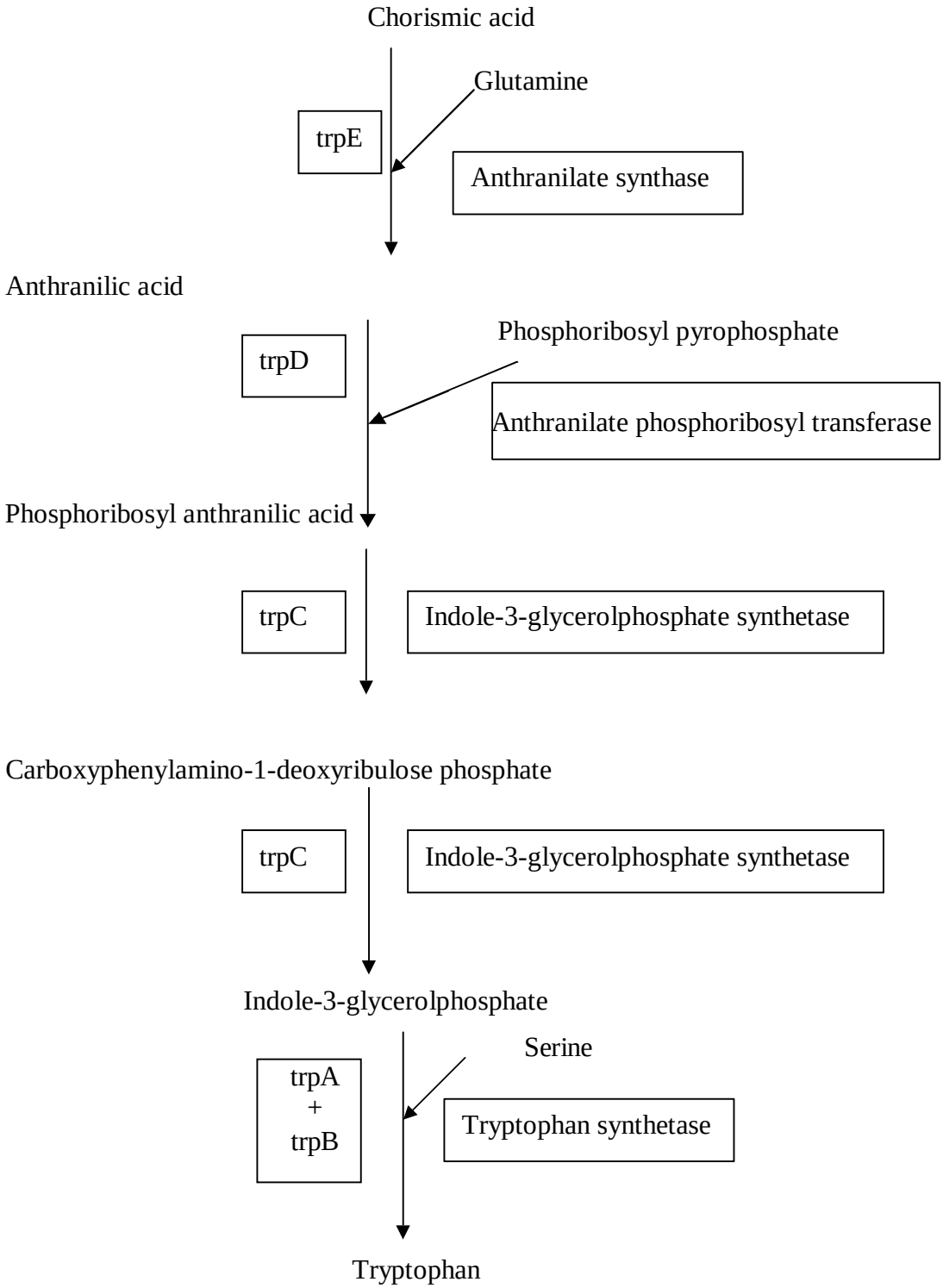
STRUCTURE:

12.1 INTRODUCTION

Organisms live and reproduce in changing environments. One way that an organism can adjust to a new environment is to alter its gene activity so that gene products appropriate to the new conditions are synthesized. As a result, the organism is optimally adjusted to grow and reproduce in that environment. In a growth medium, all the necessary amino acids many not be present to enable bacteria to assemble proteins. If an amino acid is missing, a bacterium has certain operons and other gene systems that enable the bacterium to manufacture that amino acid so that it may grow and reproduce. Each step in the biosynthetic pathway through which amino acids are assembled is catalyzed by a specific enzyme coded by a specific gene or genes.

When an amino acid is present in the medium, the genes encoding the enzymes for that amino acid's biosynthetic pathway are tuned off. If an amino acid is not present in the medium, however, those must be turned on in order for the bio synthetic enzymes to be made. Unlike the *lac* operon, where gene activity is induced when lactose is added to the medium, in this case there is a repression of gene activity when an amino acid is added. So, the amino acid biosynthesis operons controlled in this way are the repressible operons. In general, repressible operons function to turn off anabolic or biosynthetic path ways when the end productis readily available. One repressible operon in *E. coli* that has been extensively studied is the operon for the biosynthesis of the amino acid tryptophan (trp). Although the regulation of the *trp* operon shows some basic similarities to the regulation of the classical *lac* operon, some intriguing differences also appear to be common among similar, repressible, amino acid biosynthesis operons in bacteria.

12.2 BIOSYNTHETIC PATH WAY OF TRYPTOPHAN



The biosynthesis of typtophan occurs from chorismic acid. The entire biosynthetic process occurs in five enzymatic steps involving the enzymes encoded by the five structural genes of the *trp* operon. During the first step, the glutamine combines with chorismic acid to produce anthranilic acid. This step is catalyzed by the enzyme, anthranilates ynthetase coded by *trp*E gene. In the second step, the anthranilic acid combines with phosphoribosyl pyrophosphate to yield phosphoribo sylanthranilic acid under the mediation of enzyme, anthranilate phosphoribosyl transferase coded by *trp*D. In the next two steps, phosphoribosyl anthranilic acid is first converted to carboxy phenylamino-1-deoxyribulose phosphate and

then to indole-3-gycerolphosphate. Both the steps are catalyzed by the same enzyme, indole-3-glycerolphosphate synthetase coded by *trp*C gene. During the last step, indole-3-glycerolphosphate combines with serine to yield Tryptophan. This last step is mediated by the enzyme, tryptophan synthetase which is jointly coded by *trp*A and *trp*B genes.

12.3 TRYPTO PHANOPERON IN *E.COLI*: REPRESSION AND ATTENUATION

The trp operon in *E. coli* controls the synthesis of the enzymes that catalyze the biosynthesis of the amino acid typtophan. The functions of the five structural genes and the adjacent regulatory sequences of the *trp* operon (Fig.12.1) have been analyzed in detail by Charles Yanofsky and colleagues. The five structural genes (A-E) encode five enzymes that convert chorismic acid to tryptophan in five steps. The genes encoding these enzymes in *E. coli* are adjacent to one another in the same order as their use in the biosynthetic pathway; they are translated from a single polycistronicm RNA and called *trp*E, *trp*D, *trp*C, *trp*B, and *trp*A. The *trp*E gene is the first one translated. The promoter and operator regions are closely integrated in the DNA and are upstream from the *trp*E gene. Between the promoter-operator region and *trp*E is a 162-base pair region called *trp*L, or the leader region. Within *trp*L, relatively close to *trp*E, is an attenuator site (att) that plays an important role in the regulation of the tryptophan operon. The repressor gene *trp*R is located very far from this gene cluster, like the repressor in the *gal* operon. The entire tryptophan operon is approximately 7,000 base pairs long. Transcription of the operon results in the production of a polygenic mRNA for the five structural genes. Each of these transcripts is translated to an equal extent. The expression of the *trp*operon is regulated at two levels – repression, which controls the initiation of transcription, and attenuation, which governs the frequency of premature transcript termination.

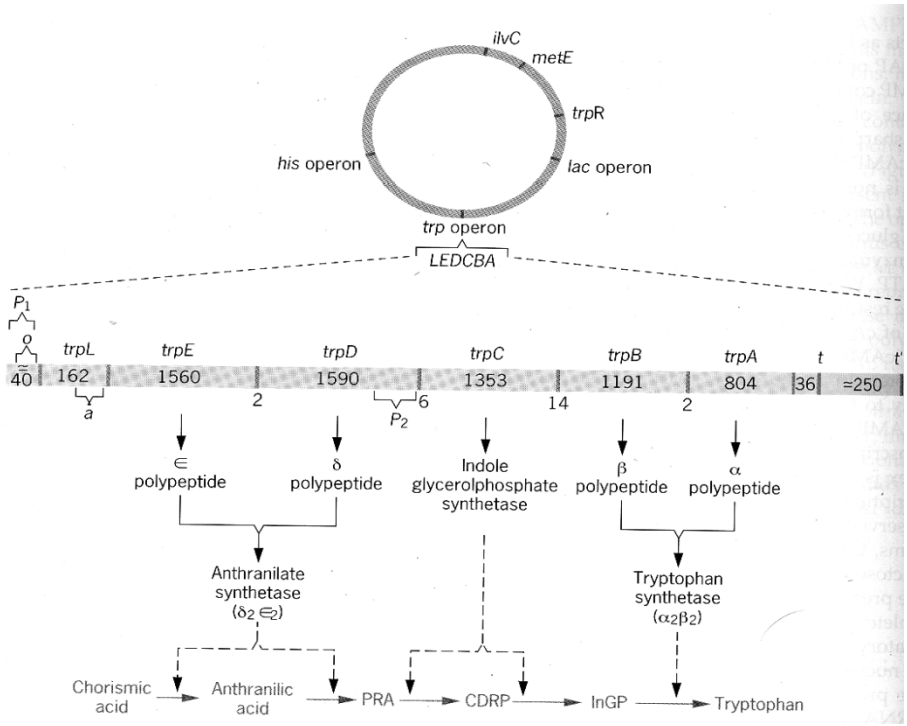


Fig-12.1: Structure of trp Operon

12.3.1 Repression

The *trp* operon of *E. coli* is probably the best known repressible operon. The operator (O) region of the *trp* operon lies within the primary promoter (P_1) region. There is also a weak promoter (P_2) at the operator-distal end of the *trp*D gene. The P_2promoter results in a somewhat increased basal level of transcription of the *trp*C, *trp*B, and *trp*A genes. Two transcription termination sequences (t and t′) are located downstream from *trp*A. The *trp*L region specifies a 162-nucleotide long mRNA leader sequence. The product of the regulatory gene, *trp*R, is an aporepressor protein, which alone cannot bind to the operator. When tryptophan is abundant in the growth medium, it binds to the aporepressor and converts it to an active repressor. The active repressor binds to the operator and prevents the initiation of transcription of the *trp* operon protein-coding genes by RNA polymerase. As a result, the tryptophan biosynthesis enzymes are not produced. In the absence of tryptophan, RNA polymerase binds to the promoter region and transcribes the structural genes of the operon.

The rate of transcription of the *trp* operon in the derepressed state i.e. the absence of typtophan, is 70 times the rate that occurs in the repressed state i.e. in the presence of tryptophan. In *trp*R mutants that lack functional repressor, the rate of synthesis of the tryptophan biosynthetic enzymes is still reduced about tenfold by the addition of tryptophan to the medium. This additional reduction in *trp* operon expression is caused by attenuation.

12.3.2 Attenuation

Deletions that remove part of the *trp*L region result in increased rates of expression of the *trp* operon. However, these deletions have no effect on the repressibility of the *trp* operon; that is, repression and derepression occur just as in *trp*L^+ strains. These results indicate that the synthesis of the trptophan biosynthetic enzymes is regulated at a second level by a mechanism that is independent of the state of repression or derepression of the *trp* operon and requires nucleotide sequences present in the *trp*L region of the *trp* operon. This second level of regulation of the *trp* operon is called attenuation, and the sequence within *trp*L that controls this phenomenon is called the attenuator. Attenuation occurs by control of the termination of transcription at a site near the end of the mRNA leader sequence. This premature termination of *trp*operon transcription occurs only in the presence of tryptophan-charged $tRNA^{Trp}$. When this premature termination or attenuation occurs, a truncated *trp* transcript of 140 nucleotides is produced.

The second regulatory mechanism is involved in the expression of the *trp* operon under conditions of tryptophan starvation or tryptophan limitation. Under severe tryptophan starvation, the *trp* genes are expressed maximally, while underless severe starvation conditions, the *trp* genes are expressed at less than maximal levels. This is accomplished by a mechanism that controls the ratio of the transcripts that include the five *trp* structural genes to those that are terminated before the structural genes.

The attenuator region has a nucleotide-pair sequence essentially identical to the transcription-termination signals found at the ends of most bacterial operons. These termination signals contain a G:C rich palindrome followed by several A:T base pairs. Transcription of these termination signals yields a nascent RNA with the potential to form a hydrogen-bonded hairpin structure followed by several Us. When a nascent transcript forms this hairpin structure, it is believed to cause a conformational change in the associated RNA polymerase, resulting in termination of transcription within the following, more weakly hydrogen-bonded region of DNA- RNA base pairing. The nucleotide sequence of the attenuator therefore explains its ability to terminate trp operon transcription prematurely.

In prokaryotes, the transcription and translation mechanisms are coupled, that is, ribosomes begin translating mRNAs while they are still being produced by transcription. Thus events that occur during translation may also affect transcription. The 162 nucleotide-long leader sequence of the *trp* operon mRNA contains sequences that can base pair to form alternate stem-and-loop or hairpin structures (Fig. 12.2). The four leader regions that can base pair to form these structures are – (1) nucleotides 59-67, (2) nucleotides 71-79, (3) nucleotides 110-121, and (4) nucleotides 126-134. The nucleotide sequences of these four regions are such that region 1 can base pair with region 2, region 2 can base pair with region 3, and region 3 can base pair with region 4. Region 2 can base pair with either region 1 or region 3, but, obviously, it can pair with only one of these regions at any given time. Thus there are two possible secondary structures for the trp leader sequence – (1) region 1 paired with region 2 and region 3 paired with region 4 or (2) region 2 paired with region3, leaving regions 1and 4 unpaired.Thepairingofregions3 and4 produces the transcription-termination hairpin. If region3 is base paired with region 2, it cannot pair with region 4, and the transcription-termination hairpin cannot form. The presence or absence of tryptophan determines the formation of these alternative structures.

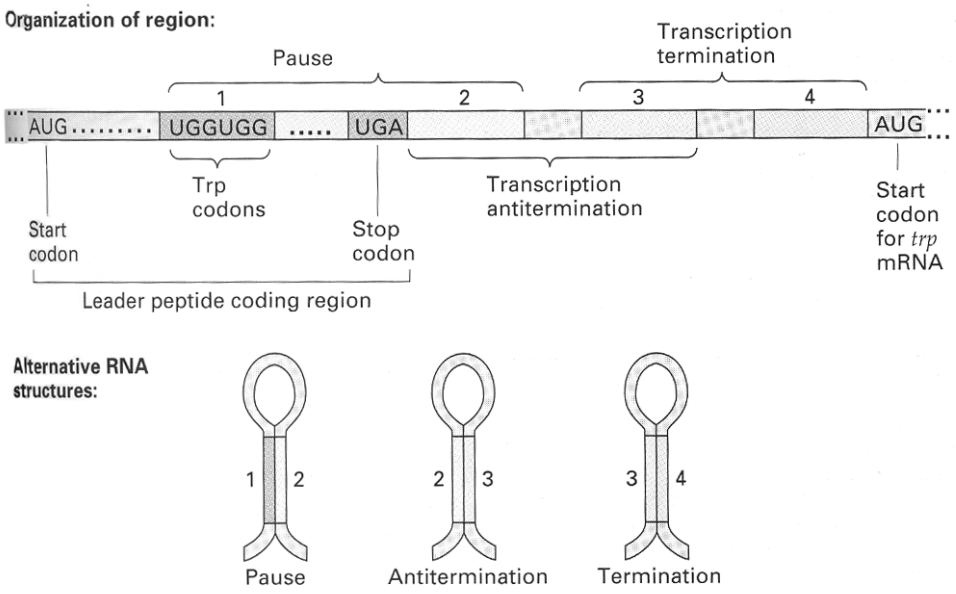


Fig-12.2: Leader Region of trp Operon and Stem and Loop Structures

The leader sequence contains an AUG translation-initiation codon, followed by 13 codons for amino acids, followed, in turn, by a UGA translation-termination codon. In addition, the trp leader sequence contains an efficient ribosome-binding site located in the appropriate position for the initiation of translation at the leader AUG initiation codon. All the available evidence indicates that a 14 amino acid 'leader peptide' is synthesized. But the synthesis of this putative leader peptide in vivo has not been verified directly, may be because of their unstable nature. The leader peptide contains two contiguous tryptophan residues. The two Trp codons are positioned such that in the absence of tryptophan, the ribosome will stall before it encounters the base paired structure formed by leader regions 2 and 3 (Fig.12.3). Because the pairing of regions 2 and 3 precludes the formation of the transcription-termination hairpin by the base pairing of regions 3 and 4, transcription will continue past the attenuator into the trpE gene in the absence of tryptophan.

In the presence of tryptophan, the ribosome can translate past the Trp codons to the leader- peptide termination codon. In the process, it will disrupt the base pairing between leader regions 2 and 3. This disruption leaves region 3 free to pair with region 4, forming the transcription- termination hairpin. Thus, in the presence of typtophan, transcription frequently terminates at the attenuator, reducing the amount of mRNA for the trp structural genes. The transcription of the *trp* operon can be regulated over a range of almost 700-fold by the combined effects of repression (up to 70-fold) and attenuation (up to 10-fold).

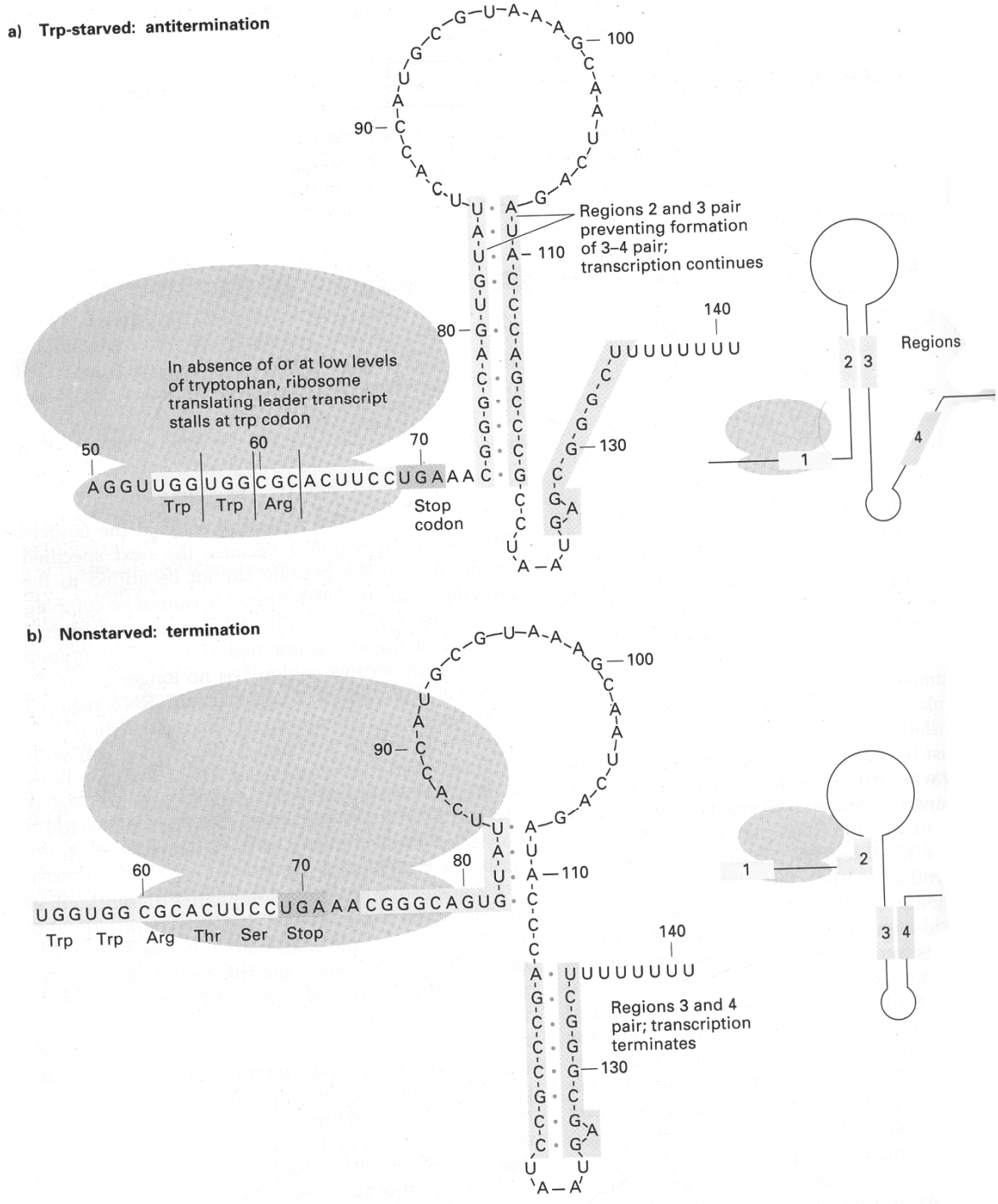


Fig.12.3 Mechanism of Attenuation Mode of Regulation

12.4 FINE-SCALE REGULATION OF THE TRPOPERON

In *E. coli*, in addition to the regulation at the gene expression level in the trp operon, another mechanism also operates to give an overall regulatory system for gene expression. This mechanism is the degradation of the polygenic mRNA transcript of the protein-coding genes. As with the case of most other prokaryotic genes, the transcripts of trp genes have a relatively short half-life, so mRNA must continually be made in order for the enzymes to be produced. The short half-life allows the cell to change its physiological state when the environment changes. When tryptophan is added to the medium, transcription termination at the attenuator and repression at the operator block further transcription of the structural genes, and the structural gene mRNAs already made are rapidly degraded. In this way the cell does not make unnecessary materials. Another part of the overall regulatory system is a fine-tuning system. Gene regulation involves control processes that function in response to long-term changes in the environment. Cells must also respond to short-term changes that occur. For example, if tryptophan being made by the tryptophan biosynthetic pathway accumulates at a rate faster than it is being used, a mechanism called feedback inhibition or end-product inhibition is triggered.

In feedback inhibition the end product of a biosynthetic pathway can often be recognized by the first enzyme in the biosynthetic pathway. When too much tryptophan is produced, for example, the tryptophan binds to the first enzyme in the tryptophan biosynthetic pathway, thereby altering the enzyme's three-dimensional conformation. When the shape of the enzyme is changed, the function of the enzyme is impaired, because the substrate for the enzyme can no longer bind to the enzyme. As a result, the tryptophan pathway is temporarily turned off at the first step in the biosynthetic pathway. The enzyme becomes nonfunctional until the tryptophan level in the cell drops and the tryptophan dissociates from the enzyme. This step reverses the feedback inhibition of the enzyme's function, and the pathway is restarted.

12.5 SUMMARY

The genes for a number of bacterial amino acid biosynthesis pathways are also arranged in operons. Expression of these operons is accomplished by a repressor-operator system or, in some cases, through attenuation at a second controlling site called an attenuator, or both. The tryptophan operon is an example of an operon with both types of transcription regulation systems. The repressor-operator system functions essentially like that of the *lac* operon, except that the addition of tryptophan to the cell activates the repressor, thereby turning the operon off. The attenuator is located downstream from the operator in a leader region that is translated. The attenuator is a transcription termination site that allows only a fraction of RNA polymerases to transcribe the rest of the operon. Attenuation involves a coupling between transcription and translation, and the formation of particular RNA secondary structures that signal whether or not transcription can continue.

Key to the attenuation phenomenon is the presence of multiple copies of codons for the amino acid synthesized by the enzymes encoded by the operon. When enough of the amino acid is present in the cell, enough charged tRNAs are produced so that the ribosome can translate the key codons in the leader region, and this causes the RNA that is being made by the RNA polymerase ahead of the ribosome to assume a secondary structure which signals transcription to stop. However, when the cell is starved for that amino acid, there are

insufficient charged tRNAs to be used at the key codons so that the ribosome stalls at that point. As a result the RNA ahead of that point assumes a secondary structure that permits continued transcription into the structural genes. The combination of repressor - operator regulation and attenuation control permits a fine degree of control of transcription of the operon.

12.6 TECHNICAL TERMS:

Tryptophan, Trp operon, Attenuation, Regulation of Transcription.

12.7 SELF ASSESSMENT QUESTIONS:

1) Give an account on the mechanism of *trp* operon regulation.
2) Describe the biosynthetic pathway of tryptophan and Attenuation mechanism of trp regulation.

12.8 SUGGESTED READINGS:

1) Freifelder, D. Molecular Biology (1990) - Narosa Publishing House, New Delhi
2) Gardner, E.J., Simmons, M.J. and Snustad, D.P. Principles of Genetics (2001) - John Wiley & Sons, Inc., New York
3) Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4) Russel, P.J. Genetics (1998) – The Benjamin / Cummings Publishing Company, Inc., California
5) Singh, B.D. Fundamentals of Genetics (2001) – Kalyani Publishers, New Delhi.

Prof. V. Uma Maheswara Rao

LESSON-13

GENETICS OF NITROGEN FIXATION

13.0 OBJECTIVE:

- This lesson plan is intended mainly to enlighten the students by emphasizing the genes involve in nitrogen fixation and nodule formation.

STRUCTURE:

13.1 INTRODUCTION

Nitrogen, one of the important elements, in its gaseous form occupies 80% of the atmosphere. But it is in accessible for utilization by many organisms unless it is turned into organic form. Nitrogen in its organic form is a major component of all living beings as proteins, nucleic acids, vitamins and numerous other vital molecules contain it as a part of them. The utilization of nitrogen gas as a source of nitrogen is called nitrogen fixation. This inert nitrogen gas is made available for utilization, in nature, by only certain prokaryotes. A variety of prokaryotes, both aerobic and anaerobic, reduces or fix the atmospheric nitrogen to an utilizable nitrogen source namely ammonia, which can be used by plants and other micro organisms as building block for the synthesis of amino acids and other nitrogenous compounds. So far, no eukaryotic organism has been reported that can fix the nitrogen. The phenomenon of making or converting the gaseous nitrogen to utilizable nitrogen form is called as 'nitrogen fixation'. As it is carried out by living prokaryotic organisms, it is called as 'biological nitrogen fixation'. Some prokaryotes fix the nitrogen in free-living state and the phenomenon is called as 'non-symbiotic nitrogen fixation'. Some other species, fix the nitrogen in association with specific host plants by forming specialized structures called 'nodules' on plant roots. These species are called as 'symbiotic nitrogen fixers' and the phenomenon is called as 'symbiotic nitrogen fixation'.

13.2 NITROGENASE ENZYME COMPLEX

In the fixation process, N_2 is reduced to ammonium and the ammonium converted into organic form. The reduction process is catalyzed by the enzyme complex nitrogenase., which consists of two separate proteins called dinitrogenase and dinitrogenase reductase. Both components contain iron, and dinitrogenase contains molybdenum as well. The molecular weight of the dinitrogenase range from 2,20,000 to 2,70,000 daltons. The Fe containing protein or the dinitrogenase reductase possess a molecular weight within the range of 55,000 – 66,800 daltons. The molybdenum and iron in dinitrogenase are contributed by a cofactor known as FeMoCo, and the actual reduction of N_2 involves participation of this iron-molybdenum center. Active nitrogenase can be reconstituted by the addition of purified Mo-Fe and Fe proteins of different microorganisms. For examples, proteins of *Klebsiella pneumonia* and *Bacillus polymyxa* and those of blue-green algae and photosynthetic bacteria have been combined to reconstitute active nitrogenases.

The Mo-Fe protein has been designated as dinitrogenase because nitrogen binds to the protein moiety whereas the Fe protein has been referred to as the dinitrogen reductase since the second moiety serves the specific function of reducing the Mo-Fe protein. Some nitrogen-fixing bacteria can synthesize more than one nitrogenase under certain growth conditions, and these so- called alternative nitrogenases do not contain molybdenum but instead contain either vanadium (or iron) or iron only. Alternative nitrogenases are not synthesized when sufficient molybdenum is present, as the molybdenum nitrogenase is generally the main nitrogenase in the cell. Alternative nitrogenases presumably serve as a back-up mechanism to ensure that N_2fixation can still occur when molybdenum is limiting in the habitat. The nitrogenase has been isolated from different genera of free-living nitrogen-fixing microorganisms - *Clostridium*, *Bacillus, Klebsiella, Chloropseudomonas, Chromatium, Rhodospirillum, Anabaena, Gloeocapsa, Plectonema, Azotobacter* and *Mycobacterium.*

Owing to the stability of the N≡N triple bond, N_2 is extremely inert and its activation is a very energy demanding process. Six electrons must be transferred to reduce N_2 to $2NH_3$, and several intermediate steps might be visualized; but since no intermediates have ever been isolated, it is now assumed that the three successive reduction steps occur with the intermediates firmly bound to nitrogenase. Nitrogen fixation is highly reductive in nature and the process is inhibited by oxygen since dinitrogenase reductase is rapidly and irreversibly inactivated by O_2. in aerobic bacteria, N_2 fixation occurs in the presence of O_2 in whole cells, but not in purified enzyme preparations, and nitrogenase in such organisms is protected from O_2 inactivation either by removal of O_2 by respiration, the production of O_2-retarding slime layers, or by compartmentalizing nitrogenase in a special type of cell.

13.3 GENETICS OF N_2-FIXATION IN *KLEBSIELLA PNEUMONIAE*

The genetics of N_2-fixation is best studied in *Klebsiella pneumoniae,* a non-symbiotic nitrogen fixer. Genes that involve in nitrogen fixation are called '*nif* genes' and form a complex regulon called '*nif* regulon' or '*nif* complex' or '*nif* gene cluster'. The regulon refers to a large stretch of genetic material with multiple operons. The *nif* regulon includes nitrogenase structural genes, genes for FeMoCo factor, genes controlling the electron transport proteins and a number of regulatory genes. In *K. pneumoniae* the *nif* region constitutes a cluster of chromosomal genes next to the genes regulating the biosynthesis of histidine. By genetic procedures involving the isolation of mutants lacking *nif* genes, complementation analysis, cloning of *nif* genes, identification of *nif*- coded polypeptides and DNA sequencing, the genes involved in nitrogen fixation and their organization in the

chromosome have been deciphered in*K. pneumoniae*. Barring the *nif* J, the entire *nif* cluster of *K. pneumoniae* has now been sequenced. The complete *nif* cluster constitutes 21 genes *nif* JCHDKTYENXUSVWZMFLABQ, of which T, W, and Z are the three potential new genes. These 21 genes are arranged in 7 or 8 transcriptional units. The *nif* gene cluster of *K. pneumoniae* possesses a molecular weight of 18 x 10^6daltons. Thus the physical map of *nif* clusterin *K. pneumoniae* is well understood and all the genes have been cloned into various vectors, which has facilitated the screening for *nif* homologyin other nitrogen-fixing microorganisms. This has resulted in the analysis of *nif* genes in *Azotobacter, Azospirillum, Rhizobium, Enterobacter, Frankia*, cyano bacteria and other species.

The properties and functions of some of the *nif* gene products of *K. pneumoniae* have been fairly well understood. These relate to the nitrogenase enzyme, the electron transfer system and the regulatory functions. There are many other gene products where properties and functions are not clear. The Mo-Fe protein moiety of nitrogenase is encoded by *nif* H. The nucleotide sequence of *nif* H and *nif* D is not well established. Five genes (*nif* BNEVQ) are involved in the formation of functional Mo-Fe protein. The genes, *nif* BNE are involved either in synthesis or insertion of FeMoCO, a catalytic cofactor of nitrogenase. The gene *nif* V plays a role in substrate specificity while *nif* Q operates under conditions of Mo deficiency. The genes *nif* M is believed to be necessary in the processing of the Fe protein. The roles of six genes *nif* TYXUSW are presently unknown. Components of the specific electron transfer are the products of *nif* F and *nif* J which contain acid labile sulphur. In the electron chain to nitrogenase, the physiological donor is pyruvate. Electrons are carried from the *nif* J protein having pyruvate flavodoxinoxydo reductase activity to the *nif* Fprotein which is the substrate of the Feprotein. The genes *nif* Hinvolves inthe synthesis of Fe-protein. The *nif* genes D and K involve in the synthesis of α and β subunits of MoFe-protein, respectively. The product of *nif* A gene acts as a positive regulator and *nif* L gene product acts as negative regulator.

Ammonia (NH_4) totally represses *nif* gene product biosynthesis. The genes involved in glutamine synthetase, an enzyme which regulates NH_4 as simulation are referred to as *Gln* while *ntr* denotes genes whose products regulate nitrogen assimilation. The genes which determine uptake hydrogenase activity are known as *hup* genes. In *K. pneumoniae*, *ntr* BC are linked to *gln*A, the structural gene for glutamine sythetase whereas *ntr*A is unlinked. *Gln*A and *ntr* BC are organized in one or two operons transcribed from two promoters in the order P1glyAP2ntr BC. P1 promotes transcription under conditions of nitrogen limitation where as P2 acts under nitrogen enriched situations. In this way, when nitrogen is the limiting factor, the biosynthesis of glutamine synthetase is derepressed including those operons under the control of *nif* which includes the *nif* regulator also. The product of *ntr*C acts as a general activator of all these operons which is again dependent on *ntr*A product. The organization of the *nif* cluster of *K. pneumoniae* is given in Fig. 13.1.

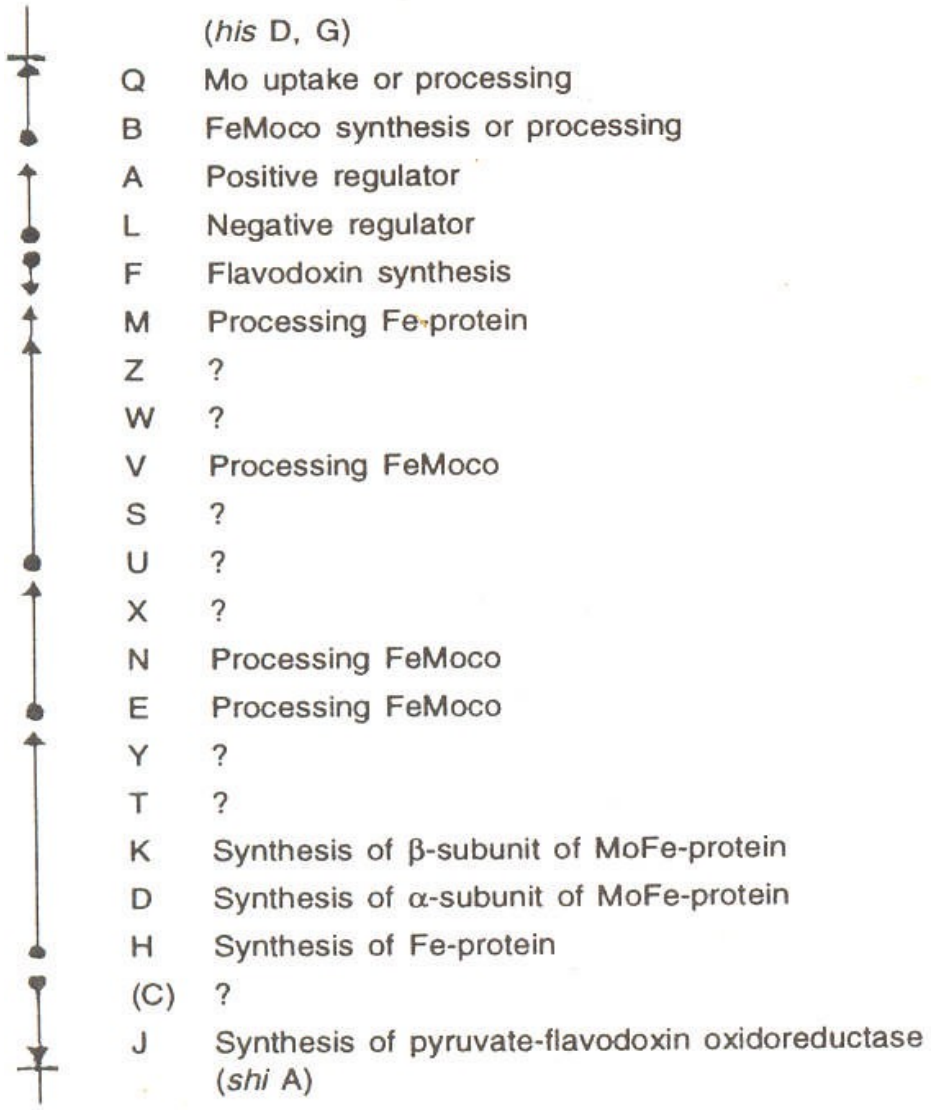


Fig-13.1: nif genecluster of *K.pneumoniae*

From the current information on the organization and functioning of the *nif* gene in some N_2fixing species, it can be reasonably concluded that a basic group of *nif* genes is perhaps common to all diazotrophs including *nif HDK* genes which are highly conserved. These are the genes concerned with the processing of the metal clusters of the nitrogenase plus the regulatory genes. Genetic analysis of *Azotobacter vinelandii* and *Azotobacter chroococcum* have revealed that a major *nif* cluster comparable to that of one in *K. pneumonia* and comprising of genes *nif* HDKTY, *nif* ENX, *nif* USV, *nif* WZM, *nif* F, also occur in *Azotobacter*. The *nif* ABQ genesare also linked to each in a separate cluster and have been sequenced in *Azotobacter*.*A. chroococcum*is capable of producing two nitrogenases – a Mo containing one in N-free molybdenum added medium, a Va containing one in the presence of vanadium and absence of molybdenum. *A. vinelandii*c an produce three enzymes–a Mo nitrogenase, a Va nitrogenase and one without either of these two metals in a medium devoid of those metals. Even though *nif* regulation in *Azotobacter* appears to be considerably similar to *K. pneumoniae*, the situation is complicated because of the unknown mechanisms through which the three types of nitrogenases are synthesized by the bacteria.

13.4 REGULATION OF *NIF* GENES

Nitrogen fixation is regulated by the expression or repression of transcription of *nif* structural genes. A positive regulatory protein encoded by *nif* A gene involves in positive regulation (expression of transcription) and *nif* L protein involves in negative regulation (repression of transcription). In the presence of a suitable source of fixed nitrogen, such as ammonia, transcription of the *nif* genes is shut down or repressed. As a result of positive regulation *nif* genes are transcripted and N_2 is reduced to ammonia. However, this ammonia produced by nitrogenase does not repress enzyme synthesis as it is incorporated into organic form and used in biosynthesis as soon as it is produced. In certain nitrogen-fixing bacteria, especially phototrophic bacteria, nitrogenase activity is also regulated by ammonia, a phenomenon called the 'ammonia-switch off effect'. In this case, excess of ammonia catalyzes a covalent modification of nitrogenase reductase, which results in a loss of enzyme activity. Again when ammonia becomes limiting, modified protein is converted back to the active form and N_2fixation resumes.

The *nif* gene regulation is very complex, involving both local control by genes within the *nif* complex and more global control by regulatory genes located elsewhere in the genome. Like the other genes, *nif* genes for their expression require an RNA polymerase enzyme to transcribe the DNA into messenger RNA. The regulatory genes that involve in global control are *ntr*A, *ntr*B, *ntr*C and glnA genes. During initiation of transcription, RNA polymerase bindsto the promoters that present at the beginning of the genes. The nucleotide sequences of the *nif* gene promoters are very different from those of other genes, such that of *E. coli*. So, the RNA polymerase that recognizes the *nif* gene promoters is somewhat different from the enzymes that bind to the promoters of other genes. The *ntr* A genecodes for a protein called as 'sigma factor' which confers on RNA polymerase the ability to recognize the *nif* gene promoter. Without this protein, the *nif* genes including the regulatory *nif* A and *nif* L could not be transcribed and nitrogen fixation would not occur. So, mutations in *ntr* A gene completely abolish nitrogen fixation.

The *nif* A gene, from within the *nif* gene complex, makes a protein that is necessary for inducing *nif* gene transcription. However, in presence of oxygen and ammonia, this induction by *nif* A gene is prevented or inhibited by *nif*L gene product. In the absence of oxygen and other fixed nitrogen sources, the product of *ntr*C gene in concert with *ntr*A product activates the *nif* A,L promoter. But in high ammonia, *ntr*B product coded by *ntr*B gene blocks the activation of *nif* A,L promoters by *ntr*C. The *ntr*B and *ntr*C genes are transcribed by either of two promoters that present, of which one (P1) also involves in the transcription of glnA gene that codes for glutamine synthetase. Thus, the expression of *nif* genes that involve in N_2-fixation is regulated or controlled locally by *nif* A and *nif* L genes that present in the *nif* gene cluster and globally by *ntr*A, *ntr*B and *ntr*C that occur as a part of global system, but located elsewhere in the genome.

13.5 NODULE FORMATION IN SYMBIOTIC N_2-FIXATION

Nitrogen can also be fixed or reduced to ammonia in a symbiotic association between certain bacteria and higher plants. The best known and well-studied example is Legume-*Rhizobium* symbiosis. The reduction of N_2 to ammonia in this process is called as symbiotic nitrogen fixation and the organisms involved are known as symbiotic nitrogen fixers. The bacteria that involve in legume-*Rhizobium* symbiosis include species of genera, *Rhizobium* and *Bradyrhizobium*. Both of them are Gram-negative motile rods. *Rhizobium* consists fast growing species and the genus *Bradyrhizobium* includes slow growing species. Both of them are able to form nodules on roots of legume plants as a result of interaction or association.

Initially, a specific adhesion protein called Ricadhes in which is present on surfaces all species of *Rhizobium* and *Bradyrhizobium* binds to the calcium complexes on root hair surface. Lectins are the other substances identified on the surfaces of both plant roots and bacteria and so involve in interaction. In this first step, the bacterium infects the growing root hair at its tip. As a response the root hair curls and an infection thread is produced by plant. Then bacterium invades into infection thread, multiply and induces the formation of incipient nodule. As the nodule develops, the bacterial cells transform into swollen, misshapen cells called 'bacteroids'. These bacteroids, either singly or in small groups, surrounded by plant produced cell membrane called 'peribacteroid membrane'. Nitro genase is localized within the bacteroids themselves and is not released into plant cytosol. The bacteria in this form of bacteroids turn on the nif genes and carryout the nitrogen fixation. *Rhizobium* needs O_2 to generate energy for N_2 –fixation. This needed O_2 is supplied to bacteroids in nodule by the O_2-binding protein called 'leghemoglobin'. This leghemoglobin is a red, iron containing protein which gives the characteristic pink coloration to the healthy nodules. The heme portion of the leghemoglobin is synthesized by bacterium and globin portion is by plant. This leghemoglobin not only supply the necessary $O_2$to bacteroids, but acts as oxygen buffer and involves in the maintenance of low but constant levels of oxygen within the nodule to avoid the inactivation of nitrogenase enzyme.

13.6 GENETICS OF NODULE FORMATION AND REGULATION

Genes directing specific steps in nodulation of a legume by a strain of *Rhizobium* have been called '*nod* genes'. Many *nod* genes from different *Rhizobium* species are highly conserved and are borne on large plasmids called 'Sym plasmids'. In addition to *nod* genes which direct specific nodulation events, Sym plasmids contain specificity gens, which restrict a strain of *Rhizobium* to a particular host plant. In the Sym plasmid of *Rhizobium leguminosarum*, *nod* genes are located between two clusters of genes for nitrogen fixation, the *nif* genes. The arrangement of *nod* gens in the *R. leguminosarum* Sym plasmid is shown in figure 13.2. In this species, ten nod genes have been identified. The entire *nod* region has been sequenced and the functions of many nod proteins are known, the genes *nod*ABC direct root hair curling by production of specific polysaccharide signal molecules called NOD factors. The *nod*D gene encodes a regulatory protein that controls expression of other nod genes. In the presence of specific nodulation signals, the *nod*D gene product controls the transcription of other nod genes.

In most of the rhizobial species the NodD protein is constitutive and regulates transcription of other nod functions by a simple induction type of control. The NodF protein is involved in the synthesis of a lipid or polysaccharide needed for effective nodulation and is thought to be involved in the transport of substances needed to begin nodulation. The *nod*E and *nod*L genes function in determining host range. The *nod*M gene encodes for glucosamine synthase involves in nod factor synthesis. The two other genes, *nod*I and *nod*J encode membrane proteins that involves in the exportation of Nod factors from bacterial cells. Cloning and sequencing of nod genes in *R. leguminosarum* suggests that they are arranged in at least three operons.

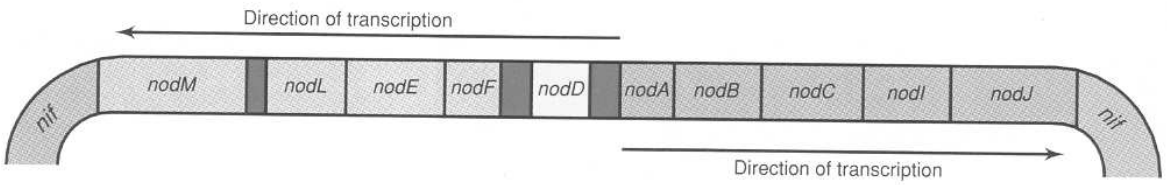


Fig-13.2: nod gene cluster in *R.leguminosarum*

As a regulatory protein, the NodD protein binds to a highly conserved 35 base pair DNA sequence which precedes each operon. These regions, referred to as nod boxes, are presumably operator genes to which the NodD protein can bind. Following the interaction with inducer molecules, the conformation of the NodD protein changes, causing it to fall off the nod box DNA. This event presumably exposes a promoter site for RNA polymerase binding which initiates transcription of nod structural genes. Several inducer molecules have been identified. In most cases, these are plant flavonoids, complex organic molecules that are widespread plant products. These flavonoids have many functions in plants, including growth regulation and attraction of pollinating animals. But in the case of leguminous plants, their roots secrete large amounts of flavonoids, presumably to trigger nod gene expression in nearby rhizobial cells in the soil. Interestingly, some flavonoids that are structurally very closely related to *nod*D inducers strongly inhibit induction of nod genesin certain rhizobial species, suggesting that part of the specificity observed between plant and bacterium in the *Rhizobium*-legume symbiosis could lie in the chemical nature of the flavonoids excreted by a particular plant.

The *nod* D gene regulates the expression of other nod genes by encoding a regulatory protein which binds to the promoters of other nod genes. The *nod*D gene can also regulate its own transcription. The expression of other nod genes by *nod*D protein increases to several folds in the presence of inducers such as flavones or flavanones that present in the plant exudates. The *nod*D gene, at least in *R. leguminosarum*, is subject to negative regulation by its product, as the high concentrations of the product repress transcription of the *nod*D gene. The *nod*D gene product activates the other genes for expression. Mutations in *nod*D, A,B,C genes completely abolish the nodule formation. However, mutations in other genes delay the onset of nodulation and reduce the number of nodules per plant. Despite the differences in the host range specificities of various *Rhizobium* species, their corresponding nod genes are similar sequence, location and function.The regulatory circuitry that affects *nif* gene expression in *Rhizobium* species is clearly very similar to that in *K.pneumoniae*. The sequences of regulatory nifA gene and *nif* promoters of both *Rhizobium* and *Klebsiella* are also found to be similar. Regulatory mechanism of nod genes by nod D is given in the figure 13.3 below.

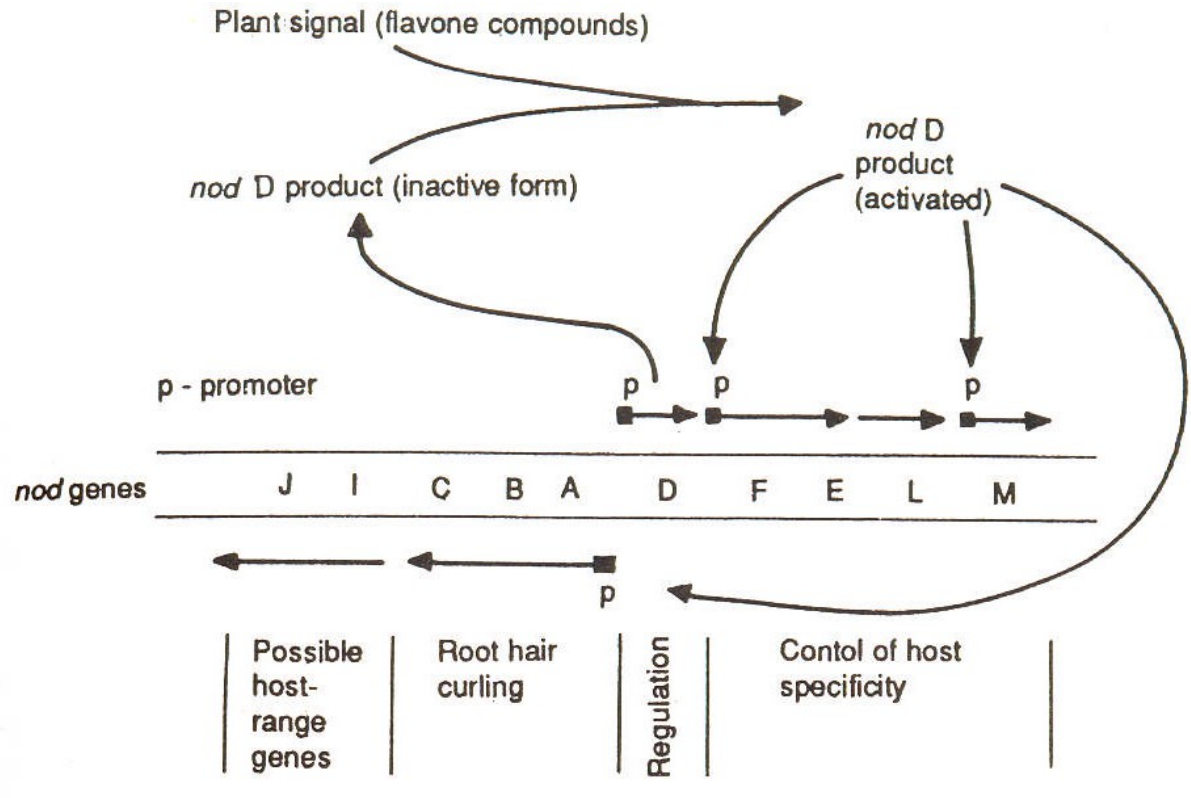


Fig-13.3 Regulation of nod genes in *R. leguminosarum*

13.7 SUMMARY:

The conversion of gaseous nitrogen to utilizable nitrogen source is referred as nitrogen fixation. This event can be carried out only by certain prokaryotes, either free living and symbiotic nitrogen fixers. A number of genes involve in the nitrogen fixation and are called as nif genes. The nitrogen fixation is catalyzed by the enzyme complex called nitrogenase. The transcription of these nif genes is regulated locally and globally by the regulatory genes. In the case of symbiotic nitrogen fixation, nodules are formed by the bacterium on the roots of plants. The bacteria are present in the root nodules in the form of bacteroids which contain the nitrogenase enzyme and act as the sites for the nitrogen fixation. The nodule formation is catalyzed by a number of protein products coded by the genes called nod genes.

13.8 TECHNICAL TERMS:

Rhizobium, *Bradyrhizobium*, Nif genes, nod genes, Symbiotic association.

13.9 SELF ASSESSMENT QUESTIONS:

1) What is nif regulon? Write an account on the role of nif genes and their regulation.
2) Write an essay on nod genes and their role in nodulation process in *R. leguminosarum*.

13.10 SUGGESTED READINGS:

1) Brock, T.D., Madigan, M.T., Martinko, J.M. and Parker, J. Biology of Micro organisms (1994) – Prentice Hall, New Jersey.

2) Subba Rao, N.S. Soil Microbiology (2001) – Oxford & IBH Publishing Co. Pvt. Ltd., New Delhi.

3) Atlas, R.M., and Bartha, R. Microbial Ecology – Fundamentals and Applications (2000)- Benjamin / Cummings Science Publishing, California.

Prof. V. Uma Maheswara Rao

LESSON-14

TRANSPOSABLE ELEMENTS IN BACTERIA

14.0 OBJECTIVE:

- Students will be enriched with the knowledge on different transposable elements in bacteria.

STRUCTURE:

14.1 INTRODUCTION

Eukaryotic genomes contain a substantial amount of repeated DNA, and some of these repeated sequences are capable of moving. Transposable elements (TEs) are defined as DNA sequences that can move from one location to another in the genome. TEs have been identified in all organisms, prokaryotic and eukaryotic, and can occupy a high proportion of a species' genome. For example, transposable elements comprise approximately 10% of the genomes of several fish species, 12% of the *C. elegans* genome, 37% of the mouse genome, 45% of the human genome, and up to more than 80% of the genomes of some plants, such as maize. From bacteria to humans, transposable elements have accumulated over time and continue to shape genomes through their mobilization.

TEs were discovered by Barbara McClintock (Fig. 14.1) during experiments conducted in 1944 on maize. Since they appeared to influence phenotypic traits, she named them controlling elements. However, her discovery was met with less than enthusiastic reception by the genetic community. Her presentation at the 1951 Cold Spring Harbor Symposium was not understood and was at least not very well received. She had no better luck with her follow-up publications, and after several years of frustration, decided not to publish on the subject for the next two decades. Not for the first time in the history of science, an unappreciated discovery was brought back to life after another discovery had been made. In this case, it was the discovery of insertion sequences (IS) in bacteria by the Szybalski

group in the early 1970s. In the original paper, they wrote: “Genetic elements were found in higher organisms which appear to be readily transposed from one to another site in the genome. Such elements, identifiable by their controlling functions, were described by McClintock in maize. They might be somehow analogous to the presently studied IS insertions”. The importance of McClintock’s original work was eventually appreciated by the genetic community with numerous awards, including 14 honorary doctoral degrees and a Nobel Prize in 1983 “for her discovery of mobile genetic elements”.

Fig-14.1: Barbara McClintock (1902-1992)

The mobilization of TEs is termed transposition or retro transposition, depending on the nature of the intermediate used for mobilization. There are several ways in which the activity of TEs can positively and negatively impact a genome; for example, TE mobilization can promote gene inactivation, modulate gene expression or induce illegitimate recombination. Thus, TEs have played a significant role in genome evolution. For example, DNA transposons can inactivate or alter the expression of genes by insertion within introns, exons, or regulatory regions. In addition, TEs can participate in genome reorganization by mobilizing non-transposon DNA or by acting as substrates for recombination. This recombination occurs through homology between two sequences of a transposon located on the same or different chromosomes, which may be the origin of various types of chromosome alterations. Indeed, TEs can contribute to the loss of genomic DNA through internal deletions or other mechanisms.

14.2 TYPES OF TRANSPOSABLE ELEMENTS IN BACTERIA

Some salient features of transposable elements are:

1) These are the DNA sequences that code for enzymes which result in self-duplication and insertion into a new DNA site.
2) Transposons are involved in transposition events which include both recombination and replication, which usually generates two copies of the original transposable elements. One of the copies remains at the parent site, whereas the other one reaches the target site on the host chromosome.

3) The integrity of the target genes of these elements is invariably disrupted by the presence of those elements.
4) Because transposons carry the genes for initiation of RNA synthesis, some previously dormant genes might be activated.
5) A transposable element doesn't have a site for the origin of replication. As a result, it cannot replicate without the host chromosome as plasmids or phages.
6) There is no homology between the transposon and its target site for insertion. These elements can insert at almost any position in the host chromosome or a plasmid. Some transposons might seem likely to enter at some specific positions (hot spots), they barely insert at base-specific target sites.

14.2.1 Insertion Sequences (IS elements)

The insertion sequences (IS) or simple transposons (Fig.14.2) are shorter sequences (800 to less than 2,500 base pairs (bp).These sequences carry the genetic information necessary for their transposition. Insertion sequences have been identified in bacteriophages, in F factor plasmid and many bacteria. IS elements are the simplest type of bacterial transposable sequences that can insert at different location of bacterial chromosome and plasmid through illegitimate recombination. They are typically short sequences and contain only one gene that encodes the enzyme for transposition.IS1 first identified in *E.coli*'s galactose operon is 768 bp long and is present with 4-19 copies in the *E.coli* chromosome. Ends of all known IS elements show inverted terminal repeats (ITRs).They are typically short sequences and contain only one gene that encodes the enzyme for transposition. IS elements were first identified as spontaneous insertion in certain lac operon mutations of E. coli which inactivate the gene and inhibit transcription and translation. The mutation of Lac operon gene was found to be unstable and molecular analysis reveals the presence of extra copies of DNA sequences near the lac gene. When the mutated E. coli undergoes reverse mutation, the extra DNA sequence is lost. A bacterial chromosome may contain several copies of a particular type of IS element. For example, 6 to 10 copies of IS1 are found in the *E. coli* chromosome.

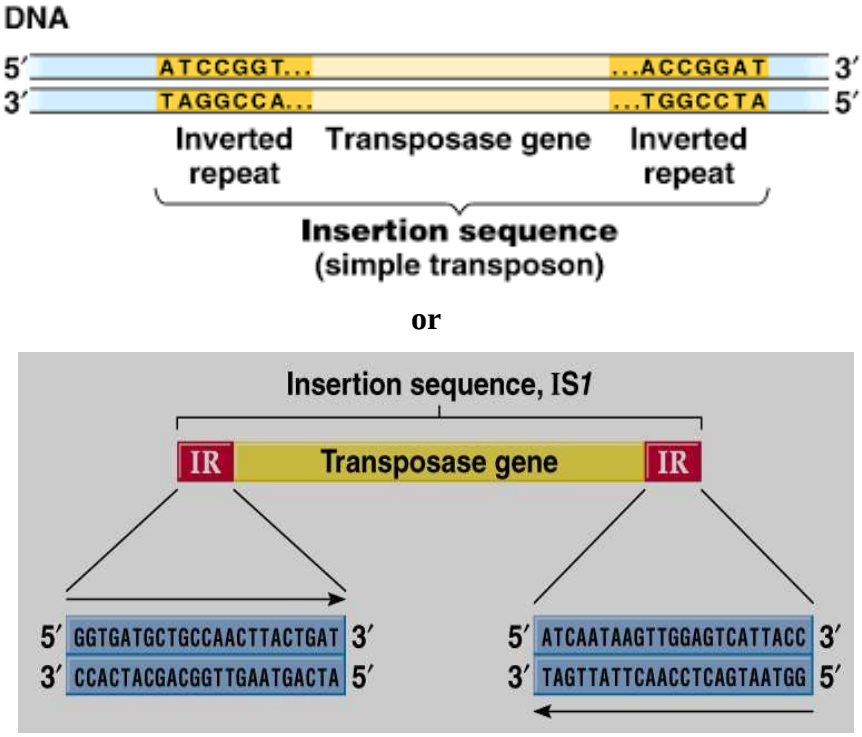


Fig-14.2: Insertion Sequences (IS Element)

Characteristics of Insertion Sequences (IS elements)

IS elements are compactly organized and containing about 1000 nucleotide pairs and contain only genes (open reading frame) which encode for enzyme for regulating transposition. Many distinct types of IS elements have been identified. The smallest IS element is IS*1* which is 768 nucleotide pairs long. Each type of IS element contains inverted terminal repeats at both end and a transposon sequence in between those inverted repeats. Transposon is the only gene that code for transposition of IS element. The inverted terminal repeats is 9-40 base pair long and is the characteristics of most IS element IS element have the capacity to duplicate the inserted sequence at the site of insertion; known as target site duplication.

Transposition of insertion sequence in bacteria

IS element contains single open reading frame (ORF) which encodes for the enzyme transposase, catalyzing its own transposition. The enzyme transposase is like restriction endonuclease which binds to terminal inverted repeats (IR) of IS element which is the restriction site. Then the enzyme cut and excise IS elements from chromosome or plasmid. The excised IS element is mobile in nature and moves along the length of chromosome to recognizes the target site for insertion on same or different chromosome or plasmid. Once recognizing the target site, it generates staggered cleavage (cut the single strand of DNA) generating sticky and itself get inserted. As IS element get inserted, the proofreading mechanism of DNA results in duplication of the DNA sequence at the target site of the insertion such that one copy of target DNA is located on each side of IS element. Thus IS elements helps in target site duplication (Fig.14.3).

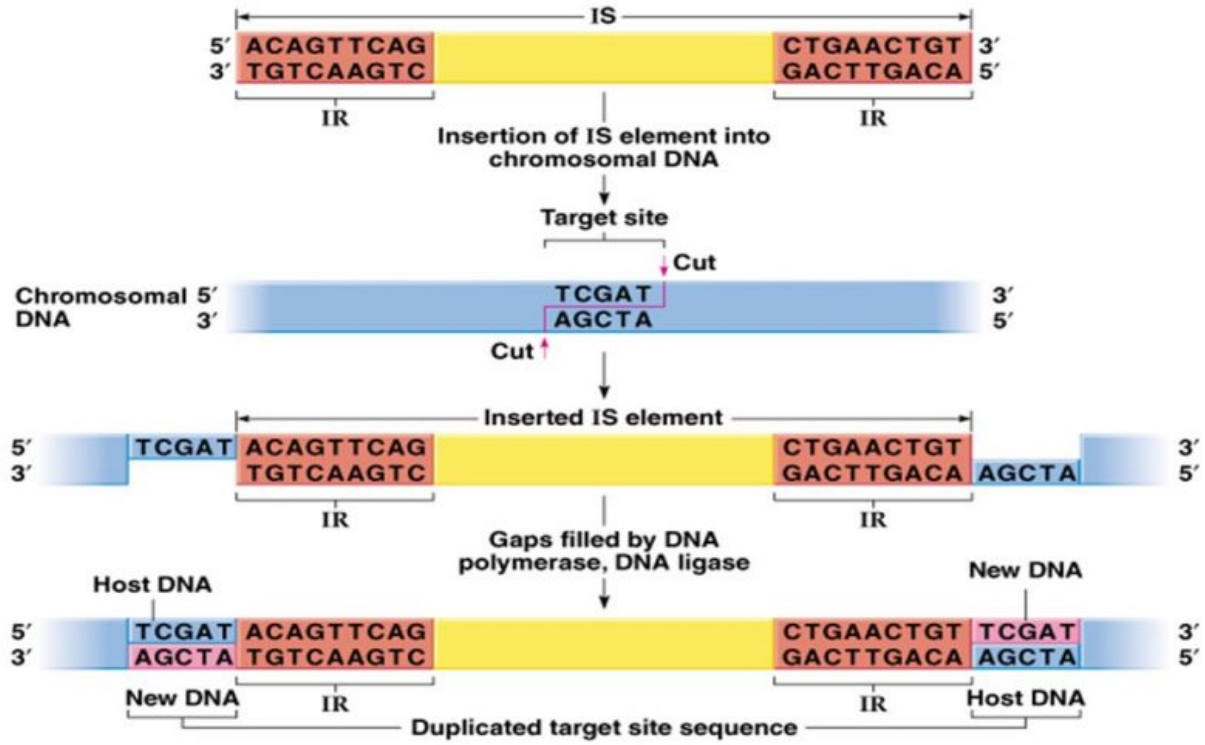


Fig-14.3: Transposition of Insertion Sequence in Bacteria

14.2.2 Composite Transposons

Prokaryotic Transposons are similar to IS element but they are larger and also contain other genes (mostly antibiotic resistance gene) in addition to gene that encode transposase. Transposons are several thousand base pairs long and contain inverted terminal repeats.There are two types of prokaryotic transposons - composite and non-composite transposons.The composite transposons and Tn3-like elements are more complex than IS elements, containing some genes that encode products unrelated to the transposition process. Composite transposons are created when two IS elements insert near each other and the region between the two IS elements can then be transposed when the elements act jointly. For example: Tn10 is a composite transposon of 9.3kbp which contains 1.4 kbp terminal inverted repeats and in between them is gene for transposase and gene for antibiotic resistance (Fig.14.4). These Tn elements may be 1000 bp long and have a complex structure with a central region containing genes that confer resistance to antibiotics. They are flanked by IS elements of same type on both sides which are called ISL for the left one and ISR for the right one. The ISs themselves have terminal inverted repeats in addition to terminal inverted repeats of the composite transposon. So, the ISL and ISR may be in the same or inverted orientation, relative to each other. Here, the IS elements help in the transposition of the composite transposons. The transposase is supplied by one or both IS elements and recognise the inverted repeats of the IS elements at the two ends of the transposon to initiate transposition. Like IS elements, they produce target site duplication after transposition, which is 9 bp long in case of Tn10. Transposition of Tn10 is rare, occurring once in 10^7 cell generations.

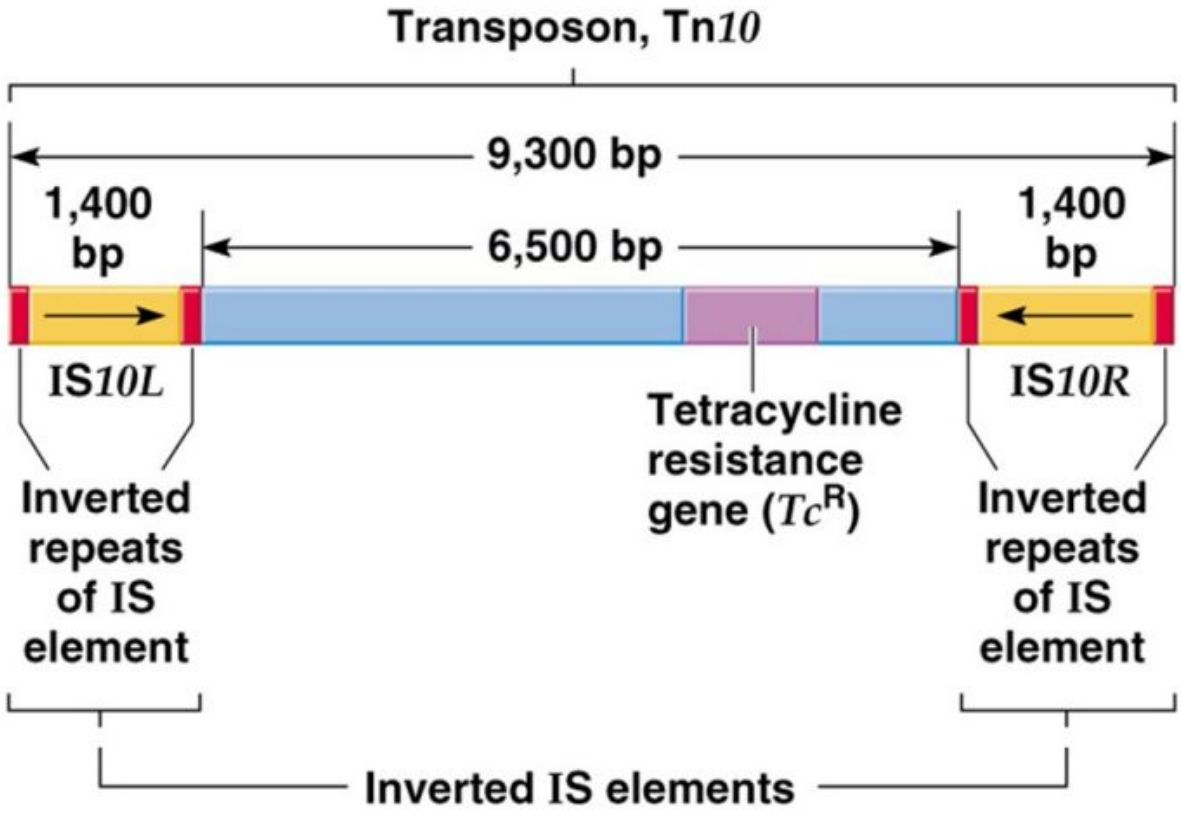


Fig.14.4: Structure of Transposon Tn10

14.2.3 Tn3 Transposons

Unlike the composite transposon, non-composite transposons do not contain IS elements at their ends, but has the repeated sequences at their ends that are required for transposition. The non-composite transposons are a sequence of DNA containing gene for trasnposase and multiple other genes in between terminal inverted repeats. Unlike composite transposons, it does not contains IS elements at each end but instead it contains simple inverted repeats 0f 38-40 nucleotide pairs at each end. For example, Tn3 (Fig.14.5) is non-composite transposon of 5kbp which contains three genes for beta-lactamase (bla), transposase (tnpA) and resolvase (tnpB). The beta lactamase provide resistance to the antibiotic ampicillin, and the other two enzymes play important roles in transposition and recombination.

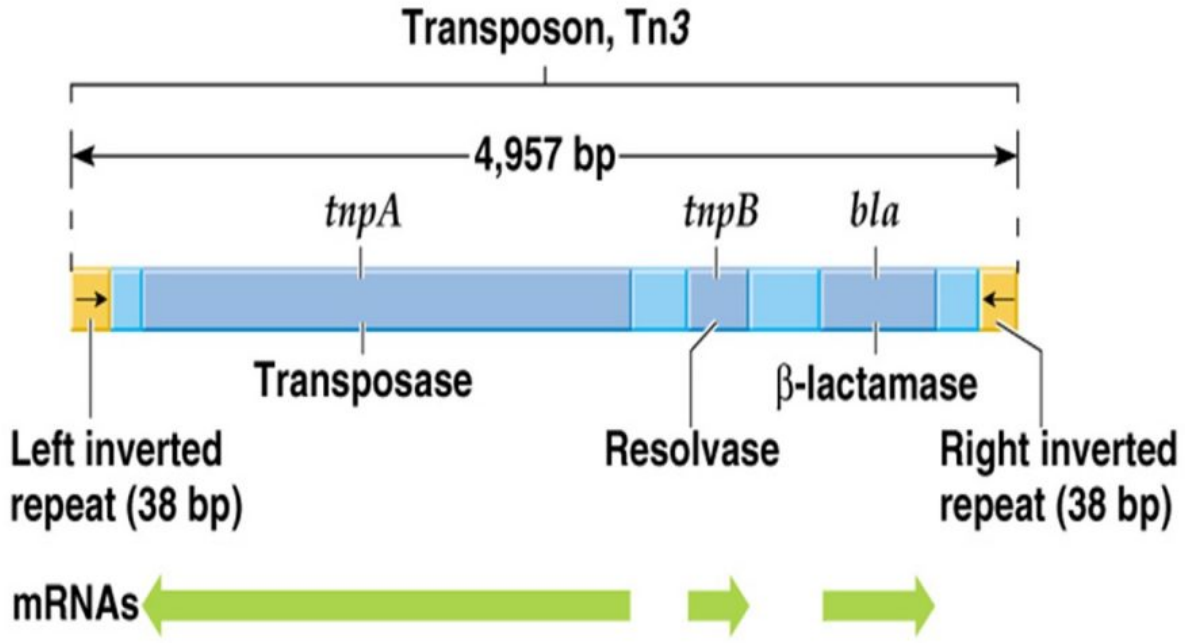


Fig-14.5: Structure of Non-Composite Transposon Tn3

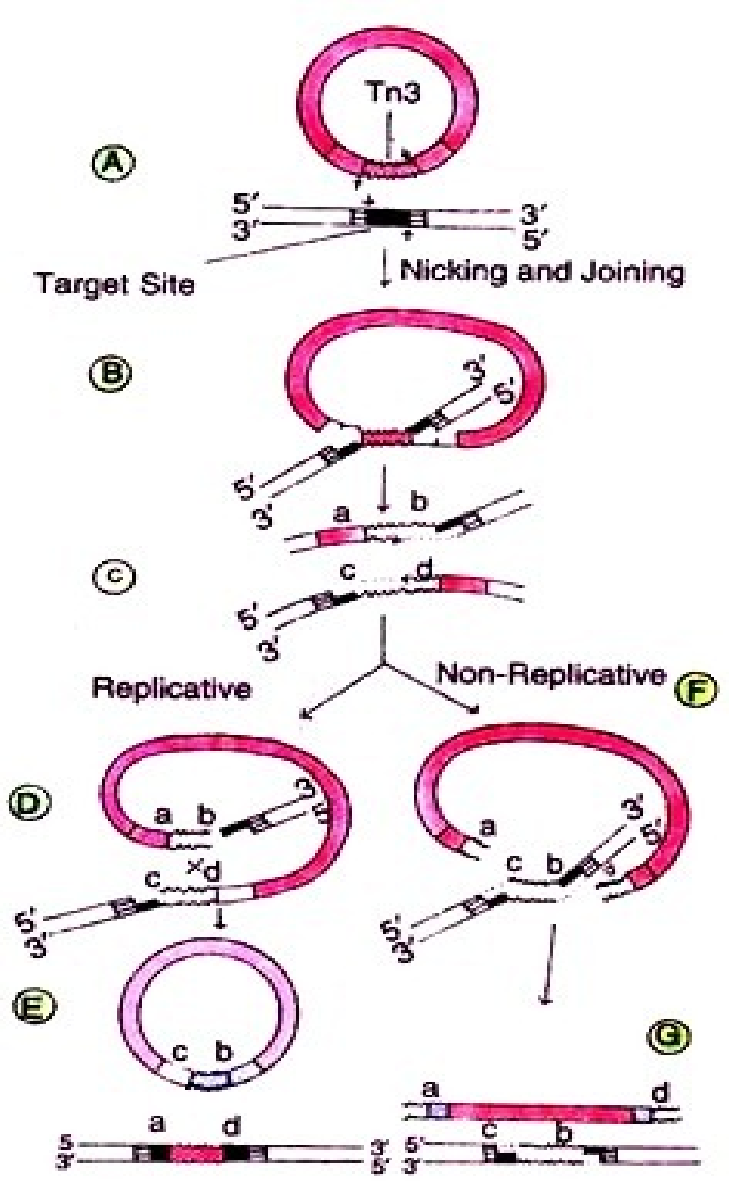
14.3 MECHANISM OF TRANSPOSITION

The phenomenon of moving genetic segments from one location to the other in a genome is known as transposition. There are different type of transposable elements depending upon structure and mechanism. On this basis, there are three different mechanism of transposition (Replicative, conservative and Retro-transposition). Most prokaryotes and eukaryotes employed one of two-either replicative or conservative mechanism of transposition whereas retro-transposition is only employed by eukaryotes. The bacterial transposon Tn3 has been extensively studied. Analysis of DNA sequences and its junction with target DNA provides some clue to the mechanism of transposition.

Movement of transposons occurs only when the enzyme transposase recognises and cleaves at either 5′ or 3′ of both ends of transposon, and catalyses at either 5 or 3′ of both ends of transposon and catalyses a staggered cut at the target site. Depending on transposon, a duplication of 3-12 bases of target DNA occurs at the site where insertion is to be done. One copy remains at each end of the transposon sequence. After attachment of both ends of

transposon to the target site, two replication forks are immediately formed (Fig. 1). From this stage there start two paths for carrying out onward processes.

The first model is the replication path where the transposon replicates and the replicated DNA sealed to flanking sequences generating a co-integrate (D). Co-integrate is resolved by the genetic exchange between the two copies of transposon resulting in a simple insertion and regeneration of donor replicon (E). This model explains the transposition of only TnA family but not explain completely for IS elements of Mu. The second model is the non-replicative path that generates simple insertions without formation of co-integrate. At the prime termini in the target DNA, repair synthesis occurs. The displaced single strand that attaches the transposon to the donor replicon is broken. This forms a simple insertion (G). It is likely that both the pathways can be used but the frequency of simple insertion and co-integrate formation varies. Thus, for transposition the two enzymes, transposase and resolvase coded by tnpA and tnpR respectively are required. Transposase recognises the ends of transposon and connects them to the target site. Resolvase provides a site-specific recombination function.



A Model Transposition Mechanism of Transposon Tn3

Genetics of Transposition:

The genes of transposase and resolvase i.e. tnpA and tnpR are identified by recessive mutations. The above enzymes accomplish the two stages of TnA mediated transposition. Like IS type elements the transposition stage involves the ends of the elements. A unique feature of TnA family is that a specific internal site is required for resolution. The mutants of tnpA cannot transpose because the enzyme transposase will not be encoded. However, transposase recognises the ends of elements and binds to 25 bp long sequence located within 38 bp of the inverted terminal repeat. Transposase also makes the staggered 5 bp breaks in target DNA where transposon is to be inserted.

These are Two Types:

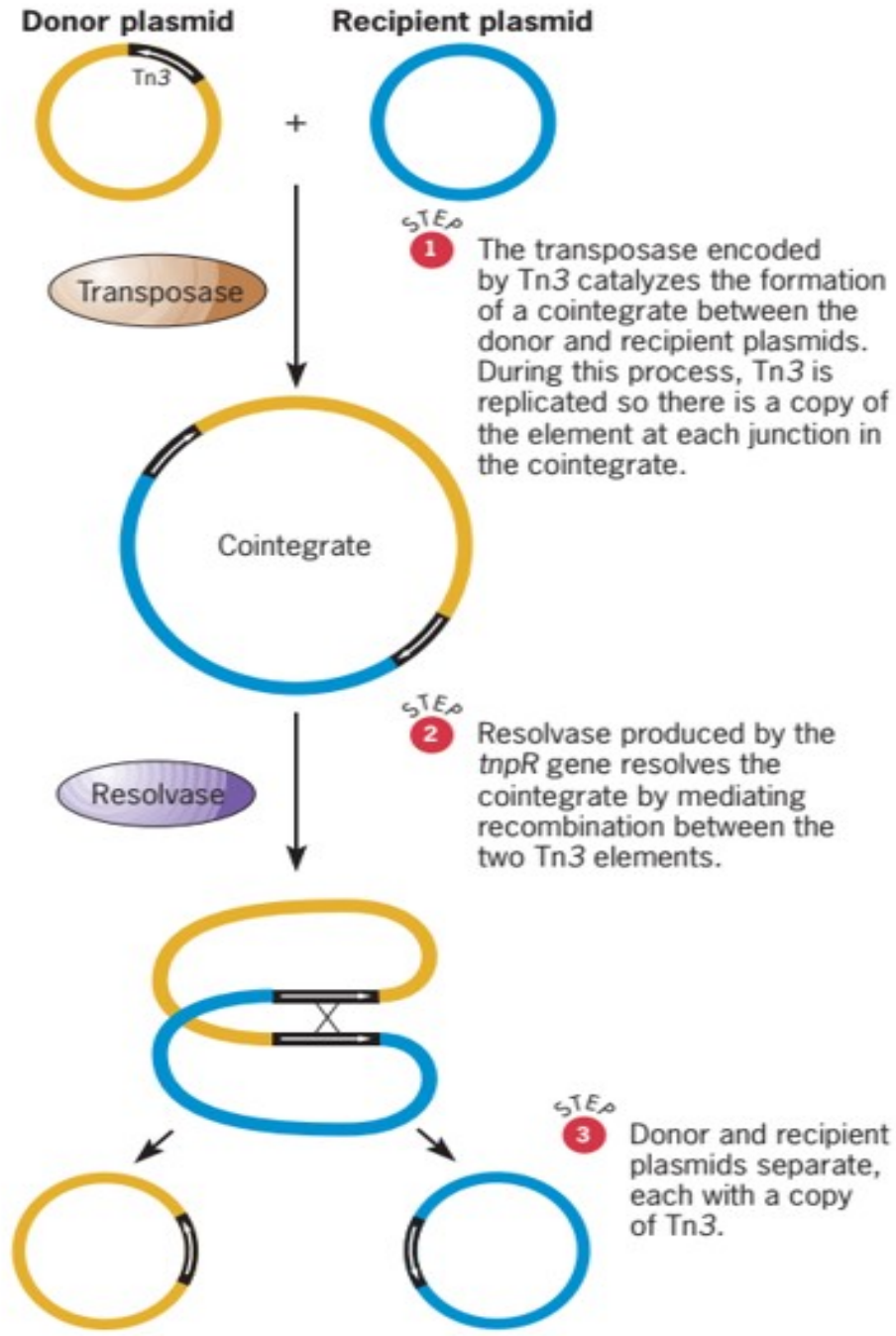
Replicative transposition and Conservative or non-replicative transposition.

Replicative Transposition or Copy Paste Transposition:

The replicative transposition involves the events of both replication and recombination processes generating the two daughter copies of the original transposable elements, one remaining at the parental site and the other at the target site. Replicative transposons are those transposons which at first replicate itself and then insert one copy of it into new position. Therefore, during replicative transposition, the transposable elements are replicated and one copy is inserted into new position while one copy remains at original position. This transposition is catalyzed by an enzyme transposase which is encoded by the transposable element itself. It is also known as copy and paste transposition. Replicative transposition is present only in prokaryotic organisms.

Example:

Non-composite transposons (Tn3) are replicative transposons that undergo transposition in two stage process. In the first stage, two plasmid- (one containing Tn3 transposons; donor plasmid) and the other recipient plasmid undergoes fusion catalyzed by transposase enzymes giving rise to a structure called co-integrate. During the formation co-integrate, Tn3 is replicated, and one copy is inserted at each point where the two plasmids have fused. In the second stage of transposition, the tnpR-encoded enzyme resolvase which mediates a site-specific recombination between the two Tn3 copies at the resolution site, and when it is completed, co-integrate is resolved into its two constituent plasmids, each with a copy of Tn3.



Transposition of Tn3 via the Formation of a Co-Integrate

Conservative or Non-Replicative: Cut and Paste Transposition

The conservative transposons are those which changes position by excision from one position and then insertion into another position within a chromosomes. Therefore, during conservative transposition, the transposable element is physically cut from its original position and paste into another position in a chromosome. It is also known as cut and paste transposition**.** This type of transposition is catalyzed by an enzyme called transposase which is coded by such transposable element itself. Conservative transposition is present in both prokaryotic and eukaryotic organisms.

14.4 USES OF TRANSPOSONS

1) **Mutagenesis:** Since it is clear that transposons transpose to new sites at certain frequencies, one might expect that you "mutagenize" a strain by taking a strain carrying the transposon (in a site considered "uninteresting" in terms of the system being studied) and looking for cases where the Tn has moved and generated an "interesting" mutation. Such a situation is achieved if the Tn is introduced into a recipient cell on a non-replicating vector. The vector should also not be able to easily integrate itself into a replicon, as would be the case for a specialized phage. In this case, the only way for a "stably" drug-resistant cell to exist is for the Tn to become associated with a replicon by transposition.
2) **Selection for a Mutant Phenotype:** Most mutations leading to the loss of gene function do not have a selectable phenotype. This means that they cannot easily be moved from one strain to another using the gene transfer. A mutation caused by a transposon therefore has a distinct advantage, since it has two phenotypes: the loss of function of the affected gene and the drug-resistance of the transposon. The latter is selectable, but brings the former along.
3) **Cleanliness:** As a mode of mutagenesis, transposons are relatively clean; that is, in contrast to other forms of mutagenesis they do not cause a large number of alterations in the genome.
4) **Organism Specificity:** Because most MGEs borrow some host machinery during transposition, they tend to be limited to hosts with "compatible" machinery. Most MGEs found in Gram-negative bacteria therefore tend to be specific for that group and fail to function in Gram- positives.
5) **Genetically Altered Transposons:** An increasing number of genetically altered transposons are being produced with useful and amusing properties.
6) **Physically Detectable Homology:** When a transposon is used to generate a mutant, the mutated region can be physically isolated using probes for the transposon sequence. The mutated region can then be used in turn as a probe for the wild-type sequence.
7) **Role in Antibiotic Resistance:** As many bacterial transposons also carry genes for antibiotic resistance apart from gene for transposase and by the nature of transposable element, the genes move or translocate from one DNA to another and from chromosome to plasmid and vice versa. This results in genetic flux of antibiotic resistance genes in bacterial population. When the transposons undergo recombination with plasmid vector within a bacterial cell then it can be transformed horizontally or vertically to other bacteria, spreading the drug resistance gene in bacterial population. This creates multidrug resistance pathogenic bacteria such that diseases become difficult to control.

14.5 SUMMARY:

Transposable elements (TEs) are defined as DNA sequences that can move from one location to another in the genome. TEs have been identified in all organisms, prokaryotic and eukaryotic, and can occupy a high proportion of a species' genome. TEs were discovered by Barbara McClintock during experiments conducted in 1944 on maize. Since they appeared to influence phenotypic traits, she named them controlling elements. However, her discovery was met with less than enthusiastic reception by the genetic community. IS elements are compactly organized and containing about 1000 nucleotide pairs and contain only genes (open reading frame) which encode for enzyme for regulating transposition. Many distinct

types of IS elements have been identified. The smallest IS element is IS*1* which is 768 nucleotide pairs long. Prokaryotic Transposons are similar to IS element but they are larger and also contain other genes (mostly antibiotic resistance gene) in addition to gene that encode transposase. Transposons are several thousand base pairs long and contain inverted terminal repeats. There are two types of prokaryotic transposons- composite and non-composite transposons. Unlike the composite transposon, non-composite transposons do not contain IS elements at their ends, but has the repeated sequences at their ends that are required for transposition. The non-composite transposons are a sequence of DNA containing gene for trasnposase and multiple other genes in between terminal inverted repeats.

14.6 TECHNICAL TERMS:

Transposable elements (TEs), Insertion sequences (IS elements), Prokaryotic Transposons, Composite transposons, Non-composite transposons.

14.7 SELF ASSESSMENT QUESTIONS:

1) Define Transposable elements? Explain in detail about Transposable elements of Bacteria.
2) Explain about Insertion sequences (IS elements).
3) Describe the Composite transposons.
4) Discuss about Tn3 transposons.

14.8 SUGGESTED READINGS:

1) Freifelder, D. Molecular Biology (1990) – Narosa Publishing House, New Delhi.
2) Gardner, E.J., Simmons, M.J. and Snustad,D.P. Principles of Genetics (2001) - John Wiley & Sons, Inc., New York
3) Snustad, D.P., Simmons, M.J.and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4) Russel, P.J. Genetics (1998) – The Benjamin / Cummings Publishing Company, Inc., California
5) Singh, B.D. Fundamentals of Genetics (2001) – Kalyani Publishers, New Delhi.

Prof. K. Mallikarjuna

LESSON-15

TRANSPOSABLE ELEMENTS IN EUKARYOTES

15.0 OBJECTIVE:

- This lesson deals with concept of general account and mechanism of diffident types of Transposable elements in eukaryotes.

STRUCTURE:

15.1 INTRODUCTION

Transposable elements (TEs) are defined as DNA sequences that can move from one location to another in the genome. TEs have been identified in all organisms, prokaryotic and eukaryotic, and can occupy a high proportion of a species' genome. For example, transposable elements comprise approximately 10% of the genomes of several fish species, 12% of the *C. elegans* genome, 37% of the mouse genome, 45% of the human genome, and up to more than 80% of the genomes of some plants, such as maize. From bacteria to humans, transposable elements have accumulated over time and continue to shape genomes through their mobilization. TEs were discovered by Barbara McClintock during experiments conducted in 1944 on maize. Since they appeared to influence phenotypic traits, she named them controlling elements. However, her discovery was met with less than enthusiastic reception by the genetic community. There are two classes of transposons in eukaryotes. Class I (RNA) transposons move (i.e., jump) by transcription of RNA at one locus, followed by reverse transcription and integration of the cDNA back into genomic DNA at a different location, called as retrotransposons. They may be derived from (or be the source of) retroviruses, since active retroviruses excise from and integrate into DNA much like retrotransposons.

Class II (DNA) transposons move by either of two mechanisms. In the cut-and-paste pathway, the transposon leaves one locus and integrates at another. In the replicative pathway, the original transposon remains in place, but new copies are mobile. The table confirms that bacteria contain relatively few transposons, in contrast to eukaryotes, which

vary widely in transposon load (transposons as a percentage of genomic DNA). Transposon load can range from as low as 4% to more than 70% in different organisms. DNA transposons move from one place to another in one of two ways. In Cut and Paste Transposition, the element excises and moves to another location in the genome. In Replicative Transposition, DNA transposons are copied and the copy transposes to a new location, leaving the original element in place. Retrotransposons are active, if their transcripts are reverse transcribed into cDNAs as well as translated into the enzymes required to integrate their cDNA copies into genomic DNA. Many transposons are inactive, having been silenced by mutation or other factors. Active eukaryotic Class I or Class II transposons are either autonomous or non-autonomous. Autonomous transposons have all the structural features necessary for transposition (e.g., the maize Ac element). Non-Autonomous transposons can have all the structural elements of autonomous transposons (e.g., inverted repeats and other DNA necessary for transposition), except they lack or can't transcribe one or more of the genes for enzymes needed for mobility (e.g., the maize Ds element). Nevertheless, they can be mobilized with the assistance of an actively transposing autonomous element that can provide the missing enzymes.

15.2 TYPES OF TRANSPOSABLE ELEMENTS IN EUKARYOTES

15.2.1 Ac and Dc Elements in Maize

In 1940's Barbara McClintock discovered changes in maize genome during somatic cell division. The changes were genetically controlled aberrations, such as, deficiencies, duplications, inversions, translocations and ring chromosomes. These changes were found to be caused by a genetic system named Dissociation-Activator (Ds-Ac) system. McClintock termed these genetic elements Ds and Ac as controlling elements in 1956. Since then, several systems of controlling elements have been discovered in maize. These elements are classified into two groups: autonomous and non-autonomous. The controlling elements which have the ability of their own excision and transposition are called autonomous elements, e.g., Activator (Ac), Suppressor mutator (Spm) and Enhancer (En).The Non-autonomous elements do not have the ability of transposition. Non-autonomous elements have originated from autonomous elements through the loss of transacting functions which are required for transposition. A single type of autonomous element and different non-autonomous elements derived from it form a family.

Deletions of different lengths and different regions from an autonomous element give rise to different types of non-autonomous elements. Such elements change their position in response to an autonomous element of the same family present in the genome. Non-autonomous element is activated in trans by its related autonomous element. Examples of non-autonomous elements are Dissociation (Ds), defective suppressor mutator (dSpm), and Inhibitor (I).

Activator (Ac) Elements:

Activator (Ac) elements are 4563 bp long and autonomous in action. It has 11 bp inverted repeats at its both ends (Fig.16.1). The target site for Ac insertion is 8 bp long; this target sequence is duplicated during the insertion as direct repeats. The Ac element has 5 exons; transcription produces and mRNA of 3500 bases which has a coding sequence for 807 codons. This element has two open reading frames.

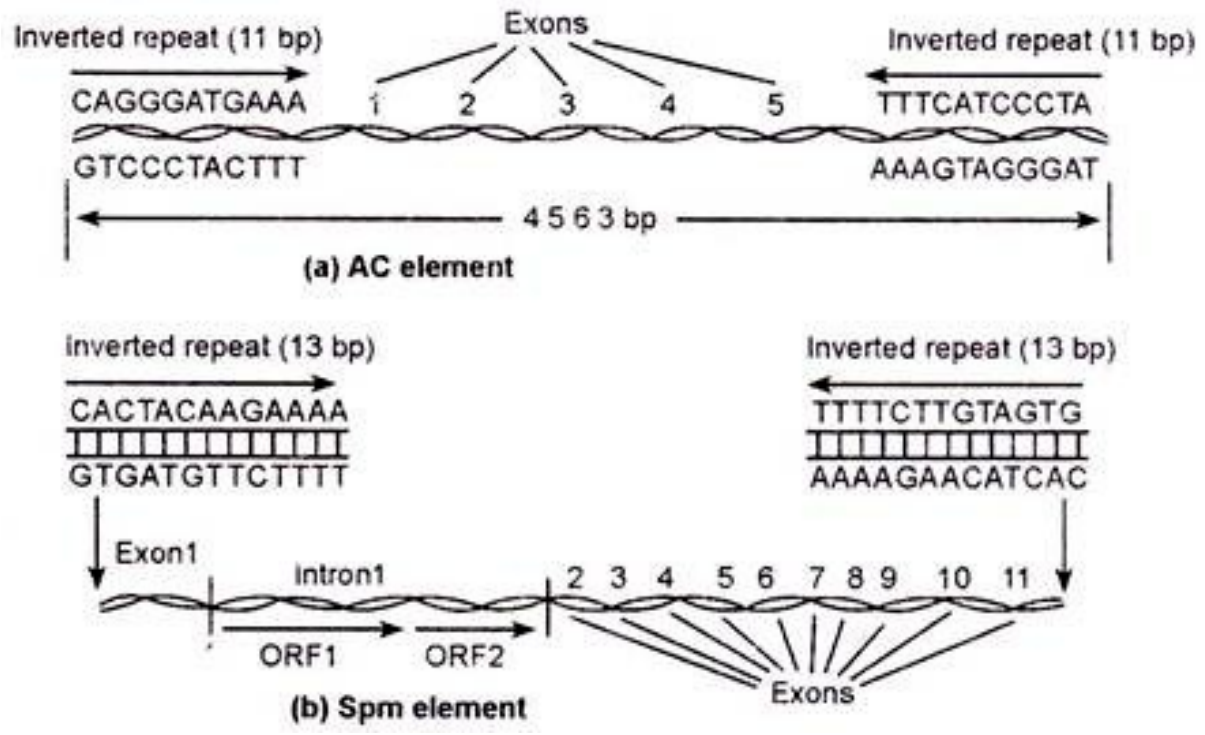


Fig-15.1: Ac and Spm Elements in Maize

Dissociation (Ds) Elements:

These elements are non-autonomous and are produced through interstitial deletions in the Ac element (Fig.15.2a). Based on the length and the region of deletion, Ds elements are grouped into several types as, Ds1, Ds2, Ds6, Ds9 Ds 2dl, and Ds2d2 etc. (Fig.15.2b). All the Ds elements contain the 11 bp inverted repeats at their ends. The Ds1 element represents an extreme case in that it has a large interstitial deletion so that only the terminal 11bp inverted repeats are retained. Ds6 element possesses 1000 bp from each end of the Ac, the rest portion being deleted. Ds9 on the other hand, represents a very short deletion of about 194 bp. Further changes may also occur in the non-autonomous elements leaving them incapable of transposition, i.e., they become permanently stabilized. Autonomous elements may also be subject to changes. During the different developmental periods of an individual these elements may undergo cycles of active and inactive phases; the phase changes are brought about by methylation of their DNA. A methylation in the target sequence CAG to GTC of an element leads to a reversible inactivation of the element.

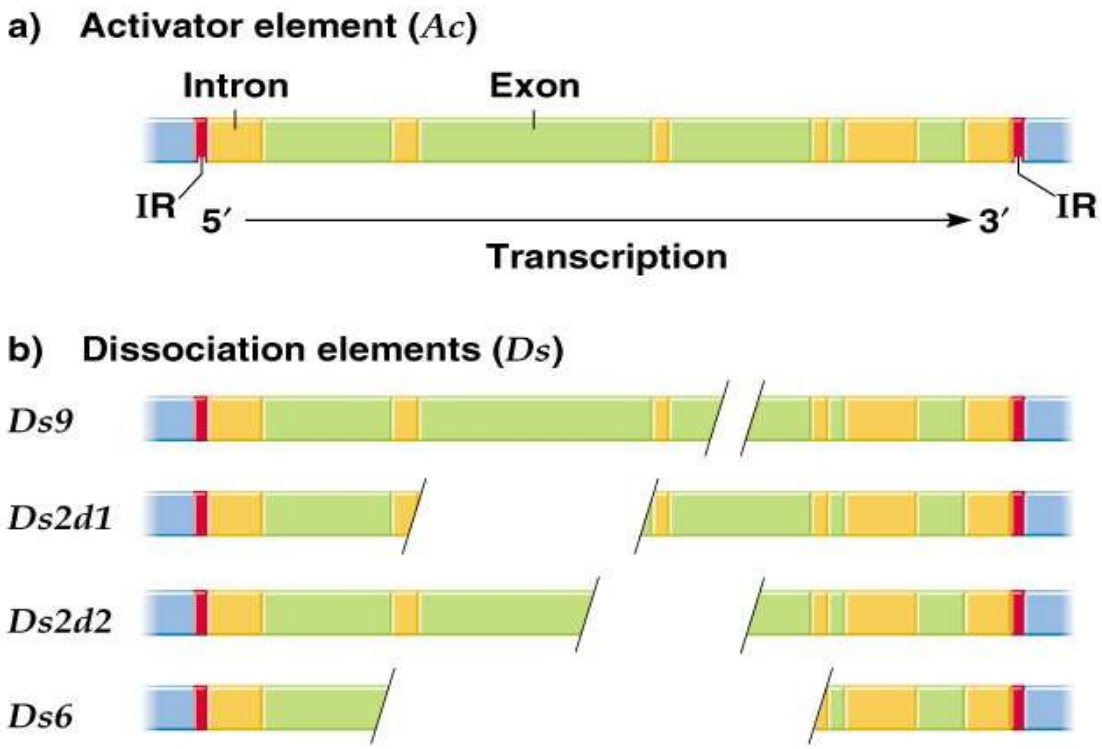


Fig-15.2: Structure of *Ac* autonomous and *Ds* non-autonomous transposable elements in corn

The mechanism of the Ds-Ac system in maize

i) An Ac element can exist in a number of states similar to other genes, and it controls the activity and time of action of the Ds element.

ii) Ac and Ds, both exhibit inter-chromosomal as well as intra-chromosomal movements (transposition). The movement occurs through excision of these elements from one site and their insertion at a new site.

iii) Ds element is unstable in the presence of the Ac element in the same nucleus. When both the elements are present, loss (deletion) of a part of the chromosome 9 occurs if the chromosome 9 carries the Ds element. The deletion is caused by breakage of the chromosome at the site of Ds.

iv) The genes lying adjacent to the Ds become inactivated.

v) The number of Ac elements present in a genome has a negative relationship with the time of Ds action during the development. Therefore, the presence of Ac in a greater number delays the transposition of Ds during the development. This can be well explained in the maize endosperm which is triploid.

In maize endosperm, the number of Ac element may vary from 0 to 6. The dominant allele I at the C locus on chromosome 9 inhibits colour formation in the aleurones of kernels so that the kernels having the I allele are colourless. In the presence of recessive allele i, colour develops normally in the aleurone. Therefore, an ii secondary nucleus fertilized by a pollen carrying i allele will produce coloured aleurone iii. But an ii secondary nucleus fertilized by pollen carrying the dominant allele I will produce colourless endosperm Iii. If both, Ac and Ds are present in the above I pollen, and Ds occupies a place within or near the C locus, coloured spots would be observed in many kernels. The coloured spots develop due

to the transposition of the Ds element from the allele I during the stages of seed development which permits the c locus to produce aleurone colour. An increase in the number of Ac elements delays the dissociation of Ds. Thus variegation pattern in the kernel will differ according to the number of Ac. (Table 1).

Table 1. Effect of number of Ac elements on the time of Ds dissociation leading to chromosome breakage and kernel colour in maize

Genotype of endosperm	Time of dissociation of DS	Degree of colour spotting in kernel
Iii - DS	No dissociation	Colourless kernel
Iii-DS AC	Early dissociation	Large colour spots
Iii- DS AC AC	Dissociation at the stage	Small colour spots
Iii – DS AC AC AC	Dissociation at very late stage	Very few tiny colour spots

Effects of Transposition of Ds Elements

Transposition of Ds elements causes breakage in the chromosome at the site from which the Ds element moves out. The mechanism of transposition is non-replicative. Following breakage, the acentric chromosome fragment is lost. If the chromosome carrying Ds has dominant alleles, e.g. A, B, C and its homologue carries the recessive alleles a, b, c, the transposition of Ds will lead to breakage and loss of the fragment carrying the dominant alleles (Fig.15.3). In the progeny, cells as a result only the recessive alleles a, b, c will be expressed.

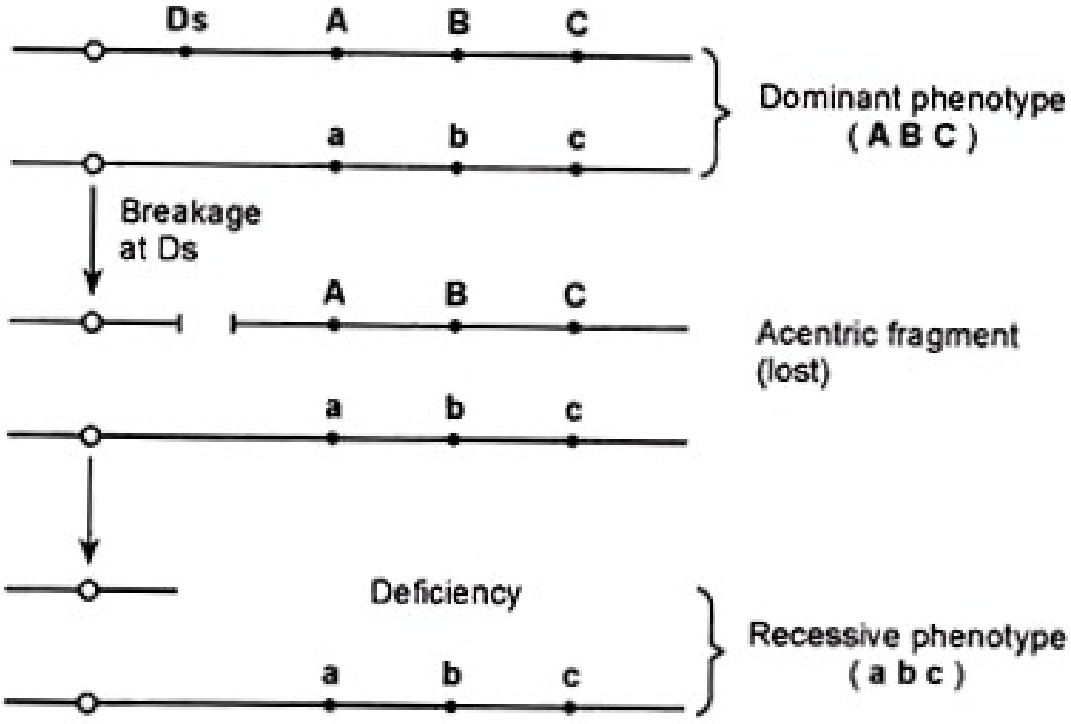


Fig-15.3: Breakage at Ds Site Produces an Acentric Fragment which is lost

Alternatively, the broken ends of the two sister chromatids may join together as they are produced through chromosome replication. The acentric fragment, as a result, will form an U- shaped structure which would be lost. The centric fragment, on the other hand, will form a dicentric chromatid bridge during anaphase. As a result of the tension created due to the centromere movement, the chromatid bridge will break at some point between them producing two dissimilar chromatids. In the next cycle, the broken chromatid ends will again fuse during chromosome replication. This will, as earlier lead to the formation of dicentric chromatid bridges in both the daughter cells. Thus a chromatid-fusion-bridge-breakage cycle is generated (Fig.15.4).

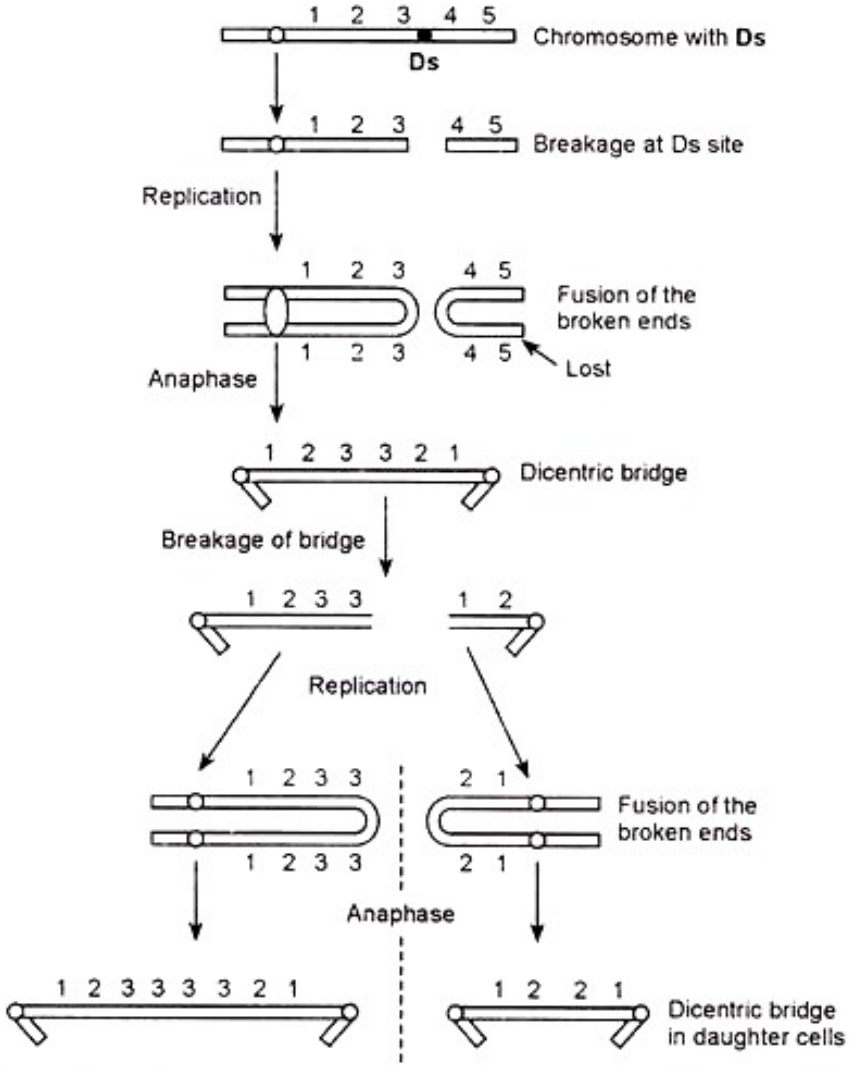


Fig-15.4: Breakage-fusion-bridge cycle generated by chromosome breakage caused by the transposition of Ds elements

15.2.2 Ty Elements in Yeast

In yeast, the transposable element is known as Ty element which is about 5-9 kb long and includes two directly repeated terminal sequences called long terminal repeats (LTR) or deltas (δ) (Fig.15.5). Each delta contains a promoter and sequences that are recognized by transposase. The Ty element encodes a single 5,700 nucleotide long mRNA. *Ty* elements in yeast*Ty* (for transposon yeast) elements are a family of common retrotransposons found in yeast; many yeast cells have 30 copies of *Ty* elements.

Ty elements belong to the retrotransposons group. The abbreviation "Ty" stands for "Transposons of yeast". *Saccharomyces cerevisiae* retrotransposons and retroviruses are often compared because of the similarity between their life cycles and their mechanism of integrating cDNA into host genomes. Ty elements generate more copies of themselves for inserting in the host cell genome. Ty genome contains two genes: TYA1 and TYB1, which correspond to the gag and pol genes of retroviruses, respectively. As with certain retroviral pol genes, TYB1 expression requires programmed ribosomal frame shifting. Ty mRNA is transcribed and processed in the nucleus and then transported to the cytoplasm, where it is translated into Gag and Gag-Pol proteins. The Ty elements of *Saccharomyces cerevisiae* produce virus-like particles (VLPs), which never leave the cell. During the assembly process, Ty1 RNA is packed within the VLPs and subsequently reverse-transcribed into a full-length cDNA. In the final step of transposition the cDNA is integrated into a new site in the host genome, and the cycle can begin anew by transcribing the newly transposed element.

Features of Ty Elements

1) About 35 copies in the haploid yeast gene
2) It has an about 340 bp sequence at both ends in a direct orientation; these are called long terminal repeats or LTRs.
3) It resembles eukaryotic retroviruses; because they lack some of the retroviral functions they are considered to be primitive retroviruses; because of their similarities they are called retrotransposons.
4) The transposition involves an RNA intermediate that is generated by transcription of the *TY* element; a reverse transcriptase (encoded by the *TyB* gene of the element) makes a DNA copy of the element which is then inserted into a new site in the yeast genome.

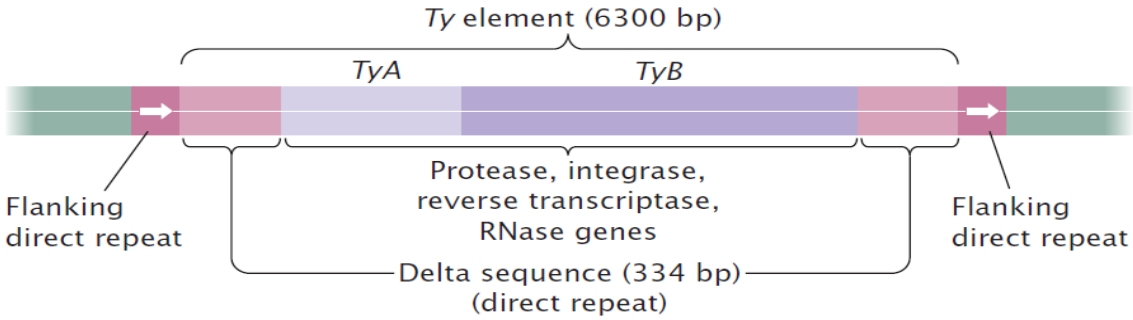


Fig-15.5: Ty Transposable Element of Yeast

This mRNA begins with a promoter in the delta at the 5′ end of the element. This mRNA has two open reading frames (ORFs), designated as TyA and TyB which encode two different proteins required for transposition. In different strains of yeast, the number of Ty element varies and average number is 35 per strain. In fact, the organisation of yeast Ty element is very similar to the retrovirus. When a retrovirus infects a cell, its RNA genome is copied by viral reverse transcriptase, producing a double stranded DNA. This DNA integrates into host's chromosome where it undergoes transcription to produce progeny RNA viral genomes and mRNAs for viral proteins.

It is assumed that the Ty element transposes in a similar way like that of retroviruses that is not by DNA mechanism. It makes an RNA copy of the integrated DNA sequence and then creates a new Ty element by reverse transcriptase. This new element would then integrate at a site in the new chromosome. Evidences in favour of such mechanism came from the experiments in which an intron was placed into the Ty element. This Ty element was then monitored from its initial position through the transposition event. The Ty element has no intron at its new location, indicating that its transposition occurred via an RNA intermediate. The imposed intron had been removed by normal splicing processes from the pre- mRNAs. However, later on it was proved that Ty element encodes reverse transcriptase which shows retrovirus-like activity in yeast cells. Since then, Ty element is also known as retrotransposons because of their similarity to retroviruses in this regard, and their transposition is called retrotransposition.

Ty element is a LTR retro-transposons present in yeasts which is about 5.9 kbp long sequence having LTRs of about 340 base pairs long at both end. Most yeast strains have about 35 copies of the Ty1 element in their genome. Ty1 elements have only two genes, TyA and TyB, which are homologous to the gag and pol genes of the retroviruses. Transposition of Ty1 element involves reverse transcriptase enzyme. After RNA is synthesized from Ty1 DNA, a reverse transcriptase encoded by the TyB gene uses it as a template to make double stranded DNA. Then the newly synthesized DNA is transported to the nucleus and inserted in targeted site in the genome, creating a new Ty1 element (Fig.15.6).

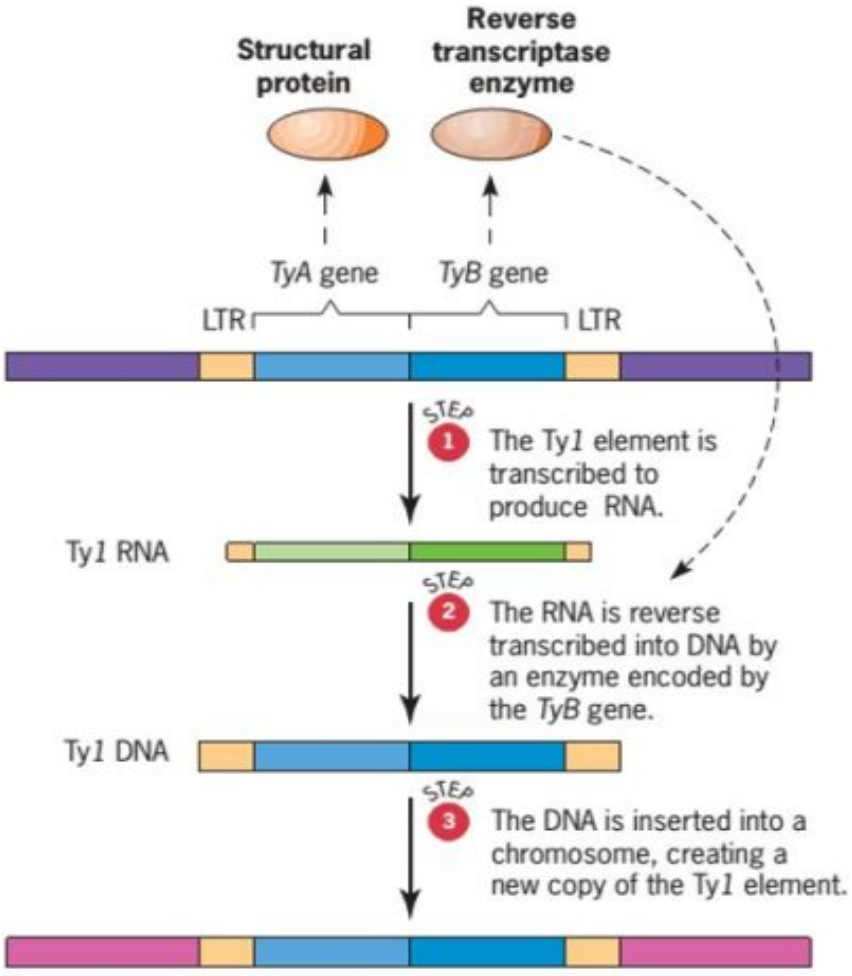


Fig-15.6: Transposition of the Yeast Ty1 Element

15.2.3 Transposons in *Drosophila*

P element in *Drosophila*:

P transposable elements were discoveredby Margaret Kidwell in *Drosophila* as the causative agents of a syndrome of genetic traits called hybrid dysgenesis. The P elements vary in size, ranging from 0.5 to 2.9 kb in length. These are small transposons with terminal 31-bp inverted repeats, and the element generates 8-bp direct repeats of target DNA sequences upon insertion. The complete element is 2907 bp and is autonomous because it encodes a functional transposase. The first DNA transposons to be molecularly characterized were the P elements in *Drosophila*. P element present in *Drosophila* resembles to Insertion sequence (IS element) of bacteria, which possesses a short-inverted terminal repeats at both end and a single open reading frame which encode single protein (transposase).

Dysgenesis in *Drosophila*

A phenomenon known as dysgenesis is present in Drosophila when laboratory reared female flies were crossed with wild male flies (Fig.15.7). In such crosses, the laboratory stocks are said to possess an M cytotype (cell type), and the natural stocks are said to possess a P cytotype. Laboratory flies do not have P element. In such cross of M (female) × P (male), the F1 progeny show a range of surprising phenotypes characters in germ line including sterility, a high mutation rate, and a high frequency of chromosomal aberration and nondisjunction. These F1 hybrid progeny are biologically deficient and are known as dysgenic. The phenomenon is called dysgenesis However in reciprocal cross between P (female) × M (male), no dysgenesis is observed in F1 progeny. To explain the phenomenon of dysgenesis, the simple answer is that P-element transposition. In all adult P cytotype flies, they possess P-element which contains a transposase genes are silenced. However, the genes are activated in the F1generation resulting in dysgenesis when crossed M (female) with P (male). In the cross between M cytotype (female flies, no P elements) × P cytotype (male flies, P elements), each parent contribute respective gametes with respective cytotype. Male gamete provides P elements in the newly formed zygote which is in a silencing-free environment, since female gamete is M cytotype. The P elements derived from the male genome can now transpose throughout the zygote genome, causing a variety of damage as they insert into genes and all F1 progeny developed from such zygote are dysgenic. On the other hand, in reciprocal cross between P cytotype (female flies, P element) ×M cytotype (male flies, no P-element) no dysgenesis is observed. It is because, in this case female gamete provide P element in the newly formed zygote but the P-element is already silenced in female gamete due to presence of some component in cytoplasm of female gamete. Therefore, zygote developed from such gamete with already silenced P element prevents dysgenesis.

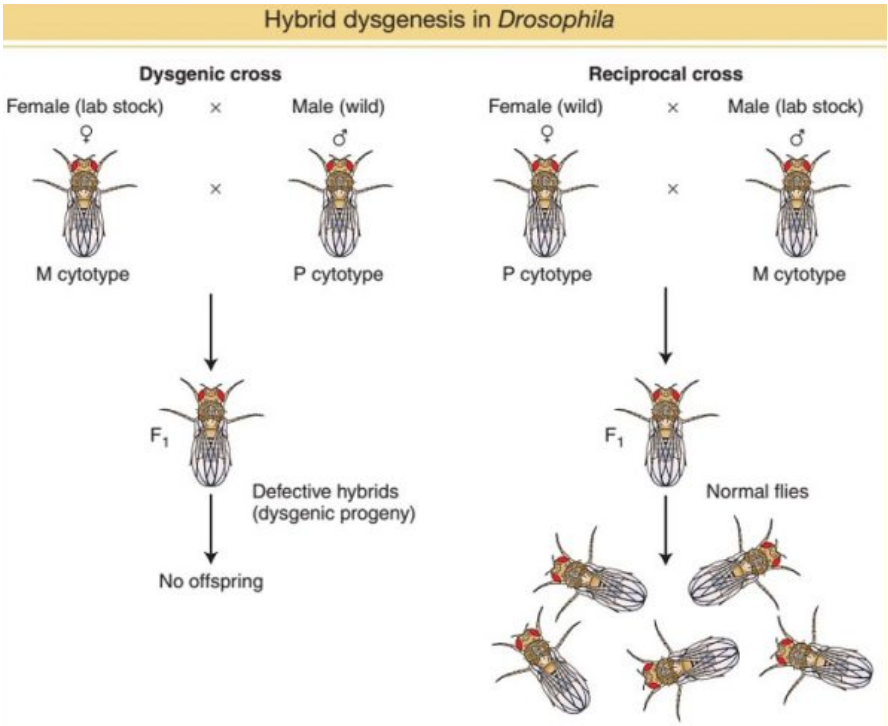


Fig-15.7: Hybrid Dysgenesis phenomena in *Drosophila*

15.2.4 Retrotransposons

Retro-transposons are those transposable elements that utilize reverse transcriptase enzyme to convert RNA intermediate into DNA and then transpose into new position by itself (Fig.15.8).

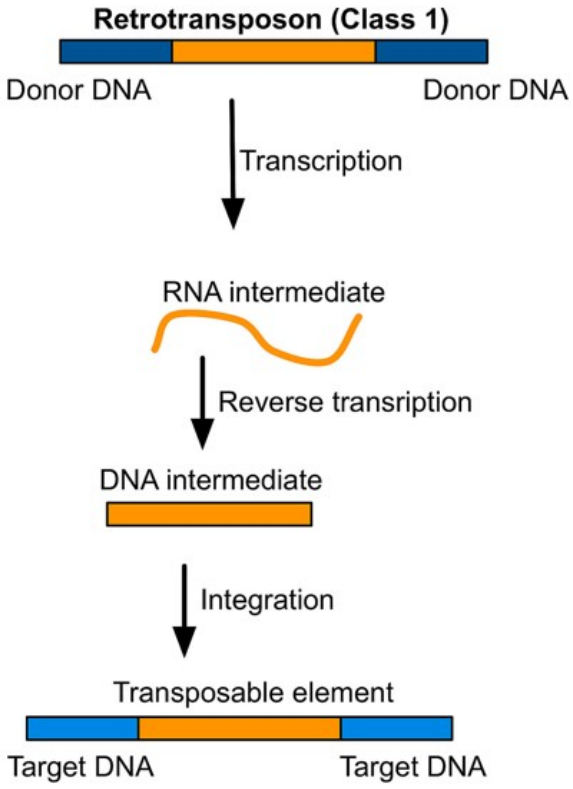


Fig-15.8: Insertion of Retrotransposon

Retrotransposons are of Two Types:

LTR Retro-Transposons or Retrovirus like Elements

All LTR retro-transposons have the same basic structure – a central coding region flanked by long terminal repeats (LTR) which are oriented in the same direction. The LTR are typically a few hundred nucleotide pairs long sequence having short inverted repeats at both ends like present in other types of transposons. Due to the characteristic LTR sequence which is similar to that of retrovirus, these transposons are known as LTR retrotransposons. The property of LTR retrotransposons is similar to retrovirus infection. LTR retrotransposons utilizes reverse transcriptase enzyme from retro virus that convert transposons RNA into DNA in the first stage and in in second stage the synthesized DNA is transposed into new location in the chromosome.

Examples: Ty elements of yeast, and Copia elements of *Drosophila*

Non-LTR Retrotransposons (Retroposons):

The non-LTR retrotransposons are common retrotransposons present in eukaryotes. They are simply known as retroposons or non-viral transposons. Non-LTR retrotransposons lacks long terminal repeats (LTRs) at the terminal end. However, they have homogeneous sequence of A:T base pairs at one end which is derived from post transcriptional modification (poly A tail). These non-LTR retrotransposons are transposed via RNA intermediate which is reverse transcribed into double stranded DNA by reverse transcriptase enzyme encoded by the element itself and then transposed to the targeted site (Fig. 16.8).

Examples: Transposable element of humans - LINEs andSINEs

Human Retrotransposons:

At least 44 % of human genome is repetitive which is derived from transposable elements, including retrovirus-like elements (8%), retroposons (33 %), and several families of elements that transpose by a cut-and-paste mechanism (3 %). Classes of human transposons: LINEs (long interspersed nuclear elements), and SINEs (short interspersed nuclear elements)

Long Interspersed Nuclear Elements (LINEs):

L1 retroposon is the principal human transposable element which belongs to long interspersed nuclear elements (LINEs) class. Two other LINEs sequences, L2 and L3 LINEs are also present in human genome, however both are transpositionally inactive. Complete L1 element is about 6 kb long sequence with an internal promoter that is recognized by RNA polymerase II, and two open reading frames (ORFs); ORF1 encodes a nucleic acid-binding protein, ORF2 encodes a protein with endonuclease and reverse transcriptase activities. The human genome contains about 3000 to 5000 copies of complete L1elements.

Transposition of L1 element involves transcription of a complete L1 element into RNA in the first step and the reverse transcription of this RNA into DNA in next step. Both transcription and reverse transcription processes take place in the nucleus. However, before the L1 RNA is reverse transcribed, it travels to the cytoplasm where it is translated into polypeptides (nucleic acid binding proteins, endonuclease and reverse transcriptase). The polypeptides remain attached with RNA when it returns to the nucleus. The polypeptide encoded by ORF2 possesses an endonuclease activity that catalyses the cleavage of one

strand of the DNA duplex at targeted site in a chromosome. The polyA tail of polypeptide juxtaposed with 5'end of cleaved DNA and the reverse transcriptase activity results into synthesis of single DNA strand. Utilizing 3' end as a primer for DNA synthesis, the newly synthesized L1 DNA is subsequently made double-stranded. Therefore, new L1 element is placed in new location.

Short Interspersed Nuclear Elements (SINEs):

The short interspersed nuclear elements (SINES), are the second most abundant class of transposable elements in the human genome. SINEs are less than 400 base pairs long and have an internal promoter but do not encode proteins. Like all retroposons, SINEs do not have terminal repeats but instead they have a sequence of A: T base pairs at one end. Transposition of SINEs involves transcription of a SINEs DNA into RNA which is then reversed transcribed into DNA. However, SINEs do not have any open reading frame to encode any enzymes. Therefore, it depends upon LINEs-type element for reverse transcriptase enzyme and other enzymes. Thus, the SINEs in depend on the LINEs to multiply and insert within the genome. Three families of SINEs are present in Human genome-Alu, MIR, and Ther2/MIR3 elements.

`

15.3 SUMMARY:

Transposable elements **(**TEs) were discovered by Barbara McClintock during experiments conducted in 1944 on maize. Since they appeared to influence phenotypic traits, she named them controlling elements. However, her discovery was met with less than enthusiastic reception by the genetic community. There are two classes of transposons in eukaryotes. Class I (RNA) transposons move (i.e., jump) by transcription of RNA at one locus, followed by reverse transcription and integration of the cDNA back into genomic DNA at a different location and called as retrotransposons, they may be derived from (or be the source of) retroviruses, since active retroviruses excise from and integrate into DNA much like retrotransposons. Class II (DNA) transposons move by either of two mechanisms. In the cut-and-paste pathway, the transposon leaves one locus and integrates at another. In the replicative pathway, the original transposon remains in place, but new copies are mobile. The table confirms that bacteria contain relatively few transposons, in contrast to eukaryotes, which vary widely in transposon load (transposons as a percentage of genomic DNA).

In yeast the transposable element is known as Ty element which is about 5-9 kb long and includes two directly repeated terminal sequences called long terminal repeats (LTR) or deltas (δ). Each delta contains a promoter and sequences that are recognized by transposase. The Ty element encodes a single 5,700 nucleotide long mRNA. *Ty* elements in yeast *Ty* (for transposon yeast) elements are a family of common retrotransposons found in yeast; many yeast cells have 30 copies of *Ty* elements. A phenomenon known as dysgenesis is present in Drosophila when laboratory reared female flies were crossed with wild male flies. In such crosses, the laboratory stocks are said to possess an M cytotype (cell type), and the natural stocks are said to possess a P cytotype. Laboratory flies do not have P element. The short interspersed nuclear elements (SINES), are the second most abundant class of transposable elements in the human genome. SINEs are less than 400 base pairs long and have an internal promoter but do not encode proteins. From bacteria to humans, transposable elements have accumulated over time and continue to shape genomes through their mobilization.

15.4 TECHNICAL TERMS:

Ac and Dc Elements, Ty elements, Transposons of yeast, P element in *Drosophila,* retro-transposons, short interspersed nuclear elements (SINES), long interspersed nuclear elements (LINEs), open reading frames (ORFs).

15.5 SELF ASSESSMENT QUESTIONS:

1) Explain the structure and mechanism involved in Ac and Dc Elements of Maize?
2) Write an essay on Ty elements in Yeast?
3) Explain importance of P element in *Drosophila*?
4) Define Dysgenesis? Explain the mechanism of Dysgenesis in Drosophila?
5) What a note on short interspersed nuclear elements (SINES)?
6) Explain briefly about long interspersed nuclear elements (LINEs)?

15.6 SUGGESTED READINGS:

1) Verma PS and Agarwal VK (2005). Cell Biology, Genetics, Molecular Biology, Evolution, and Ecology. Multicolored Edition.
2) McGee, David & Coker, Christopher & Harro, Janette & Mobley, Harry. (2001). Bacterial Genetic Exchange. Doi: 10.1038/npg.els.0001416.
3) Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4) Russel, P.J. Genetics (1998) – The Benjamin / Cummings Publishing Company, Inc., California
5) Singh, B.D. Fundamentals of Genetics (2001) - Kalyani Publishers, New Delhi.

Prof. K. Mallikarjuna

LESSON-16

MATING TYPE SWITCHING IN YEAST, DIVERSITY IN Ig MOLECULES AND REGULATION OF GENOME ACTIVITY DURING SPORULATION

16.0 OBJECTIVE:

- This lesson deals with the concept of Genome rearrangements and regulation of genome activity.

STRUCTURE:

16.1 INTRODUCTION

The mating behaviour of a yeast cell is specified by the set of genes that it contains at its mating-type (MAT) locus. In almost all ascomycete yeast species, typified by *Saccharomyces cerevisiae*, there are three types of cells, defined by their mating behaviour: two types of haploid cells with genotypes MATa and MATα and one type of diploid cell with genotype MATa/α. Each of these cell types can replicate to produce another cell of the same type vegetatively by mitosis and budding. This asexual cell cycle is the main way that yeast cells proliferate, but many species can also reproduce sexually in a cycle that involves mating followed by meiosis and sporulation.

The B lymphocytes are generated in the bone marrow through Hematopoiesis. After their maturation, the B cells migrate from the bone marrow to blood. If the B cell receptor present on a B cell binds to an epitope present on an antigen, the B cell gets activated and converted to plasma cells and memory cells. Plasma cells secrete immunoglobulins. In the absence of antigen-induced activation, naive B cells in the periphery have a short life span, dying within a few weeks by apoptosis.

16.2 GENOME REARRANGEMENTS

Mating-Type Switching in Yeast

The mating behaviour of a yeast cell is specified by the set of genes that it contains at its mating-type (MAT) locus. In almost all ascomycete yeast species, typified by *Saccharomycescerevisiae*, there are three types of cell, defined by their mating behaviour: two types of haploid cell with genotypes MATa and one type of diploid cell with genotype MATa/a. Each of these cell types can replicate to produce another cell of the same type by mitosis and budding. This asexual cell cycle is the main way that yeast cells proliferate, but many species can also reproduce sexually in a cycle that involves mating followed by meiosis and sporulation. Mating is the process by which two haploid cells of opposite mating types (MATa and MATa) fuse to form a diploid cell (MATa/a). Conversely, meiosis followed by sporulation is the process by which a diploid cell turns into four spores that germinate into haploid cells (two with genotype MATa and two with MATa) (Fig.16.1).

About half of the known budding yeast species have a known sexual cycle. These are called teleomorphic species. Their predominant vegetative form can be either haploid or diploid, and the species is termed haplontic or diplontic accordingly. For example, most natural isolates of *S. cerevisiae* are diploid. The difference between these two lifestyles depends on how they use a nutrient starvation signal (usually a lack of nitrogen) as a control point in the sexual cycle: some species such as *S. cerevisiae* use starvation as a trigger for meiosis and are therefore diplontic, whereas other species such as *K. lactis* use starvation as a trigger for mating and are therefore haplontic. Thus, the diploid phase is stable in *S. cerevisiae* (diploid cells will replicate vegetatively for generation after generation until they are starved and sporulate in response), usually sporulate as soon as they are formed by mating, without undergoing diploid vegetative cell cycles). Conversely, the haploid phase is unstable in *S. cerevisiae* (haploid cells mate as soon as they detect a haploid cell of the opposite mating type, without requiring any other signal). The other half of the known budding yeast species have no known true sexual cycle and are called anamorphic species. By a true sexual cycle, we mean a complete cycle that includes meiosis as well as mating.

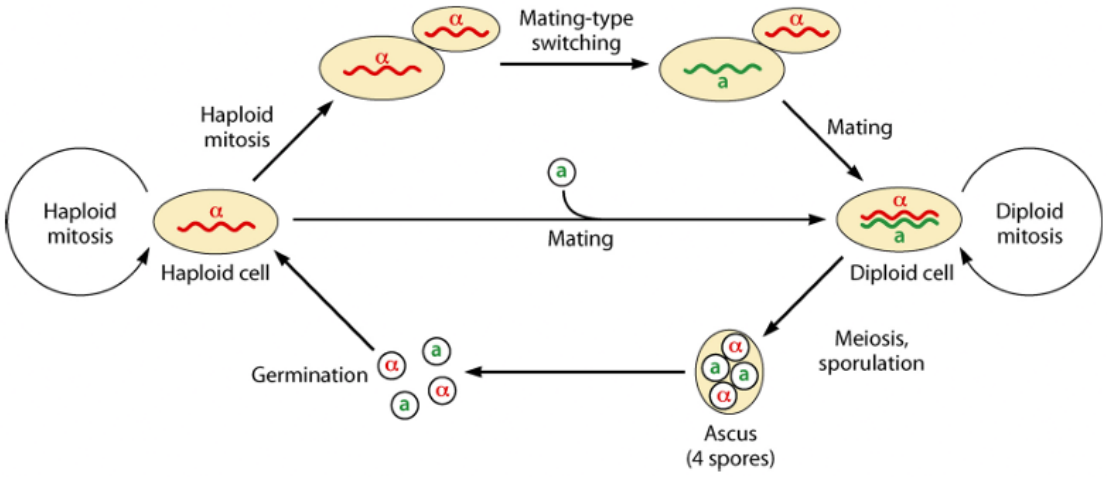


Fig-16.1: Mating type switching in the life cycle of *S. cerevisiae*

Structure of the MatLocus and its Four Canonical Genes

The yeast MAT locus is not a typical genetic locus in many respects. At most loci in any genome, different alleles have very similar DNA sequences and usually only differ by a few nucleotides, but this is not the case at the MAT locus. The genes contained in the MATa and MATa "alleles" are highly dissimilar from each other. For this reason, the two alternative sequences at the MAT locus are often called idiomorphs rather than alleles. Each of the MAT idiomorphs typically contains two genes-one coding for a transcription factor (a2ora1) that acts in haploids and one coding for one subunit of a dimeric transcription repressor (the a1/a2 heterodimer) that acts in diploids (Fig.16.2). The nomenclature convention is that MAT is used in the gene names (MATa1, MATa2, MATa1, and MATa2) and the names of the idiomorphs (MATa and MATa), but not in the corresponding protein names (a1, a2, a1, and a2). The a2 protein is a MATa-specific transcription factor that activates expression of the a-specific genes (asgs)—thatis, genes whose expression is specifically required to enable MATa cells to mate, such as MFA1, encoding the a-factor pheromone, and STE2, encoding the a-factor receptor. Similarly, the a1 protein is a MATa-specific transcription factor that activates expression of the a-specific genes such as MFa1, encoding the a-factor pheromone, and STE3,encoding the a-factor receptor. In addition to expressing asgs or asgs, haploid cells also need to express many haploid-specific genes (hsgs) that are shared by the mating pathways in both types of haploid cell, such as GPA1, STE4,andSTE18, the genes coding for subunits of the G-protein complex that transmits the signal that a pheromone has been detected on the cell surface.

First, *S. cerevisiae* has no MATa2 gene and no a2 transcription activator. Instead of using a2 to activate transcription of its asgs, *S. cerevisiae* instead expresses asgs by default in haploid cells, unless they are repressed by a complex consisting of Mcm1 and two molecules of a2; this complex is present in a-cells and in a/a diploids, where asgs need to be repressed. The loss of MATa2 in the *S. cerevisiae* lineage occurred as part of an evolutionary rewiring of the cell-type determination pathway, which also involved the gain of a new interaction between a2 and Mcm1, at approximately the same time as the whole-genome duplication event (WGD). All post-WGD budding yeast species lack MATa2, whereas all non-WGD species have it. Second, S. cerevisiae contains two other MAT-like loci, called HML and HMR, and uses these loci to replace the DNA at its MAT locus, converting MATa cells into MATa cells or vice versa in a process called mating-type switching. However, since (i) most yeast species cannot switch mating types and (ii) most of those that are able to switch do not use the same switching mechanism as *S. cerevisiae*, we will first introduce non-switching species and species with simpler switching systems before describing the "prototype," *S. cerevisiae*. Presenting the species in this order provides a better evolutionary perspective on how the complex switching system used by *S. cerevisiae* arose.

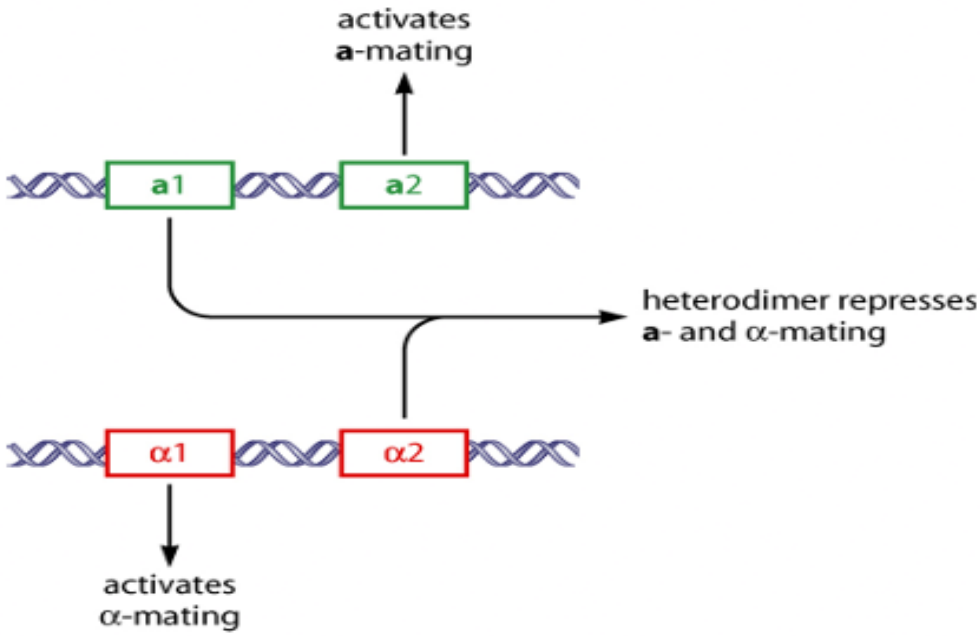


Fig-16.2: The Four Canonical MAT Genes in a Typical Budding Yeast Species

16.2.2 Diversity in Ig Molecules

The B lymphocytes are generated in the bone marrow through Hematopoiesis. After their maturation, the B cells migrate from the bone marrow to blood. If the B cell receptor present on a B cell binds to an epitope present on an antigen, the B cell gets activated and converted to plasma cells and memory cells. Plasma cells secrete immunoglobulins. In the absence of antigen-induced activation, naive B cells in the periphery have a short life span, dying within a few weeks by apoptosis.

Isotype Switching:

Each secreted immunoglobulin is composed of two heavy chains and two light chains (H2L2). The amino-terminal part of both the chains consisting of 100–110 amino acids, show great sequence variation and therefore called the variable (V) region. Detailed map of the variable region shows that three small regions inside the V region carry maximum variation and are called hypervariable regions or complementarity determining regions (CDRs). The remaining part of the heavy chain protein revealed five basic sequence patterns, corresponding to five different heavy-chain constant (C) regions. Each of these five different heavy chains is called an isotype. The length of the constant regions is approximately 330 amino acids for α, γ and δ, and 440 amino acids for μ and ε. The constant region (110amino acid) of light chain has two sequences, lambda or kappa.

The isotype or class of a given antibody molecule is determined based on the constant region sequence: IgM(m), IgG(g), IgA(a), IgD(d), or IgE(e). Each isotype can have either lambda or kappa light chains. Depending on the minor differences in the amino acid sequences of the α, γ heavy chains, they are further classified into subisotypes that determine the subclass of antibody molecules they constitute. In humans, there are two subisotypes of heavy chains and thus two subclasses, IgA1 and IgA2 and four subisotypes of g heavy chains and therefore four subclasses, IgG1, IgG2, IgG3, and IgG4.

Molecular Mechanism of Formation of Ig Chains

The first direct evidence that separate genes encode the V and C regions of immunoglobulins and that the genes are rearranged in the course of B-cell differentiation came from the work of S. Tonegawa and N. Hozumi in 1976. They suggested that during differentiation of lymphocytes from the embryonic state to the fully differentiated plasma-cell stage, the V and C genes undergo rearrangement. In the embryo, a large DNA segment is present between the V and C genes. During differentiation, the V and C genes are brought closer together and the large DNA sequence is eliminated.

The kappa and lambda light-chain families contain V, J, and C gene segments (Fig.16.3). The rearranged VJ segments encode the variable region of the light chains. The heavy-chain family contains V, D, J, and C gene segments; the rearranged VDJ gene segments encode the variable region of the heavy chain. In each gene family, C gene segments encode the constant regions. Each V gene segment is preceded at its 5' end by a small exon that encodes a short signal or leader (L) peptide that guides the heavy or light chain through the endoplasmic reticulum.

The VH gene segment was found to encode amino acids 1 to 94; amino acids 95 to 97 was found to encoded by D gene segments and the JH gene segment was found to encode amino acids 98 to 113. The D gene segment encoded amino acids fall within the third complementarity - determining region (CDR3) and responsible for creating more diversity. Thus it is called D gene segment for *diversity* because of its contribution to the generation of antibody diversity.

The heavy-chain multigene family in the mouse germ line contains 134 variable (V) gene segments, 13 D gene segments and four joining (J) gene segments, and four constant (C) gene segments. In humans, the lambda locus is more complex. The heavy-chain multigene family on human chromosome 14 has been shown to carry 51 functional V gene segments located upstream of a cluster of 27 functional D gene segments followed by 6 J gene segments, and 7 C segments. In additional to the functional gene segments, the human lambda complex contains many V, J, and C pseudogenes.

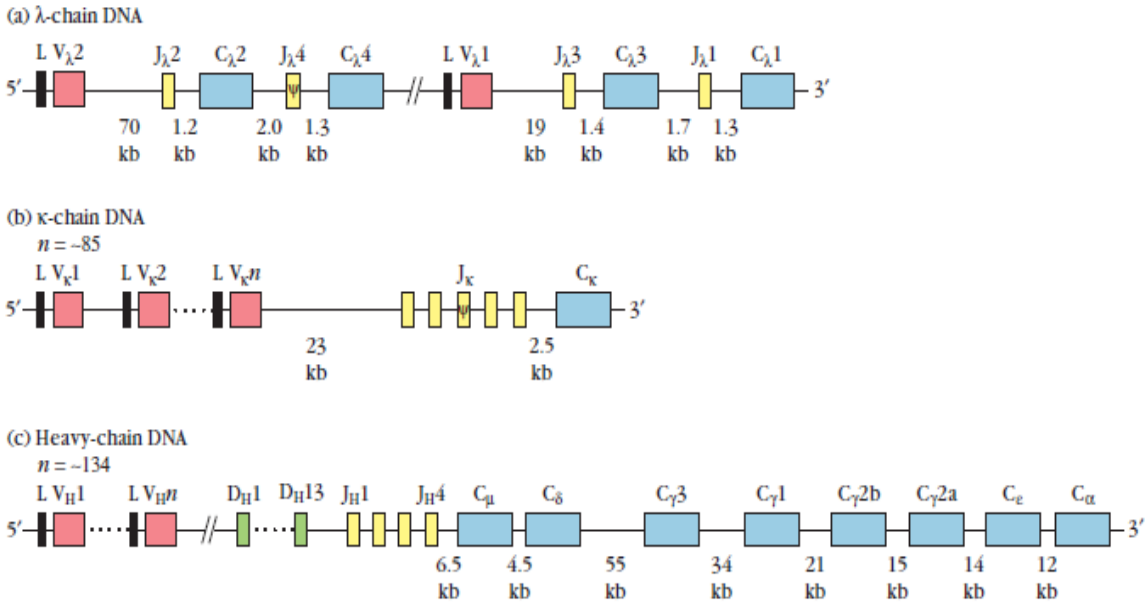


Fig-16.3: The Gene Segments of Lambda, Kappa and Heavy Chain of Mouse Genome

Two separate rearrangement events within the variable region generate the functional immunoglobulin heavy chain. A DH gene segment first joins to a JH segment and the resulting DHJH segment then joins a VH segment to generate a VHDHJH unit that encodes the entire variable region. In heavy-chain DNA, variable-region rearrangement produces a rearranged gene consisting of the following sequences, starting from the 5' end: a short L exon, an intron, a joined VDJ segment, another intron, and a series of C gene segments.

As with the light-chain genes, a promoter sequence is located a short distance upstream from each heavy-chain leader sequence. Once heavy-chain gene rearrangement is accomplished, RNA polymerase transcribes the entire heavy-chain gene, including the introns. Initially, both C(μ) and C(δ) gene segments are transcribed. Differential polyadenylation and RNA splicing remove the introns and process the primary transcript to generate mRNA including either the C (μ) or the C(δ) transcript. These two mRNAs are then translated, and the leader peptide of the resulting nascent polypeptide is cleaved, generating finished μorδ chains. The production of two different heavy-chain mRNAs allows a mature, immunocompetent B cell to express both IgM and IgD with identical antigenic specificity on its surface (Fig.16.4 and 16.5).

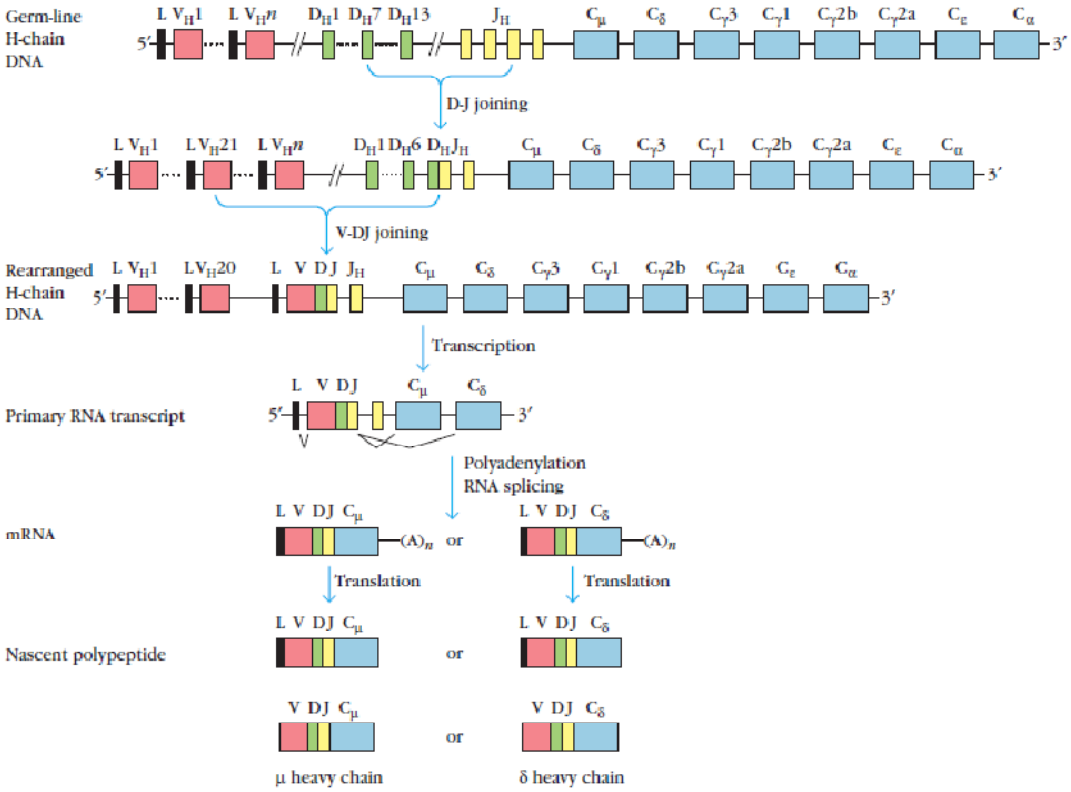


Fig-16.4: Recombination in the Heavy Chain

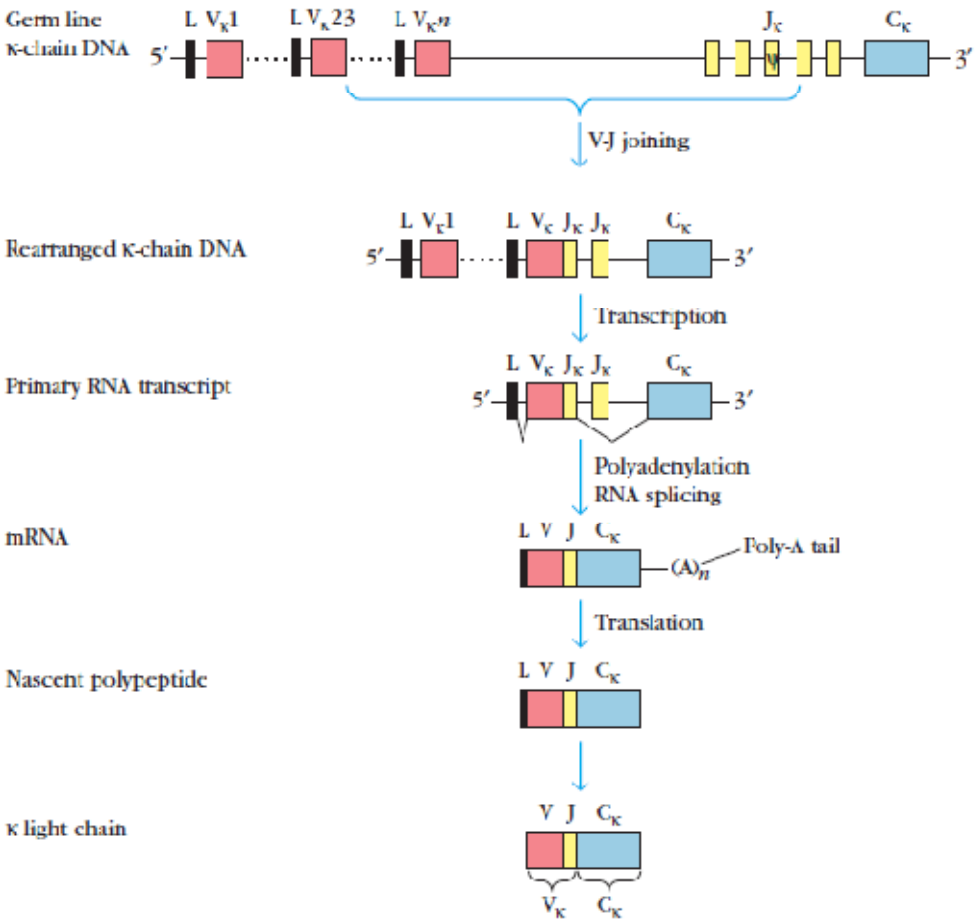


Fig-16.5: Recombination in The Kappa Light Chain

Mechanism of Class Switching:

Exposure to antigen activates a B cell, which undergoes a number of changes. The heavy-chain DNA can undergo a further rearrangement in which the VHDHJH unit can combine with any constant region gene segment. This process is called class switching or isotype switching (Fig.16.6). Class switching depends upon the interplay of three elements: switch regions, a switch recombinase, and the cytokine signals that dictate the isotype to which the B cell switches.

The exact mechanism is unclear but involves DNA flanking sequences (called switch regions) located 2–3 kb upstream from each CH segment (except C δ). These 2-10kb long switch regions are composed of multiple copies of short repeats (GAGCT and TGGGG). The switch recombinase recognizes these repeats and carries out the DNA recombination those results in class switching. The recombination is initiated by a B cell-specific enzyme, activation-induced cytidinedeaminase (AID), which deaminatescytosines in both the donor and acceptor switch regions. AID activity converts several deoxy cytosine bases to dU bases in each S region, and the dU bases are then excised by the uracil DNA glycosylase (UNG). The resulting abasic sites are nicked by apurinic / apyrimidinic endonuclease (APE). AID attacks both strands of transcriptionally active S regions, but how transcription promotes AID targeting is not entirely clear. Mismatch repair proteins then convert the resulting single-strand DNA breaks to double-strand breaks with DNA ends appropriate for end-joining recombination. Several other proteins are also required for this recombination.

Intercellular regulatory proteins known as cytokines act as “switch factors” and play major roles in determining the particular immunoglobulin class that is expressed as a consequence of switching. Interactions of numerous cytokines with B cells generate signals required for proliferation and class switching during the differentiation of B cells into plasma cells. Binding of the proliferation cytokines, which are released by activated TH cells, provides the progression signal needed for proliferation of activated B cells. Similar or identical effects may be mediated by cytokines beyond the ones shown.

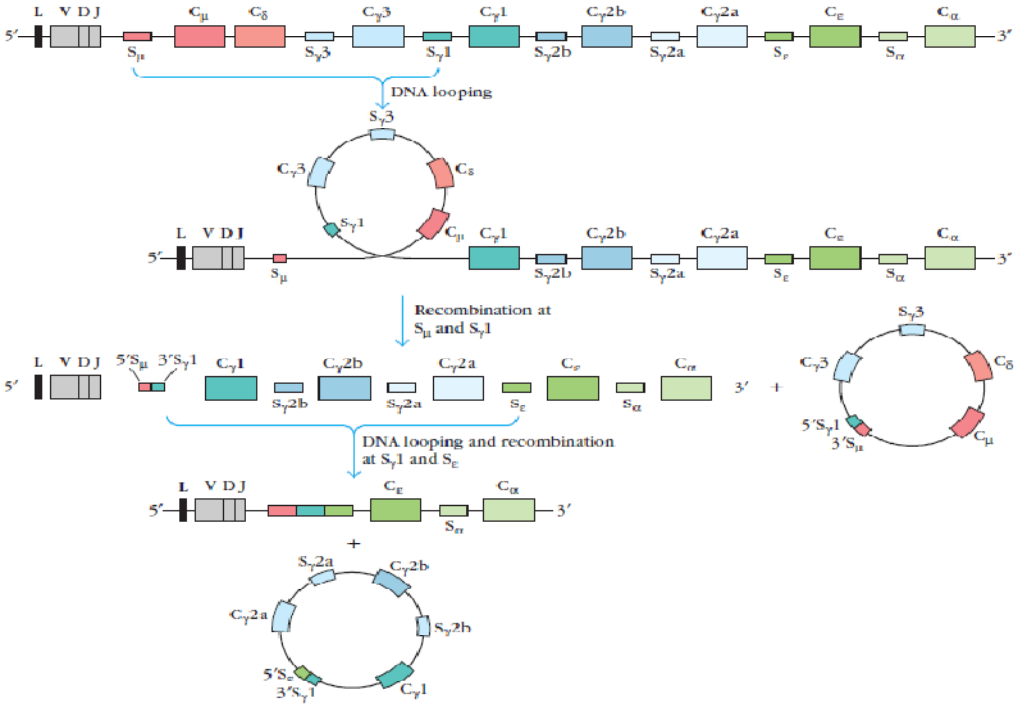


Fig-16.6: Mechanism of Class Switching

Class switching in the response to thymus-dependent antigens also requires the CD40/CD40L interaction. In some cases, IL-4 has been observed to induce class switching in a successive manner: first from C μ to C γ1 and then from C γ1 to C ε. Examination of the DNA excision products produced during class switching showed that a circular excision product containing other Constant regions was generated.

Synthesis and Release of Immunoglobulins From Activated B Cells:

Immunoglobulin heavy-and light-chain mRNAs are translated on separate polyribosomes of the rough endoplasmic reticulum. The amino-terminal leader sequence guides the chains into the lumen of the RER, where the signal sequence is then cleaved. The light (L) and heavy (H) chains join together through disulfide linkages glycosylated inside the RER. The complete immunoglobulin molecule is transported to the Golgi apparatus and then into secretory vesicles, which fuse with the plasma membrane. The nature and order of chain assembly is different for different immunoglobulin classes. In the case of IgM, the heavy and light chains assemble within the RER to form half-molecules and then two half-molecules assemble to form the complete molecule.

In the case of IgG, two heavy chains join first then it joins to a light chain and finally the complete H2L2 molecule is formed. Interchain disulfide bonds are formed, and the polypeptides are glycosylated as they move through the Golgi apparatus. Before the

activation of B cell the molecule contains the transmembrane sequence and is anchored in the membrane of a secretory vesicle. As the vesicle fuses with the plasma membrane, the immunoglobulin molecule is exposed outside. After activation of the B cell, the immunoglobulin molecules secreted contain the hydrophilic sequence and is transported as a free molecule in a secretory vesicle and is released from the cell when the vesicle fuses with the plasma membrane.

16.3 REGULATION OF GENOME ACTIVITY DURING SPORULATION BY SIGMA FACTOR

The first developmental pathway that was examined is the formation of spores by the bacterium *Bacillus subtilis*. Strictly speaking, this is not a developmental pathway, merely a type of cellular differentiation, but the process illustrates two of the fundamental issues that have to be addressed when genuine development in multicellular organisms is studied. These issues are how a series of changes in genome activity over time is controlled, and how signalling establishes coordination between events occurring in different cells. The advantages of *Bacillus* as a model system are that it is easy to grow in the laboratory and is amenable to study by genetic and molecular biology techniques such as analysis of mutants and sequencing of genes.

Bacillus is one of several genera of bacteria that produce endospores in response to unfavorable environmental conditions. Resistance is due to the specialized nature of the spore coat, which is impermeable to many chemicals and to biochemical changes that retard the decay of DNA and other polymers and enable the spore to survive a prolonged period of dormancy. In the laboratory, sporulation is usually induced by nutrient starvation. This causes the bacteria to abandon their normal vegetative mode of cell division, which involves synthesis of a septum (or cross-wall) in the centre of the cell. Instead the cells construct an unusual septum, one that is thinner than normal, at one end of the cell. This produces two cellular compartments, the smaller of which is called the prespore and the larger the mother cell. As sporulation proceeds, the prespore becomes entirely engulfed by the mother cell. By now the two cells are committed to different but coordinated differentiation pathways, the prespore undergoing the biochemical changes that enable it to become dormant, and the mother cell constructing the resistant coat around the spore and eventually dying.

Special σ Subunits Control Genome Activity during Sporulation:

Changes in genome activity during sporulation are controlled largely by the synthesis of special σ subunits that change the promoter specificity of the *Bacillus* RNA polymerase. Recall that the σ subunit is the part of the RNA polymerase that recognizes the bacterial promoter sequence, and that replacement of one σ subunit with another with a different DNA-binding specificity can result in a different set of genes being transcribed.

During bacterial sporulation, sigma factors regulate genome activity by activating specific transcriptional programs in different cell compartments. A cascade of sporulation-specific sigma factors, including σF, σE, σG, and σK in *B. subtilis*, replaces the vegetative sigma factors (σA and σH) and directs gene expression for the different stages of spore formation in both the mother cell and the forespore. Activation of these sigma factors occurs at different times and in a cell-type-specific manner, ensuring the correct sequence of events, such as engulfment and cortex formation, occurs at the right time and place.

It is also the key to the changes in genome activity that occur during sporulation. The standard *B. subtilis* σ subunits are called σ^{A} and σ^{H}. These subunits are synthesized in vegetative cells and enable the RNA polymerase to recognize promoters for all the genes it needs to transcribe in order to maintain normal growth and cell division. In the prespore and

mother cell these subunits are replaced by σ^F and σ^E, respectively, which recognize different promoter sequences and so result in large-scale changes in gene expression patterns. The master switch from vegetative growth to spore formation is provided by a protein called SpoOA, which is present in vegetative cells but in an inactive form. This protein is activated by phosphorylation, the protein kinases that phosphorylate it responding to various extracellular signals that indicate the presence of an environmental stress such as lack of nutrients. Activated SpoOA is a transcription factor that modulates the expression of various genes transcribed by the vegetative RNA polymerase and hence recognized by the regular σ^A and σ^H subunits. The genes that are switched on include those for σ^F and σ^E, resulting in the switch to prespore and mother cell differentiation (Fig.16.7).

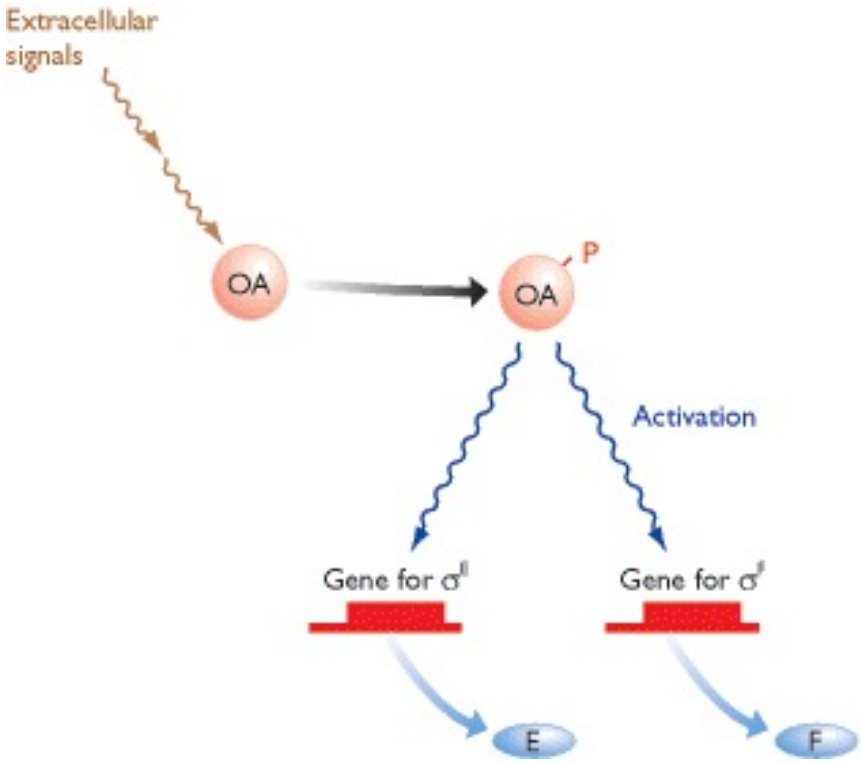


Fig-16.7: Role of SpoOA in *Bacillus* Sporulation

Initially, both σ^F and σ^E are present in each of the two differentiating cells. This is not exactly what is wanted because σ^F is the prespore-specific subunit and so should be active only in this cell, and σ^E is mother-cell specific.

σ^F is activated by release from a complex with a second protein, SpoIIAB:

This is controlled by a third protein, SpoIIAA, which, when unphosphorylated, can also attach to SpoIIAB and prevent the latter from binding to σ^F. If SpoIIAA is unphosphorylated then σ^F is released and is active; when SpoIIAA is phosphorylated then σ^F remains bound to SpoIIAB and so is inactive. In the mother cell, SpoIIAB phosphorylates SpoIIAA and so keeps σ^F in its bound inactive state. But in the prespore, SpoIIAB's attempts to phosphorylate SpoIIAA are antagonized by yet another protein, SpoIIE, and so σ^F is released and becomes active. SpoIIE's ability to antagonize SpoIIAB in the prespore but not the mother cell derives from the fact that SpoIIE molecules are bound to the membrane on the surface of the septum. Because the prespore is much smaller than the mother cell, but the septum surface area is similar in both, the concentration of SpoIIE is greater in the prespore, and this enables it to antagonize SpoIIAB.

σ^{E} *is Activated by Proteolytic Cleavage of a Precursor Protein:*

The protease that carries out this cleavage is the SpoIIGA protein, which spans the septum between the prespore and mother cell. The protease domain, which is on the mother-cell side of the septum, is activated by binding of SpoIIR to a receptor domain on the prespore side. It is a typical receptor-mediated signal transduction system (Fig.16.8). SpoIIR is one of the genes whose promoter is recognized specifically by σ^{F}, so activation of the protease, and conversion of pre-σ^{E} to active σ^{E}, occurs once σ^{F}-directed transcription is underway in the prespore.

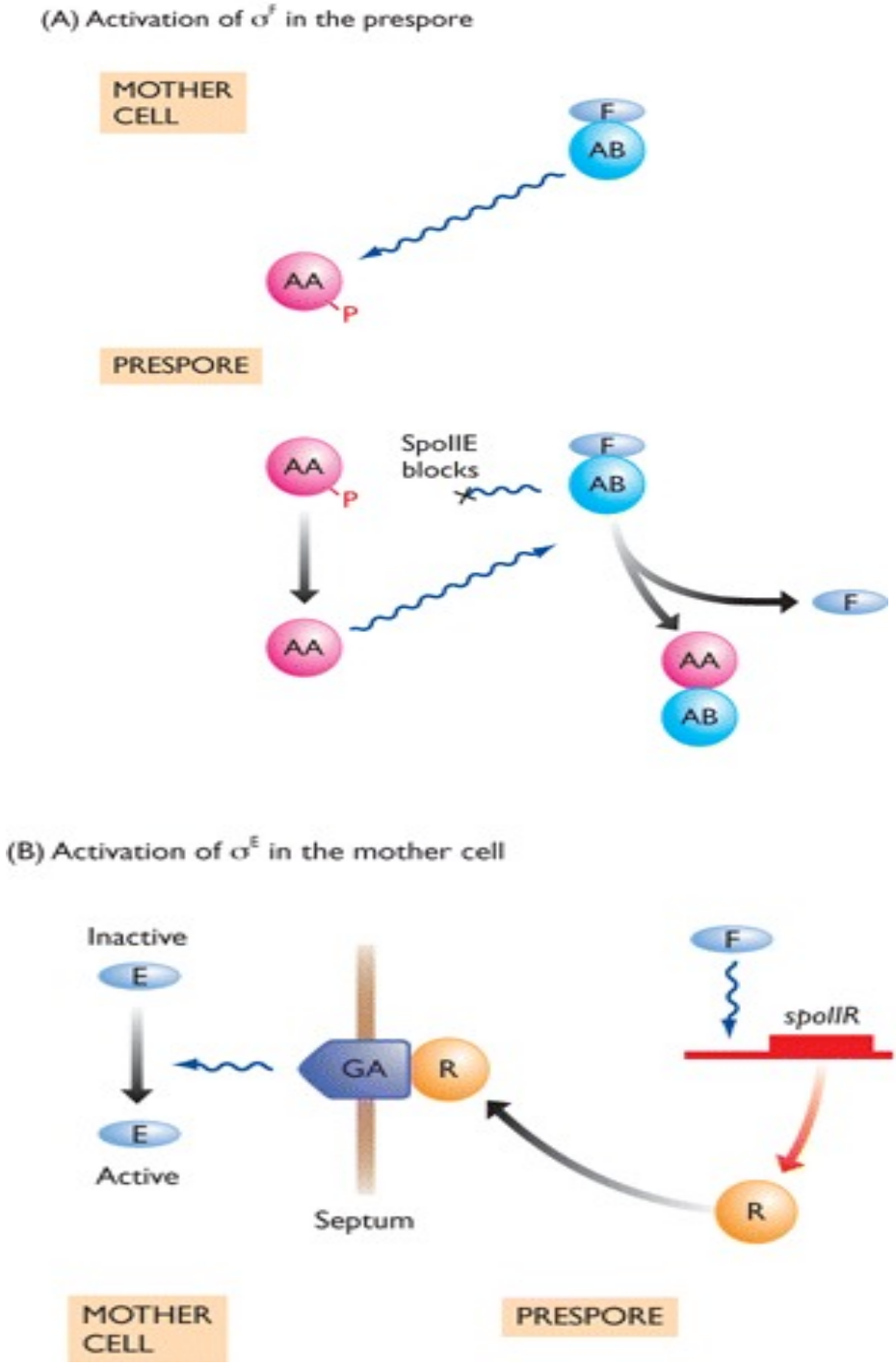


Fig-16.8: Activation of the prespore- and mother-cell-specific σ subunits during *Bacillus* sporulation

Activation of σ^F and σ^E is just the beginning of the story. In the prespore, about 1 hour after its activation, σ^F responds to an unknown signal (possibly from the mother cell) which results in a slight change in genome activity in the spore. This includes transcription of a gene for another σ subunit, σ^G, which recognizes promoters upstream of genes whose products are required during the later stages of spore differentiation. One of these proteins is SpoIVB, which activates another septum-bound protease, SpoIVF (Fig.16.9). This protease then activates a second mother cell σ subunit, σ^K, which is coded by a σ^E-transcribed gene but retained in the mother cell in an inactive form until the signal for its activation is received from the prespore. σ^K directs transcription of the genes whose products are needed during the later stages of the mother-cell differentiation pathway.

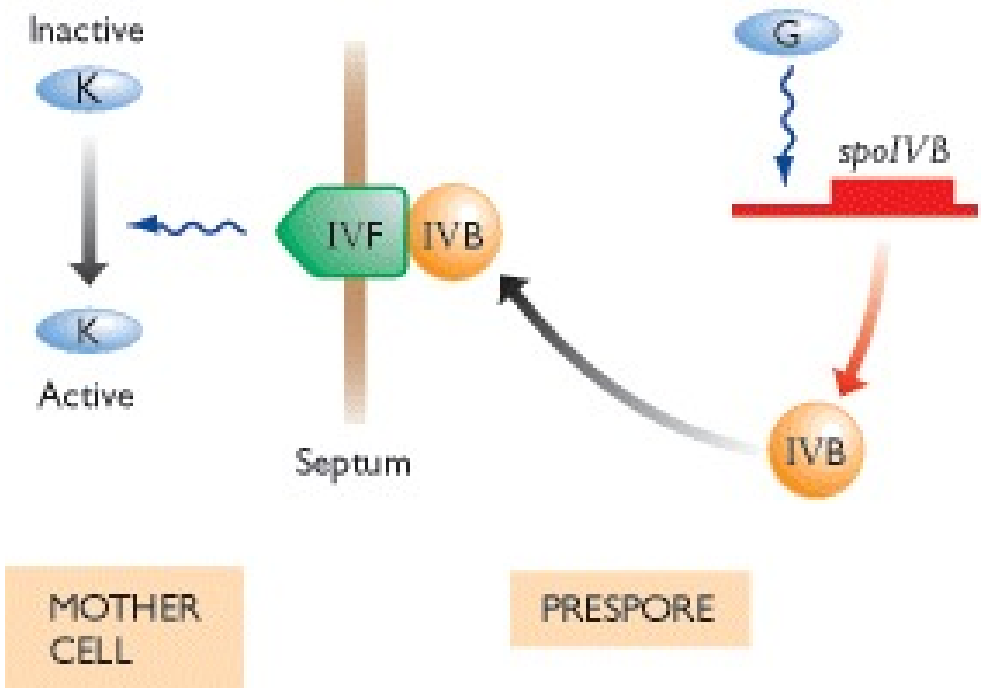


Fig-16.9: Activation of σ^K during *Bacillus* Sporulation

16.4 SUMMARY:

The mating behaviour of a yeast cell is specified by the set of genes that it contains at its mating-type (MAT) locus. In almost all ascomycete yeast species, typified by *Saccharomyces cerevisiae*, there are three types of cells, defined by their mating behaviour: two types of haploid cell with genotypes MATa and MATa and one type of diploid cell with genotype MATa/a. Each of these cell types can replicate to produce another cell of the same type vegetatively by mitosis and budding. This asexual cell cycle is the main way that yeast cells proliferate, but many species can also reproduce sexually in a cycle that involves mating followed by meiosis and sporulation. Mating is the process by which two haploid cells of opposite mating types (MATa and MATa) fuse to form a diploid cell (MATa/a). Conversely, meiosis followed by sporulation is the process by which a diploid cell turns into four spores that germinate into haploid cells (two with genotype MATa and two with MATa).

The isotype or class of a given antibody molecule is determined based on the constant region sequence: IgM(m), IgG(g), IgA(a), IgD(d), or IgE(e). Each isotype can have either

lambda or kappa light chains. Depending on the minor differences in the amino acid sequences of the α, γ heavy chains, they are further classified into sub-isotypes that determine the subclass of antibody molecules they constitute. In humans, there are two sub-isotypes of heavy chains and thus two subclasses, IgA1 and IgA2 and four sub-isotypes of g heavy chains and therefore four subclasses, IgG1, IgG2, IgG3, and IgG4.

Changes in genome activity during sporulation are controlled largely by the synthesis of special σ subunits that change the promoter specificity of the *Bacillus* RNA polymerase. Recall that the σ subunit is the part of the RNA polymerase that recognizes the bacterial promoter sequence, and that replacement of one σ subunit with another with a different DNA-binding specificity can result in a different set of genes being transcribed. During bacterial sporulation, sigma factors regulate genome activity by activating specific transcriptional programs in different cell compartments. A cascade of sporulation-specific sigma factors, including σF, σE, σG, and σK in *B. subtilis*, replaces the vegetative sigma factors (σA and σH) and directs gene expression for the different stages of spore formation in both the mother cell and the forespore. Activation of these sigma factors occurs at different times and in a cell-type-specific manner, ensuring the correct sequence of events, such as engulfment and cortex formation, occurs at the right time and place.

16.5 TECHNICAL TERMS:

IgM(m), IgG(g), IgA(a), IgD(d), or IgE(e), Sigma Factors, σF, σE, σG, σK.

16.6 SELF-ASSESSMENT QUESTIONS:

1) Explain in detail about the mechanism involved in mating type switching in Yeast.
2) Write an essay on diversity in Ig molecules.
3) Mechanism of class switching process in Ig diversity.
4) Give an account on Regulation of genome activity during sporulation by Sigma subunits.

16.7 SUGGESTED READINGS:

1) Verma PS and Agarwal VK (2005). Cell Biology, Genetics, Molecular Biology, Evolution, and Ecology. Multi colored Edition.
2) McGee, David & Coker, Christopher & Harro, Janette & Mobley, Harry. (2001). Bacterial Genetic Exchange. Doi: 10.1038/npg.els.0001416.
3) Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4) Russel, P.J. Genetics (1998) – The Benjamin / Cummings Publishing Company, Inc., California
5) Singh, B.D. Fundamentals of Genetics (2001) – Kalyani Publishers, New Delhi.

Prof. K. Mallikarjuna