Authors

Dr Geetika Singh, Assistant Professor, Department of Botany, Govt. College for Girls, Panjab University, Chandigarh, India
Units (1.2, 3, 4, 11)

Dr Richa Sharma, Assistant Professor, Department of Microbiology, Shaheed Rajguru College of Applied Science for Women, University of Delhi

Dr Rekha Mehrotra, Associate Professor, Department of Biology, Shaheed Rajguru College of Applied Science for Women, University of Delhi
Units (7 & 8)

Dr. Sushil Kumar Upadhyay, Assistant Professor, Department of Biotechnology, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala (Haryana), India
Units (2 & 5)

Meena Kumar, M.Sc Botany, B.Ed., TGT, Senior Science Teacher, Delhi Government School
Units (6, 12, 13, 14)

Dr. Helianthous Verma, Assistant Professor, Department of Zoology, Ramjas College, University of Delhi, Delhi
Units (9.2-9.3)

Dr. Pradeep Kumar, Research Associate-III, Department of Pediatrics, Army Hospital Research & Referral, New Delhi
Units (9.4-9.5, 10.2)

Vikas® Publishing House: Units (1.0 - 1.1, 1.3 to 1.9, 9.0-9.1, 9.6-9.10, 10.0-10.1, 10.3 to 10.8)

“The copyright shall be vested with Alagappa University”

All rights reserved. No part of this publication which is material protected by this copyright notice may be reproduced or transmitted or utilized or stored in any form or by any means now known or hereinafter invented, electronic, digital or mechanical, including photocopying, scanning, recording or by any information storage or retrieval system, without prior written permission from the Alagappa University, Karaikudi, Tamil Nadu.

Information contained in this book has been published by Vikas® Publishing House Pvt. Ltd. and has been obtained by its Authors from sources believed to be reliable and are correct to the best of their knowledge. However, the Alagappa University, Publisher and its Authors shall in no event be liable for any errors, omissions or damages arising out of use of this information and specifically disclaim any implied warranties or merchantability or fitness for any particular use.

Vikas® is the registered trademark of Vikas® Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: A-27, 2nd Floor, Mohan Co-operative Industrial Estate, New Delhi 1100 44
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

<table>
<thead>
<tr>
<th>Syllabi</th>
<th>Mapping in Book</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BLOCK I: MUTATION, DNA DAMAGE AND REPAIR</strong></td>
<td></td>
</tr>
<tr>
<td>Unit 1: Mutation and its Types, Mutation Rate and its Determination and</td>
<td>Unit 1: Mutation (Pages I-23);</td>
</tr>
<tr>
<td>Mutagenesis.</td>
<td></td>
</tr>
<tr>
<td>Unit 2: Mutagens-Types, Physical Mutagens, DNA Reactive Chemicals,</td>
<td>Unit 2: Mutagens and its Types (Pages 24-52);</td>
</tr>
<tr>
<td>Base Analogs, Intercalating Agents, Metals and Biological Agents.</td>
<td></td>
</tr>
<tr>
<td>Unit 3: DNA Damages-Deamination of Bases, Alkylation, Damage Due to</td>
<td>Unit 3: DNA Damages (Pages 53-70);</td>
</tr>
<tr>
<td>Reactive Oxygen, UV Induced Damage.</td>
<td></td>
</tr>
<tr>
<td>Unit 4: Repair Pathways (Methyl-Directed Mismatch Repair, Nucleotide</td>
<td>Unit 4: Repair Pathways (Pages 71-91);</td>
</tr>
<tr>
<td>Excision Repair, Base Excision Repair, Recombinational Repair, SOS</td>
<td></td>
</tr>
<tr>
<td>Inducible Repair, Specific Repair for Oxidative DNA Damage, Pyrimidine</td>
<td></td>
</tr>
<tr>
<td>Dimers and Alkylation Induced Damage and Adaptive Response).</td>
<td></td>
</tr>
<tr>
<td><strong>BLOCK II: BACTERIAL RECOMBINATION</strong></td>
<td></td>
</tr>
<tr>
<td>Unit 5: Recombination- Types of Recombinations, Models for Homologous</td>
<td>Unit 5: Recombination and its Types (Pages 92-123);</td>
</tr>
<tr>
<td>Recombination, Molecular Mechanism of Homologous Recombination, Homologous</td>
<td></td>
</tr>
<tr>
<td>Recombination in Eukaryotes, Mating Type Switching, Molecular Mechanism</td>
<td>Unit 6: Conjugation and F-Factor (Pages 124-141);</td>
</tr>
<tr>
<td>for Site-Specific Recombination. Biological Roles of Site Specific</td>
<td></td>
</tr>
<tr>
<td>Recombination.</td>
<td></td>
</tr>
<tr>
<td>Unit 6: Conjugation- Conjugation by <em>E. coli</em> F Factor, Structure of</td>
<td>Unit 7: Transformation (Pages 142-157);</td>
</tr>
<tr>
<td>F-Factor, Regulation of F-Factor Fertility, Establishment of Cell</td>
<td></td>
</tr>
<tr>
<td>Contact, DNA Mobilization, Transfer and Separation of Mating Pair, Hfr</td>
<td>Unit 8: Transduction and its Types (Pages 158-171);</td>
</tr>
<tr>
<td>Conjugation and Chromosomal Transfer, Interrupted Mating and</td>
<td></td>
</tr>
<tr>
<td>Conjugational Mapping.</td>
<td></td>
</tr>
<tr>
<td>Unit 7: Transformation- Mechanism of Natural Competence and Transformation in <em>Bacillus Subtilis</em> and <em>Streptococcus Pneumoniae</em>. Transformation by Inducing Artificial Competence, Gene Linkage and Mapping by Transformation.</td>
<td></td>
</tr>
<tr>
<td>Unit 8: Transduction- Generalized and Specialized Transduction.</td>
<td></td>
</tr>
<tr>
<td><strong>BLOCK III: OPERON CONCEPT AND EXTRA CHROMOSOMAL MATERIALS</strong></td>
<td></td>
</tr>
<tr>
<td>Unit 9: Gene Concept- Regulation of Bacterial Gene Expression, Lactose</td>
<td>Unit 9: Gene Concept and Lactose System (Pages 172-205);</td>
</tr>
<tr>
<td>System, Coordinate Regulation, Lac Components, Positive, Negative</td>
<td></td>
</tr>
<tr>
<td>Regulation and Catabolite Repression.</td>
<td></td>
</tr>
<tr>
<td>Unit 10: Tryptophan Operon - Attenuation, Arabinose Operon and its</td>
<td>Unit 10: Tryptophan Operon and Arabinose Operon (Pages 206-219);</td>
</tr>
<tr>
<td>Regulation.</td>
<td></td>
</tr>
<tr>
<td>Unit 11: Plasmids- Types of Plasmids - F, R &amp; Col Plasmids. Properties</td>
<td>Unit 11: Plasmids: Types and its Properties (Pages 220-242);</td>
</tr>
<tr>
<td>of Plasmids - Sex Factors, Drug Resistant, Colicinogenic, Agrobacterium</td>
<td></td>
</tr>
<tr>
<td>Ti and Broad Host Range Plasmid.</td>
<td></td>
</tr>
<tr>
<td>BLOCK IV: TRANPOSABLE ELEMENTS AND EPIGENETICS</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Unit 12:</strong> Detection and Purification of Plasmid DNA. Transfer of Plasmid DNA. Replication of Plasmid. Control of Copy Number, Plasmid Amplification, Curing and Incompatibility.</td>
<td></td>
</tr>
<tr>
<td><strong>Unit 13:</strong> Transposable Elements- Types of Transposable Elements, Structure, Genetic Organization and Mechanism of Transposition of Tn5, Tn3 and Related Transposons, Bacteriophage Mu, Tn7 and IS911, Integrons, Retrotransposons.</td>
<td></td>
</tr>
<tr>
<td><strong>Unit 14:</strong> Epigenetics- Definition, Molecular basis, Mechanisms, Functions and Epigenetics in Bacteria.</td>
<td></td>
</tr>
<tr>
<td><strong>Unit 12:</strong> Detection, Purification and Transfer of Plasmid DNA, and Plasmid Replication (Pages 243-265); <strong>Unit 13:</strong> Transposable Elements (Pages 266-290); <strong>Unit 14:</strong> Epigenetics (Pages 291-310)</td>
<td></td>
</tr>
</tbody>
</table>
# CONTENTS

## INTRODUCTION

### BLOCK 1: MUTATION, DNA DAMAGE AND REPAIR

#### UNIT 1 MUTATION

1.0 Introduction  
1.1 Objectives  
1.2 Mutation and its Types  
1.3 Mutation Rate and its Determination  
1.4 Mutagenesis  
1.5 Answers to Check Your Progress Questions  
1.6 Summary  
1.7 Key Words  
1.8 Self Assessment Questions and Exercises  
1.9 Further Readings

#### UNIT 2 MUTAGENS AND ITS TYPES

2.0 Introduction  
2.1 Objectives  
2.2 Mutagens: History and Causes  
2.2.1 History of Mutagens  
2.2.2 Causes of Mutagens  
2.3 Types of Mutagens  
2.3.1 Physical Mutagens  
2.3.2 Chemical Mutagens  
2.3.3 Biological Mutagens  
2.4 Tests System for Mutagens  
2.5 Consequences of Mutagens Exposures  
2.6 Answers to Check Your Progress Questions  
2.7 Summary  
2.8 Key Words  
2.9 Self Assessment Questions and Exercises  
2.10 Further Readings

#### UNIT 3 DNA DAMAGES

3.0 Introduction  
3.1 Objectives  
3.2 DNA Damages  
3.2.1 Deamination of Bases  
3.2.2 Alkylation  
3.2.3 Damage Due to Reactive Oxygen  
3.2.4 UV Induced DNA Damage  
3.3 Answers to Check Your Progress Questions  
3.4 Summary  
3.5 Key Words  
3.6 Self Assessment Questions and Exercises  
3.7 Further Readings
UNIT 4 REPAIR PATHWAYS 71-91
4.0 Introduction
4.1 Objectives
4.2 DNA Damage and Repair Pathways
   4.2.1 An SOS Repair System or Recombinational Repair
   4.2.2 Methyl Directed Mismatch Repair
   4.2.3 Nucleotide Excision Repair
   4.2.4 Recombinational Repair
   4.2.5 SOS Inducible Repair
   4.2.6 Specific Repair for Oxidative DNA Damage
   4.2.7 Pyrimidine Dimers
   4.2.8 Alkylation Induced Damage and Adaptive Response
4.3 Answers to Check Your Progress Questions
4.4 Summary
4.5 Key Words
4.6 Self Assessment Questions and Exercises
4.7 Further Readings

BLOCK II: BACTERIAL RECOMBINATION

UNIT 5 RECOMBINATION AND ITS TYPES 92-123
5.0 Introduction
5.1 Objectives
5.2 Recombination and its Types
   5.2.1 Types of Recombination
5.3 Homologous Recombination
   5.3.1 Models for Homologous Recombination
   5.3.2 Molecular Mechanism of Homologous Recombination
   5.3.3 Enzymes of Homologous Recombination
   5.3.4 Role of Reca Protein in Homologous Genetic Recombination
   5.3.5 Homologous Recombination in Prokaryotes and Eukaryotes
   5.3.6 In Vitro Homologous Recombination
5.4 Site-Specific Recombination
   5.4.1 Types of Site-Specific Recombination
   5.4.2 Molecular Mechanism of Site-Specific Recombination
   5.4.3 Biological Roles of Site-Specific Recombination
5.5 Answers to Check Your Progress Questions
5.6 Summary
5.7 Key Words
5.8 Self Assessment Questions and Exercises
5.9 Further Readings

UNIT 6 CONJUGATION AND F-FACTOR 124-141
6.0 Introduction
6.1 Objectives
6.2 Conjugation
   6.2.1 Conjugation by Escherichia coli F Factor
   6.2.2 Fertility Factor or F Factor
   6.2.3 Hfr Conjugation and Chromosomal Transfer
   6.2.4 The F’ (F Prime) Factor
   6.2.5 Interrupted Mating and Conjugational Mapping
UNIT 10  TRYPTOPHAN OPERON AND ARABINOSE OPERON 206-219
10.0 Introduction
10.1 Objectives
10.2 Tryptophan Operon
10.3 Arabinoise Operon and its Regulation
10.4 Answers to Check Your Progress Questions
10.5 Summary
10.6 Key Words
10.7 Self Assessment Questions and Exercises
10.8 Further Readings

UNIT 11  PLASMIDS: TYPES AND ITS PROPERTIES 220-242
11.0 Introduction
11.1 Objectives
11.2 Plasmids: Types and Properties
   11.2.1 Types of Plasmids
   11.2.2 Properties and Characteristics
11.3 Answers to Check Your Progress Questions
11.4 Summary
11.5 Key Words
11.6 Self Assessment Questions and Exercises
11.7 Further Readings

BLOCK IV: TRANSPOSABLE ELEMENTS AND EPIGENETICS
UNIT 12  DETECTION, PURIFICATION AND TRANSFER OF PLASMID DNA, AND PLASMID REPLICATION 243-265
12.0 Introduction
12.1 Objectives
12.2 Plasmid DNA
12.3 Detection and Purification of Plasmid DNA
   12.3.1 Isolation and Purification of Plasmid DNA
12.4 Transfer of Plasmid DNA
12.5 Replication of Plasmid
12.6 Plasmid Copy Number
12.7 Answers to Check Your Progress Questions
12.8 Summary
12.9 Key Words
12.10 Self Assessment Questions and Exercises
12.11 Further Readings

UNIT 13  TRANSPOSABLE ELEMENTS 266-290
13.0 Introduction
13.1 Objectives
13.2 Transposable Elements
   13.2.1 Types of Transposable Elements
13.3 Genetic Organization and Mechanism of Transposition
   13.3.1 Tn3 Transposon
   13.3.2 Tn5 Transposon
13.3.3 Bacteriophage Mu
13.3.4 Insertion Sequence–Tn7 and IS911
13.3.5 Integrons and Retrotransposons
13.4 Answers to Check Your Progress Questions
13.5 Summary
13.6 Key Words
13.7 Self Assessment Questions and Exercises
13.8 Further Readings

UNIT 14 EPIGENETICS 291-310
14.0 Introduction
14.1 Objectives
14.2 Epigenetics
  14.2.1 Definitions of Epigenetics
  14.2.2 Molecular Basis of Epigenetic
  14.2.3 Epigenetics Mechanisms
  14.2.4 Functions and Consequences
  14.2.5 Epigenetics in Bacteria
14.3 Answers to Check Your Progress Questions
14.4 Summary
14.5 Key Words
14.6 Self Assessment Questions and Exercises
14.7 Further Readings
Microbial genetics is a subject area within microbiology and genetic engineering. Microbial genetics studies microorganisms for different purposes. The microorganisms that are observed are bacteria, and archaea. Some fungi and protozoa are also subjects used to study in this field. The studies of microorganisms involve studies of genotype and expression system. Genotypes are the inherited compositions of an organism. Genetic Engineering is a field of work and study within microbial genetics. The usage of recombinant DNA technology is a process of this work. The process involves creating recombinant DNA molecules through manipulating a DNA sequence. That DNA created is then in contact with a host organism. Cloning is also an example of genetic engineering.

After the discovery of DNA transfer in bacteria, bacteria became objects of great interest to geneticists because their rate of reproduction and mutation is higher than in larger organisms, i.e., a mutation occurs in a gene about one time in 10,000,000 gene duplications, and one bacterium may produce 10,000,000,000 offspring in 48 hours. Conjugation, transformation, and transduction have been important methods for mapping the genes on the chromosomes of bacteria.

Microbial genetics is principally a branch of genetics concerned with the transmission of hereditary characters in microorganisms. Within the usual definition, microorganisms include prokaryotes like bacteria, unicellular or mycelial Eukaryotes, for example yeasts and other fungi, and viruses, notably bacterial viruses (bacteriophages). Because of their relative simplicity, microbes are ideally suited for combined biochemical and genetic studies, and have been successful in providing information on the genetic code and the regulation of gene activity. Microbial genetics has played a unique role in developing the fields of molecular and cell biology and also has found applications in medicine, agriculture, and the food and pharmaceutical industries. Microbial genetics also has applications in being able to study processes and pathways that are similar to those found in humans, such as drug metabolism.

This book, Microbial Genetics, is divided into four blocks that are further divided into fourteen units which will help to understand the basic concepts of mutation, DNA damage and repair, mutagenesis, mutagens, DNA reactive chemicals, DNA damages, repair pathways, bacterial recombination, models for homologous recombination, conjugation, F factor, Hfr conjugation and chromosomal transfer, transformation, transformation by inducing artificial competence, gene linkage and mapping by transformation, transduction - generalized and specialized transduction, gene concept, lactose system, tryptophan operon, arabinose operon, plasmids, transposable elements and epigenetics.

The book follows the self-instruction mode or the SIM format wherein each unit begins with an ‘Introduction’ to the topic followed by an outline of the ‘Objectives’. The content is presented in a simple and structured form interspersed with ‘Check Your Progress’ questions and answers for better understanding. A list of ‘Key Words’ along with a ‘Summary’ and a set of ‘Self Assessment Questions and Exercises’ is provided at the end of the each unit for effective recapitulation.
UNIT 1 MUTATION

Structure
1.0 Introduction
1.1 Objectives
1.2 Mutation and its Types
1.3 Mutation Rate and its Determination
1.4 Mutagenesis
1.5 Answers to Check Your Progress Questions
1.6 Summary
1.7 Key Words
1.8 Self Assessment Questions and Exercises
1.9 Further Readings

1.0 INTRODUCTION

At the simplest level, a mutation is a change or transformation. In biology, mutations refer to changes in chromosomes and genes, which typically manifest physically. The effect of a mutation can depend on the region in which the sequence of genetic material has been changed. The simplest and the most harmless are substitutions of a single base pair with another, with no effect on protein sequence. At the other end are insertion or deletion mutations that lead to non-functional gene products. Mutations can also occur on a large scale, with long stretches of DNA (or RNA when it is the genetic material) being inverted, inserted, duplicated, deleted, transposed or translocated.

The result of a mutation could be harmful, beneficial, neutral or even silent. Mutation can lead to the loss or gain of a specific function, to changes to the expression levels, or in extreme cases, even embryonic lethality. Mutations can be classified in various ways depending on the cause of the mutation, its effect on the function of the gene product or the kind of changes to the structure of the gene itself.

In genetics, the mutation rate is the frequency of new mutations in a single gene or organism over time. Mutation rates are not constant and are not limited to a single type of mutation, therefore there are many different types of mutations. Basically, the mutation rates are given for specific classes of mutations. Point mutations are a class of mutations which are small or large scale insertions or deletions. The rate of these types of substitutions can be further subdivided into a mutation spectrum which describes the influence of the genetic context on the
mutation rate. Mutagenesis is a process by which the genetic information of an organism is changed, resulting in a mutation. It may occur spontaneously in nature, or as a result of exposure to mutagens. It can also be achieved experimentally using laboratory procedures. In nature mutagenesis can lead to cancer and various heritable diseases, but it is also a driving force of evolution.

In this unit, you will study about the mutation and its types, mutation rate and its determination, and mutagenesis.

1.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss what mutation is
- Explain radiation induced mutations
- Analyze practical applications of mutations
- Describe gene mutations and its types
- Explain about the mutation rate
- Determine the mutation rate
- Understand mutagenesis

1.2 MUTATION AND ITS TYPES

The term mutation refers both to the change in the genetic material and to the process by which the change occurs. An organism exhibiting a novel phenotype as a result of the presence of mutation is referred to as mutant. Mutation refers to any sudden, heritable change in the genotype of an organism not explainable by recombination of preexisting genetic variability. Such genotypic changes includes change in chromosome number (euploidy and aneuploidy), gross changes in the structure of chromosomes (chromosomes aberration) and changes in individual genes. The term mutation refers to changes in individual genes.

Many mutations involve changes in single base pairs, the substitution of one basepair for another, or duplication or deletion of single base pairs, such mutation is referred as point mutation.

Mutation is the ultimate source of all genetic variation, it provides the raw material for evolution,(Refer Figure 1.1). Recombination (independent assortment plus recombination of genetic variability present in individual chromosomes) rearranges this genetic variability into new combinations and natural selection preserves the combinations adapted to existing environmental conditions. Without mutation, all genes would exist on only one form. Alleles would not exist and genetic analysis would not be possible. Organisms would not be able to evolve and adapt to environmental changes.
Mutations are inherited alterations in the DNA sequence. DNA is a highly stable molecule that is replicated with amazing accuracy but changes in DNA structure and errors of replication do occur. (Refer Figure 1.2). A mutation is defined as an inherited change in genetic information, the descendants may be cells or organisms.

Fig. 1.2 DNA Incoming UV Structure Before and After Incoming UV Photon

The appearance of a new mutation is a rare event. Most mutations that were originally studied occurred spontaneously. This class of mutation is termed spontaneous mutations. Historically, geneticists recognized these in nature. The mutations were collected, and the inheritance of these mutations were analyzed. But these mutations clearly represent only a small number of all possible mutations. To genetically dissect a biological system further, new mutations were created by scientists by treating an organism with a mutagenizing agent. These mutations are called induced mutations.

Mutagens are chemical compounds or forms of radiation (such as, UltraViolet (UV) light or X-rays) that cause irreversible and heritable changes (mutations) in the cellular genetic material, DeoxyriboNucleic Acid (DNA). In this way, mutagenesis becomes a cumulative process, stretching over the lifetime of an organism.
The biological consequences of a mutation depend upon many critical factors, such as the target loci, size of the mutation, timing during the cell cycle, and compounding effects of preexisting mutations. Mutagenic lesions persist when they escape detection by protective cellular DNA repair mechanisms, when mistakes occur in the repair process, or when repair mechanisms are overwhelmed by extensive damage. Upon subsequent cellular replication, these mutations become fixed in the genome and are inherited by all daughter cells. Thus, a mutagenic event occurring in a nonfunctional area of DNA will have no effect (silent mutation), whereas a similar change in an actively transcribed region may profoundly affect gene expression and phenotype or even lead to cell death (lethal mutation).

There are many chemical mutagens which can react with nucleus and the DNA, for example, base analogues which are derivatives of normal bases. This can cause direct mutagenesis. Similarly alkylation agents like methylmethanesulphonate, ethylmethanesulphonate and ethylnitrosourea are capable of producing DNA lesions which are repaired to prevent any disruption to process of transcription and replication. Nitrous acid deaminates cytosine to produce Uracil (U) which can base pair with adenine creating transitions. Similarly, 5 bromouracil can lead to mutations. Transition is where a pyrimidine changes to another Pyrimidine or Purine to another Purine while in transversion change involves Purine to Pyrimidine or Pyrimidine to Purine.

The spontaneous mutation rate varies. Large gene provide a large target and tend to mutate more frequently. A study of the five coat color loci in mice showed that the rate of mutation ranged from $2 \times 10^{-6}$ to $40 \times 10^{-6}$ mutations per gamete per gene. Data from several studies on Eukaryotic organisms shows that in general the spontaneous mutation rate is $2-12 \times 10^{-6}$ mutations per gamete per gene. Given that the human genome contains 100,000 genes, we can conclude that we would predict that 1-5 human gametes would contain a mutation in some gene. Spontaneous mutations occur infrequently although the observed frequencies vary from gene to gene and from organism to organism.

**Phenotypic Effects of Mutation**

Mutations cause some detectable phenotypic change for their presence to be recognized. The effects of mutations on phenotype range from alterations so minor that they can be detected only by special genetic or biochemical techniques to gross modifications of morphology. A gene is a specific sequence of nucleotide pairs coding for a particular polypeptide. Any mutation occurring within a given gene will produce a new form. Because of the degeneracy of the genetic code, some base pair changes do not change the protein products encoded by the genes in any way. Genes containing mutations with small effects can be recognized by special techniques called isospecies. Mutations may be either recessive or dominant. In haploid organisms like viruses or bacteria, both recessive and dominant mutations can be recognized by their effects on the phenotype of the organism in which they originated. In diploid organisms, recessive mutation will be recognized only when
present in the homozygous condition. Most recessive mutations in diploids will not be recognized at the time of their occurrence, since they are present in heterozygous state. Sex linked recessive mutations are an exception, since they will be expressed in hemizygous state in the heterogametic sex. The most useful mutations for genetic analyses of many biological processes are conditional lethal mutations.

**Somatic and Germinal Mutations**

Mutation occur in any cell and at any state in the cell cycle. The immediate effect of mutation and its ability to produce phenotypic change are determined by its dominance, the type of cell in which it occurs and when it happens relative to life cycle of the organism.

If the mutation occurs in a somatic cell, which can produce cell like itself but not the whole organism, the mutant change will be perpetuated only in somatic cells that descend from original cell in which mutation occurred. The Delicious apple and the navel orange were mosaics in somatic tissues. Changes that give these two fruits their desirable qualities followed spontaneous mutation in single cells. If dominant mutation occur in germ cells, their effects may be expressed immediately in progeny. If mutations are recessive, their effects are obscured in diploids. Germinal mutations may occur at any stage in the reproductive cycle of the organism, but they are common during some stages than others.

**Back Mutation and Suppressor Mutation**

The mutation of a wild type gene to form that results in a mutant phenotype is referred as forward mutation. In a population composed entirely of brown eyed individual, the allele for blue eyes is thought as mutant allele. Mutation events are reversible. Mutation may occur that restores the original wild type phenotype. This is referred to as back mutation, reverse mutation or reversion. Reversion may occur by true back mutation at the same site in the gene as the original mutation, restoring the wild type nucleotide sequence or by the occurrence of a second mutation at a different location in the genome, which compensates for first mutation, these mutations are called suppressor mutations. It may occur at distinct sites in the same gene as the original mutation or in different genes, even in different chromosomes.

**Radiation Induced Mutations**

That portion of the electromagnetic spectrum containing wavelengths that are shorter and of higher energy than the visible light can be divided into ionizing radiation (X rays, gamma rays, cosmic rays) and non-ionizing radiations (UV light). Ionizing radiations are of high energy and are useful for medical diagnosis because they penetrate the living tissue. UV rays having lower energy penetrate only the surface layer of cells in higher plants and animals and do not induce ionizations. UV rays dissipate their energy to atoms that they encounter raising the electrons in the outer orbitals to higher energy levels referred as excitation. Molecules containing
atoms in their ionic or excited forms are chemically more reactive than those containing atoms in their normal stable state. The increasing reactivity of atoms present in DNA molecules is the basis of the mutagenic effects of UV light and ionizing radiations. UV radiations are readily absorbed by substances like purines and pyrimidines which enter a more reactive or excited state. The two major products of UV absorption by pyrimidines appear to be pyrimidine hydrates and pyrimidine dimmers.

**Chemically Induced Mutations**

The first chemical mutagen discovered was mustard gas. When C. Auerbach and her associates discovered the mutagenic effects of mustard gas and related compounds during World War II. These compounds are examples of the large class of chemical mutagens that transfer alkyl group to the bases in DNA, called alkylating agents.

Chemical mutagens are divided into two classes: those that are mutagenic to both replicating and non-replicating DNA such as the alkylating agent and nitrous acid, those that are mutagenic only to replicating DNA, which mainly includes acridine dyes, which bind to DNA and increase the probability of mistakes during replication and base analogs, purines and pyrimidines with structure similar to normal bases of DNA. The base analogs must be incorporated into DNA chains in place of normal bases during replication to exert their mutagenic effects.

**Base Analogs**: The base analogs that are mutagenic have structures similar to the normal bases so that they are incorporated into DNA during replication. The two most commonly used base analogs are 5 Bromo Uracil (U) and 2 Amino Purine.

**Nitrous Acid**: It is a very potent mutagen that acts directly on either replicating or non-replicating DNA by oxidative deamination of the bases that contain amino groups—Adenine (A), Guanine (G) and Cytosine (C). Conversion of the amino groups to keto groups changes the hydrogen bonding potential of the bases.

**Acridines**: The positively charged acridine intercalate between the stacked base pairs in DNA. They increase the rigidity and alter the conformation of the double helix, causing kinks in the molecule. When DNA molecules containing intercalated acridines replicate, addition and deletions of one or few base pairs occur. These small additions and deletions of single base pair result in reading frameshifts for the portion of the gene distal to the mutation.

**Alkylating and Hydroxylating Agents**: Alkylating agents, such as Nitrogen and Sulphur mustards, Methyl and Ethyl methanesulphonate have several effects on DNA. Nitrosoguanidine, one of the most potent chemical mutagens has been found to induce clusters of closely linked mutations in the segment of the chromosome that is replicating during mutagenic treatment. The hydroxylamine has a very specific mutagenic effects. It induces only GC-AT transitions.
Practical Applications of Mutations

Mutations are invaluable to the process of evolution since they provide the raw material required for its occurrence. Mutation provide the alleles required for various types of genetic analysis, from Mendel's two factor crosses to chromosome mapping to studies of genetic structures of populations. Even though most mutations make the organism less efficient and disadvantageous, the possibility of developing new desirable traits through induced mutations has intrigued many plant breeders. Plant breeders have reported induced mutants in barley, wheat, oat, soybeans, tomatoes, fruit trees that improve cultivated strains. Barley mutants have been obtained that provide increased protein content and hull less seeds. One application of induced mutations came from efforts to improve the yield of penicillin by the mold penicillium. When penicillin was first discovered, yield was low and production was limited. Then millions of spores were irradiated and few of the surviving colonies produced more penicillin than the average. Such mutant overproducers of penicillin have proven invaluable in the commercial production of this important antibiotic.

Mutations and Humans

Purposeful artificial selection is not practiced in humans and therefore the possible advantages cited for domestic animals and plants do not apply to humankind. Variations do exist in populations, however and presumably they originated through past mutations. Since mutations are detrimental, it would seem advantageous from the stand point of short term effects for humans to avoid excessive exposure to mutagenic agents. In case of acute irradiation, two types of dangers are considered. i.e. the immediate damage to the exposed person, more damage to the DNA in his or her reproductive cells, which would affect future generations. The immediate damage is indicated by burns and other direct or secondary effects on body tissues. When doses are on the order of 50 mr or lower, no immediate damage can be detected, although some harmful effects such as induction of leukemia and shortening
of life span may occur. Effects of the second type of damage will be observed in only future generations. This is a reason to believe that exposure to high energy radiations at any dosage level is potentially harmful. In the few results to date, mutations have been proportional to the dosage and the effects have been cumulative.

The relation between dosage and effects cannot be accurately measured in humans at present because of the complexity of the subject and difficulties of dealing with the genetics of humans. Preliminary reports including data on children born to parents who survived the Nagasaki bombing have revealed a significant increase in incidence of leukemia. The normal sex ratio has been altered, and changes have been interpreted as resulting from induced sex linked lethals. Most of the data available on mutation rates and the nature of mutations have come from other organisms and only by inference are they applied to humans. However the genetic material in humans is DNA, just as it is in most other organisms. It should not be surprising that the effects of irradiation are comparable to most organisms.

Mutations can be induced by several methods. The three general approaches used to generate mutations are radiation, chemical and transposon insertion. The first induced mutations were created by treating Drosophila with X-rays. Using this approach Mueller to induce lethal mutations. In addition to X-rays, other types of radiation treatments that have proven useful include gamma rays and fast neutron bombardment. These treatments can induce point mutations (changes in a single nucleotide) or deletions (loss of a chromosomal segment).

Chemical mutagens work mostly by inducing point mutations. Point mutations occur when a single base pair of a gene is changed. These changes are classified as transitions or transversions. Transitions occur when a purine is converted to a purine (A to G or G to A) or a pyrimide is converted to a pyrimidine (T to C or C to T). A transversion results when a purine is converted to a pyrimidine or a pyrimidine is converted to a purine. A transversion example is adenine being converted to a cytosine.

Two major classes of chemical mutagens are routinely used. These are alkylating agents and base analogs. Each has a specific effect on DNA. Alkylating agents (such as, ethyl Methane Sulphonate (EMS), Ethyl Ethane Sulphonator (EES) and Mustard Gas) can mutate both replicating and non-replicating DNA. By contrast, a base analog (5-bromouracil and 2-aminopurine) only mutate DNA when the analog is incorporated into replicating DNA. Each class of chemical mutagen has specific effects that can lead to transitions, transversions or deletions.

Scientists are now using the power of transposable elements to create new mutations. Transposable elements are mobile pieces of DNA that can move from one location in a genome to another. Often when they move to a new location, the result is a new mutant. The mutant arises because the presence of a piece of DNA in a wild type gene disrupts the normal function of that gene. As more and more is being learned about genes and genomes, it is becoming apparent that transposable elements are a power source of or creating insertional mutants.
Gene Mutations and its Types

Germ line mutations arise in cells that ultimately produce gametes. A germ line mutation can be passed to future generations producing individual organisms that carry the mutation in all their somatic and germ line cells. Mutations have been partitioned into those that affect a single gene, called gene mutations and those that affect the number or structure of chromosomes called chromosome mutations.

Base Substitutions: The simplest type of gene mutation is a base substitution, the alteration of a single nucleotide in the DNA. Base substitution are of two types. In transition, a purine is replaced by different purine or pyrimidine is replaced by a different pyrimidine. In transversion, a purine is replaced by a pyrimidine or a pyrimidine is replaced by a purine.

Insertions and Deletions: It mainly includes the addition or the removal of one or more nucleotide pairs. Insertions and deletions within sequences that encode proteins may lead to frameshift mutations, changes in the reading frame of the gene. It alter all amino acids encoded by nucleotides following mutation and have drastic effects on the phenotype.

Expanding Nucleotide Repeats: Mutations in which the number of copies of a set of nucleotides increases in number are called expanding nucleotide repeats. This type of mutation was observed in 1991 in a gene called FMR-1 which causes fragile X syndrome, the most common cause heredity cause of mental retardation. Expanding nucleotide repeats are found in almost 30 human diseases. The number of copies of the nucleotide repeat correlates with the severity or age of onset of disease.

The detailed knowledge of the structure and function of transposable elements is now being applied in the pursuit of new mutations. Stocks are created in which a specific type of element is present. This stock is then crossed to a genetic stock that does not contain the element. Once that element enters the virgin stock, it can begin to move around that genome. By carefully observing the offspring, new mutants can be discovered. This approach to developing mutants is termed insertional mutagenesis.

1.3 MUTATION RATE AND ITS DETERMINATION

In genetics, the mutation rate is the frequency of new mutations in a single gene or organism over time. Mutation rates are not constant and are not limited to a single type of mutation, therefore there are many different types of mutations. Mutation rates are given for specific classes of mutations. Point mutations are a class of mutations which are small or large scale insertions or deletions. There are also Missense and Nonsense mutations, which are variations of point mutations. The rate of these types of substitutions can be further subdivided into a mutation spectrum which describes the influence of the genetic context on the mutation rate.
There are several natural units of time for each of these rates, with rates being characterized either as mutations per base pair per cell division, per gene per generation, or per genome per generation. The mutation rate of an organism is an evolved characteristic and is strongly influenced by the genetics of each organism, in addition to strong influence from the environment. The upper and lower limits to which mutation rates can evolve is the subject of ongoing investigation. However, the mutation rate does vary over the genome. Over DNA, RNA or a single gene, mutation rates are changing.

When the mutation rate in humans increases certain health risks can occur, for example, cancer and other hereditary diseases. Having knowledge of mutation rates is vital to understanding the future of cancers and many hereditary diseases.

**Measurement**

An organism’s mutation rates can be measured by a number of techniques.

1. One way to measure the mutation rate is by the fluctuation test, also known as the Luria–Delbrück experiment. This experiment exhibits in bacteria mutations that occur in the absence of selection instead of the presence of selection.

2. This is very important to mutation rates because it proves experimentally mutations can occur without selection being a component. Therefore, mutations occur at random in bacteria and other organisms.

**Substitution Rates**

Many sites in an organism’s genome may admit mutations with small fitness effects. These sites are typically called neutral sites. Theoretically mutations under no selection become fixed between organisms at precisely the mutation rate. Fixed synonymous mutations, i.e., synonymous substitutions, are changes to the sequence of a gene that do not change the protein produced by that gene. They are often used as estimates of that mutation rate, despite the fact that some synonymous mutations have fitness effects. As an example, mutation rates have been directly inferred from the whole genome sequences of experimentally evolved replicate lines of *Escherichia coli* B.

**Mutation Accumulation Lines**

A particularly labor-intensive way of characterizing the mutation rate is the mutation accumulation line.

Mutation accumulation lines have been used to characterize mutation rates with the Bateman-Mukai Method and direct sequencing of, for example intestinal bacteria, round worms, yeast, fruit flies, small annual plants.
Variation in Mutation Rates

Mutation rates differ between species and even between different regions of the genome of a single species. These different rates of nucleotide substitution are measured in substitutions (fixed mutations) per base pair per generation. For example, mutations in intergenic, or non-coding, DNA tend to accumulate at a faster rate than mutations in DNA that is actively in use in the organism (gene expression). That is not necessarily due to a higher mutation rate, but to lower levels of purifying selection. A region which mutates at predictable rate is a candidate for use as a molecular clock.

If the rate of neutral mutations in a sequence is assumed to be constant (clock-like), and if most differences between species are neutral rather than adaptive, then the number of differences between two different species can be used to estimate how long ago two species diverged. In fact, the mutation rate of an organism may change in response to environmental stress. For example, UV light damages DNA, which may result in error prone attempts by the cell to perform DNA repair.

The human mutation rate is higher in the male germ line (sperm) than the female (egg cells), but estimates of the exact rate have varied by an order of magnitude or more. This means that a human genome accumulates around 64 new mutations per generation because each full generation involves a number of cell divisions to generate gametes. Human mitochondrial DNA has been estimated to have mutation rates of $3 \times 10^{-5}$ or $2.7 \times 10^{-5}$ per base per 20 year generation (depending on the method of estimation); these rates are considered to be significantly higher than rates of human genomic mutation at $2.5 \times 10^{-8}$ per base per generation. Using data available from whole genome sequencing, the human genome mutation rate is similarly estimated to be $1.1 \times 10^{-8}$ per site per generation.

The rate for other forms of mutation also differs greatly from point mutations. An individual microsatellite locus often has a mutation rate on the order of $10^{-4}$, though this can differ greatly with length.

Some sequences of DNA may be more susceptible to mutation. For example, stretches of DNA in human sperm which lack methylation are more prone to mutation.

Figure 1.4 illustrates the reported estimates of the human genome-wide mutation rate. The human germ line mutation rate is approximately $0.5 \times 10^{-9}$ per base pair per year.
In general, the mutation rate in unicellular Eukaryotes (and Bacteria) is roughly 0.003 mutations per genome per cell generation. However, some species, especially the ciliate of the genus *Paramecium* have an unusually low mutation rate. For instance, *Paramecium tetraurelia* has a base-substitution mutation rate of \(~2 \times 10^{-11}\) per site per cell division. This is the lowest mutation rate observed in nature so far, being about 75 × lower than in other eukaryotes with a similar genome size, and even 10 × lower than in most prokaryotes. The low mutation rate in *Paramecium* has been explained by its transcriptionally silent germ line nucleus, consistent with the hypothesis that replication fidelity is higher at lower gene expression levels.

The highest per base pair per generation mutation rates are found in viruses, which can have either RNA or DNA genomes. DNA viruses have mutation rates between \(10^{-6}\) to \(10^{-8}\) mutations per base per generation, and RNA viruses have mutation rates between \(10^{-3}\) to \(10^{-5}\) per base per generation.

**Mutational Spectrum**

The mutation spectrum of an organism is the rate at which different mutations occur at different sites. Typically two sites are considered, each of which may have three mutations, resulting in six total rates for most mutation spectra. The two sites are the two correct pairs possible in DNA, the A:T pairs and the C:G pairs. Figure 1.5 illustrates the Transitions (Alpha) and Transversions (Beta).
There is a systematic difference in rates for transitions (Alpha) and transversions (Beta).

**Spontaneous Mutations**

Spontaneous mutations are mutations that occur in the absence of exogenous agents. They may be due to errors made by DNA polymerases during replication or repair, errors made during recombination, the movement of genetic elements, or spontaneously occurring DNA damage. The rate at which spontaneous mutations occur can yield useful information about cellular processes.

The mutation rate is the expected number of mutations that a cell will sustain during its lifetime. The mutant fraction or frequency is the proportion of cells in a population that are mutant. Even though mutant frequencies can be adequate indicators of the rate at which mutations are induced by DNA damaging agents, they are inadequate indicators of spontaneous mutation rates. This is because the population of mutants is composed of clones, each of which arose from a cell that sustained a mutation.

**Determination of Mutation Rate**

Broadly, there are two methods for determination of the mutation rate: Mutation Accumulation and Fluctuation Analysis. Mutant accumulation methods have the advantage that they are very accurate, but they are complicated and time-consuming to perform because the culture is sampled at multiple time points. The methodology depends on growing bacteria exponentially until probability dictates that a mutant will be present. If the assumption is made that the growth rates of wild-type and mutant bacteria are the same, then the proportion of mutants will increase linearly with time. Furthermore, if the number of mutants and the total number of bacterial cells are known at each time point, then the mutation rate (m) can be calculated from the slope of the line describing the relationship between the number of mutants against the generation number. The mutation rate can be determined by using the equation,

\[
\mu = \frac{(r_2/N_2) - (r_1/N_1)}{\ln (N_2/N_1)} = \left( f_2 - f_1 \right) \times \ln (N_2/N_1)
\]
Where \( r_1 \) is the observed number of mutants at time point 1, \( r_2 \) is the observed number of mutants at the next time point, and \( N_1 \) and \( N_2 \) are the numbers of cells at time points 1 and 2, respectively, while \( f_1 \) and \( f_2 \) are the mutant frequencies at points 1 and 2.

For this method to be accurate, a very large difference in the total cell number is required between \( N_1 \) (the number of cells at the first time point) and \( N_2 \) (the number of cells at the second time point). Serial dilutions can help in estimating this, but this includes sampling errors. Alternatively, the continuous culture can be used for the selection of waves of bacteria, each suited than the generation before to take over the culture.

Following is the description of the terms used in the determination of mutation rate:

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m )</td>
<td>Number of Mutations Per Culture</td>
</tr>
<tr>
<td>( \mu )</td>
<td>Mutation Rate; Probability of Mutation Per Cell Per Division or Generation</td>
</tr>
<tr>
<td>( N )</td>
<td>Number of Cells</td>
</tr>
<tr>
<td>( N_0 )</td>
<td>Initial Number of Cells in a Culture</td>
</tr>
<tr>
<td>( N_1 )</td>
<td>Final Number of Cells in a Culture</td>
</tr>
<tr>
<td>( r )</td>
<td>Observed number of Mutants in a Culture</td>
</tr>
<tr>
<td>( r_0 )</td>
<td>Median Number of Mutants in a Culture</td>
</tr>
<tr>
<td>( f )</td>
<td>Mutant Fraction or Frequency = ( \frac{r}{N} )</td>
</tr>
<tr>
<td>( V )</td>
<td>Volume of a Culture</td>
</tr>
<tr>
<td>( C )</td>
<td>Number of Cultures in Experiment</td>
</tr>
<tr>
<td>( p_0 )</td>
<td>Proportion of Cultures without Mutants</td>
</tr>
<tr>
<td>( z )</td>
<td>Dilution Factor or Fraction of a Culture Plated</td>
</tr>
<tr>
<td>( p_r )</td>
<td>Proportion of Cultures with ( r ) Mutants</td>
</tr>
<tr>
<td>( C_r )</td>
<td>Number of Cultures with ( r ) Mutants</td>
</tr>
<tr>
<td>( P_r )</td>
<td>Proportion of Cultures with ( r ) or More Mutants</td>
</tr>
<tr>
<td>( Q_{0.25} )</td>
<td>Value of ( r ) at 25% of the Ranked Series of ( r )</td>
</tr>
<tr>
<td>( Q_{0.50} )</td>
<td>Value of ( r ) at 50% of the Ranked Series of ( r ), the Median</td>
</tr>
<tr>
<td>( Q_{0.75} )</td>
<td>Value of ( r ) at 75% of the Ranked Series of ( r )</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>( CL )</td>
<td>Confidence Limit = ((1 - \alpha) \times 100 )</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>Level of Statistical Significance, usually 0.05 or 0.01</td>
</tr>
</tbody>
</table>

It is significant to distinguish between \( m \) which is the mean number of mutations that occur during the growth of a culture, and \( \mu \) which is the mutation rate, specifically the mean number of mutations that occur during the lifetime of a cell. Almost all methods to calculate mutation rates start by determining \( m \) and then obtaining \( \mu \) by dividing \( m \) by some measure of the number of cell-lifetimes at
risk for mutation, usually $N_t$. It is also significant to distinguish between the number of mutations per culture, $m$, and the number of mutants per culture, $r$, divided by $N_t$ is the mutant fraction or mutant frequency, $f$.

Calculating the Mutation Rate

The mutation rate, $\mu$, is the mean number of mutations, $m$, normalized to some measure of the number of cells at risk for mutation. Three such measures are used, each of which is based on different assumptions about the underlying mutational process. If the probability of mutation is constant over the cell cycle, then $m$ should be divided by the number of cell divisions that have taken place. Since the final number of cells in a culture, $N_f$, arose from $N_t$ – 1 divisions, the mutation rate is (Luria and Delbrück, 1943):

$$\mu = \frac{m}{(N_t - 1) H} = \frac{m}{N_f}$$

1.4 MUTAGENESIS

Mutagenesis is a process by which the genetic information of an organism is changed, resulting in a mutation. It may occur spontaneously in nature, or as a result of exposure to mutagens. It can also be achieved experimentally using laboratory procedures. In nature mutagenesis can lead to cancer and various heritable diseases, but it is also a driving force of evolution. Mutagenesis as a science was developed based on work done by Hermann Muller, Charlotte Auerbach and J. M. Robson in the first half of the 20th century.

Mutagenesis may occur endogenously, for example, through spontaneous hydrolysis, or through normal cellular processes that can generate reactive oxygen species and DNA adducts, or through error in replication and repair. Mutagenesis may also arise as a result of the presence of environmental mutagens that induce changes to the DNA. The mechanism by which mutation arises varies according to the causative agent, the mutagen, involved. Most mutagens act either directly, or indirectly via mutagenic metabolites, on the DNA producing lesions. Some, however, may affect the replication or chromosomal partition mechanism, and other cellular processes.

Mutagenesis may also be self-induced by unicellular organisms when environmental conditions are very restrictive, for instance, in presence of toxic substances like antibiotics or, in yeasts, in presence of an antifungal agent or in absence of a nutrient.

Many chemical mutagens require biological activation to become mutagenic. An important group of enzymes involved in the generation of mutagenic metabolites is cytochrome P450. Other enzymes that may also produce mutagenic metabolites include glutathione S-transferase and microsomal epoxide hydrolase. Mutagens that are not mutagenic by themselves but require biological activation are called promutagens.
Many mutations arise as a result of problems caused by DNA lesions during replication, resulting in errors in replication. In bacteria, extensive damage to DNA due to mutagens results in single-stranded DNA gaps during replication. This induces the SOS response, an emergency repair process that is also error-prone, thereby generating mutations. In mammalian cells, stalling of replication at damaged sites induces a number of rescue mechanisms that help bypass DNA lesions, but which also may result in errors. The Y family of DNA polymerases specializes in DNA lesion bypass in a process termed TransLesion Synthesis (TLS) whereby these lesion-bypass polymerases replace the stalled high-fidelity replicative DNA polymerase, transit the lesion and extend the DNA until the lesion has been passed so that normal replication can resume. These processes may be error-prone or error-free.

Transposons and viruses may insert DNA sequences into coding regions or functional elements of a gene and result in inactivation of the gene.

Mutagenesis in the laboratory is the significant technique whereby DNA mutations are deliberately engineered to produce Mutant Genes, Proteins, or Strains of Organism. Various constituents of a gene, such as its control elements and its gene product, may be mutated so that the functioning of a gene or protein can be examined in detail. The mutation may also produce mutant proteins with interesting properties, or enhanced or novel functions that may be of commercial use.

Early methods of mutagenesis produced entirely random mutations; however, later methods of mutagenesis may produce site-specific mutation.

**Adaptive Mutagenesis Mechanisms**

Adaptive mutagenesis has been defined as mutagenesis mechanisms that enable an organism to adapt to an environmental stress. For instance, in bacteria, while modulation of the SOS response and endogenous prophage DNA synthesis has been shown to increase *Acinetobacter baumannii* resistance to Ciprofloxacin. Resistance mechanisms are presumed to be linked to chromosomal mutation untransferable via horizontal gene transfer in some members of *Enterobacteriaceae* family, such as *Escherichia coli*, *Salmonella* spp., *Klebsiella* spp., and *Enterobacter* spp. Chromosomal events are also relevant to this adaptive mutagenesis in bacteria.

**Check Your Progress**

1. Explain the term mutation.
2. What is silent mutation and lethal mutation?
3. What is a gene?
4. What is phenotypic effects of mutation?
5. Define the term germ line mutation. Differentiate between gene mutations and chromosome mutations.
6. Explain mutation rate with reference to genetics.
7. Define mutation spectrum of an organism.
8. What is mutagenesis? How it occurs?
1.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The term mutation refers both to the change in the genetic material and to the process by which the change occurs. An organism exhibiting a novel phenotype as a result of the presence of mutation is referred to as mutant. Mutation refers to any sudden, heritable change in the genotype of an organism not explainable by recombination of pre-existing genetic variability. Such genotypic changes includes change in chromosome number (euploidy and aneuploidy), gross changes in the structure of chromosomes (chromosomes aberration) and changes in individual genes.

2. A mutagenic event occurring in a nonfunctional area of DNA will have no effect (silent mutation), whereas a similar change in an actively transcribed region may profoundly affect gene expression and phenotype or even lead to cell death (lethal mutation).

3. A gene is a specific sequence of nucleotide pairs coding for a particular polypeptide.

4. Mutations cause some detectable phenotypic change for their presence to be recognized. The effects of mutations on phenotype range from alterations so minor that they can be detected only by special genetic or biochemical techniques to gross modifications of morphology. A gene is a specific sequence of nucleotide pairs coding for a particular polypeptide. Any mutation occurring within a given gene will produce a new form.

5. Germ line mutations arise in cells that ultimately produce gametes. A germ line mutation can be passed to future generations producing individual organisms that carry the mutation in all their somatic and germ line cells. Mutations have been partitioned into those that affect a single gene, called gene mutations and those that affect the number or structure of chromosomes called chromosome mutations.

6. In genetics, the mutation rate is the frequency of new mutations in a single gene or organism over time. Mutation rates are not constant and are not limited to a single type of mutation, therefore there are many different types of mutations. Point mutations are a class of mutations which are small or large scale insertions or deletions. The rate of these types of substitutions can be further subdivided into a mutation spectrum which describes the influence of the genetic context on the mutation rate.

There are several natural units of time for each of these rates, with rates being characterized either as mutations per base pair per cell division, per gene per generation, or per genome per generation. However, the mutation rate does vary over the genome. Over DNA, RNA or a single gene, mutation rates are changing.
7. The mutation spectrum of an organism is the rate at which different mutations occur at different sites. Typically two sites are considered, each of which may have three mutations, resulting in six total rates for most mutation spectra. The two sites are the two correct pairs possible in DNA, the A:T pairs and the C:G pairs.

8. Mutagenesis is a process by which the genetic information of an organism is changed, resulting in a mutation. It may occur spontaneously in nature, or as a result of exposure to mutagens. It can also be achieved experimentally using laboratory procedures. In nature mutagenesis can lead to cancer and various heritable diseases, but it is also a driving force of evolution. Mutagenesis may occur endogenously, for example, through spontaneous hydrolysis, or through normal cellular processes that can generate reactive oxygen species and DNA adducts, or through error in replication and repair. Mutagenesis may also arise as a result of the presence of environmental mutagens that induce changes to the DNA. The mechanism by which mutation arises varies according to the causative agent, the mutagen, involved.

1.6 SUMMARY

- The term mutation refers both to the change in the genetic material and to the process by which the change occurs.
- An organism exhibiting a novel phenotype as a result of the presence of mutation is referred to as mutant.
- Mutation refers to any sudden, heritable change in the genotype of an organism not explainable by recombination of pre-existing genetic variability. Such genotypic changes includes change in chromosome number (euploidy and aneuploidy), gross changes in the structure of chromosomes (chromosomes aberration) and changes in individual genes.
- Sudden, heritable changes in the genetic material are called mutations.
- It refers to the process by which such changes are produced.
- It may occur spontaneously or may be induced by agents that interact with DNA and RNA.
- At the simplest level, a mutation is a change or transformation. In biology, mutations refer to changes in chromosomes and genes, which typically manifest physically.
- The result of a mutation could be harmful, beneficial, neutral or even silent.
- Mutation can lead to the loss or gain of a specific function, to changes to the expression levels, or in extreme cases, even embryonic lethality.
- Mutations can be classified in various ways depending on the cause of the mutation, its effect on the function of the gene product or the kind of changes to the structure of the gene itself.
A mutation is defined as an inherited change in genetic information, the descendants may be cells or organisms. The appearance of a new mutation is a rare event. Most mutations that were originally studied occurred spontaneously. Mutagens are chemical compounds or forms of radiation (such as ultraviolet (UV) light or X-rays) that cause irreversible and heritable changes (mutations) in the cellular genetic material, deoxyribonucleic acid (DNA). The biological consequences of a mutation depend upon many critical factors such as the target loci, size of the mutation, timing during the cell cycle, and compounding effects of pre-existing mutations. Mutations cause some detectable phenotypic change for their presence to be recognized. Mutation occur in any cell and at any state in the cell cycle. The immediate effect of mutation and its ability to produce phenotypic change are determined by its dominance, the type of cell in which it occurs and when it happens relative to life cycle of the organism. The portion of the electromagnetic spectrum containing wavelengths that are shorter and of higher energy than the visible light can be divided into ionizing radiation (X rays, gamma rays, cosmic rays) and non-ionizing radiations (UV light). The first chemical mutagen discovered was mustard gas. When C. Auerbach and her associates discovered the mutagenic effects of mustard gas and related compounds during world war II. The positively charged acridine intercalate between the stacked base pairs in DNA. They increase the rigidity and alter the conformation of the double helix, causing kinks in the molecule. Mutations are invaluable to the process of evolution since they provide the raw material required for its occurrence. Mutation provide the alleles required for various types of genetic analysis, from Mendel’s two factor crosses to chromosome mapping to studies of genetic structures of populations. The relation between dosage and effects cannot be accurately measured in humans at present because of the complexity of the subject and difficulties of dealing with the genetics of humans. Germ line mutations arise in cells that ultimately produce gametes. A germ line mutation can be passed to future generations producing individual organisms that carry the mutation in all their somatic and germ line cells. In genetics, the mutation rate is the frequency of new mutations in a single gene or organism over time.
Mutation rates are not constant and are not limited to a single type of mutation, therefore there are many different types of mutations.

- Mutation rates are given for specific classes of mutations. Point mutations are a class of mutations which are small or large scale insertions or deletions.
- There are several natural units of time for each of these rates, with rates being characterized either as mutations per base pair per cell division, per gene per generation, or per genome per generation.
- The mutation rate of an organism is an evolved characteristic and is strongly influenced by the genetics of each organism, in addition to strong influence from the environment.
- The upper and lower limits to which mutation rates can evolve is the subject of ongoing investigation. However, the mutation rate does vary over the genome. Over DNA, RNA or a single gene, mutation rates are changing.
- Mutation rates differ between species and even between different regions of the genome of a single species. These different rates of nucleotide substitution are measured in substitutions (fixed mutations) per base pair per generation. For example, mutations in intergenic, or non-coding, DNA tend to accumulate at a faster rate than mutations in DNA that is actively in use in the organism (gene expression).
- The mutation spectrum of an organism is the rate at which different mutations occur at different sites. Typically two sites are considered, each of which may have three mutations, resulting in six total rates for most mutation spectra. The two sites are the two correct pairs possible in DNA, the A:T pairs and the C:G pairs.
- The mutation rate is the expected number of mutations that a cell will sustain during its lifetime. The mutant fraction or frequency is the proportion of cells in a population that are mutant.
- Mutagenesis is a process by which the genetic information of an organism is changed, resulting in a mutation. It may occur spontaneously in nature, or as a result of exposure to mutagens. It can also be achieved experimentally using laboratory procedures.
- In nature mutagenesis can lead to cancer and various heritable diseases, but it is also a driving force of evolution.
- Mutagenesis as a science was developed based on work done by Hermann Muller, Charlotte Auerbach and J. M. Robson in the first half of the 20th century.
- Mutagenesis may occur endogenously, for example, through spontaneous hydrolysis, or through normal cellular processes that can generate reactive oxygen species and DNA adducts, or through error in replication and repair.
Mutagenesis may also arise as a result of the presence of environmental mutagens that induce changes to the DNA. The mechanism by which mutation arises varies according to the causative agent, the mutagen, involved.

Most mutagens act either directly, or indirectly via mutagenic metabolites, on the DNA producing lesions. Some, however, may affect the replication or chromosomal partition mechanism, and other cellular processes.

Adaptive mutagenesis has been defined as mutagenesis mechanisms that enable an organism to adapt to an environmental stress.

1.7 KEY WORDS

- **Mutation**: It is a permanent alteration of the nucleotide sequence of a genome.
- **Ionizing radiation**: These are radiations that consist of X-rays, gamma rays with sufficient energy to cause ionization in a medium.
- **Mutagens**: These are either physical or chemical agents that causes genetic mutation.
- **Somatic mutations**: It is a change that appears in somatic cells and it shows mutant change.
- **Germinal mutations**: If mutations originate in a gamete, if dominant mutations occur in germ cells, effect is expressed immediately in the progeny.
- **Spontaneous mutations**: It is a kind of mutations that occurs without any cause.
- **Induced mutations**: It arise due to exposure of mutagenic agents, such as ionizing radiations or various chemicals.
- **Mutation rate**: In genetics, the mutation rate is the frequency of new mutations in a single gene or organism over time.
- **Mutation spectrum**: The mutation spectrum of an organism is the rate at which different mutations occur at different sites.
- **Mutagenesis**: It is a process by which the genetic information of an organism is changed, resulting in a mutation, it may occur spontaneously in nature, or as a result of exposure to mutagens.

1.8 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What are the phenotypic effects of mutation?
2. Write a note on somatic and germinal mutations.
3. Discuss back mutation and suppressor mutation.
4. What are chemically induced mutations?
5. Write a note on mutations and humans.
6. What is radiation induced mutations?
7. List a few practical applications of mutations.
8. Discuss gene mutations and its types.
9. Explain the significance of mutation rate.
10. What is mutation spectrum?
11. How mutagenesis occurs?

Long Answer Questions

1. What do you mean by somatic and germinal mutations? Also discuss the effects of mutation.
2. From your learning of the text, differentiate between back mutation and suppressor mutation.
3. How are radiation induced mutations different from chemical induced mutation? Discuss.
4. “Purposeful artificial selection is not practiced in humans and therefore the possible advantages cited for domestic animals and plants do not apply to humankind.” Discuss.
5. Briefly discuss the concept of mutation rate.
6. Explain the concept of mutation spectrum giving example.
7. “The mutation rate is the expected number of mutations that a cell will sustain during its lifetime. The mutant fraction or frequency is the proportion of cells in a population that are mutant.” Discuss.
8. Explain the method of determining the mutation rate with the help of an example.
9. Discuss the concept of mutagenesis giving examples.
10. Why mutagenesis is the significant technique in the laboratory? Explain.

1.9 FURTHER READINGS


UNIT 2  MUTAGENS AND ITS TYPES

2.0  INTRODUCTION

A mutagen is defined as any chemical that can cause changes in the DNA sequence of an organism. These changes are called mutations. There are various types of well-known mutagens, for example UV radiation, intercalating agents, metals, base analogs, benzene, X-rays, 2,4 – diaminoanisole, radioactivity, most aromatic amines, alkylating agents, etc. Thus mutagens may be of physical, chemical or biological origin. They may act directly on the DNA, causing direct damage to the DNA, and most often result in replication error, or mutations. In genetics, a mutagen is a physical or chemical agent that changes the genetic material, usually DNA, of an organism and thus increases the frequency of mutations above the natural background level. As many mutations can cause cancer, mutagens are therefore also likely to be carcinogens, although not always necessarily so.

All mutagens have characteristic mutational signatures with some chemicals becoming mutagenic through cellular processes. Not all mutations are caused by mutagens, the so-called ‘spontaneous mutations’ occur due to spontaneous hydrolysis, errors in DNA replication, repair and recombination.

In this unit, you will study about history and introduction of mutagens, types of mutagens and effects of exposure of mutagens on the survival of organisms in detail.
2.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss the history and introduction of mutagens
- Understand the types of mutagens
- Analyse the effects of exposure of mutagens on the survival of organisms

2.2 MUTAGENS: HISTORY AND CAUSES

Mutagens are chemical compounds or forms of radiation, such as UltraViolet (UV) light or X-rays that cause irreversible and heritable changes (mutations) in the cellular genetic material, DeoxyriboNucleic Acid (DNA). Mutagens are not necessarily carcinogens, and vice versa. Sodium azide for example may be mutagenic (and highly toxic), but it has not been shown to be carcinogenic.

Mutagenic lesions persist when they escape detection by protective cellular DNA repair mechanisms, when mistakes occur in the repair process, or when repair mechanisms are overwhelmed by extensive damage. Upon subsequent cellular replication, these mutations become fixed in the genome and are inherited by all daughter cells. In this way, mutagenesis becomes a cumulative process, stretching over the lifetime of an organism. In addition, the accumulation of mutations over time, leading to gradually less efficient cellular repair capabilities, has been linked to the aging process and associated degenerative diseases. While the health of a particular individual can be affected by mutations in somatic cells, mutagenic events in germ cells (germline mutations) lead to the transmission of genetic diseases to subsequent generations. Genetic predispositions to certain breast and colon cancers, cystic fibrosis, and Huntington’s disease are included among many examples of such inheritable diseases. Although many dietary and environmental agents have been classified as mutagens, cells are constantly subjected to a barrage of spontaneous DNA damage.

UV has enough energy to destroy and break molecular bonds. As a result, if a material is exposed to UV radiation, there always exists the possibility that the chemical structure of that material is changed. This fact, couples with the fact that DNA strongly absorbs UV radiation of some frequencies (especially the 260–280 nm and below 100 nm bands, means that in any living organism exposed to UV, there will be a relatively chance of DNA damage occurring during exposure. There are multiple types of UV-induced DNA damage, such as base loss where the base is severed from the deoxyribose, resulting in an abasic site, thymidine dimers where two adjacent or near-adjacent thymidine molecules are bonded together, single-strand and double-strand breaks, etc. Figure 2.1 shows different types of mutagens and plant materials used for mutagenic treatment.
We are regularly exposed to another example of a DNA-damaging agent. It is called Ethidium Bromide (EtBr). It is used as a DNA stain due to its enormous affinity towards DNA. EtBr works efficiently because it is capable of inserting itself in between two adjacent DNA bases, a so-called ‘intercalating agent’. When it inserts itself like that, it can fluoresce intensely under UV or green light, making it a near-ideal agent for visualizing DNA. However, this intercalation means that whenever DNA Polymerase tries to replicate DNA, it malfunctions at the site of EtBr intercalation and can insert an inappropriate base, or skip that base altogether or ‘jam’ in place, not only failing to synthesize any more DNA, but also halting DNA replication in the middle of the strand until it is removed by other repair proteins or the protein p53 (the shepherd of the genome) commits the cell to apoptosis.
One mutagen that is either essential or deadly is the ‘Oxygen’. It is one of the most potent mutagens known, in the form of Reactive Oxygen Species or ROSs. These ROSs arise from multiple sources but the primary sources are all related to energy metabolism. One particularly dangerous ROS is superoxide, \( \text{O}_2^- \) where the ‘-’ in parentheses is the charge. It can attack almost all organic substances in the cell and has a very high potency to cause damage to DNA by producing hydroxyl radicals, which can then attack DNA bases and damage them. Luckily the cells have a line of defense against superoxide in the form of SuperOxide Dismutase (SOD). SOD can take two superoxide molecules and oxidize one to molecular oxygen and reduce the other to peroxide, \( \text{O}_2 \cdot \) (-2). While peroxide itself is dangerously toxic, it is far less toxic than superoxide. Peroxide is then destroyed by catalase, which converts it to water and molecular oxygen.

The biological consequences of a mutation depend upon many critical factors, such as the target loci, size of the mutation, and timing during the cell cycle, and compounding effects of preexisting mutations. Thus, a mutagenic event occurring in a nonfunctional area of DNA will have no effect (silent mutation), whereas a similar change in an actively transcribed region may profoundly affect gene expression and phenotype or even lead to cell death (lethal mutation). The influence of mutations in human health is underscored by several human disease states caused by mutations that disrupt regulatory regions or gene coding sequences, resulting in altered gene expression and protein function. For example, mutations in genes that promote or inhibit growth and cellular replication (proto-oncogenes and tumor suppressor genes, respectively) or code for components of DNA repair pathways are important contributors to the multistage development of cancer.

Many natural constituents of food are mutagenic and are produced by plants as defense agents. However, additional food borne mutagens can be present as residues of compounds used during food production or leached from packaging materials. Mutagenic compounds can also be produced during food cooking and preparation. Foods also contain many compounds that can modulate the activity of mutagens. In order to minimize exposure to mutagens, the majority of developed countries have instituted regulatory protocols governing the introduction of new foods and food-associated chemicals before they are accepted into the marketplace. With advances in the understanding of mutagenic processes, existing foods and food contact items may also be reassessed for acceptability.

### 2.2.1 History of Mutagens

The first mutagens to be identified were carcinogens, substances that were shown to be linked to cancer. Tumors were described more than 2,000 years before the discovery of chromosomes and DNA in 500 B.C., the Greek Physician Hippocrates named tumors resembling a crab *karkinos* (from which the word ‘Cancer’ is derived via Latin), meaning crab. In 1761, J. Hill made the first direct link of cancer to chemical substances by noting that excessive use of snuff may
cause nasal cancer. In 1775, Sir P. Pott wrote a paper on the high incidence of scrotal cancer, and suggested chimney soot as the cause of scrotal cancer. In 1915, Yamagawa and Ichikawa showed that repeated application of coal tar to rabbit’s ears produced malignant cancer. Subsequently, in the 1930s the carcinogen component in coal tar was identified as a PolyAromatic Hydrocarbon (PAH), benzo[a]pyrene.

The association of exposure to radiation and cancer had been observed as early as 1902, six years after the discovery of X-ray radioactivity. The mutagenic property of mutagens was first demonstrated in 1927, when H. Muller discovered that X-rays can cause genetic mutations in fruit flies, *Drosophila*, producing phenotypic mutants as well as observable changes to the chromosomes, visible due to presence of enlarged “polynene chromosomes” in fruit fly salivary glands. Similar work related to UV radiation showed the mutational effect on maize in 1936. The effect of sunlight had previously been noted in the nineteenth century where rural outdoor workers and sailors were found to be more prone to skin cancer. Chemical mutagens were not demonstrated to cause mutation until the 1940s, when C. Auerbach and J.M. Robson found that mustard gas can cause mutations in fruit flies. A large number of chemical mutagens have since been identified, especially after the development of the Ames test in the 1970s by B. Ames that screens for mutagens and allows for preliminary identification of carcinogens.

2.2.2 Causes of Mutagens

Mutagens can cause changes to the DNA and are therefore genotoxic. They can affect the transcription and replication of the DNA, which in severe cases can lead to cell death. The mutagen produces mutations in the DNA, and deleterious mutation can result in aberrant, impaired or loss of function for a particular gene, and accumulation of mutations may lead to cancer. Mutagens may therefore be also carcinogens. However, some mutagens exert their mutagenic effect through their metabolites, and therefore whether such mutagens actually become carcinogenic may be dependent on the metabolic processes of an organism, and a compound shown to be mutagenic in one organism may not necessarily be carcinogenic in another.

Different mutagens act on the DNA differently. Powerful mutagens may result in chromosomal instability, causing chromosomal breakages and rearrangement of the chromosomes, such as translocation, deletion and inversion. Such mutagens are called cladogens. Mutagens may also modify the DNA sequence; the changes in nucleic acid sequences by mutations include substitution of base pairs, insertions and deletions of one or more nucleotides in DNA sequences. Although some of these mutations are lethal or cause serious disease, many have minor effects as they do not result in residue changes that have significant effect on the structure and function of the proteins.
Chemical mutagens, such as Ethyl-MethaneSulfonate (EMS) and X-rays can both be used to induce pseudorandom mutations into the genome of the fly. According to Greenspan, EMS typically causes single base-pair changes while X-rays often result in deletions and gross chromosomal rearrangements. EMS is highly toxic and mutagenic for humans and all manipulations should be performed in a fume hood with the appropriate protection. Offspring of the EMS- or X-ray-mutagenized males are tested for the presence of relevant mutations using the screening protocol and found to significant.

When chemical mutagens or radiation damage DNA during G1 phase of cell cycle, DNA replication is postponed until the damage is repaired. The key molecular species that operates this checkpoint is a protein called p53. The p53 is a transcription factor that induces the expression of DNA-repair genes. It also acts indirectly to inhibit the activity of the cyclin–CDK complex that normally drives the cell from G1 to S. Mutations in p53 interfere with its checkpoint function, leading to chromosomal rearrangements and gene amplification. These rearrangements predispose to cancer. Furthermore, in the presence of p53, irreparably damaged cells commit suicide via apoptosis. In the absence of p53, damaged cells may proliferate, thereby being another predisposing factor toward malignancy.

Some mutagens can produce modified DNA bases that appear ‘uninformative’ to the polymerase. These may be bypassed in an error-prone manner with a random selection of a base opposite the modified base. The ultimate example of this class is an abasic site in which the template base has been completely removed. Polymerase can synthesize over such lesions, albeit inefficiently. In Escherichia coli, there appears to be a preference for the insertion of an Adenine opposite the uninformative lesion, but this is not absolute, while in mammalian cells such a preference is not evident.

According to Bishop and Shelby (1990), Andler et al., (2007) many mutagens (DNA damaging agents) have been tested for effects in germ cells, although the database for effects on male germ cells is much larger than for oocytes. The dominant lethal test for germ cell mutagenicity can be applied to males or females. The resulting pregnant females are then examined late in gestation for numbers of living or dead fetuses, or implantation sites of embryos that died shortly after implantation, as well as numbers of Corpora lutea to be used as a measure of the number of potential embryos (oocytes ovulated). Pre- and post-implantation loss can then be calculated. Although this test cannot distinguish between embryo mortality due to genetic defect in the oocyte and pregnancy loss due to maternal toxicity, non-genetic oocyte toxicity, or even failed fertilization, it can be used in conjunction with other measures of oocyte function (such as chromosome analysis, zygote morphology and cleavage, or blastocyst formation) to characterize acute effects of chemicals on the oocyte and its developmental potential.
However, many mutagens are highly toxic to proliferating cells, and they are often used to destroy cancer cells. Alkylating agents, such as cyclophosphamide and cisplatin, as well as intercalating agent, such as daunorubicin and doxorubicin may be used in chemotherapy. However, due to their effect on other cells which are also rapidly dividing, they may have side effects, such as hair loss and nausea. Research on better targeted therapies may reduce such side-effects. Ionizing radiations are used in radiation therapy.

Check Your Progress

1. What is mutagen?
2. Expand EtBr. How EtBr works efficiently?
3. From where ROSs arises from?
4. What are biological consequences of a mutation?

2.3 TYPES OF MUTAGENS

Mutagenesis is the process of inducing mutation by a number of physical, chemical or biological agents. The agents that cause mutations are called as mutagens. So that mutagens may be of physical, chemical or biological in their origin. They may act directly on the DNA, causing direct damage to the DNA, and most often result in replication error. Some however may act on the replication mechanism and chromosomal partition. Many mutagens are not mutagenic by themselves, but can form mutagenic metabolites through cellular processes, for example through the activity of the cytP450 system and other oxygenases, such as cyclooxygenases. Such mutagens are called promutagens. Mutation induced by mutagens is called induced mutation. Sometime mutation occurs spontaneously due to error during DNA replication. However, mutagens increase the chances of mutation.

There are three types of mutagens (Refer Figure 2.2).

- Physical Mutagens
- Chemical Mutagens
- Biological Mutagens
2.3.1 Physical Mutagens

Physical mutagen is a mutation agent which is in the form of physical substances, such as short waves, particles and radioactive elements. Physical mutagens are either ionizing which includes particulates, such as γ-rays, β-rays, fast neutrons, ionic elements/metals (Cobalt-60 and Cesium-137) and non-particulate matters (γ-rays, cosmic rays and X-rays) or non-ionizing UV-rays/UV light. The ionizing radiations have capacity to ionize water of the cell to release hydroxyl free radical (OH). The hydroxyl radical is a powerful oxidizing agent and oxidizes the phosphodiester bond of DNA. Higher dose of X-rays can even causes death of an organism. However, the non-ionizing physical mutagens lead the formation of Thymine dimer (Pyrimidine dimer). The occurrence of two Thymines together in one strand of DNA, UV light causes fusion to form Thymine dimer. Nitrogenous bases absorb UV lights and the absorption is maximum at 260 nm. At the site of Thymine dimer confirmation of DNA is changed, so rate of error during DNA replication is high. The physical mutagens may forward as radioactive decay, such as 14C in DNA which decays into nitrogen and affect the normal biological activities.

Radiations as Mutagen

Radiation is the most important among the physical mutagens and damaging the DNA molecules fall in the wavelength range below 340 nm and photon energy above 1 electro-Volt (eV). The destructive radiation consists of UltraViolet (UV) rays, X-rays, γ-rays, alpha (α) rays, beta (β) rays, cosmic rays, neutrons, etc. (Refer Figure 2.3).
Mutagens and its Types

NOTES

Self-Instructional Material

Radiation induced damage can be categorized into the three broad types: lethal damage (killing the organisms), potentially lethal damage (can be lethal under certain ordinary conditions) and sub-lethal damage (cells do not die unless radiation reaches to a certain threshold value). The effect of damage is at molecular level (proteins, lipoproteins, DNA, carbohydrates, etc.) in a live cell is caused directly by ionization/excitation, or indirectly through highly reactive free radicals produced by radiolysis of cellular water. DNA stores genetic information so any damage to it assumes great dimension. It can perpetuate genetic effects and, therefore, the cellular repair system is largely devoted to its welfare.

When the Bacteria are exposed to radiation they gradually lose the ability to develop colonies. This gradual loss of viability can be expressed graphically by plotting the surviving colonies against the gradually increasing exposure time. This dose-response graph is called survival curve. The survival curve is analyzed by a simple mathematical theory called hit theory. According to this theory each organism possesses at least one sensitive site which is known as target site. Radiation photons (particles of light) damage or hit the target site and inactivate the organisms. One can derive the equation based on this theory. The equations help to calculate the survival curve for many kinds of populations of N identical organisms exposed to dose D of radiation causing damage. The number dN damaged by a dose dD is proportional to the initial population that has not received radiation; hence dN = KN, where, K is the constant which measures the effectiveness of dose (Refer Figure 2.4).
UltraViolet (UV) Radiations as Mutagen

Ionization usually occurs because the radiation source has very large energy, for example, radiation from radioactive substances and cosmic ray, if hit the DNA molecules, the DNA chain will lose so that the DNA chain cannot function in protein synthesis. Ultraviolet ray generally do not cause ionization, however, the energy from UltraViolet ray will be absorbed by purine and pyrimidine so that the atom becomes more reactive (the electron undergoes excitation). Consequently, DNA double-helix becomes in disorder and inhibits replication, one of the effects caused by ultraviolet ray is skin cancer.

UV radiation causes damage in the DNA duplex of the Bacteria and Phages. The UV rays are absorbed and cause excitation of macromolecules. The absorption maxima of nucleic acid (280 nm) and protein (260 nm) are more or less similar. The DNA molecule is the target molecule for UV rays but not the proteins. However, absorption spectrum of RNA is quite similar to that of DNA. The excited DNA leads to cross-linking, single strand breaks and base damage as minor lesion and generation of nucleotide dimer as a major one. Purines are generally more radio-resistant than the pyrimidine of the latter, Thymine is more reactive than Cytosine. Hence, the ratio of Thymine-Thymine (TT), Thymine-Cytosine (TC) and Cytosine-Cytosine (CC) dimer is 10:3:3, respectively. A few dimers of TU and UU also appear. The initial step in pyrimidine dimerization is known to be hydration of their 4: 5 bonds (Refer Figure 2.5).
NOTES

Mutagens and its Types

Fig. 2.5 UV Radiation induced Formation of Pyrimidine Dimer

Formation of Thymine-Thymine (TT) dimer causes distortion of DNA helix because the Thymines are pulled towards one another. The distortion results in weakening of hydrogen-bonding to Adenines in the opposing strand. This structural distortion inhibits the advance of replication fork.

X-Rays as Mutagen

The X-rays cause breaking of phosphate ester linkages in the DNA. This breakage occurs at one or more points. Consequently, a large number of bases are deleted or rearranged in the DNA molecule. The X-rays may break the DNA either in one or both strands. If breaks occur in both strands, it becomes lethal. The DNA segment between the two breaks is removed resulting in deletion. Since both the X-rays and UV rays bring about damage in DNA molecule, they are used in sterilization of Bacteria and Viruses.

2.3.2 Chemical Mutagens

Chemical mutagens are compounds that increase the frequency of some types of mutation. Among them are alkylators, including Ethyl-MethaneSulfonate (EMS), Methyl MethaneSulfonate (MMS), DiEthylSulfate (DES), and Nitrosoguanidine which tend to de-acidify the acidic properties of DNA. They also have the property of inducing error-prone DNA repair. This stimulation of error-prone repair allows all sorts of mutation types to occur. Nitrous Oxide can also cause mutations in the DNA affecting protein synthesis. Thus chemical mutagens are standard tools for mutagenesis in a variety of organisms, and they are a primary means of creating mutations in phenotype-based screens in most genetic systems. Although varied in the experimental design, all whole animal screens involve the generation of lines harboring mutated chromosomes followed by the examination of the resulting phenotypes in the heterozygous or homozygous state. In contrast, gene-based
screens rely on the identification of lines that carry mutations in specific genes, prior to any phenotypic examination. The ability to perform gene-based screens has been made possible by recent technological advances in the high-throughput detection of subtle mutations.

**Table 2.1 Various Classes of Chemicals as Mutagenic Agents for Living Beings**

<table>
<thead>
<tr>
<th>Class of Chemical</th>
<th>Chemical Mutagens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acriderin</td>
<td>Ethylmethanesulphonate (EMS)</td>
</tr>
<tr>
<td>Mustard</td>
<td>Ethylmethanesulphonate (EMS)</td>
</tr>
<tr>
<td>Nitrosomes</td>
<td>Ethylmethanesulphonate (EMS)</td>
</tr>
<tr>
<td>Epoxide</td>
<td>Ethylmethanesulphonate (EMS)</td>
</tr>
<tr>
<td>Alkyl sulfobetates</td>
<td>Ethylmethanesulphonate (EMS)</td>
</tr>
<tr>
<td>Others</td>
<td>Ethylmethanesulphonate (EMS)</td>
</tr>
</tbody>
</table>

Chemical mutagens have also been used successfully in the mutagenesis of mouse Embryonic Stem (ES) cells. This approach holds a variety of advantages over mutagenesis in the whole animal. Cell culture–based mutagenesis is distinguished from whole animal mutagenesis in its ability to use a variety of mutagens, to monitor and modulate mutation load, and to screen a large number of lines. The possibility of generating and characterizing mutations in cell culture enables the rapid identification of many mutations in genes, which can be transmitted to a mouse line. In this unit, we describe the merits and methods of gene-based screens to identify chemically induced mutation in ES cells. Several chemical mutagens have been successfully used to mutagenize mouse ES cells, and the potential exists for the use of almost any mutagen.

**Base Analogs as Chemical Mutagen**

These chemicals are morphologically similar to those of normal nitrogen bases. So during replication these molecules are incorporated instead of normal nitrogen bases and hence cause mutation, for example, 2-aminopurine is a analogue to Adenine, and 5-bromouracil is an analogue to thymine. So, base analog mutagens are chemicals that mimic bases to such an extent that they can be incorporated into DNA in place of one of the normal bases but in doing so lead to an increase in the rate of mutation. To be mutagenic, a base analog must mispair more frequently than the normal base it replaced. This mispairing can occur either during the initial incorporation into DNA, or during subsequent rounds of replication when the base analog is used as a template. Most of these mutagens typically induce only
base pair substitutions (and not other types of mutation). They are usually not highly toxic nor do they increase rates of recombination (Refer Figure 2.6).

**Fig. 2.6** Schematic Representation of Base Analogs Insertion during Replication

Thus a base analogue is a chemical compound similar to one of the four bases of DNA. It can be incorporated into a growing polynucleotide chain when normal process of replication occurs. These compounds have base pairing properties different from the bases. They replace the bases and cause stable mutation. A very common and widely used base analogue is 5-BromoUracil (5-BU) which is an analogue of thymine. The 5-BU functions like Thymine and pairs with Adenine. The 5-BU undergoes tautomeric shift from keto form to Enol form caused by bromine atom. The Enol form can exist for a long time for 5-BU than for Thymine. If 5-BU replaces a Thymine, it generates a Guanine during replication which in turn specifies Cytosine causing G: C pair (Refer Figure 2.7).

**Fig. 2.7** Base Analog (5-BromoUracil) Induced Mutagenesis, Keto Form of 5-BU Pairs with Adenine, however Tautomerised Enol with Guanine
During the replication, keto form of 5-BU substitutes for T and the replication of an initial AT pair becomes an A: BU pair. The rare Enol form of 5-BU that pairs with G is the first mutagenic step of replication. In the next round of replication G pairs with C. Thus, the transition is completed from AT→GC pair. The 5-BU can also induce the conversion of GC to AT. The Enol form infrequently acts as an analogue of Cytosine rather than thymine. Due to error, GC pair is converted into a G: BU pair which in turn becomes an AT pair. Due to such pairing properties 5-BU is used in chemotheraphy of Viruses and cancer. Because of pairing with guanine it disturbs the normal replication process in microorganisms. The 5-bromodeoxyuridine (5-BDU) can replace thymidine in DNA molecule. The 2-amino-purine (2-AP) and 2, 6-di-aminopurine (2, 6-DAP) are the purine analogues. The 2-AP normally pairs with Thymine but it is able to form a single hydrogen bond with Cytosine resulting in transition of AT to GC. The 2-AP and 2, 6-DAP are not as effective as 5-BU and 5-BDU (Refer Figure 2.8).

![Fig. 2.8 Replication during 5-BU induced Mutagenesis; (A). AT-GC Replication and; (B). GC-AT Replication](image)

**Chemical Altering the Hydrogen Specificity as Mutagen**

There are many chemicals that after incorporation into DNA change the specificity of hydrogen-bonding. Those which are used as mutagens are nitrous oxide (HNO₂), Hydroxyl Amines (HA) and Ethyl-Methane Sulphonate (EMS).

- **Nitrous Oxide (HNO₂):** Nitrous oxide converts the amino group of bases into keto group through oxidative deamination. The order of frequency of deamination (removal of amino group) is Adenine > Cytosine > Guanine.
- **Deamination of Adenine:** Deamination of Adenine results in formation of hypoxanthine, the pairing behaviour of which is like Guanine. Hence, it pairs with Cytosine instead of Thymine replacing AT pairing by GC pairing (Refer Figure 2.9).
Mutagens and its Types

NOTES

Deamination of Cytosine: Deamination of Cytosine results in formation of Uracil by replacing – NH₂ group with -OH group. The affinity for hydrogen bonding of Uracil is like Thymine; therefore, C-G pairing is replaced by U-A pairing (Refer Figure 2.10).

Deamination of Guanine: Deamination of Guanine results in formation of Xanthine, the later is not mutagenic. Xanthine behaves like Guanine because there is no change in pairing behaviour. Xanthine pairs with Cytosine. Therefore, G-C pairing is replaced by X-C pairing (Refer Figure 2.11).

Fig. 2.9 Nitrous Oxide induced Deamination of Adenine to Hypoxanthine

Fig. 2.10 Nitrous Oxide Induced Deamination of Cytosine to Uracil

Fig. 2.11 Nitrous Oxide Induced Deamination of Guanine to Xanthine and 6-Methylcytosine to Thymine
- **Hydroxylamine (NH₂OH):** It hydroxylates the C₄ nitrogen of Cytosine and converts into a modified base via deamination which causes to base pairs like Thymine. Therefore, GC pairs are changed into AT pairs.

**Alkylation Agents as Chemical Mutagen**

Addition of an alkyl group to the hydrogen bonding oxygen of Guanine (N₇ position) and Adenine (at N₃ position) residues of DNA is done by alkylation agents. As a result of alkylation, possibility of ionization is increased with the introduction of pairing errors. Hydrolysis of linkage of base-sugar occurs resulting in gap in one chain. This phenomenon of loss of alkylated base from the DNA molecule (by breakage of bond joining the nitrogen of purine and deoxyribose) is called **depurination.** Depurination is not always mutagenic. The gap created by loss of a purine can effectively be repaired.

Following are some of the important widely used alkylating agents:

- Dimethyl Sulphate (DMS)
- Ethyl-MethaneSulphonate (EMS)
- Ethyl-EthaneSulphonate (EES)

EMS has the specificity to remove Guanine and Cytosine from the chain and results in gap formation. Any base (A, T, G, C) may be inserted in the gap. During replication chain without gap will result in normal DNA. In the second round of replication gap is filled by suitable base. If the correct base is inserted, normal DNA sequence will be produced. Insertion of incorrect bases results in transversion or transition mutation. Another example is methyl nitrosoguanidine that adds methyl group to Guanine causing it to mispair with Thiamine. After subsequent replication, GC is converted into AT transition.

**Intercalating Agents as Mutagen**

The chemical intercalate or slip in between two base pair in double stranded DNA helix and hence alter the morphology of DNA at that position. The chances of error during replication are higher at this position causing mutation. There are certain dyes, such as acridine orange, proflavine and acriflavine which are three ringed molecules of similar dimensions as those of purine pyrimidine pairs. In aqueous solution these dyes can insert themselves in DNA (i.e., intercalate the DNA) between the bases in adjacent pairs by a process called intercalation (Refer Figure 2.12).

![Fig. 2.12 Chemical Structure of Two Mutagenic Acridine Derivatives](image)
Therefore, the dyes are called intercalating agents. The acridines are planer (flat) molecules which can be intercalated between the base pairs of DNA; distort the DNA and results deletion or insertion after replication of DNA molecule. Due to deletion or insertion of intercalating agents, there occur frameshift mutations (Refer Figure 2.13).

**Fig. 2.13** Intercalation of an Acridine Molecule during Replication

**DNA Reactive Chemicals as Mutagen**

These chemical mutagens react directly with the nitrogenous bases of DNA and chemically modify the DNA causing mutation. A large number of chemicals may interact directly with DNA. However, many PAHs, aromatic amines, benzene are not necessarily mutagenic by themselves, but through metabolic processes in cells they produce mutagenic compounds.

- **Reactive Oxygen Species (ROS):** These may be superoxide, hydroxyl radical and hydrogen peroxide, and large number of these highly reactive species are generated by normal cellular processes, for example as a by-products of mitochondrial electron transport, or lipid peroxidation. As an example of the latter, 15-hydroperoxicosatetraenoic acid, a natural product of cellular cyclooxygenases and lipoxygenases, breaks down to form 4-hydroxy-2(E)-nonenal, 4-hydroperoxy-2(E)-nonenal, 4-oxo-2(E)-nonenal, and cis-4,5-epoxy-2(E)-decanal; these bifunctional electrophils are
mutagenic in mammalian cells and may contribute to the development and/or progression of human cancers. A number of mutagens may also generate these ROS. These ROS may result in the production of many base adducts, as well as DNA strand breaks and crosslinks.

- **EthylNitrosourea**: The ethylnitrosourea compounds transfer methyl or ethyl group to bases or the backbone phosphate groups. Guanine when alkylated may be mispaired with thymine. Some may cause DNA crosslinking and breakages. Nitrosamines are an important group of mutagens found in tobacco, and may also be formed in smoked meats and fish via the interaction of amines in food with nitrites added as preservatives. Other alkylating agents include mustard gas and vinyl chloride.

- **Aromatic Amines**: The aromatic amines and amides have been associated with carcinogenesis since 1895 when German physician L. Rehn observed high incidence of bladder cancer among workers in German synthetic aromatic amine dye industry. The 2-acetylaminofluorene originally used as a pesticide but may also be found in cooked meat, may cause cancer of the bladder, liver, ear, intestine, thyroid and breast.

- **Alkaloids**: The alkaloids from plants, such as those from *Vinca* species, may be converted by metabolic processes into the active mutagen or carcinogen.

- **Others**: Bromine and some compounds that contain bromine in their chemical structure, sodium azide, psoralen with UV radiation and benzene may also acts as potent mutagenic agents when exposed to living beings.

**Metals as Mutagen**

Many metals, and their compounds may be mutagenic, but they may act, however, via a number of different mechanisms. Arsenic, chromium, iron, and nickel may be associated with the production of ROS, and some of these may also alter the fidelity of DNA replication. Nickel may also be linked to DNA hypermethylation and histone deacetylation, while some metals, such as cobalt, arsenic, nickel and cadmium may also affect DNA repair processes, such as DNA mismatch repair, base and nucleotide excision repair. There are large numbers of metals for mutagenic in Bacteria and phases for example, compounds of copper, manganese, molybdenum, platinum and selenium. The compounds containing aluminum, antimony, arsenic, cadmium, copper, lead, mercury, and tellurium have been shown to induce chromosomal aberrations or abnormal cell divisions in animal and plant cells. The genetic evidence suggests that arsenic, chromium, and molybdenum compounds may influence the accuracy of DNA repair processes in microorganisms.

Several metals are known mutagens and carcinogens. These metals effectively displace acridine orange from DNA when measured by fluorescence polarization. The displacement of 50% of the acridine orange is obtained with less than 0.5mM concentrations of lead, manganese, cobalt, zinc, cadmium, nickel, iron, copper,
and cis platinum. In contrast, greater than 80mM concentrations of lithium, sodium, and potassium are required to displace an equivalent amount of acridine orange from calf thymus DNA. Although cis platinum shows the best DNA reactivity in this assay, the interaction between this metal and DNA does not occur immediately, as it does for the other metals tested. These results indicate that the acridine orange displacement assay provides a relative measure of the interaction of metals with DNA, and this DNA reactivity shows a positive correlation with mutagenic/carcinogenic potential.

2.3.3 Biological Mutagens

Biological mutagen is a mutation agent in the form of Virus or Bacteria which can induce mutation in every living organism. When cell divides, Virus will change the genetic material (DNA) composition of the attacked cell in order to damage the cell and tissues, toxin produced by bacteria can also cause disorder or damage on genetic material or certain cell and tissues. Hepatitis, chickenpox, measles, yellow fever, or food poisoning (botulism) may begin from genetical material change induced either by Virus or Bacteria. The horticulturists achieve a specific mutation in plant cuttings with the application of colchicine, to the reproductive parts of the flowers. The colchicine is an extract from plants of the Autumn crocus. The mutation which occurs is to double the number of chromosomes, which results in plant with parts twice as large as normal. In this way a plain wild flower can be turned into a showy flower twice as large as the original. Another biological mutagen is simply aging of the egg or sperm. Damage to DNA in both plants and animals causes mutations, such as mongoloidism in humans when the egg has begun to deteriorate from age. There are many biological organism and parts of organism that acts as mutagenic agents given below:

- **Virus:** Virus DNA may be inserted into the genome and disrupts genetic function. Infectious agents have been suggested to cause cancer as early as 1908 by Vilhelm Ellermann and Oluf Bang, and 1911 by Peyton Rous who discovered the Rous sarcoma Virus.

- **Bacteria:** Some Bacteria, such as *Helicobacter pylori* cause inflammation during which oxidative species are produced, causing DNA damage and reducing efficiency of DNA repair systems, thereby increasing mutation.

- **Transposon and Insertion Sequence (IS):** Transposon is a section of DNA that undergoes autonomous fragment relocation/multiplication. Its insertion into chromosomal DNA disrupts functional elements of the genes. These elements (Transposon and IS) are small sequence of DNA that moves from one site to another along DNA strand and causes mutation. They are known as jumping gene. These sequences contain gene which codes the enzyme transposase which helps in transposition of these sequence from one site to other. IS element are simplest type of transposable element. They are short DNA about 1000 nucleotide long and typically contain inverted repeats of
10-50 base pair. The only gene they possess is the gene for enzyme transposase. IS element are found in both chromosome and plasmid of Bacteria and archaea as well as some bacteriophage. The transposons are larger than IS element but similar as IS element because it also has two essential component; contains inverted repeats and encodes transposases. The transposons and Insertion Sequence (IS) elements are biological mutagens, for example; mutator gene, bacteriophage MU, etc.

Check Your Progress
5. Define the term mutagenesis?
6. Name the types of mutagens.
7. What is physical mutation?

2.4 TESTS SYSTEM FOR MUTAGENS

Many different systems for detecting mutagen have been developed. Animal systems may more accurately reflect the metabolism of human, however; they are expensive and time-consuming (may take around three years to complete), they are therefore not used as a first screen for mutagenicity or carcinogenicity.

I. Bacterial Tests

- **Ames Test**: This is the most commonly used test, and *Salmonella typhimurium* strains deficient in histidine biosynthesis are used in this test. The test checks for mutants that can revert to wild-type. It is an easy, inexpensive and convenient initial screen for mutagens.

- **Resistance to 8-Azaguanine in *Salmonella typhimurium***: Similar to Ames test, but instead of reverse mutation, it checks for forward mutation that confer resistance to 8-azaguanine in a histidine revertant strain.

- **Escherichia coli Systems**: Both forward and reverse mutation detection system have been modified for use in *Escherichia coli*. Tryptophan deficient mutant is used for the reverse mutation, while galactose utility or resistance to 5-methyltryptophan may be used for forward mutation.

- **DNA Repair**: *Escherichia coli* and *Bacillus subtilis* strains deficient in DNA repair may be used to detect mutagens by their effect on the growth of these cells through DNA damage.

II. Yeast Tests

Systems similar to Ames test have been developed in yeast. *Saccharomyces cerevisiae* is generally used. These systems can check for forward and reverse mutations, as well as recombinant events.
Mutagens and its Types

III. Drosophila for Sex-Linked Recessive Lethal Test

Males from a strain with yellow bodies are used in this test. The gene for the yellow body lies on the X-chromosome. The fruit flies are fed on a diet of test chemical, and progenies are separated by sex. The surviving males are crossed with the females of the same generation, and if no males with yellow bodies are detected in the second generation, it would indicate a lethal mutation on the X-chromosome has occurred.

IV. Plant Assays

Plants such as Zea mays, Arabidopsis thaliana and Tradescantia have been used in various test assays for mutagenecity of chemicals.

V. Cell Culture Assay

Mammalian cell lines such as Chinese hamster V79 cells, Chinese Hamster Ovary (CHO) cells or mouse lymphoma cells may be used to test for mutagenesis. Such systems include the HRT assay for resistance to 8-azaguainine or 6-thioguanine, and OUA/bain-resistance (OUA) assay. Rat primary hepatocytes may also be used to measure DNA repair following DNA damage. Mutagens may stimulate unscheduled DNA synthesis that results in more stained nuclear material in cells following exposure to mutagens.

VI. Chromosome Check Systems

These systems check for large scale changes to the chromosomes and may be used with cell culture or in animal test. The chromosomes are stained and observed for any changes. Sister chromatids exchange is a symmetrical exchange of chromosome material between sister chromatids and may be correlated to the mutagenic or carcinogenic potential of a chemical. In micronucleus test, cells are examined for micronuclei, which are fragments or chromosomes left behind at anaphase, and is therefore a test for clastogenic agents that cause chromosome breakages. Other tests may check for various chromosomal aberrations such as chromatid and chromosomal gaps and deletions, translocations, and ploidy.

VII. Animal Test Systems

Rodents are usually used in animal test. The chemicals under test are usually administered in the food and in the drinking water, but sometimes by dermal application, by gavage, or by inhalation, and carried out over the major part of the life span for rodents. In tests that check for carcinogens, maximum tolerated dosage is first determined, then a range of doses are given to around 50 animals throughout the notional lifespan of the animal of two years. After death the animals are examined for sign of tumors. Differences in metabolism between rat and human however means that human may not respond in exactly the same way to mutagen, and dosages that produce tumors on the animal test may also be unreasonably high for
a human, i.e., the equivalent amount required to produce tumors in human may far exceed what a person might encounter in real life.

Mice with recessive mutations for a visible phenotype may also be used to check for mutagens. Females with recessive mutation crossed with wild-type males would yield the same phenotype as the wild-type, and any observable change to the phenotype would indicate that a mutation induced by the mutagen has occurred. Mice may also be used for dominant lethal assays where early embryonic deaths are monitored. Male mice are treated with chemicals under test, mated with females, and the females are then sacrificed before parturition and early fetal deaths are counted in the uterine horns.

Check Your Progress
8. What is Ames test?
9. Write a short note on Escherichia coli systems.
10. How can DNA repair be defined?
11. What is cell culture assay?

2.5 CONSEQUENCES OF MUTAGENS EXPOSURES

Mutagenesis arises due to mutagens not only from the errors in the replication but also from damage to the DNA. Some damage is caused by environmental factors, such as radiation and chemicals like mutagens, which are chemical agents that increase the rate of mutation. DNA also undergoes spontaneous damage from the action of water. The mechanism which causes DNA damage and mutagenesis include: hydrolysis (deamination and depurination), alkylation, oxidation, radiation reactions, and base analogue and intercalating agents.

Consequences Due to Radiation Exposure
- Radiation damages the spermatogonia and the damaged germ cells could occur for a very long time, perhaps a lifetime.
- Radiation also induces recessive and dominant point mutations.
- Sometimes gross chromosomal damage may also occur.
- Majority of the mutations after radiation exposure will be of recessive type and, therefore, not affect the phenotype in the first generation.
- Mature oocytes are more susceptible regarding the radiation induced mutation than spermatogonia.
- If conception has taken place shortly after radiation exposure it will be more dangerous.
Consequences Due to Chemical Exposure

The effects of chemical mutagens are less easy to generalize about than those of the radiation mutagen. One important thing is to be remembered that chemical mutagen is very stage specific and, accordingly, chemical mutagens can be classified into two classes:

- Chemicals which are mutagenic to both replicating and non-replicating DNA.
- Chemicals which are mutagenic only to replicating DNA. It has long been recognized that most of the strongly chemical mutagenic agents are also carcinogenic agents, because most of the geneticists agree that somatic mutation can cause cancer. There are a number of chemicals which can act as a potential mutagenic agent in humans and some of these chemicals are also used as drugs for curing some diseases.

Most of cytostatic, antimetabolite, hallucinogenic drugs and some antibiotics also act as potential mutagenic agents in normal therapeutic doses. Therefore, if a patient is treated with a high dose of an unusual or potentially dangerous drug, the doctor must take some careful measures like the recommendation of the use of contraceptives during the period of therapy and at least 8-10 weeks after the therapy, etc.

Check Your Progress

12. From where does mutagenesis arises?
13. Name the mechanism which causes DNA damage and mutagenesis.
14. What are the consequences that are caused due to radiation exposure?
15. In how many classes chemical mutagens can be classified?

2.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Mutagens are chemical compounds or forms of radiation, such as UltraViolet (UV) light or X-rays that cause irreversible and heritable changes (mutations) in the cellular genetic material, DeoxyriboNucleic Acid (DNA).
2. Ethidium Bromide (EtBr) works efficiently because it is capable of inserting itself in between two adjacent DNA bases, a so-called ‘intercalating agent’.
3. Reactive Oxygen Species (ROSs) arise from multiple sources but the primary sources are all related to energy metabolism.
4. The biological consequences of a mutation depend upon many critical factors such as the target loci, size of the mutation, and timing during the cell cycle, and compounding effects of preexisting mutations.
5. Mutagenesis is the process of inducing mutation by a number of physical, chemical or biological agents.

6. There are three types of mutagens:
   - Physical Mutagens
   - Chemical Mutagens
   - Biological Mutagens

7. Physical mutagen is a mutation agent which is in the form of physical substances, such as short waves, particles and radioactive elements.

8. Ames test is the most commonly used test, and Salmonella typhimurium strains deficient in histidine biosynthesis are used in this test. The test checks for mutants that can revert to wild-type. It is an easy, inexpensive and convenient initial screen for mutagens.

9. Escherichia coli Systems, both forward and reverse mutation detection system have been modified for use in Escherichia coli. Tryptophan deficient mutant is used for the reverse mutation, while galactose utility or resistance to 5-methyltryptophan may be used for forward mutation.

10. Escherichia coli and Bacillus subtilis strains deficient in DNA repair may be used to detect mutagens by their effect on the growth of these cells through DNA damage.

11. Cell Culture Assay: Mammalian cell lines such as Chinese hamster V79 cells, Chinese Hamster Ovary (CHO) cells or mouse lymphoma cells may be used to test for mutagenesis. Such systems include the HRT assay for resistance to 8-azaguanine or 6-thioguanine, and OUAbain-resistance (OUA) assay. Rat primary hepatocytes may also be used to measure DNA repair following DNA damage. Mutagens may stimulate unscheduled DNA synthesis that results in more stained nuclear material in cells following exposure to mutagens.

12. Mutagenesis arises due to mutagens not only from the errors in the replication but also from damage to the DNA.

13. The mechanism which causes DNA damage and mutagenesis include: hydrolysis (deamination and depurination), alkylation, oxidation, radiation reactions, and base analogue and intercalating agents.

14. Consequences due to radiation exposure are as follows:
   - Radiation damages the spermatogonia and the damaged germ cells could occur for a very long time, perhaps a lifetime.
   - Radiation also induces recessive and dominant point mutations.
   - Sometimes gross chromosomal damage may also occur.
   - Majority of the mutations after radiation exposure will be of recessive type and, therefore, not affect the phenotype in the first generation.
Mutagens and its Types

- Mature oocytes are more susceptible regarding the radiation induced mutation than spermatogonia.
- If conception has taken place shortly after radiation exposure it will be more dangerous.

15. Chemical mutagens can be classified into two classes:
- Chemicals which are mutagenic to both replicating and non-replicating DNA.
- Chemicals which are mutagenic only to replicating DNA. It has long been recognized that most of the strongly chemical mutagenic agents are also carcinogenic agents, because most of the geneticists agree that somatic mutation can cause cancer. There are a number of chemicals which can act as a potential mutagenic agent in humans and some of these chemicals are also used as drugs for curing some diseases.

2.7 SUMMARY

- Mutagens are chemical compounds or forms of radiation, such as UltraViolet (UV) light or X-rays that cause irreversible and heritable changes (mutations) in the cellular genetic material, DeoxyriboNucleic Acid (DNA).
- Mutagens are not necessarily carcinogens, and vice versa. Sodium azide for example may be mutagenic (and highly toxic), but it has not been shown to be carcinogenic.
- Mutagenic lesions persist when they escape detection by protective cellular DNA repair mechanisms, when mistakes occur in the repair process, or when repair mechanisms are overwhelmed by extensive damage.
- Mutagenesis becomes a cumulative process, stretching over the lifetime of an organism.
- In addition, the accumulation of mutations over time, leading to gradually less efficient cellular repair capabilities, has been linked to the aging process and associated degenerative diseases.
- Genetic predispositions to certain breast and colon cancers, cystic fibrosis, and Huntington’s disease are included among many examples of such inheritable diseases.
- Although many dietary and environmental agents have been classified as mutagens, cells are constantly subjected to a barrage of spontaneous DNA damage.
- In genetics, a mutagen is a physical or chemical agent that changes the genetic material, usually DNA, of an organism and thus increases the frequency of mutations above the natural background level.
Ethidium Bromide (EtBr) is used as a DNA stain due to its enormous affinity towards DNA.

EtBr works efficiently because it is capable of inserting itself in between two adjacent DNA bases, a so-called ‘intercalating agent’.

One mutagen that is either essential or deadly is the ‘oxygen’. It is one of the most potent mutagens known, in the form of Reactive Oxygen Species or ROSs.

ROSs arise from multiple sources but the primary sources are all related to energy metabolism.

One particularly dangerous ROS is superoxide, O₂ (-) where the ‘-’ in parentheses is the charge.

O₂ (-) can attack almost all organic substances in the cell and has a very high potency to cause damage to DNA by producing hydroxyl radicals, which can then attack DNA bases and damage them.

The first mutagens to be identified were carcinogens, substances that were shown to be linked to cancer.

Tumors were described more than 2,000 years before the discovery of chromosomes and DNA in 500 B.C., the Greek.

Physician Hippocrates named tumors resembling a crab *karkinos* (from which the word ‘cancer’ is derived via Latin), meaning crab.

Many mutagens are highly toxic to proliferating cells, and they are often used to destroy cancer cells.

Alkylating agents such as cyclophosphamide and cisplatin, as well as intercalating agent such as daunorubicin and doxorubicin may be used in chemotherapy.

Research on better targeted therapies may reduce such side effects. Ionizing radiations are used in radiation therapy.

Many metals, and their compounds may be mutagenic, but they may act, however, via a number of different mechanisms.

Arsenic, chromium, iron, and nickel may be associated with the production of ROS, and some of these may also alter the fidelity of DNA replication.

Nickel may also be linked to DNA hypermethylation and histone deacetylation, while some metals such as cobalt, arsenic, nickel and cadmium may also affect DNA repair processes such as DNA mismatch repair, base and nucleotide excision repair.

There are large numbers of metals for mutagenic in Bacteria and phases for example, compounds of copper, manganese, molybdenum, platinum and selenium.
Mutagens and its Types

- The compounds containing aluminum, antimony, arsenic, cadmium, copper, lead, mercury, and tellurium have been shown to induce chromosomal aberrations or abnormal cell divisions in animal and plant cells.
- The genetic evidence suggests that arsenic, chromium, and molybdenum compounds may influence the accuracy of DNA repair processes in microorganisms.
- Biological mutagen is a mutation agent in the form of Virus or Bacteria which can induce mutation in every living organism.
- When cell divides, Virus will change the genetic material (DNA) Composition of the attacked cell in order to damage the cell and tissues toxin produced by Bacteria can also cause disorder or damage on genetic material or certain cell and tissues.
- Transposon is a section of DNA that undergoes autonomous fragment relocation/multiplication. Its insertion into chromosomal DNA disrupts functional elements of the genes.
- Transposon and IS elements are small sequence of DNA that moves from one site to another along DNA strand and causes mutation, also known as jumping gene.
- Mutagenesis arises due to mutagens not only from the errors in the replication but also from damage to the DNA.
- Some damage is caused by environmental factors, such as radiation and chemicals like mutagens, which are chemical agents that increase the rate of mutation. DNA also undergoes spontaneous damage from the action of water.
- The mechanism which causes DNA damage and mutagenesis include: hydrolysis (deamination and depurination), alkylation, oxidation, radiation reactions, and base analogue and intercalating agents.

2.8 KEY WORDS

- **Mutagen**: A mutagen is defined as any chemical that can cause changes in the DNA sequence of an organism.
- **Biological mutagen**: Biological mutagen is a mutation agent in the form of Virus or Bacteria which can induce mutation in every living organism.
- **Transposon**: Transposon is a section of DNA that undergoes autonomous fragment relocation/multiplication.
2.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What are mutagens? How is it caused?
2. Explain what physical mutagens are.
3. Write a note on base analogs as chemical mutagen.
4. Explain DNA reactive chemicals as mutagen.
5. What are biological mutations?
6. Name the different tests systems for mutagens.
7. What are the consequences due to radiation exposure?

Long Answer Questions

1. What is mutagenesis? Discuss in details about the mechanism of mutagenesis based on exposure of physical mutagens.
2. In what ways the base analogs act as mutagen. Discuss in detail with necessary diagrams.
3. What are chemical mutagens? Give detail accounts of any two chemicals with mutagenic efficiency studied by you.
4. What are the major consequences of physical and chemical mutagens?
5. Discuss in details about the different types of tests for the assessment of mutagenic potential of various mutagens.

2.10 FURTHER READINGS

Mutagens and its Types

NOTES


3.0 INTRODUCTION

DNA damage is distinctly different from mutation, although both are types of error in DNA. DNA damage is an abnormal chemical structure in DNA, while a mutation is a change in the sequence of standard base pairs. DNA damages cause changes in the structure of the genetic material and prevents the replication mechanism from functioning and performing properly. DNA damage and mutation have different biological consequences. While most DNA damages can undergo DNA repair, such repair is not 100% efficient. Unrepaired DNA damages accumulate in non-replicating cells, such as cells in the brains or muscles of adult mammals, and can cause aging. In replicating cells, such as cells lining the colon, errors occur upon replication past damages in the template strand of DNA or during repair of DNA damages. These errors can give rise to mutations or epigenetic alterations. Both of these types of alteration can be replicated and passed on to subsequent cell generations. These alterations can change gene function or regulation of gene expression and possibly contribute to progression to cancer.

Throughout the cell cycle there are various checkpoints to ensure the cell is in good condition to progress to mitosis. The three main checkpoints are at G1/s, G2/m, and at the spindle assembly checkpoint regulating progression through anaphase. G1 and G2 checkpoints involve scanning for damaged DNA. During S phase the cell is more vulnerable to DNA damage than any other part of the cell cycle. G2 checkpoint checks for damaged DNA and DNA replication completeness. DNA damage is an alteration in the chemical structure of DNA, such as a break in a strand of DNA, a base missing from the backbone of DNA, or a chemically changed base such as 8-OHdG. DNA damage can occur naturally or via environmental factors. The DNA Damage Response (DDR) is a complex signal transduction pathway which recognizes when DNA is damaged and initiates the cellular response to the damage.
In this unit, you will study about DNA damages, deamination of bases, alkylation, damage due to reactive oxygen and UV induced damage in detail.

### 3.1 OBJECTIVES

After going through this unit, you will be able to:
- Explain various DNA damages
- Discuss about deamination of bases and alkylation
- Understand the damage caused due to reactive oxygen
- Analyse UV induced damage

### 3.2 DNA DAMAGES

Charles’s Darwin theory (1859) was based on slow variations exhibited by individuals of same race. Variations increase the adaptability of the individuals and make them better fitted in the struggle for existence and play role in natural selection.

All variations were called as Darwin’s continuous variations. There are two sources of variations in populations, i.e., recombination and mutation. The most significant consequence of oxidative stress in the body is thought to be damage to DNA. DNA may be modified in a variety of ways, which can ultimately lead to mutations and genomic instability. This could result in the development of a variety of cancers including colon, breast, and prostate. Here we discuss the various types of damage to DNA, including oxidative damage, hydrolytic damage, DNA strand breaks, and others. Oxidative DNA damage refers to the oxidation of specific bases. 8-hydroxydeoxyGuanosine (8-OHdG) is the most common marker for oxidative DNA damage and can be measured in virtually any species. It is formed and enhanced most often by chemical carcinogens. A similar oxidative damage can occur in RNA with the formation of 8-OHG (8-hydroxyGuanosine), which has been implicated in various neurological disorders.

Hydrolytic DNA damage involves deamination or the total removal of individual bases. Loss of DNA bases, known as AP (APurinic/APyrimidinic) sites, can be particularly mutagenic and if left unrepaired they can inhibit transcription. Hydrolytic damage may result from the biochemical reactions of various metabolites as well as the overabundance of reactive Oxygen species. Ultraviolet and other types of radiation can damage DNA in the form of DNA strand breaks. This involves a cut in one or both DNA strands; double-strand breaks are especially dangerous and can be mutagenic, since they can potentially affect the expression of multiple genes. UV-induced damage can also result in the production of Pyrimidine Dimers, where covalent cross-links occur in Cytosine and Thymine residues. The most common Pyrimidine Dimers are Cyclobutane Pyrimidine Dimers (CPD) and Pyrimidine - (6-4) Pyrimidine Photoproducts (6-
DNA Damages

4PP). CPD and 6-4PP are the most frequent DNA mutations found in the p53 protein in skin cancers. Pyrimidine dimers can disrupt polymerases and prevent proper replication of DNA.

DNA damage may also result from exposure to Polycyclic Aromatic Hydrocarbons (PAHs). PAHs are potent, ubiquitous atmospheric pollutants commonly associated with oil, coal, cigarette smoke, and automobile exhaust fumes. A common marker for DNA damage due to PAHs is Benzo(a)Pyrene Diol Epoxide (BPDE). BPDE is found to be very reactive, and known to bind covalently to Proteins, Lipids, and Guanine residues of DNA to produce BPDE adducts. If left unrepaired, BPDE-DNA adducts may lead to permanent mutations resulting in cell transformation and ultimately tumor development.

DNA is a highly stable and versatile molecule. Though sometimes the damage is caused to it, it is able to maintain the integrity of information contained in it. The perpetuation of genetic material from generation to generation depends upon keeping the rates of mutation at low level. DNA has many elaborate mechanisms to repair any damage or distortion. The most frequent sources of damage to DNA are the inaccuracy in DNA replication and chemical changes in DNA. Malfunction of the process of replication can lead to incorporation of wrong bases, which are mismatched with the complementary strand.

The damage causing chemicals break the backbone of the strand and chemically alter the bases. Alkylation, oxidation and methylation cause damage to bases. X-rays and gamma radiations cause single or double stranded breaks in DNA. A change in the sequence of bases if replicated and passed on to the next generation becomes permanent and leads to mutation. At the same time mutations are also necessary which provide raw material for evolution. Without evolution, the new species, even human beings would not have arisen. Therefore a balance between mutation and repair is essential.

Types of Damage

Damage to DNA includes any deviation from the usual double helix structure. **Simple Mutations:** Simplest mutations are switching of one base for another base. In transition one pyrimidine is substituted by another Pyrimidine and Purine with another Purine. Transversion involves substitution of a Pyrimidine by a Purine and Purine by a Pyrimidine, such as T by G or A and A by C or T. Other simple mutations are detection, insertion of a single nucleotide or a small number of nucleotides. Mutations which change a single nucleotide are called point mutations.

**Deamination:** The common alteration of form or damage includes deamination of Cytosine (C) to form Uracil (U) which base pairs with Adenine (A) in next replication instead of Guanine (G) with which the original Cytosine (C) would have paired. As Uracil (U) is not present in DNA, Adenine (A) base pairs with Thymine (T). Therefore C-G pair is replaced by T-A in next replication cycle. Similarly, hypoxanthine results from Adenine (A) deamination.
DNA Damages

**NOTES**

**Missing Bases:** Cleavage of N-Glycosidic bond between Purine and sugar causes loss of Purine base from DNA. This is called Depurination. This Apurinic site becomes non-coding lesion.

**Chemical Modification of Bases:** Chemical modification of any of the four bases of DNA leads to modified bases. Methyl groups are added to various bases. Guanine (G) forms 7-Methylguanine, 3-Methylguanine. Adenine (A) forms 3-Methyladenine. Cytosine (C) forms 5-Methylcytosine.

Replacement of Amino group by a Keto group converts 5-Methylcytosine to Thymine (T). Following is Thymine (T) Dimer in one strand.

**Formation of Pyrimidine Dimers (Thymine (T) Dimers):** Formation of Thymine (T) dimers is very common in which a Covalent Bond (Cyclobutyl Ring) is formed between adjacent Thymine (T) bases. This leads to loss of base pairing with opposite stand. A bacteria may have thousands of dimers immediately after exposure to ultraviolet radiations.

**Strand Breaks:** Sometimes phosphodiester bonds break in one strand of DNA helix. This is caused by various chemicals like peroxides, radiations and by enzymes like DNases. This leads to breaks in DNA backbone. Single strand breaks are more common than double strand breaks. Following is the double strand breaks.

Sometimes X-rays, electronic beams and other radiations may cause phosphodiester bonds breaks in both strands which may not be directly opposite to each other. This leads to double strand breaks. Some sites on DNA are more susceptible to damage. These are called hot-stops.

**Repair Mechanisms**

Most kinds of damage create impediments to replication or transcription. Altered bases mispairing can cause permanent alteration to DNA sequence after replication. In order to maintain the integrity of information contained in it, the DNA has various repair mechanisms.

**Direct Repair:** The damage is reversed by a repair enzyme which is called photoreactivation. This mechanism involves a light dependent enzyme called DNA
**Photolyase.** The enzyme is present in almost all cells from bacteria to animals. It uses energy from the absorbed light to cleave the C-C bond of Cyclobutyl Ring of the Thymine (T) dimers. In this way Thymine (T) dimers are monomerized.

**Excision Repair:** It includes base excision repair and nucleotide excision repair. Base excision repair system involves an enzyme called N-Glycosylase which recognizes the abnormal base and hydrolyses Glycosidic bond between it and sugar.

Another enzyme, an endonuclease cleaves the DNA backbone on the 5'-side of the abnormal base. Then the DNA polymerase by its exonuclease activity removes the abnormal base. DNA polymerase then replaces it with normal base and DNA ligase seals the region.

Nucleotide repair system includes three steps, incision, excision and synthesis. Incision is done by endonuclease enzyme precisely on either side of the damaged patch of the strand. In this way damaged portion of the strand is cleaved.

Endonuclease enzymes involved are UvrA, UvrB which recognizes the damaged stretch of the strand. UvrC makes two cuts (incision) on either side. Exonuclease removes the damaged strand. Enzyme involved is UvrD.

Later, DNA polymerase synthesizes the new strand by using complementary strand as a template. DNA ligase forms phosphodiester bonds which seal the ends on newly synthesized strand.

**Mismatch Base Repair:** Sometimes wrong bases are incorporated during replication process, G-T or C-A pairs are formed. The wrong base is always incorporated in the daughter strand only. Therefore in order to distinguish the two strands for the purpose of repair, the Adenine (A) bases of the template strand are labelled or tagged by Methyl groups. In this way the newly replication DNA helix is hemimethylated. The excision of wrong bases occur in the non-methylated or daughter strand.

**Recombination Repair or Retrieval System:** In Thymine (T) dimer or other type of damage, DNA replication cannot proceed properly. A gap opposite to Thymine (T) dimer is left in the newly synthesized daughter strand. The gap is repaired by recombination mechanism or retrieval mechanism also called sister strand exchange.

During replication of DNA two identical copies are produced. Replicating DNA molecule has four Strands A, B, C and D. Strands A and C have same DNA, sequence. Strands B and D also have same sequence as they are identical. A Thymine (T) dimer is present in strand A. The replication fork passes the dimer as it cannot form hydrogen bonds with incoming Adenine (A) bases, thus creating a gap in the newly synthesized Strand B.

In recombination repair system a short identical segment of DNA is retrieved from Strand D and is inserted into the gap of Strand B. But this creates a gap in Strand D which is easily filled up by DNA polymerase using normal Strand C as a template.
template. This event is dependent on the activity of a special Protein RecA. The RecA protein plays its role in retrieving a portion of the complementary strand from the other side of the replication fork to fill the gap. RecA is a strand exchange protein.

After filling both gaps, Thymine (T) is monomerised. So in this repair mechanism a portion of DNA strand is retrieved from the normal homologous DNA segment. This is also known as daughter strand gap repair mechanism.

**SOS Repair Mechanism:** Sometimes the replicating machinery is unable to repair the damaged portion and bypasses the damaged site. This is known as translesion synthesis also called bypass system and is an emergency repair system. This mechanism is catalyzed by a special class of DNA polymerases called Y-family of DNA polymerases which synthesized DNA directly across the damaged portion.

There are different types of DNA damage and therefore different molecular pathways of DNA repair to correct them, including non-homologous end joining, homologous recombination, mismatch repair and nucleotide excision repair. Numerous proteins and pathways have been involved in these processes. The ATM and ATR kinases, as well as DNA-PK, are key for detection of the DNA lesions. Chromatin remodelers participate in making sure that the chromatin environment is accessible to the DNA repair apparatus, and key repair factors, such as RPA, Rad51 and the Fanconi Anemia proteins directly act in repairing the DNA lesions. Fanconi anemia is a rare disease passed down through families (inherited) that mainly affects the bone marrow. It results in decreased production of all types of blood cells. Importantly the p53 network, the RAS GTPase superfamily, and the ubiquitin system are also involved in different aspects of the DNA damage response (Refer Figure 3.1).
Direct DNA damage can occur when DNA directly absorbs a UVB photon, or for numerous other reasons. UVB light causes Thymine (T) base pairs next to each other in genetic sequences to bond together into Pyrimidine Dimers, a disruption in the strand, which reproductive enzymes cannot copy. It causes sunburn and it triggers the production of Melanin.

Other names for the ‘Direct DNA Damage’ are:
- Thymine Dimers
- Pyrimidine Dimers
- Cyclobutane Pyrimidine Dimers (CPDs)
- UV Endonuclease Sensitive-Sites (UPESS)

Due to the excellent photochemical properties of DNA, this nature made molecule is damaged by only a tiny fraction of the absorbed photons. DNA transforms more than 99.9% of the photons into harmless heat (but the damage from the remaining < 0.1% is still enough to cause sunburn). The transformation of excitation energy into harmless heat occurs via a photochemical process called internal conversion. In DNA, this internal conversion is extremely fast, and therefore efficient. This ultrafast (subpicosecond) internal conversion is a powerful photo-protection provided by single nucleotides. However, the Ground-State Recovery is much slower (picoseconds) in G–C–DNA duplexes and hairpins. It is presumed to be even slower for double-stranded DNA in conditions of the nucleus.

The absorption spectrum of DNA shows a strong absorption for UVB radiation and a much lower absorption for UV A radiation. Since the action spectrum of sunburn is indistinguishable from the absorption spectrum of DNA, it is generally accepted that the direct DNA damages are the cause of sunburn.

While the human body reacts to direct DNA damages with a painful warning signal, no such warning signal is generated from indirect DNA damage.

### 3.2.1 Deamination of Bases

Cytosine (C) susceptible to hydrolysis deaminated to Uracil (U). If left uncorrected, the conversion of Cytosine (C) to Uracil (U) mutations migration occurs. It is a base for foreign DNA, Uracil (U) will change back enzyme specific Cytosine (C) to Uracil (U) DNA Glycosylase or UDG.

Deamidation is by removing the Amino groups of the molecule. Enzyme that catalyzes this reaction, known as deaminase. The human body, deamination takes place mainly in the liver, but glutamic acid, is deaminated in the kidney. If there is an excess of proteins, is a process when the deamination amino acid is split. It is removed from an Amino acid, and an Amino group is converted to Ammonia. Consists essentially of Carbon and Hydrogen, the rest of the Amino Acid sequence is Oxidized or recycled as energy. Ammonia is toxic to the human body, the enzyme converts the uric acid and urea by adding a molecule (which is
DNA Damages

Release of Ammonia in the process, deamination spontaneous reaction is a hydrolysis of Cytosine (C) to Uracil (U). Instead of the bisulfite to convert Cytosine (C), and can be prepared by in vitro of using the 5’-Methylcytosine. This property has been allowed for researchers to distinguish between Methylated Cytosine (C) (shown as Uracil (U)) Unmethylated Cytosine (C) to sequence methylated DNA. The DNA, the deamination spontaneous is adjusted removal of Uracil (U). DNA glycosylase generator of abasic (AP) site (rather than Cytosine (C) deamination, the product of the portion of the DNA) by. By replacing Cytosine (C) another basic sites obtained is recognized by the (AP endonuclease) enzymes that degrade to allow repair of the lesion, the phosphodiester bonds of DNA. DNA polymerase is able to fill the reaction of the polymerase activity by nick translation terminal excision reaction due to the 3’ → 5’ exonuclease activity, to perform this substitution followed. DNA ligase to form a phosphodiester bond sealing the two strand nicks resulting product contains a new, correct Cytosine 5-methylcytosine is a methylated form of the DNA base Cytosine (C), which can be incorporated in the regulation of gene transcription.

5-Methylcytosine is a methylated form of the DNA base Cytosine (C) that may be involved in the regulation of gene transcription. When Cytosine (C) is methylated, the DNA maintains the same sequence, but the expression of methylated genes can be altered. 5-Methylcytosine is incorporated in the nucleoside 5-methylcytidine. In 5-methylcytosine, a methyl group is attached to the 5th atom in the 6-atom ring, counting counterclockwise from the NH nitrogen at the six o’clock position, not the 2 o’clock. This methyl group distinguishes 5-methylcytosine from Cytosine (C).

When trying to separate the bacterial toxins responsible for tuberculosis, separating the nucleic acids called Tuberculinic Acid in 1898 by a new Mycobacterium tuberculosis. It was first identified by a German chemist W.G. Ruppel in 1898. The nucleic acid, in addition to Methylated Nucleotides, Guanine (G), Cytosine (C) and Thymine (T), can be included in it. Coghill and Johnson (1925), successfully detected a minor amount of a Methylated Cytosine derivative as a product of hydrolysis of Tuberculinic Acid with Sulfuric Acid. This report was criticized for the identification of those based on the optical properties of crystalline picric acid exclusively rapidly, other scientists did not have the ability to reproduce the same results. Using a paper chromatography to establish a Methylated Cytosine (C) unique completely different Uracil (U) and Cytosine (C) conventional, when it is divided into the nucleic acid of calf thymus DNA, and finally, Stapler, facts 1948. After 70 years, the precise role is unclear, but also, it should be a common feature of the RNA molecule in a variety was found.

Life, seems to be designed to minimize the error. The universal nature of proof-reading/repair machine, to optimize the genetic code, parity sign of the G/
C/T will all agree on this point. However, in spite of the logic of this design, Cytosine (C), is particularly sensitive to deamination, such as removal of the exocyclic Amino group is interesting, but there, it will be a base, which is normally found in RNA Uracil (U). Typically the Uracil (U) in DNA is not present, it can be removed by detecting the repair enzymes if effectively. If it is detected, it is not deleted, base pair with Adenine (A), which means that you specify the Adenine (A) during DNA replication can be it. Thymine (T) in set in subsequent rounds of replication of Adenine (A) in order. The conclusion is that the deamination spontaneous Cytosine (C), there is a possibility that base substitutions called transition, is replaced by (being substituted with the other strand of the DNA and G) T is C occurs. The conversion to C half-life single-stranded DNA of Cytosine (C) for specific about 200 years certain Cytosine (C) deamination degree, since such mutations are a common occurrence at 37 degrees C. In fact, the high rate of these deamination, led researchers from the pool. Al complain of "Confusion Cytosine".

3.2.2 Alkylation

Alkylation lesions in DNA and RNA result from endogenous compounds, environmental agents and alkylating drugs. Simple methylating agents, for example methyl nitrosourea, tobacco-specific nitrosamines and drugs like Temozolomide or Streptozotocin, form adducts at N- and O-atoms in DNA bases. These lesions are mainly repaired by direct base repair, base excision repair, and to some extent by Nucleotide Excision Repair (NER). The identified carcinogenicity of O (6)-methylGuanine (O (6)-meG) is largely caused by its miscoding properties. Mutations from this lesion are prevented by O (6)-alkylG-DNA alkyltransferase (MGMT or AGT) that repairs the base in one step. However, the genotoxicity and cytotoxicity of O (6)-meG is mainly due to recognition of O(6)-meG/T (or C) mispairs by the MisMatch Repair (MMR) system and induction of futile repair cycles, eventually resulting in cytotoxic double-strand breaks. Therefore, inactivation of the MMR system in an AGT-defective background causes resistance to the killing effects of O (6)-alkylating agents, but not to the mutagenic effect. Bifunctional alkylating agents, such as Chlorambucil or Carmustine (BCNU), are commonly used anti-cancer drugs. DNA lesions caused by these agents are complex and require complex repair mechanisms. Thus, primary chloroethyl adducts at O (6)-G are repaired by AGT, while the secondary highly cytotoxic Interstrand Cross-Links (ICLs) require nucleotide excision repair factors for incision and homologous recombination to complete repair. Recently, Escherichia coli protein AlkB and human homologues were shown to be oxidative demethylases that repair cytotoxic 1-methylAdenine (1-meA) and 3-methylCytosine (3-meC) residues. Numerous AlkB homologues are found in Viruses, Bacteria and Eukaryotes, including eight human homologues (hABH1–8). These have distinct locations in subcellular compartments and their functions are only starting to be understood. Surprisingly, AlkB and hABH3 also repair RNA. An evaluation of the biological effects of environmental mutagens, as well as understanding the mechanism of
action and resistance to alkylating drugs require a detailed understanding of DNA repair processes (Refer Figure 3.2).

### 3.2.3 Damage Due to Reactive Oxygen

About 100 Oxidatively generated base lesions and 2-Deoxyribose modifications, including initially formed Thymidine Hydroperoxides and Diastereomeric nucleosides, have been isolated and identified in model studies. The number of products detected in cellular DNA is much lower, owing to several limitations and difficulties. These include, among others, the lack of sensitivity of available methods for detecting lesions produced in low yields, instability of some modifications, such as base hydroperoxides, optimization of assays that may require the synthesis of internal standards labeled with stable isotopes, and finally, artefactual oxidation of overwhelming normal nucleosides during DNA extraction and subsequent workup.

**Hydroxyl Radical**

The hydroxyl radical (•OH) is a highly Reactive Oxygen Species (ROS) that efficiently reacts with nearby biomolecules at diffusion-controlled rates of reaction. The reaction volume of •OH is less than 2 nm in cells and tissues; thus, it reacts essentially at the site of generation. The most likely source of •OH in cells is the Fenton reaction which involves the reaction of reduced redox active metal ions, such as Ferrous and Cuprous Ions, with metabolically produced H₂O₂. For this reason, the main lines of defense against ROS by aerobic organisms include metal-binding chelators and proteins (for example, Ferritin) to minimize the concentration of labile metal ions, together with catalase and peroxidases to minimize the
concentration of H$_2$O$_2$. The generation of ‘OH by Fenton-like reactions is believed to take place in a site-specific manner, for example, involving metal ions in close proximity or bound to DNA. ‘OH is also generated by the radiolysis of water molecules according to the so-called indirect effect of ionizing radiation.

**Thymine**

Two main reactions mediated by ‘OH have been shown to take place with Thymine (T) nucleobases in cellular DNA: addition across the 5,6-pyrimidine bond and H-atom abstraction from the methyl group. Model studies have shown that ‘OH preferentially adds to C5 and to a lesser extent to C6, giving rise to reducing C6-yl and oxidizing C5-yl radicals, respectively. In the case of nucleoside Thymidine (T), O$_2$ rapidly adds to the radical site, giving rise to the corresponding hydroperoxyl radicals that subsequently convert into eight cis and trans diastereomers of 5-hydroxy-6-hydroperoxy-5,6-dihydrothymidin and 6-hydroxy-5-hydroperoxy-5,6-dihydrothymidin. The major radiation-induced base degradation products so far detected in cellular DNA are the cis and trans diastereomers of 5,6-dihydroxy-5,6-dihydrothymine. These products may be explained by stereospecific reduction of intermediate Thymine Hydroperoxides. Typically, the Thymine Hydroperoxides may also decomposed by Pyrimidine Ring cleavage to 5-hydroxy-5-methylhydantoin derivatives (Hyd-Thy), which was recently detected in irradiated cells. The second major pathway of ‘OH-mediated decomposition of Thymine (T) and its derivatives, including DNA in solution, involves H-atom abstraction from the methyl group. This leads to the 5-(Uracilyl) methyl radical, which is readily converted into the corresponding peroxyl radical after O$_2$ addition and hydroperoxide after subsequent reduction and protonation. In turn, hydroperoxide decomposes by reduction and competitive dehydration to 5-HydroxyHethylUracil (5-HmUra) and 5-FormylUracil (5-FoUra) derivatives, respectively. The latter products are major oxidation products detected in cellular DNA by HPLC coupled to Electrospray Ionization-Tandem Mass Spectrometry (ESI-MS/MS). The 5-(Uracilyl) methyl radical can also react with neighboring Guanine (G) and Adenine (A) bases to produce intrastrand or possibly interstrand cross-links connected between the methyl group of Thymine (T) and the C8 position of either Guanine (G) or Adenine (A) DNA intrastrand cross-link.

**3.2.4 UV Induced DNA Damage**

Induction of DNA damage by solar UVR is a key event that drastically influences the normal life processes of all organisms. A number of endogenous factors, such as free radicals generated during metabolic processes as well as exogenous factors, such as UV or ionizing radiations are known to interfere with genome integrity (Refer Figure 3.3).

DNA damage results in:

- Misincorporation of bases during replication process.
DNA Damages

- Hydrolytic damage, which results in deamination of bases, depurination, and depyrimidination.
- Oxidative damage, caused by direct interaction of Ionizing Radiations (IR) with the DNA molecules, as well as mediated by UV radiation-induced free radicals or reactive oxygen species.
- Alkylating agents that may result in modified bases.

The hydrolytic deamination (loss of an Amino group) can directly convert one base to another; for example, deamination of Cytosine (C) results in Uracil (U) and at much lower frequency Adenine (A) to hypoxanthine. In depurination/depyrimidination, there are complete removals of Purine/Pyrimidine bases, leaving the deoxyribose sugar depurinated/depyrimidinated that may cause breakage in the DNA backbone. The exposure of UVR, IR, and certain genotoxic chemicals may result in single as well as double DNA strand breaks. Among different types of damages, DNA Double Strand Breaks (DSBs) are the most deleterious, since they affect both strands of DNA and can lead to the loss of genetic material. At high concentrations oxygen-free radicals or, more frequently, Reactive Oxygen Species (ROS) can induce damage to cell structure, lipids, proteins as well as DNA and results in oxidative stress which has been implicated in a number of human diseases. The hydroxyl radicals (OH°) can damage all components of DNA molecules, such as Purine and Pyrimidine bases and also the deoxyribose backbone, inhibiting the normal functions of the cell.

Fig. 3.3 UV Induced DNA Damage
Check Your Progress

1. What does hydrolytic DNA damage involves?
2. Where can DNA damage result?
3. What is simple mutations?
4. Define deamination.
5. Write in short about direct repair.
6. Give the other names for the ‘direct DNA damage’.
7. What are DNA damage results?

3.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Hydrolytic DNA damage involves deamination or the total removal of individual bases. Loss of DNA bases, known as AP (APurinic/APyrimidinic) sites, can be particularly mutagenic and if left unrepaired they can inhibit transcription. Hydrolytic damage may result from the biochemical reactions of various metabolites as well as the overabundance of reactive oxygen species.

2. DNA damage may also result from exposure to Polycyclic Aromatic Hydrocarbons (PAHs). PAHs are potent, ubiquitous atmospheric pollutants commonly associated with oil, coal, cigarette smoke, and automobile exhaust fumes.

3. Simple mutations are switching of one base for another base. In transition one Pyrimidine is substituted by another Pyrimidine and Purine with another Purine. Transversion involves substitution of a Pyrimidine by a Purine and Purine by a Pyrimidine, such as T by G or A and A by C or T. Other simple mutations are detection, insertion of a single nucleotide or a small number of nucleotides. Mutations which change a single nucleotide are called point mutations.

4. The common alteration of form or damage includes deamination of Cytosine (C) to form Uracil (U) which base pairs with Adenine (A) in next replication instead of Guanine (G) with which the original Cytosine (C) would have paired.

5. The damage is reversed by a repair enzyme which is called photoreactivation. This mechanism involves a light dependent enzyme called DNA photolyase. The enzyme is present in almost all cells from bacteria to animals. It uses energy from the absorbed light to cleave the C-C bond of Cyclobutyl Ring of the Thymine (T) dimers. In this way Thymine (T) dimers are monomerized.
6. Other names for the 'Direct DNA Damage' are:
   - Thymine Dimers
   - Pyrimidine Dimers
   - Cyclobutane Pyrimidine Dimers (CPDs)
   - UV Endonuclease Sensitive-Sites (UVESS)

7. DNA damage results in:
   - Misincorporation of bases during replication process.
   - Hydrolytic damage, which results in deamination of bases, depurination, and depyrimidination.
   - Oxidative damage, caused by direct interaction of Ionizing Radiations (IR) with the DNA molecules, as well as mediated by UV radiation-induced free radicals or Reactive Oxygen Species (ROS).
   - Alkylation agents that may result in modified bases.

3.4 SUMMARY

- Charles’s Darwin theory (1859) was based on slow variations exhibited by individuals of same race.
- Variations increase the adaptability of the individuals and make them better fitted in the struggle for existence and play role in natural selection. All variations were called as Darwin’s continuous variations.
- The most significant consequence of oxidative stress in the body is thought to be damage to DNA.
- DNA may be modified in a variety of ways, which can ultimately lead to mutations and genomic instability.
- Oxidative DNA damage refers to the oxidation of specific bases. 8-hydroxydeoxyguanosine (8-OHdG) is the most common marker for oxidative DNA damage and can be measured in virtually any species.
- A similar oxidative damage can occur in RNA with the formation of 8-OHG (8-hydroxyguanosine), which has been implicated in various neurological disorders.
- Hydrolytic DNA damage involves deamination or the total removal of individual bases.
- Loss of DNA bases, known as AP (APurinic/APyrimidinic) sites, can be particularly mutagenic and if left unrepaired they can inhibit transcription.
- Hydrolytic damage may result from the biochemical reactions of various metabolites as well as the overabundance of Reactive Oxygen Species (ROS).
DNA Damages

- Ultraviolet and other types of radiation can damage DNA in the form of DNA strand breaks. This involves a cut in one or both DNA strands; double-strand breaks are especially dangerous and can be mutagenic, since they can potentially affect the expression of multiple genes.
- UV-induced damage can also result in the production of Pyrimidine dimers, where covalent cross-links occur in Cytosine (C) and Thymine (T) residues.
- The most common pyrimidine dimers are Cyclobutane Pyrimidine Dimers (CPD) and Pyrimidine -(6-4) Pyrimidine Photoproducts (6-4PP).
- CPD and 6-4PP are the most frequent DNA mutations found in the p53 protein in skin cancers.
- Pyrimidine dimers can disrupt polymerases and prevent proper replication of DNA.
- DNA damage may also result from exposure to Polycyclic Aromatic Hydrocarbons (PAHs).
- PAHs are potent, ubiquitous atmospheric pollutants commonly associated with oil, coal, cigarette smoke, and automobile exhaust fumes.
- A common marker for DNA damage due to PAHs is Benzo(a)Pyrene Diol Epoxide (BPDE). BPDE is found to be very reactive, and known to bind covalently to proteins, lipids, and guanine residues of DNA to produce BPDE adducts.
- The perpetuation of genetic material from generation to generation depends upon keeping the rates of mutation at low level. DNA has many elaborate mechanisms to repair any damage or distortion.
- The most frequent sources of damage to DNA are the inaccuracy in DNA replication and chemical changes in DNA.
- Malfunction of the process of replication can lead to incorporation of wrong bases, which are mismatched with the complementary strand.
- The damage causing chemicals break the backbone of the strand and chemically alter the bases.
- Alkylation, Oxidation and Methylation cause damage to bases. X-rays and gamma radiations cause single or double stranded breaks in DNA.
- A change in the sequence of bases if replicated and passed on to the next generation becomes permanent and leads to mutation.
- Simplest mutations are switching of one base for another base. In transition one Pyrimidine is substituted by another Pyrimidine and Purine with another Purine.
- Transversion involves substitution of a Pyrimidine by a Purine and Purine by a Pyrimidine, such as T by G or A and A by C or T. Other simple
DNA Damages

mutations are detection, insertion of a single nucleotide or a small number of nucleotides.

- Mutations which change a single nucleotide are called point mutations.
- The common alteration of form or damage includes deamination of Cytosine (C) to form Uracil (U) which base pairs with Adenine (A) in next replication instead of Guanine (G) with which the original Cytosine (C) would have paired.
- Cleavage of N-Glycosidic bond between purine and sugar causes loss of purine base from DNA. This is called depurination. This apurinic site becomes non-coding lesion.
- Chemical modification of any of the four bases of DNA leads to modified bases. Methyl groups are added to various bases. Guanine forms 7-Methylguanine, 3-Methylguanine.
- The damage is reversed by a repair enzyme which is called photoreactivation. This mechanism involves a light dependent enzyme called DNA Photolyase.
- Another enzyme, an endonuclease cleaves the DNA backbone on the 5' side of the abnormal base. Then the DNA polymerase by its exonuclease activity removes the abnormal base.
- DNA polymerase then replaces it with normal base and DNA ligase seals the region.
- Endonuclease enzymes involved are UvrA, UvrB which recognize the damaged stretch of the strand. UvrC makes two cuts (incision) on either side.
- The absorption spectrum of DNA shows a strong absorption for UVB radiation and a much lower absorption for UVA radiation.
- Deamidation is by removing the amino groups of the molecule. Enzyme that catalyzes this reaction, known as deaminase. The human body, deamination takes place mainly in the liver, but glutamic acid, is deaminated in the kidney.
- Two main reactions mediated by 'OH have been shown to take place with Thymine nucleobases in cellular DNA: addition across the 5,6 - pyrimidine bond and H-atom abstraction from the methyl group.
- Induction of DNA damage by solar UVR is a key event that drastically influences the normal life processes of all organisms.
- A number of endogenous factors, such as free radicals generated during metabolic processes as well as exogenous factors, such as UV or ionizing radiations are known to interfere with genome integrity.
3.5 **KEY WORDS**

- **Malfunction**: Malfunction of the process of replication that can lead to incorporation of wrong bases, which are mismatched with the complementary strand.
- **Point mutations**: Mutations which change a single nucleotide are called point mutations.

3.6 **SELF ASSESSMENT QUESTIONS AND EXERCISES**

**Short Answer Questions**

1. What is DNA damage?
2. What are the types of DNA damages?
3. How can missing bases be defined?
4. Draw the structure of Thymine (T) dimer in one strand.
5. Distinguish between direct repair and excision repair.
6. What is SOS repair mechanism?
7. Explain UV induced DNA damage.

**Long Answer Questions**

1. Discuss about DNA damages and its types in detail.
2. Write a note on formation of pyrimidine dimers.
3. What is repair mechanism? Explain.
4. Discuss about mismatch base repair.
5. Explain deamination of bases.
6. Write in detail about alkylation.
7. Explain damage due to reactive Oxygen.

3.7 **FURTHER READINGS**


UNIT 4  REPAIR PATHWAYS

4.0 INTRODUCTION

DNA repair processes are of crucial importance for the maintenance of the genetic information of all organisms. The stability of the genome is constantly endangered by environmental agents, endogenous metabolic processes, for example reactive species inside cells, and errors of cellular processes involving DNA. Modifications of DNA can lead to mutations, which alter the coding sequence of DNA and can lead to cancer in humans and other mammals. Other DNA lesions interfere with normal cellular transactions, such as DNA replication or transcription, and are deleterious to the cell. To counteract DNA damage, organisms have evolved various damage prevention and repair systems. These systems ensure the stability of DNA and accurate transmission of genetic information by protecting the genome against a large number of different chemical and structural alterations. At the same time, random changes in DNA are viewed as a main source of genetic variability, and thus a driving force for evolution. In multicellular organisms changes in the DNA sequence and structure are responsible, for example differential production of antibodies by the immune system. Therefore, DNA repair mechanisms have to balance the noxious against the beneficial effects of alterations in the genome sequence and chemical structure.

DNA repair is a very complicated process, involving many factors. For instance to date, 168 genes that encode proteins involved in DNA repair have been identified in the human genome. They are involved in diverse processes, starting from detection of a damage site in the DNA, through several steps of

4.1 Objectives

4.2 DNA Damage and Repair Pathways

4.2.1 An SOS Repair System or Recombinational Repair
4.2.2 Methyl Directed Mismatch Repair
4.2.3 Nucleotide Excision Repair
4.2.4 Recombinational Repair
4.2.5 SOS Inducible Repair
4.2.6 Specific Repair for Oxidative DNA Damage
4.2.7 Pyrimidine Dimers
4.2.8 Alkylation Induced Damage and Adaptive Response

4.3 Answers to Check Your Progress Questions

4.4 Summary

4.5 Key Words

4.6 Self Assessment Questions and Exercises

4.7 Further Readings
enzy


tomat transformation of the damaged DNA, to recombination and signaling to stop the cell cycle or initiate apoptosis. Another form of dealing with the DNA damage is lesion bypass, which facilitates continuation of replication even when irremovable modifications occur, but does not guarantee proper recreation of the original sequence and frequently leads to mutation generated by TransLesion Synthesis (TLS) polymerases.

In this unit, you will study about repair pathways, methyl directed mismatch repair, nucleotide excision repair, base excision repair, recombinational repair, SOS inducible repair, specific repair for oxidative DNA damage, pyrimidine dimers and alkylation induced damage and adaptive response in detail.

4.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss about repair pathways and its types
- Explain methyl directed mismatch repair
- Understand what nucleotide excision repair is
- Explain base excision and recombinational repair
- Describe SOS inducible repair and specific repair for oxidative DNA damage

4.2 DNA DAMAGE AND REPAIR PATHWAYS

DNA damage is a change in the basic structure of DNA that is not itself replicated when the DNA is replicated. A DNA damage can be a chemical addition or disruption to a base of DNA (creating an abnormal nucleotide or nucleotide fragment) or a break in one or both chains of the DNA strands. When DNA carrying a damaged base is replicated, an incorrect base can often be inserted opposite the site of the damaged base in the complementary strand, and this can become a mutation in the next round of replication. Also, DNA double-strand breaks may be repaired by an inaccurate repair process leading to mutations. In addition, a double strand break can cause rearrangements of the chromosome structure (possibly disrupting a gene, or causing a gene to come under abnormal regulatory control), and, if such a change can be passed to successive cell generations, it is also a form of mutation. Mutations, however, can be avoided if accurate DNA repair systems recognize DNA damages as abnormal structures, and repair the damages prior to replication (Refer Figure 4.1).
DNA Damage occurs due to:

- **Single Base Change**: Such changes take place due to conversion of one base to another, for example deamination of 5-methylcytosine to Thymine (T).
- **Structural Distortion**: Such distortion takes place due to single strand nick, removal of a base or introduction of a covalent link between bases of same or different strands, for example formation of Thymine (T) dimer due to UV.

There are several types of damage to DNA due to endogenous cellular processes (Refer Figure 4.2):

- Oxidation of bases, for example 8-oxo-7, 8-dihydroGuanine (8-oxoG) and generation of DNA strand interruptions from reactive oxygen species.
- Alkylation of bases (usually methylation), such as formation of 7-methylguanosine, 1-methyladenine, 6-O-Methylguanine.
- Hydrolysis of bases, such as deamination, depurination, and depyrimidination.
- Mismatch of bases, due to errors in DNA replication, in which the wrong DNA base is stitched into place in a newly forming DNA strand, or a DNA base is skipped over or mistakenly inserted.

![Fig. 4.1 DNA Damage](image)

![Fig. 4.2 Types of DNA Damages](image)
Damage caused by exogenous agents comes in many forms. Some examples are:

- UV-B light causes crosslinking between adjacent Cytosine and Thymine bases creating pyrimidine dimers. This is called direct DNA damage.
- UV-A light creates mostly free radicals. The damage caused by free radicals is called indirect DNA damage.
- Ionizing radiation such as that created by radioactive decay or in cosmic rays causes breaks in DNA strands. Intermediate-level ionizing radiation may induce irreparable DNA damage (leading to replicational and transcriptional errors needed for neoplasia or may trigger viral interactions) leading to pre-mature aging and cancer.
- Thermal disruption at elevated temperature increases the rate of depurination (loss of purine bases from the DNA backbone) and single-strand breaks. For example, hydrolytic depurination is seen in the thermophilic bacteria, which grow in hot springs at 40–80°C. The rate of depurination (300 purine residues per genome per generation) is too high in these species to be repaired by normal repair machinery, hence a possibility of an adaptive response cannot be ruled out.
- Industrial chemicals such as vinyl chloride and hydrogen peroxide, and environmental chemicals such as polycyclic aromatic hydrocarbons found in smoke, soot and tar create a huge diversity of DNA adducts ethane bases, oxidized bases, alkylated phosphotriesterase and crosslinking of DNA, just to name a few.

UV damage, alkylation/methylation, X-ray damage and oxidative damage are examples of induced damage. Spontaneous damage can include the loss of a base, deamination, sugar ring puckering and tautomeric shift. Constitutive (spontaneous) DNA damage caused by endogenous oxidants can be detected as a low level of histone H2AX phosphorylation in untreated cells.

**Nuclear Versus Mitochondrial**

In human cells, and eukaryotic cells in general, DNA is found in two cellular locations – inside the nucleus and inside the mitochondria. Nuclear DNA (nDNA) exists as chromatin during non-replicative stages of the cell cycle and is condensed into aggregate structures known as chromosomes during cell division. In either state the DNA is highly compacted and wound up around bead-like proteins called histones. Whenever a cell needs to express the genetic information encoded in its nDNA the required chromosomal region is unravelled, genes located therein are expressed, and then the region is condensed back to its resting conformation.

Mitochondrial DNA (mtDNA) is located inside mitochondria organelles, exists in multiple copies, and is also tightly associated with a number of proteins to form a complex known as the nucleoid. Inside mitochondria, Reactive Oxygen Species (ROS) or free radicals, byproducts of the constant production of Adenosine TriPhosphate (ATP) via oxidative phosphorylation, create a highly oxidative environment that is known to damage mtDNA. A critical enzyme in counteracting...
the toxicity of these species is superoxide dismutase, which is present in both the mitochondria and cytoplasm of Eukaryotic cells.

**Senescence and Apoptosis**

Senescence, an irreversible process in which the cell no longer divides, is a protective response to the shortening of the chromosome ends. The telomeres are long regions of repetitive non-coding DNA that cap chromosomes and undergo partial degradation each time a cell undergoes division. In contrast, quiescence is a reversible state of cellular dormancy that is unrelated to genome damage. Senescence in cells may serve as a functional alternative to apoptosis in cases where the physical presence of a cell for spatial reasons is required by the organism, which serves as a 'last resort' mechanism to prevent a cell with damaged DNA from replicating inappropriately in the absence of pro-growth cellular signaling. Unregulated cell division can lead to the formation of a tumor, which is potentially lethal to an organism. Therefore, the induction of senescence and apoptosis is considered to be part of a strategy of protection against cancer.

**DNA Repair System**

The mechanism of DNA repair in different species exhibit similarities as well as differences. There are many methods for DNA repair. Every method repairs a particular type of damage. Some of the repair systems are as follows (Refer Figure 4.3):

![Fig. 4.3 DNA Repair Mechanism](image)
Photoreactive Repair or Direct Repair: In direct repair of DNA damage no fresh synthesis of DNA is required and process of DNA damage is reversed. Albert Kelner (1949) reported the direct repair of UV induced DNA damage in *Escherichia coli*. He observed that DNA damage process can be reversed if damaged cells were exposed to light in blue range of visible spectrum. The photoreactivation is due to enzyme photo-reactivation enzyme and temperature controlled. The enzyme cleaves the bond between Thymine dimers and due to this effect of UV radiations on DNA is reversed.

Excision Repair: It is light independent repair system by Paul Howard Flanders in 1960’s. A group of genes called Uvr genes (UvrA, UvrB, and UvrC) code for the components of repair enzyme endonuclease which is involved in the process of excision repair. Another enzyme UvrD is also needed for helicase activity. The mechanism involves the following steps:

- **Recognition and Cleavage**: The distorted or affected of strand caused by UV induced dimer is recognized and enzymatically clipped out by endonuclease enzyme that cleaves phosphodiester bond. This excision may include several nucleotides adjacent to dimer as well and leaves a gap in the helix.
- **Gap Filling**: DNA polymerase I fills this gap by inserting the ribonucleotides complementary to those on the intact strands. The enzyme adds these bases to the 3OH end of the clipped DNA.
- The enzyme DNA ligase seals the final nick that remains at the 3OH end of the last base inserted thus closing the gap. The process of excision repair can be activated in response to any damage to DNA that distorts the helix provided the distortion may be recognized, for example DNA glycosylases recognizes the presence of Uracil when it is a part of DNA.

Excision repair system is absent in some human beings causing rare disease Xeroderma Pigmentosum (XP). Such persons cannot tolerate sunlight.

Mismatch Repair: To cope up with the errors that remain after ‘Proof Reading’ another mechanism called mismatch repair was proposed by Robin Holliday like other DNA lesions. It has following features:

- Detection of Mismatch or Alteration.
- Removal of Incorrect Nucleotide.
- Replacement with Correct Base.

But a special problem exists in the recognition of correct template, which contains the mismatch, for example there is mismatch in base pair as GC-GT then a repair system may give rise to GC or AT, wild type or mutant type. The repair system has to distinguish the new strand and to restore the wild type.

In *Escherichia coli*, this process has been elucidated and is based on the process called DNA methylation. The bacteria contain an enzyme called Adenine (A) Methylase that recognizes DNA sequence as substrate. Upon recognition, a
methyl group is added to each of the Adenine (A) residues. Following replication, the newly synthesized strand remains unmethylated for a short period and at this point, the repair enzyme recognizes the mismatch and starts the process of repair.

4.2.1 An SOS Repair System or Recombinational Repair

It was first discovered in an excision defected strain of *Escherichia coli*. This system is thought to respond when damaged DNA escaped repair and process of replication is interrupted. About 20 different gene products are involved in this repair mechanism of which LexA product is of particular interest. The SOS repair response is initiated by the interaction of RecA protein with LexA repressor coded by gene RecA and LexA, respectively. It is also called recombinational repair (Refer Figure 4.4).

Following steps are involved in SOS repair system or recombinational repair:
- DNA damage activates RecA proteins (a protease).
- RecA interacts with LexA and causes it to undertake autocatalytic cleavage. This inactivates the repression function of LexA and all the operon bound to LexA are induced.
- Since LexA represses its own synthesis its own synthesis and that of RecA, its proteolytic degradation leads to amplification of both RecA and LexA proteins.
- When damage signal is removed, RecA loses induction and free LexA protein quickly accumulates in the uncleaved forms and turn off SOS genes.

![Fig. 4.4 SOS Repair System](image-url)
4.2.2 Methyl Directed Mismatch Repair

DNA mismatch repair is a system for recognizing and repairing erroneous insertion, deletion, and misincorporation of bases that can arise during DNA replication and recombination, as well as repairing some forms of DNA damage.

Mismatch repair is strand-specific. During DNA synthesis the newly synthesized (daughter) strand will commonly include errors. In order to begin repair, the mismatch repair machinery distinguishes the newly synthesized strand from the template (parental). In Gram-Negative Bacteria, transient hemimethylation distinguishes the strands (the parental is methylated and daughter is not). However, in other prokaryotes and eukaryotes, the exact mechanism is not clear. It is suspected that, in eukaryotes, newly synthesized lagging-strand DNA transiently contains nicks (before being sealed by DNA ligase) and provides a signal that directs mismatch proofreading systems to the appropriate strand. This implies that these nicks must be present in the leading strand, and evidence for this has recently been found. Recent work has shown that nicks are sites for RFC-dependent loading of the replication sliding clamp PCNA, in an orientation-specific manner, such that one face of the donut-shape protein is juxtaposed toward the 3'-OH end at the nick. Oriented PCNA then directs the action of the MutL alpha endonuclease to one strand in the presence of a mismatch and MutS alpha or MutS beta. Any mutational event that disrupts the superhelical structure of DNA carries with it the potential to compromise the genetic stability of a cell. The fact that the damage detection and repair systems are as complex as the replication machinery itself highlights the importance evolution has attached to DNA fidelity.

Examples of mismatched bases include a G/T or A/C pairing. Mismatches are commonly due to tautomerization of bases during G2. The damage is repaired by recognition of the deformity caused by the mismatch, determining the template and non-template strand, and excising the wrongly incorporated base and replacing it with the correct nucleotide. The removal process involves more than just the mismatched nucleotide itself. A few or up to thousands of base pairs of the newly synthesized DNA strand can be removed.

DNA MisMatch Repair (MMR) is an evolutionarily conserved process that corrects mismatches generated during DNA replication and escape proofreading. MMR proteins also participate in many other DNA transactions, such that inactivation of MMR can have wide-ranging biological consequences, which can be either beneficial or detrimental. We begin this review by briefly considering the multiple functions of MMR proteins and the consequences of impaired function. We then focus on the biochemical mechanism of MMR replication errors. Emphasis is on structure-function studies of MMR proteins, on how mismatches are recognized, on the process by which the newly replicated strand is identified, and on excision of the replication error.
4.2.3 Nucleotide Excision Repair

In both prokaryotes and eukaryotes, a major cellular mechanism for the removal of DNA damage is nucleotide excision repair (excision repair), an enzymatic pathway that recognizes and corrects a wide spectrum of structural anomalies (DNA lesions) ranging from bulky, helix-distorting adducts to non-helix-distorting lesions. The modifications that transform normal bases into damaged bases corrected by nucleotide excision repair are so diverse that it is unlikely that a specific chemical structure is recognized. Rather, it appears that any abnormal DNA structure that destabilizes (denatures) the double helix is recognized as damage both in *Escherichia coli* and human cells.

The primary function of nucleotide excision repair is removal of bulky adducts generated by chemicals or UV radiation, while base excision repair is the major pathway for correction of non-helix-distorting lesions, such as those introduced by ionizing radiation or cellular metabolic events. Additional pathways exist for direct reversal of certain types of damage (for example, photolyase and methyltransferase), correction of mismatched bases, removal of interstrand crosslinks, and repair of DNA strand breaks. Excision repair involves removal of a damaged nucleotide by dual incisions bracketing the lesion, this is accomplished by a multisubunit enzyme referred to as the excision nuclease or excinuclease (Refer Figure 4.5).

![Fig. 4.5 Nucleotide Excision Repair](image)

Fig. 4.5 Nucleotide Excision Repair
The basic mechanism of excision repair involves:

- **Damage Recognition.**
- **Subunit Assembly.**
- **Dual Incisions that result in Excision of the Damage-Containing Oligomer.**
- **Resynthesis to fill in the Gap.**
- **Ligation to Regenerate an Intact Molecule.**

**In Prokaryotes: Uvr Proteins**

A schematic representation of models for the nucleotide excision repair pathway controlled by Uvr proteins.

The process of nucleotide excision repair is controlled in *Escherichia coli* by the UvrABC endonuclease enzyme complex, which consists of four Uvr proteins – UvrA, UvrB, UvrC, and DNA Helicase II (sometimes also known as UvrD in this complex). First, UvrA-UvrB complex scans the DNA, with the UvrA subunit recognizing distortions in the helix, caused for example by pyrimidine dimers. When the complex recognizes such a distortion, the UvrA subunit leaves and an UvrC protein comes in and binds to the UvrB monomer and, hence, forms a new UvrBC dimer. UvrB cleaves a phosphodiester bond 4 nucleotides downstream of the DNA damage, and the UvrC cleaves a phosphodiester bond 8 nucleotides upstream of the DNA damage and created 12 nucleotide excised segment. DNA Helicase II (sometimes called UvrD) then comes in and removes the excised segment by actively breaking the hydrogen bonds between the complementary bases. The resultant gap is then filled in using DNA Polymerase I and DNA Ligase. The basic excision process is very similar in higher cells, but these cells usually involve many more proteins – *Escherichia coli* is a simple example.

**TC-NER, i.e., Transcription-Coupled Nucleotide Excision Repair** also exists in bacteria, and is mediated by the TRCF, i.e., Transcription Repair Coupling Factor (Mfd, i.e., Mutation frequency decline) protein. TRCF is an SF2 ATPase that uses ATP hydrolysis to translocate on dsDNA upstream of the transcription bubble and forward translocate RNA polymerase, thus initiating dissociation of the RNA polymerase ternary elongation complex. TRCF also recruits the UvrABC nucleotide excision repair machinery by direct physical interaction with the UvrA subunit.

**4.2.4 Recombinational Repair**

In cases where DNA is severely damaged, a cell will engage in a phenomenon called the SOS response in an effort to salvage a functioning set of genetic information. This response, also called error prone repair, represents a last ditch response to salvage a chromosomal information system. In addition, recombinational repair systems act to allow one copy of the replicating DNA at a replication fork to supply information to the other daughter chromosome. Recombinational repair is a way of using one copy of the cell’s information to ensure that the overall information store remains intact.
The biochemical process of recombination occurs by breaking and rejoining DNA strands. The key reaction is strand displacement initiated at a nick in the chromosome. Then a protein called RecA (which stands for recombination; rec bacteria are unable to recombine their DNA information and therefore are abnormally sensitive to UV radiation) binds to a single stranded DNA fragment and catalyzes its exchange with the same sequence of the duplex. RecA protein is a strand displacement protein.

RecA preferentially binds to single stranded DNA in a cooperative fashion; this cooperativity means that RecA will cover an entire single stranded DNA molecule rather than bind to several molecules partially. RecA then aligns homologous segments (those with complementary information) to form base pairs. The key reaction of RecA coated DNA is the movement of the single stranded regions of the DNA to form a joint molecule—a process called strand displacement. This reaction involves ATP hydrolysis (Refer Figure 4.6).

![Image of DNA repair pathways](image_url)

**Fig. 4.6 Homologous Recombination DNA Repair Pathway**

In homologous recombination, two double helices align and are nicked. Then RecA catalyzes the invasion of each double helix by one strand of the other. This forms a crossed structure called a Holliday junction. If the Holliday structure
were simply broken at the point where it was formed, no genetic recombination could occur because the two original DNA molecules would simply reform. Instead, the junction migrates by displacement of one strand of DNA. Finally, the displaced Holliday junction is broken and rejoined, or resolved. The exact type of recombination between the two strands depends on which of the strands is broken and rejoined.

**Note:** Each recombination event involves two breaking and rejoining events - one to initiate strand displacement and one to resolve the Holliday junction.

### 4.2.5 SOS Inducible Repair

Chromosomal DNA is exposed to continuous damage and repair. Cells contain a number of proteins and specific DNA repair systems that help maintain its correct structure. The SOS response was the first DNA repair system described in *Escherichia coli* induced upon treatment of bacteria with DNA damaging agents arrest DNA replication and cell division. Induction of the SOS response involves more than forty independent SOS genes, most of which encode proteins engaged in protection, repair, replication, mutagenesis and metabolism of DNA. Under normal growth conditions the SOS genes are expressed at a basal level, which increases distinctly upon induction of the SOS response. The SOS-response has been found in many bacterial species (for example, *Salmonella typhimurium*, *Caulobacter crescentus*, *Mycobacterium tuberculosis*), but not in Eukaryotic cells.

However, species from all kingdoms contain some SOS-like proteins taking part in DNA repair that exhibit amino acid homology and enzymatic activities related to those found in *Escherichia coli*, but are not organized in an SOS system.

### 4.2.6 Specific Repair for Oxidative DNA Damage

The integrity of DNA is constantly challenged by damaging agents and chemical modifications. Base oxidation is a frequent insult that can arise from endogenous metabolic processes as well as from exogenous sources such as ionizing radiation. At background levels, a human cell is estimated to undergo 100 to 500 such modifications per day, most commonly resulting in 8-oxo-7, 8-dihydro guanine (8-oxoG) and related products, which are then processed into repair intermediates. At steady state, up to 2400 8-oxoG sites per cell are reported. However, estimates differ widely due to differences in methodology.

Oxidative damage is processed in a two-step process through the Base Excision Repair (BER) pathway. The damaged base is first recognized and excised by 8-OxoGuanine DNA Glycosylase 1 (OGG1), leaving an Apurinic Site (AP-site). Glycohydrolysis is highly efficient, with an 8-oxoG half-life of 11 minutes. AP-sites are removed through backbone incision by AP lyase (APEX 1), and end processing through Flap-Endonuclease 1 (FEN1), and the base is subsequently replaced with an undamaged nucleotide. Alternatively, in short-patch base excision repair, replacement is dependent on polymerase beta. Other sources of AP-sites
Repair Pathways

include spontaneous depurination and excision of non-oxidative base modifications, such as Uracil. Cells are reported to typically present with a steady state of ~15,000 to ~30,000 AP-sites per cell, which includes the associated beta-elimination product. Left unrepaired, 8-oxoG can compromise transcription, DNA replication, and telomere maintenance. Also, AP-sites can lead to genomic instability and compromise genomic processes. Moreover, damaged sites provide direct and indirect routes to C-to-A mutagenesis.

Ionizing radiation is one of the most relevant exogenous sources of high-level oxidative DNA damage and DNA strand breaks. Each gray (Gy) is estimated to lead to ~10^6 ionization events in the nucleus, only ~2000 of which are supposed to target DNA directly. Most DNA damage from ionizing radiation occurs indirectly from radiolysed water and 60–70% can be prevented through radical scavenging. While absolute numbers differ throughout the literature, Lehnert estimates 1000–2000 base modifications per gray, 250 alkali labile sites, 1000 Single-Strand Breaks (SSB), and 40 double-strand breaks. Others report base modifications to be threefold more prevalent than SSBs or even several orders of magnitude increased. Interestingly, direct formation of AP-sites however has been shown not to increase more than 5% from background levels. Therefore, after ionizing radiation, most AP-sites likely arise from excision of oxidized bases, which comprise mostly of 8-oxoG and the related modification FaPy-Guanine (G). Though originally controversial, there is now broad acceptance that mutation rates vary across different genomic regions. Background mutation rates in *Escherichia coli* were shown to vary non-randomly between genes by an order of magnitude, with highly expressed genes displaying lower mutation rates. In cancer genomes, Single Nucleotide Variants (SNVs) tend to accumulate preferentially in heterochromatin. More recently, it was reported that SNV densities in cancers are lower in regions surrounding transcription factor binding but are elevated at the binding sites themselves and at sites with high nucleosome occupancy. These variabilities likely arise through a combination of regional differences in damage sensitivity and the accessibility to the DNA repair machinery.

However, since mutations represent the endpoint of mutagenesis, it is impossible to tease apart the contributions from damage and repair through resequencing alone. The role of oxidative damage in regional differences of mutagenesis remains largely unclear. Repair intermediates remain unexplored, but the genome-wide distribution of 8-oxoG has been studied through chemical enrichment and immunoprecipitation. The specificity of 8-oxoG antibodies, however, remains questionable and the studies using chemical enrichment also arrive at disparate conclusions. Both Wu *et al.*, and Ding *et al.*, find 8-oxoG enriched at telomeres in yeast and mouse embryonic fibroblasts, respectively.

However, Wu *et al.*, find 8-oxoG largely depleted at promoters, while Ding *et al.* report increased damage at these sites. Therefore, we reassessed the raw data and did not find evidence for increased 8-oxoG at promoters. Using antibodies, however, peaks of 8-oxoG accumulation under conditions of hypoxia have been...
reported in active promoters linked to specific transcription factors on a larger scale, studies found accumulation in GC and CpG island rich, early replicating DNA, but also in gene deserts and the nuclear periphery. Some of these apparently contradicting conclusions may be explained through different levels of resolution, experimental systems, and methodology. So far, ionizing radiation-induced oxidative damage has not been addressed genome wide. In addition, base modifications, which have been processed into the more persistent AP-sites remain hidden from the previously used techniques.

4.2.7 Pyrimidine Dimers

Pyrimidine dimers are molecular lesions formed from Thymine or Cytosine bases in DNA via photochemical reactions. Ultraviolet (UV) light induces the formation of covalent linkages between consecutive bases along the nucleotide chain in the vicinity of their Carbon–Carbon double bonds.

The dimerization reaction can also occur among pyrimidine bases in dsRNA (double-stranded RNA)—Uracil (U) or Cytosine (C). Two common UV products are Cyclobutane Pyrimidine Dimers (CPDs) and 6–4 photoproducts. These premutagenic lesions alter the structure and possibly the base-pairing. Up to 50–100 such reactions per second might occur in a skin cell during exposure to sunlight, but are usually corrected within seconds by photolyase reactivation or nucleotide excision repair. Uncorrected lesions can inhibit polymerases, cause misreading during transcription or replication, or lead to arrest of replication. Pyrimidine dimers are the primary cause of melanomas in humans (Refer Figure 4.7).

Fig. 4.7 Pyrimidine Dimers

4.2.8 Alkylation Induced Damage and Adaptive Response

Alkylation damage to DNA occurs when cells encounter alkylating agents in the environment or when cellular metabolism produces active alkylators. To cope
with DNA alkylation, cells have evolved genes that encode proteins with alkylation-specific DNA repair activities. In *Escherichia coli*, the main response specific for alkylation damage has been called the adaptive response. The adaptive response genes are induced upon exposure to exogenous alkylators by Ada-dependent induction, and also during stationary phase by *rpoS*-dependent gene expression, possibly to prevent accumulation of DNA damage due to increased endogenous production of alkylating agents. Recent studies of the regulatory mechanisms of Ada protein and the various responses of the individual promoters regulated by this protein has revealed a complexity of regulation not initially recognized.

Check Your Progress

1. What is DNA damage?
2. Give the causes of DNA damage.
3. What is mismatch repair? Give its features.
4. List the steps involved in SOS repair system or recombinational repair.
5. What are the basic mechanism of excision repair?
6. Define pyrimidine dimers.
7. How does alkylation damage to DNA occur?

4.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. DNA damage is a change in the basic structure of DNA that is not itself replicated when the DNA is replicated. A DNA damage can be a chemical addition or disruption to a base of DNA (creating an abnormal nucleotide or nucleotide fragment) or a break in one or both chains of the DNA strands.

2. DNA damage occurs due to:
   - Single Base Change: Such changes takes place due to conversion of one base to another, for example deamination of 5-methylcytosine to Thymine (T).
   - Structural Distortion: Such distortion take place due to single strand nick, removal of a base or introduction of a covalent link between bases of same or different strands, for example formation of Thymine (T) dimer due to UV.

3. To cope up with the errors that remain after ‘Proof Reading’ another mechanism called mismatch repair was proposed by Robin Holliday like other DNA lesions. It has following features:
   - Detection of Mismatch or Alteration.
   - Removal of Incorrect Nucleotide.
   - Replacement with Correct Base.
4. Following steps are involved in SOS repair system or recombinational repair:
   - DNA damage activates RecA proteins (a protease).
   - RecA interacts with LexA and causes it to undertake autocatalytic cleavage. This inactivates the repression function of LexA and all the operon bound to LexA are induced.
   - Since LexA represses its own synthesis and that of RecA, its proteolytic degradation leads to amplification of both RecA and LexA proteins.
   - When damage signal is removed, RecA loses induction and free LexA protein quickly accumulates in the uncleaved forms and turn off SOS genes.

5. The basic mechanism of excision repair involves:
   - Damage Recognition.
   - Subunit Assembly.
   - Dual Incisions that result in Excision of the Damage-Containing Oligomer.
   - Resynthesis to fill in the Gap.
   - Ligation to Regenerate an Intact Molecule.

6. Pyrimidine dimers are molecular lesions formed from Thymine or Cytosine bases in DNA via photochemical reactions. UltraViolet (UV) light induces the formation of covalent linkages between consecutive bases along the nucleotide chain in the vicinity of their Carbon–Carbon double bonds.

7. Alkylation damage to DNA occurs when cells encounter alkylating agents in the environment or when cellular metabolism produces active alkylators. To cope with DNA alkylation, cells have evolved genes that encode proteins with alkylation-specific DNA repair activities.

4.4 SUMMARY

- DNA damage is a change in the basic structure of DNA that is not itself replicated when the DNA is replicated.
- A DNA damage can be a chemical addition or disruption to a base of DNA (creating an abnormal nucleotide or nucleotide fragment) or a break in one or both chains of the DNA strands.
- When DNA carrying a damaged base is replicated, an incorrect base can often be inserted opposite the site of the damaged base in the complementary strand, and this can become a mutation in the next round of replication.
- A double strand break can cause rearrangements of the chromosome structure (possibly disrupting a gene, or causing a gene to come under abnormal regulatory control), and, if such a change can be passed to successive cell generations, it is also a form of mutation.
Mutations, however, can be avoided if accurate DNA repair systems recognize DNA damages as abnormal structures, and repair the damages prior to replication.

- Mismatch of bases, due to errors in DNA replication, in which the wrong DNA base is stitched into place in a newly forming DNA strand, or a DNA base is skipped over or mistakenly inserted.
- Ionizing radiation such as that created by radioactive decay or in cosmic rays causes breaks in DNA strands.
- UV damage, alkylation/methylation, X-ray damage and oxidative damage are examples of induced damage.
- In human cells, and eukaryotic cells in general, DNA is found in two cellular locations – inside the nucleus and inside the mitochondria. Nuclear DNA (nDNA) exists as chromatin during non-replicative stages of the cell cycle and is condensed into aggregate structures known as chromosomes during cell division.
- Mitochondrial DNA (mtDNA) is located inside mitochondria organelles, exists in multiple copies, and is also tightly associated with a number of proteins to form a complex known as the nucleoid.
- Inside mitochondria, Reactive Oxygen Species (ROS), or free radicals, byproducts of the constant production of Adenosine TriPhosphate (ATP) via oxidative phosphorylation, create a highly oxidative environment that is known to damage mtDNA.
- A critical enzyme in counteracting the toxicity of these species is superoxide dismutase, which is present in both the mitochondria and cytoplasm of Eukaryotic cells.
- Senescence, an irreversible process in which the cell no longer divides, is a protective response to the shortening of the chromosome ends.
- The telomeres are long regions of repetitive non-coding DNA that cap chromosomes and undergo partial degradation each time a cell undergoes division.
- Senescence in cells may serve as a functional alternative to apoptosis in cases where the physical presence of a cell for spatial reasons is required by the organism, which serves as a ‘last resort’ mechanism to prevent a cell with damaged DNA from replicating inappropriately in the absence of pro-growth cellular signaling.
- Unregulated cell division can lead to the formation of a tumor, which is potentially lethal to an organism.
- The mechanism of DNA repair in different species exhibit similarities as well as differences.
- The distorted or affected strand caused by UV induced dimer is recognized and enzymatically clipped out by endonuclease enzyme that cleaves phosphodiester bond.
- DNA Polymerase I fills this gap by inserting the ribonucleotides complementary to those on the intact strands.
- DNA mismatch repair is a system for recognizing and repairing erroneous insertion, deletion, and misincorporation of bases that can arise during DNA replication and recombination, as well as repairing some forms of DNA damage.
- During DNA synthesis the newly synthesised (daughter) strand will commonly include errors.
- In Gram-Negative Bacteria, transient hemimethylation distinguishes the strands (the parental is methylated and daughter is not).
- DNA MisMatch Repair (MMR) is an evolutionarily conserved process that corrects mismatches generated during DNA replication and escape proofreading.
- MMR proteins also participate in many other DNA transactions, such that inactivation of MMR can have wide-ranging biological consequences, which can be either beneficial or detrimental.
- Emphasis is on structure-function studies of MMR proteins, on how mismatches are recognized, on the process by which the newly replicated strand is identified, and on excision of the replication error.
- In both prokaryotes and eukaryotes, a major cellular mechanism for the removal of DNA damage is nucleotide excision repair (excision repair), an enzymatic pathway that recognizes and corrects a wide spectrum of structural anomalies (DNA lesions) ranging from bulky, helix-distorting adducts to non-helix-distorting lesions.
- The modifications that transform normal bases into damaged bases corrected by nucleotide excision repair are so diverse that it is unlikely that a specific chemical structure is recognized.
- The primary function of nucleotide excision repair is removal of bulky adducts generated by chemicals or UV radiation, while base excision repair is the major pathway for correction of non-helix-distorting lesions such as those introduced by ionizing radiation or cellular metabolic events.
- The process of nucleotide excision repair is controlled in *Escherichia coli* by the UvrABC endonuclease enzyme complex, which consists of four Uvr proteins: UvrA, UvrB, UvrC, and DNA Helicase II (sometimes also known as UvrD in this complex).
● TC-NER, i.e., Transcription-Coupled Nucleotide Excision Repair also exists in bacteria, and is mediated by the TRCF, i.e., Transcription Repair Coupling Factor (Mfd, i.e., Mutation frequency decline) protein.

● TRCF also recruits the UvrABC nucleotide excision repair machinery by direct physical interaction with the UvrA subunit.

● In cases where DNA is severely damaged, a cell will engage in a phenomenon called the SOS response in an effort to salvage a functioning set of genetic information.

● The biochemical process of recombination occurs by breaking and rejoining DNA strands.

● RecA preferentially binds to single stranded DNA in a cooperative fashion; this cooperativity means that RecA will cover an entire single stranded DNA molecule rather than bind to several molecules partially.

● Chromosomal DNA is exposed to continuous damage and repair. Cells contain a number of proteins and specific DNA repair systems that help maintain its correct structure.

● The integrity of DNA is constantly challenged by damaging agents and chemical modifications. Base oxidation is a frequent insult that can arise from endogenous metabolic processes as well as from exogenous sources such as ionizing radiation.

● Ionizing radiation is one of the most relevant exogenous sources of high-level oxidative DNA damage and DNA strand breaks.

● Pyrimidine dimers are molecular lesions formed from Thymine or Cytosine bases in DNA via photochemical reactions. Ultraviolet (UV) light induces the formation of covalent linkages between consecutive bases along the nucleotide chain in the vicinity of their Carbon–Carbon double bonds.

● The dimerization reaction can also occur among pyrimidine bases in dsRNA (double-stranded RNA)—Uracil (U) or Cytosine (C).

● Uncorrected lesions can inhibit polymerases, cause misreading during transcription or replication, or lead to arrest of replication.

● Alkylation damage to DNA occurs when cells encounter alkylating agents in the environment or when cellular metabolism produces active alkylators.

● In Escherichia coli, the main response specific for alkylation damage has been called the adaptive response.

● The adaptive response genes are induced upon exposure to exogenous alkylators by Ada-dependent induction, and also during stationary phase by rpoS-dependent gene expression, possibly to prevent accumulation of DNA damage due to increased endogenous production of alkylating agents.
4.5 KEY WORDS

- **DNA damage**: DNA damage is a change in the basic structure of DNA that is not itself replicated when the DNA is replicated.

- **DNA mismatch repair**: DNA mismatch repair is a system for recognizing and repairing erroneous insertion, deletion, and misincorporation of bases that can arise during DNA replication and recombination, as well as repairing some forms of DNA damage.

- **Senescence**: Senescence is an irreversible process in which the cell no longer divides, is a protective response to the shortening of the chromosome ends.

4.6 SELF ASSESSMENT QUESTIONS AND EXERCISES

**Short Answer Questions**

1. What is senescence and apoptosis?
2. Explain the types of DNA damages with the help of diagram.
3. What is DNA repair system?
4. Define what excision repair is.
5. Draw a well-labelled diagram to show nucleotide excision repair.
7. What are SOS inducible repair?
8. What are pyrimidine dimers? Explain with the help of example.

**Long Answer Questions**

1. Write a detailed note on DNA damage and repair pathways.
2. Draw a well-labelled diagram to show DNA damage.
3. What is photoreactive repair or direct repair? Explain.
4. Elaborate a note on SOS repair system.
5. Write in detail about methyl directed mismatch repair.
7. Discuss in detail about specific repair for oxidative DNA damage.
8. What are pyrimidine dimers? Explain briefly.
9. What are alkylation induced damage and adaptive response?
4.7 FURTHER READINGS

5.0 INTRODUCTION

Recombination, in genetics, regrouping of the maternal and paternal genes during the formation of gametes, the sex cells. Recombination occurs randomly in nature as a normal event of meiosis, the process by which gametes are produced. Recombination is enhanced by the phenomenon of crossing over, in which gene sequences called linkage groups are disrupted, resulting in an exchange of segments between paired chromosomes that are undergoing separation. Thus, although a normal daughter cell produced in meiosis always receives half of the genetic material contained in the parent cell, recombination acts to ensure constant variability: no two daughter cells are identical, nor are any identical in genetic content to the parent cell.

In eukaryotic cells, which are cells with a nucleus and organelles, recombination typically occurs during meiosis. Meiosis is a form of cell division that produces gametes, or egg and sperm cells. During the first phase of meiosis,
the homologous pairs of maternal and paternal chromosomes align. During the alignment, the arms of the chromosomes can overlap and temporarily fuse, causing a crossover. Crossovers result in recombination and the exchange of genetic material between the maternal and paternal chromosomes. As a result, offspring can have different combinations of genes than their parents. Genes that are located farther apart on the same chromosome have a greater likelihood of undergoing recombination, which means they have a greater recombination frequency.

In this unit, you will study about history and introduction of recombination, types of recombination, mechanism and effects of homologous recombination and mechanism and biological role of site-specific recombination in detail.

5.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the history and introduction of recombination
- Explain the types of recombination
- Discuss the mechanism and effects of homologous recombination
- Explain the mechanism and biological role of site-specific recombination

5.2 RECOMBINATION AND ITS TYPES

The genetic recombination involves the exchange or sharing of genetic information between multiple parental sequences to create new 'Progeny' genes. A key benefit of using recombination to generate genetic diversity is the fact that the parental sequence fragments being combined correspond to pieces of related and functional proteins (the homologues provide 'functional diversity'). Compared to random mutagenesis, where increasing the number of mutations drastically reduces the probability of obtaining a folded or functional protein, recombination allows for a higher probability of a greater number of amino acid changes (relative to a given parent sequence) to retain the protein’s ability to fold and carry traits from both parents, as well as potentially new properties. The in vitro recombination for the sake of creating gene libraries is classified as either homologous, when the parent genes share a high level of sequence similarity, or non-homologous, when the parents are not homologues or share low sequence similarity.

The genetic recombination (also known as genetic reshuffling) is the exchange of genetic material between different organisms which leads to production of offspring with combinations of traits that differ from those found in either parent. In eukaryotes, genetic recombination during meiosis can lead to a novel set of genetic information that can be passed on from the parents to the offspring. Most recombination is naturally occurring. During meiosis in eukaryotes, genetic recombination involves the pairing of homologous chromosomes. This may be
followed by information transfer between the chromosomes. The information transfer may occur without physical exchange (a section of genetic material is copied from one chromosome to another, without the donating chromosome being changed), or by the breaking and rejoining of DNA strands, which forms new molecules of DNA. The recombination may also occur during mitosis in eukaryotes where it ordinarily involves the two sister chromosomes formed after chromosomal replication. In this case, new combinations of alleles are not produced since the sister chromosomes are usually identical. In meiosis and mitosis, recombination occurs between similar molecules of DNA.

In meiosis, non-sister homologous chromosomes pair with each other so that recombination characteristically occurs between non-sister homologues. In both meiotic and mitotic cells, recombination between homologous chromosomes is a common mechanism used in DNA repair. The recombination can be artificially induced in laboratory (in vitro) settings, producing rDNA (recombinant DNA) or purposes including vaccine development. The V(D)J recombination in organisms with an adaptive immune system is a type of site-specific genetic recombination that helps immune cells rapidly diversify to recognize and adapt to new pathogens. The V(D)J recombination occurs in the primary lymphoid organs and in a nearly random fashion rearranges Variable (V), Joining (J), and in some cases, Diversity (D) gene segments.

The genetic recombination is catalyzed by many different enzymes, for example, recombinase (catalyze the strand transfer step during recombination found in *Escherichia coli*), RAD51 and DMC1 required for mitotic and meiotic recombination found in yeast. The non-homologous recombination can occur between DNA sequences that contain no sequence homology. This can cause chromosomal translocations, sometimes leading to cancer. In genetic engineering recombination can also refer to artificial and deliberate recombination of disparate pieces of DNA, often from different organisms, creating what is called rDNA. A prime example of such a use of genetic recombination is gene targeting, which can be used to add, delete or otherwise change an organism’s genes. This technique is important to biomedical researchers as it allows them to study the effects of specific genes. Techniques based on genetic recombination are also applied in protein engineering to develop new proteins of biological interest. The achiasmy is the phenomenon where autosomal recombination is completely absent in one sex of a species. Achiasmatic chromosomal segregation is well documented in male fruit fly. The heterochiasmy occurs when recombination rates differ between the sexes of a species. This sexual dimorphic pattern in recombination rate has been observed in many species. In mammals, females most often have higher rates of recombination. The ‘Haldane-Huxley Rule’ states that achiasmy usually occurs in the heterogametic sex. Thus the DNA recombination involves the exchange of genetic material either between multiple chromosomes or between different regions of the same chromosome. This process is generally mediated by homology; that is, homologous regions of chromosomes line up in preparation for exchange, and
some degree of sequence identity is required. Various cases of non-homologous recombination do exist, however (Refer Figure 5.1).

![Image](image_url)

**Fig. 5.1 Model of Meiotic Recombination Started with dsDNA Breaks followed by Pairing with Homologous Chromosomes**

The role of recombination during the inheritance of chromosomes was first demonstrated through experiments with maize. Specifically, in 1931, Barbara McClintock and Harriet Creighton obtained evidence for recombination by physically tracking an unusual knob structure within certain maize chromosomes through multiple genetic crosses. Using a strain of maize in which one member of a chromosome pair exhibited the knob but its homologue did not, the scientists were able to show that some alleles were physically linked to the knobbed chromosome, while other alleles were tied to the normal chromosome. McClintock and Creighton then followed these alleles through meiosis, showing that alleles for specific phenotypic traits were physically exchanged between chromosomes (Refer Figure 5.2). Evidence for this finding came from the fact that alleles first introduced into the cross on a knobbed chromosome later appeared in offspring without the knob; similarly, alleles initially introduced on a knobless chromosome subsequently appeared in progeny with the knob. The recombination also occurs in Prokaryotic...
Recombination and its Types

NOTES

Self-Instructional Material

Cells, and it has been especially well characterized in Escherichia coli. Although bacteria do not undergo meiosis, they do engage in a type of sexual reproduction called conjugation, during which genetic material is transferred from one bacterium to another and may be recombined in the recipient cell. As in Eukaryotes, recombination also plays important roles in DNA repair and replication in Prokaryotic organisms.

Fig. 5.2 McClintock and Creighton’s Model Showing Physical Evidences of Recombination

Recombination

Genetic recombination occurs when genetic material is exchanged between two different chromosomes or between different regions within the same chromosome. We can observe it in both Eukaryotes (like animals and plants) and Prokaryotes (like archaea and bacteria). Keep in mind that in most cases, in order for an exchange to occur, the sequences containing the swapped regions have to be homologous, or similar, to some degree. The process occurs naturally and can also be carried out in the lab. Recombination increases the genetic diversity in
Genetic recombination occurs naturally in meiosis. Meiosis is the process of cell division that occurs in eukaryotes, such as humans and other mammals, to produce offspring. In this case, it involves crossing over. What happens is that two chromosomes, one from each parent, pair up with each other. Next, a segment from one crosses over, or overlaps, a segment of the other. This allows for the swapping of some of their material, as you can see in the illustration below. What we end up with is a new combination of genes that did not exist before and is not identical to either parent’s genetic information. Note that recombination is also observed in mitosis, but it does not occur as often in mitosis as it does in meiosis (Refer Figure 5.3).

The cell can undergo recombinational repair also, for example, if it notices that there is a harmful break in the DNA which is a kind of break that occurs in both strands. What we observe is an exchange between the broken DNA and a homologous region of DNA that will fill the gaps. There are also other ways that recombination is used to repair DNA. Thus genetic recombination refers to the rearrangement of DNA sequences by some combination of the breakage, rejoining, and copying of chromosomes or chromosome segments. It also describes the consequences of such rearrangements, i.e., the inheritance of novel combinations of alleles in the offspring that carry recombinant chromosomes. Genetic recombination is a programmed feature of meiosis in most sexual organisms, where it ensures the proper segregation of chromosomes. Because the frequency of recombination is approximately proportional to the physical distance between markers, it provides the basis for genetic mapping. Recombination also serves as a mechanism to repair some types of potentially lethal damage to chromosomes.

Genetic recombination is often used as a general term that includes many types of DNA rearrangements and underlying molecular processes. Recombination in meiosis is reciprocal, because each participating chromosome receives information comparable to what it donates to the other partner, since all the information on both sides of the effective break has been exchanged.
5.2.1 Types of Recombination

Scientists have observed the following types of recombination in nature:

- **Homologous (General) Recombination:** As the name implies, this type occurs between DNA molecules of similar sequences. Our cells carry out general recombination during meiosis.

- **Non-Homologous (Illegitimate) Recombination:** Again, the name is self-explanatory. This type occurs between DNA molecules that are not necessarily similar. Often, there will be a degree of similarity between the sequences, but it is not as obvious as it would be in homologous recombination.

Techniques have been devised for the artificial transfer of DNA fragments from any source into cells of many different species, thus conferring new properties upon them. In bacteria and the yeast (Saccharomyces cerevisiae or S. cerevisiae), integration of such DNA into the genome (on which the stability of transformation generally depends) requires substantial sequence similarity between incoming DNA and the recipient site. However, cells of other fungi, higher plants, and animals are able to integrate foreign DNA into their chromosomes with little or no sequence similarity. These organisms appear to have some unidentified system that recombines the free ends of DNA fragments into chromosomes regardless of their sequences. It may have something in common with the mechanism, equally obscure, whereby broken ends of chromosomes can heal by nonspecific mutual joining.

- **Site-Specific Recombination:** This is observed between particular, very short, sequences, usually containing similarities. Bacteriophages, plasmids, bacteria, and unicellular eukaryotes provide many examples of differentiation through controlled and site-specific recombination of DNA segments. Invertebrates, a controlled series of deletions leads to the generation of the great diversity of gene sequences encoding the antibodies and T-cell receptors necessary for immune defense against pathogens. All these processes depend upon interaction and recombination between specific DNA sequences, generally but not always with some sequence similarity, catalyzed by site-specific recombinase enzymes. The molecular mechanisms may have some similarities with those responsible for general meiotic recombination, except that the latter does not depend on any specific sequence, only on similarity (homology) of the sequence recombined.

- **Mitotic Recombination:** This does not actually happen during mitosis, but occurs during inter-phase. The process is similar to that in meiotic recombination, and has its possible advantages. But it is usually harmful and can result in tumors. This type of recombination is increased when cells are exposed to radiation. Crossing-over between homologous chromosome pairs can also occur during the prophase of mitotic nuclear division. The
frequency is very much lower than in meiosis, presumably because the mitotic cell does not form the synaptic apparatus for efficient pairing of homologs. Mitotic crossing-over has been studied in the fruit fly *Drosophila melanogaster*, in the filamentous fungus *Aspergillus nidulans*, and in *Saccharomyces* yeast. In these species it is detected through the formation of homozygous clones of cells in an initially heterozygous diploid. There is a 50% chance of homozygosity in daughter cells whenever a cross-over occurs between chromatids in the interval between the marker and the centromere, in the chromosonal site attachment to the mitotic spindle. The frequency of mitotic crossing-over is greatly increased by radiation.

- **Conservative Recombination:** In conservative recombination events, the number of copies of the interacting chromosomes or DNA sequences is maintained throughout the process.

- **Non-Conservative Recombination:** In non-conservative events, two original copies are reduced to one in the product. This distinction of conservative and non-conservative events can be made for both homologous and non-homologous recombination.

- **Ectopic Recombination:** Essentially analogous events may take place between homologous sequences that are present at different locations on non-homologous chromosomes is often called ectopic recombination.

The prokaryotic cells can undergo recombination through one of these three processes:

- **Conjugation:** Conjugation is where genes are donated from one organism to another after they have been in contact. At any point, the contact is lost and the genes that were donated to the recipient replace their equivalents in its chromosome. What the offspring ends up having is a mix of traits from different strains of bacteria.

- **Transformation:** This is where the organism acquires new genes by taking up naked DNA from its surroundings. The source of the free DNA is another bacterium that has died, and therefore, its DNA was released to the environment.

- **Transduction:** Transduction is gene transfer that is mediated by viruses. Viruses called **bacteriophages** attack bacteria and carry the genes from one bacterium to another.

### Check Your Progress

1. Define the term genetic recombination.
2. What does genetic recombination involves?
3. When does genetic recombination occur?
5.3 HOMOLOGOUS RECOMBINATION

Homologous recombination can occur between homologous chromosomes or sister chromatids in meiotic cell and mitotic cells as well. The homologous recombination is also called general recombination. Smith (1989) reviewed the homologous recombination in prokaryotes. General recombination in *Escherichia coli* is guided by base pairing interactions between the complementary strands of two homologous DNA molecules. Double helix of two DNA molecules breaks and the two broken ends join to their opposite partners to reunite to form double helix. The site of exchange can occur anywhere in the homologous nucleotide sequence where a strand of one DNA molecule becomes base paired to the second strand to yield heteroduplex just between two double helices.

In the heteroduplex no nucleotide sequences are changed at the site of exchange due to cleavage and rejoining events. However, heteroduplex joints can have a small number of mismatched base pairs. In bacteria and viruses general recombination is carried out by the products of Rec genes, such as RecA Protein. The RecA Protein is very important for DNA repair, therefore, it is RecA dependent recombination (Refer Figure 5.4).

**Fig. 5.4 Formation of Heteroduplex Joint during Homologous Recombination**

5.3.1 Models for Homologous Recombination

I. Holliday Model for Homologous Recombination

Holliday (1974) presented a model to show the homologous or general recombination. According to this model recombination occurs in seven steps, such
Recombination and its Types

as strand breakage, strand pairing, strand invasion/assimilation, branch pairing, chiasma (crossing-over) formation, breakage and reunion, and mismatch repair (Refer Figure 5.5).

Figure 5.5 (A) illustrates the Holliday model for homologous or general recombination while the Figure 5.5 (B) illustrates the Holliday model for reciprocal general recombination.

Fig. 5.5 (A) Holliday Model for Homologous or General Recombination
Recombination and its Types

NOTES

Fig. 5.5 (B) Holliday Model for Reciprocal General Recombination

1. Strand Breakage: General recombination occurs through crossing over by pairing between the complementary single strands of DNA duplex (A). Two homologous regions of DNA double helix undergo an exchange reaction. The homologous region contains a long sequence of complementary base pairing between a strand from one or two original double helices and a complementary strand from the other. However, it is unknown how the homologous region of DNA recognizes each other. The RecBCD proteins of RecBCD or RecJ genes are required for recombination in *Escherichia coli*. This protein enters the DNA from one end of double helix, and travels along the DNA at double helix, at the rate of about 300 nucleotides per second. It creates a loop of ssDNA along travelling DNA (B). It uses energy derived from hydrolysis of ATP molecules. A special recognition site (a) sequence of eight nucleotides scattered throughout *Escherichia coli* chromosome (b) is nicked in the travelling loop of DNA formed by RecBCD protein.

2. Strand Pairing: The RecBCD proteins act as DNA helicase because these hydrolyze ATP and travel along DNA helix. Thus, the RecBCD proteins result in formation of single stranded whisker at the recognition site which is displaced from the helix (c). This initiates a base pairing interaction between the two complementary sequences of DNA double helix.
3. **Strand Invasion/Assimilation:** A single strand (whisker) generated from one DNA double helix invades the other double helix (d). In *Escherichia coli*, RecA gene produces RecA protein which is important for recombination between the chromosomes like Single Strand Binding (SSB) Proteins. The RecA protein binds firmly to single stranded DNA to form a nucleoprotein filament. Roca and Cox (1990) have reviewed the structure and function of RecA protein. RecA protein promotes rapid renaturation of complementary ssDNA hydrolyzing ATP in the process. RecA protein has several binding sites, therefore, it can bind to ssDNA and subsequently a dsDNA. Basically, the RecA protein binds first to ssDNA, then search for homology between the donor strand and the recipient molecule. Due to the presence of these sites RecA protein catalyses a multistep reaction (called synopsis) between the homologous region of ssDNA and a DNA double helix. *Escherichia coli* SSB protein helps the Rec protein to carry out these reactions. When a region of homology is identified by an initial base pairing between the complementary sequences, the crucial step in synopsis occurs.

The in vivo experiments have shown that several types of complexes are formed between ssDNA covered with RecA protein and a dsDNA helix. First a non-base paired complex is formed which is converted into a three stranded structure (ssDNA, dsDNA and RecA Protein) when a homologous region is found. This complex is unstable and spins out a DNA heteroduplex plus a displaced ssDNA from the original helix. Once the homologous regions are encountered and the ssDNA and dsDNA are complexed, a stable D-Loop is formed (d).

4. **Branch Migration:** The next step is the assimilation of strand and nick ligation (e). The donor strand gradually displaces the recipient strand which is called branch migration. After formation of synopsis, the heteroduplex region is enlarged through protein-directed branch migration catalysed by RecA protein. RecA protein directed branch migration proceeds at a uniform rate in one direction due to addition of more RecA protein to one end of RecA protein filament on the ssDNA. Branch migration can take place at any point where two single strands with the sequence make attempts to pair with the same complementary strand. An unpaired region of other single strand results in movement of branch point without changing total number of DNA base pairs. Special DNA helicases that catalyze protein directed branch migration are involved in recombination. In contrast, the spontaneous branch migration proceeds in both the directions almost at the same rate. Therefore, it makes a little progress over a long distance.

5. **Chiasma or Crossing Over Formation:** Exchange of a single strand between two double helices is a different step in a general recombination event. After the initial cross strand exchange, further strand exchanges between the two closely opposed helices is thought to proceed rapidly. A
nuclease cleaves and partly degrades the D-Loop at some points. At this stage possibly different organisms follow different pathways.

However, in most of the cases an important structure called cross-strand exchange (also called Holliday Juncture or chi form or chiasmas, is formed by the two participating DNA helices (g). A chi form of single stranded connections in the cross over region has also been observed under the electron microscope by Dressier and Potter (1982). The chi form of two homologous helices that initially paired and held together by mutual exchange of two of the four strands where one strand originates from each of the helices (g). The chi form has two important properties, i.e., as follows:

- The point of exchange can migrate rapidly back and forth along the helices by a double branch migration.
- It contains two pairs of strands, one pair of crossing strands and the other pair of non-crossing strands.

6. Breakage and Reunion: The chi structure can isomerizes several rotations (h). This results in alteration of two original non-crossing strands into the crossing strands, and the crossing strands into the non-crossing strands. In order to regenerate two separate DNA helices, breakage and reunion in two crossing strands are required. If breakage and reunion occur before isomerization the two crossing strands would not occur. Therefore, isomerization is required for the breakage and reunion of two homologous DNA double helices resulting from general genetic recombination.

Breakage and reunion occur either in the vertical or horizontal plane. If breakage occurs horizontally the recombinants would contain genotype AB/ab with a little change in base sequences at the inner region (i). However, if breakage occurs vertically the recombinants would contain Ab/aB (j). The RurC protein and RecG protein expressed from RuvC and RecG genes, respectively, are thought to be alternative endonucleases specific for Holliday structure.

7. Mismatch Repair (Mismatch Proof Read-ing System): It is such a repair system which corrects mismatched base pairs of unpaired regions after recombination. This system recognizes mis-matched function of DNA polymerase. The mecha-nism involves the excision of one of the other mismatched bases along with about 3,000 nucleotides. This RecFJO is involved in the repair of short mismatch either in the initial stage or at the end of recombination. The two proteins MutS and MutL are present in Bacteria and Eukaryotes. The MutS protein binds to mismatched base pair, whereas MutL scan the DNA for a nick.

When a nick is formed MutL triggers the degradation of the nicked strand all the way back through the mismatch, because the nicks are largely confined to the newly replicated strands in eukaryotes, replication errors are selectively removed.
In bacteria the mechanism is the same except that an additional protein MutH nicks the un-methylated GATC sequences and begins the process (Refer Figure 5.6).

**Fig. 5.6 Mechanism of Removal of Error in Newly Made Strand by Mismatch Repair System**

It has been demonstrated in yeast and bacteria that the same mismatch repair system which removes replication errors also interrupts the genetic recombination events between imperfectly matched DNA sequences. It is known that homologous genes in two closely related bacteria (Escherichia coli and Salmonella typhimurium or S. typhimurium) generally will not recombine, even after having 80% identical nucleotide sequences. However, when mismatch repair system is inactivated by mutation, the frequency of such interspecies recombination increases by 100-folds. This mechanism protects the bacterial genome from sequence changes that would be caused by recombination with foreign DNA molecules entering in the cell.

**II. Two Holliday Junctions for Homologous Recombination**

The double stranded breaks in both strands of one DNA molecules occur quite frequently. During DNA repair of double stranded breaks, homologous recombination’s occurs. This type of genetic recombination is called double stranded break repair mechanism. These types of breaks may be caused by ionizing radiations and various damaging agents. Degeneration of broken strands and generation of single strand tails (ss tails) with 3’ ends (Refer Figure 5.7).
Enzymes RecBCD further degrades the broken DNA strands. It generates single stranded tails with 3’2 ends. One 3’2 end tail invades the other unbroken homologous DNA molecule, displacing one of the strands. The invading 3’2 end serves as a primer for new DNA synthesis. The displaced strand serves as a template to fill the gap left in the first DNA. If different alleles are present at the site of break, they are permanently lost as regeneration involves homologous DNA as template which has different alleles. This process is called gene conversion because genes of the broken strand are replaced by genes of the homologous DNA. In double strand break repair, two Holliday junctions are created. These Holliday junctions move laterally by branch migration and the cleavage resolves and separates the two DNA molecules forming a crossover and a non-crossover structure. This is similar to the resolution in the single Holliday junction.

5.3.2 Molecular Mechanism of Homologous Recombination

The most important features of organisms are to adapt in the environment and to maintain their DNA sequence in the cells generation to generations with very little alterations. In long term survival of organisms depends on genetic variations, a key feature through which the organism can adapt to an environment which changes with time. This variability among the organisms occurs through the ability of DNA to undergo genetic rearrangements resulting in a little change in gene combination. Rearrangement of DNA occurs through genetic recombination.
Thus, recombination is the process of formation of new recombinant chromosome by combining the genetic material from two organisms. The new recombinants show changes in phenotypic characters. Most of the eukaryotes show a complete sexual life cycle including meiosis, an important event that generates new allelic combinations by recombination. It is made possible through chromosomal exchange resulting from crossing over between the two homologous chromosomes containing identical gene sequences. Much work was done on eukaryotic genetics until 1945 that laid the foundation of classical genetics. The work on bacterial genetics was done between 1945 and 1965 that advanced the understanding of microbial genetics at molecular level. Basically, there are three theories, viz., breakage and reunion, breakage and copying and complete copy choice that explain the mechanism of recombination (Refer Figure 5.8).

![Fig. 5.8 Schematic Presentation of the Three Possible Mechanisms of Recombination](image)

- **Breakage and Reunion:** Two homologous duplex of chromosome laying in paired form breaks between the gene loci 'a' and 'b', and 'a+' and 'b+' (A). The broken segments rejoin crosswise and yield recombinants containing 'a' and 'b' segment, and 'a+' and 'b' segment. This type of recombination does not require the synthesis of new DNA. This concept has been used to explain genetic recombination.

- **Breakage and Copying:** One helix of paired homologous chromosome, 'ab' and 'a' 'b' breaks between 'a' and 'b' (B). Segment 'b' is replaced by a newly synthesized segment copied from 'b+' and attached to 'a' section. Thus the recombinants contain and 'ab+' and 'a' 'b'.

- **Complete Copy Choice:** In, 1931, Belling proposed this theory for recombination of chromosome in higher animals. However, it has been questioned by several workers. Therefore, it has only historical importance. According to this theory a portion of one parental strand of homologous chromosome acts as template for the synthesis of a copy of its DNA molecule.
NOTES

5.3.3 Enzymes of Homologous Recombination

There are various proteins that catalyze various steps in the process of homologous recombination in *Escherichia coli*. Enzymes RecBCD load onto one end of DNA of double stranded break and move along DNA (Rec-recombination). In the process it unwinds DNA (helical activity) and degrades one or both DNA strands (nuclease activity). RecBCD is an endonuclease enzyme. It is encoded by three genes, RecB, RecC and RecD. RecBCD continues its degrading activity until it reaches a chi site (%). At this point activities of RecBCD are stopped. The chi site has eight nucleotides (5' GCTGGTGG3' ) which promote recombination.

The single strand DNA tails generated by BCD enzymes are coated by RecA enzyme. RecA stimulates pairing or synopsis between two homologous DNA molecules. RecA also promotes strand invasion, displacing one strand of unbroken DNA molecule and forming D-loop. The displaced strand invades the broken DNA molecule. The missing portions of DNA strands are synthesized using homologous strand as template and gaps are sealed by ligase enzyme.

RuvAB enzymes recognize and bind to Holliday junction and perform branch migration. RuvC enzyme cuts DNA strands at Holliday junction and causes separation and resolution of Holliday junction. Different biological processes like replication, recombination and repair occur in a co-ordinated manner. In this way new DNA can be synthesized, damaged DNA repaired and genetic recombination takes place. Nucleotide sequences can be replaced through heteroduplexes and gene conversion.

5.3.4 Role of RecA Protein in Homologous Genetic Recombination

In the various homologous genetic recombination models, the central features are similar in all recombination models. These include:

- Breaks or nicks in DNA molecules.
- Alignment or pairing or synapses of homologous sequences of two different DNA molecules.
- Formation of a crossover structure or Holliday junction in which DNA strand from each molecule creates short regions of heteroduplex DNA.
- Extension of heteroduplex DNA, which is called branched migration.
- Lastly, resolution of crossover junction to yield end products.

This is an extremely complex process involving the action of several different enzymes. The first event of creating breaks or nicks in DNA strands and the last event of resolution are undertaken by various enzymes like helicase, nuclease, and
Recombination and its Types

NOTES

Self-Instructional Material 109

Recombination

ligases. But the event starting from pairing of DNA molecules, formation of Holliday junction branch migration are the central features in recombination process. These events are undertaken by a special protein called RecA protein.

- RecA protein is involved in pairing, exchange of strands and branch migration. It is also known as strand exchange protein.
- RecA protein plays a major role in homologous recombination. It is a special protein a completely distinct class of enzymes.
- RecA protein binds quickly to single stranded DNA along the phosphate backbone of DNA helix.
- DNA is completely covered by RecA protein. Alongside RecA, a second protein called Single Strand Binding Protein (SSBP) is also involved.

Each RecA molecule has 352 amino acids. There is one RecA monomer every 3-4 nucleotides, of DNA. The ssDNA in duplex is aligned with homologous sequence of the other DNA molecule. Several steps occur in this process. Two types of homologous interactions occur. The first is the formation of paranemic joints in aligned homologous strands. The end second interaction involves formation of plactonemic joints. RecA protein is a DNA dependent ATPase. ATP hydrolysis is required for branch migration, in which strands are replaced and strand exchange occurs. It exhibits polarity as branch migration proceeds in 52 to 32 direction only.

5.3.5 Homologous Recombination in Prokaryotes and Eukaryotes

Homologous recombination in Prokaryotes and Eukaryotes is described below.

Homologous Recombination in Prokaryotes

Mechanics of molecular level of exchange is studied in detail in bacteria and phages. At present we will restrict our discussion to the recombination mechanism where DNA strands recognize each other by complementary strands bounded by base pairs. In bacteria, genetic recombination occurs during conjugation, transformation, transduction post-replication repair, during repair of double strand breaks in DNA, integration of phage DNA with chromosomal DNA and transposons etc. Homologous recombination can lead to gene conversion. Bacteria are haploid, therefore do not undergo meiosis. They possess only one double stranded DNA molecule or chromosome. There are several types of genetic recombination in microorganisms. The most common recombination is the reciprocal exchange between homologous DNA sequences. During genetic recombination usually only a part of the genetic material of a donor cell is transferred to a recipient cell. The DNA of the recipient cell and the donor pair with each other and reciprocally exchange DNA strands by crossing over. This gives rise to a new genetic constitution of the recipient cell with new characters. Subsequent daughter cells that contain only recombined chromosome.

Homologous Recombination in Prokaryotes

Mechanics of molecular level of exchange is studied in detail in bacteria and phages. At present we will restrict our discussion to the recombination mechanism where DNA strands recognize each other by complementary strands bonded by base pairs. In bacteria, genetic recombination occurs during conjugation, transformation, transduction post-replication repair, during repair of double strand breaks in DNA, integration of phage DNA with chromosomal DNA and transposons etc. Homologous recombination can lead to gene conversion. Bacteria are haploid, therefore do not undergo meiosis. They possess only one double stranded DNA molecule or chromosome. There are several types of genetic recombination in microorganisms. The most common recombination is the reciprocal exchange between homologous DNA sequences. During genetic recombination usually only a part of the genetic material of a donor cell is transferred to a recipient cell. The DNA of the recipient cell and the donor pair with each other and reciprocally exchange DNA strands by crossing over. This gives rise to a new genetic constitution of the recipient cell with new characters. Subsequent daughter cells that contain only recombined chromosome.
The exchange of DNA strand is an important event in post-replication repair and very prominent in *Escherichia coli*. The strand exchange plays a very prominent role in repair of DNA damage. As the advancing replication fork comes across a lesion or damaged site such as thymine dimers, it is bypassed during replication process. The damaged protein may be cleaved which may prove to be lethal. Repair of this lesion requires conversion of this DNA into double stranded DNA and this is achieved by RecA protein. RecA protein plays its role in retrieving a portion of the complementary strand from other side of the replication fork to fill the gap. This involves branch migration by RecA protein. This proves that branch migration is essential activity of the cell.

**Homologous Recombination in Eukaryotes**

Recombination of DNA takes place by mutation, exchange of DNA strands and incorporation of DNA. In this process the genetic information is rearranged between chromosomes that possess similar sequences. Homologous genetic recombination occurs in eukaryotes at the time of gamete formation during long Prophase I of Meiosis.

Each chromosome has two sister chromatids, each of which contains a duplex DNA. The homologous chromosomes (one maternal and the other paternal) pair with each other, pairing is known as synapsis and involves entire length of homologous chromosomes. The recombination occurs by crossing-over. It involves reciprocal exchange of chromosomal segments between non-sister chromatids of a homologous pair involving breakage and subsequent reunion in a new arrangement. Chiasma is formed at the site of crossing over. Enzymes like helicases, endonucleases and ligases are involved.

The genetic recombination causes re-arrangement of genes producing altogether new genotypes and phenotypes. These cause variations which lead to evolution. In humans about 30 homologous recombination events occurs during meiosis. The recombination events are much more in bacteria and even more in fungi. The studies of Meiosis in Lily plants by Herbert Stern and Yasuo Hotta have provided clinching evidence of recombination. Meiocytes of Lily flower buds divide synchronously. It has also been discovered that Endonuclease, DNA Polymerase, Ligase and other repair enzymes are present in Early Prophase.

**5.3.6 In Vitro Homologous Recombination**

The DNA shuffling is the first reported method of in vitro recombination. Here, DNase I is first used to randomly fragment the parental genes. Then, provided there is sufficient sequence similarity between overlapping regions of DNA fragments from different parents, the fragments can anneal to one another and reassemble into a full length gene using PCR. Depending on the total gene size, the degree of sequence similarity between parents, and desired number of crossovers (regions
where different parental fragments are assembled), fragment sizes isolated from DNAse I treatment for reassembly can vary from as low as 10–50 bp to greater than 1 kb. Random point mutation tend to occur at low rates during recombination even with a high-fidelity polymerase, and researchers will often intentionally employ error-prone PCR during PCR-based gene recombination to further diversify their library. In a noteworthy, early example of DNA shuffling, a family of twenty Human Interferon-a genes were shuffled followed by selection of antiviral and anti-proliferation activities in murine cells, resulting in variants having 26 000-fold increased activities. The best chimeras were composed of up to five parental genes and contained no random point mutations. In another example, shuffling of 26 homologous proteases genes generated many chimeric proteases that were significantly improved over any of the parental enzymes for four different properties assayed.

In vitro recombination methods are also often used in directed evolution, even when the only genetic diversity is introduced by random mutagenesis of a single parent gene. Here, one or more rounds of mutagenesis and screening to isolate improved variants results in a handful of mutant genes, each carrying a different set of point mutations. By shuffling these highly identical mutant DNA sequences, one can readily obtain a library containing all combinations of point mutations. Beneficial mutations can be combined and may show additive effects, while any potentially deleterious mutations that have accumulated will be eliminated by “Back-Crossing” with the wild-type sequence.

Non-Reciprocal Recombination

It is also known as gene conversion. The fundamental law of genetics is that the two partners contribute the equal amount of genes to the offspring’s. It means that the offspring’s inherit the half complete set of genes from the male and half from the female. One Diploid Cell undergoes Meiosis producing Four Haploid Cells; therefore, the number of genes contributed by male gets halved and so the genes of female. In higher animals like man it is not possible to analyse these genes taking a single cell. However, in certain organisms such as fungi it is possible to recover and analyse all the four daughter cells produced from a single cell through Meiosis. Occasionally, three copies of maternal allele and only one copy of paternal allele is formed by Meiosis. This indicates that one of two copies of parental alleles has been altered to the maternal allele. This gene alteration is of non-reciprocal type and is called gene conversion. Gene conversion is thought to be an important event in the evolution of certain genes and occurs as a result of the mechanism of general recombination and DNA repair (Refer Figure 5.9).
Recombination and its Types

Fig. 5.9 Ideal Model of Non-Reciprocal Recombination

The above Figure 5.9 shows the non-reciprocal general recombination, it starts when a nick is made in one of the strands, in which the following are represented:

Step (a): From this point DNA polymerase synthesizes an extra copy of a strand and displaces the original copy as a single strand.

Step (b): This single strand starts pairing with the homologous region as in lower duplex of DNA molecule in (b). The short unpaired strand produced in Step (b) is degraded when the transfer of nucleotide sequence is completed.

Step (c): The results are observed (in the next cycle) when DNA replication has separated the two non-matching strands.

Check Your Progress

4. When does homologous recombination occur?
5. Is nucleotide sequences changed at the site of exchange in heteroduplex? Why?
6. What is strand pairing?

5.4 SITE-SPECIFIC RECOMBINATION

Site-specific recombination events are mediated by sequence-specific recombination enzymes often encoded by viruses or transposable elements. The site specific recombination alters the relative position of nucleotide sequences in chromosome. The base pairing reaction depends on protein mediated recognition
of the two DNA sequences that will combine. Very long homologous sequence is not required. Unlike general recombination, site specific recombination is guided by a recombination enzyme that recognizes specific nucleotide sequences present on one of both recombining DNA molecules. Base pairing is not involved, however, if occurs the heteroduplex joint is only a few base pair long (Refer Figure 5.10).

It was first discovered in Phage $\lambda$ (Lambda) by which its genome moves into and out of the *Escherichia coli* chromosome. After penetration phage encoded an enzyme, lambda or $\lambda$ integrase which catalyses the recombination process. Lambda integrase binds to a specific attachment site of DNA sequence on each chromosome. It makes cuts and breaks a short homologous DNA sequences. The integrase switches the partner strands and rejoins them to form a heteroduplex joint of 7 bp long. The integrase resembles a DNA topoisomerase in rejoining the strands which have previously been broken.

Fig. 5.10 Schematic Representation of Site-Specific Recombination
5.4.1 Types of Site-Specific Recombination

Following are the types of site-specific recombination:

- **Conservative Site-Specific Recombination**: Production of a very short heteroduplex by requiring some DNA sequence that is the same on the two DNA molecules is known as conservative site-specific recombination.

- **Transpositional Site-Specific (TSS) Recombination**: There is another type of recombination system known as Transpositional Site-Specific (TSS) recombination. The TSS recombination does not produce heteroduplex and requires no specific sequences on the largest DNA.

5.4.2 Molecular Mechanism of Site-Specific Recombination

There are several mobile DNA sequences including many viruses and transposable elements that encode integrates. The enzyme integrates by involving a mechanism different from ‘Phage λ’ insert its DNA into a chromosome. Each enzyme of integrates recognizes a specific DNA sequence like ‘Phage λ’. K. Mizuuchi (1992) reviewed the mechanism of transpositional recombination based on the studies of Bacteriophage Mu and the other elements. The enzyme integrase was first purified from Mu. Similar to integrase of phage λ, the Mu integrase also carries out of its cutting and rejoining reactions without requirement of ATP. Also they do not require a specific DNA sequence in the target chromosome and do not form a joint of heteroduplex (Refer Figure 5.11).

During steps of TSS recombinational events the integrase makes a cut in one strand at each end of the viral DNA sequences, and exposes the 32'-OH group that protrudes out. Therefore, each of these 32'-OH ends directly invades a phosphodiester bond on opposite strands of a randomly selected site on a target chromosome. This facilitates to insert the viral DNA sequence into the target chromosome, leaving two short single stranded gaps on each side of recombinational DNA molecule. These gaps are filled in later on by DNA repair process (i.e., DNA polymerase) to complete the recombination process. This mechanism results in the formation of short duplication of the adjacent target DNA sequences (short repeats of 3 to 12 nucleotide long). Formation of short repeats is the hallmark of a TSS recombination.
5.4.3 Biological Roles of Site-Specific Recombination

The cells and viruses use conservative site-specific recombination for a wide variety of biological functions. Many phages insert their DNA into the host chromosome during infection using this recombination mechanism. In other cases, site-specific recombination is used to alter gene expression. For example, inversion of a DNA segment can allow two alternative genes to be expressed. Site-specific recombination is also widely used to help maintain the structural integrity of circular DNA molecules during cycles of DNA replication, homologous recombination, and cell division.

Fig. 5.11 Mechanism of Transpositional Site-Specific (TSS) Recombination; SDR, Short Direct Repeats of Target DNA Sequence
A comparison of site-specific recombination systems reveals some general themes. All reactions depend critically on the assembly of the recombinase protein on the DNA, and the bringing together of the two recombination sites. For some recombination reactions this assembly is very simple, requiring only the recombinase and its DNA recognition sequences as just described for Cre. In contrast, other reactions require accessory proteins. These accessory proteins include so-called architectural proteins that bind specific DNA sequences and bend the DNA. They organize DNA into a specific shape and thereby stimulate the recombination. Architectural proteins can also control the direction of a recombination reaction, for example, to ensure that integration of a DNA segment occurs while preventing the reverse reaction. This type of regulation is essential for a logical biological outcome. As can be seen from the examples discussed above, the same mechanism of DNA recombination can be utilized in different biological contexts to bring about integration, excision (deletion) and inversion of DNA segments. In principle, then, one should be able to adapt site-specific recombination systems to direct one or more of these types of DNA rearrangements in selected regions of a genome of interest. This expectation has been fully validated. Site-specific recombination has been utilized in promoting genetic alterations for answering fundamental questions in biology and for developing biotechnological tools.

Some of these functions are as follows:

- Lambda phage integrase promotes the integration and excision of a viral genome into the host cell chromosome.
- Tracking cell lineage during development.
- Ablating a gene function during development.
- Inducing the expression of a gene at a specific time in development.
- Site-specific recombination in biotechnological applications.
- Removal of integrated harmful viral sequences would be a potential beneficial application of site-specific recombination.
- Making the genetic screen more robust.
- Structure based mutagenesis.
- Changing the target specificity of Cre.
- Changing the target specificity of Flp.

Since Cre and Flp have extremely simple reaction requirements, these recombination systems have been reconstituted in a variety of organisms, such as Bacteria, Fungi, Plants, Nematodes, Flies and Animals. Cre and Flp can be placed under regulatable promoters for conditional or tissue-specific expression. An important first step in applying site-specific recombination in a genetic context of interest is the introduction of the target site or sites at the desired locale(s). Once this has been accomplished, the rest of the experimental steps are quite straightforward. To facilitate the rapid screen of desired variants, bacterial cells containing the library of variants were screened by fluorescence-based cell sorting.
Check Your Progress

7. How are site-specific recombination events mediated?
8. On what does base pair reaction depend?
9. Give the function of lambda integrase.
10. What happens in TSS recombinational events? Give its result.

5.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The Genetic recombination (also known as genetic reshuffling) is the exchange of genetic material between different organisms which leads to production of offspring with combinations of traits that differ from those found in either parent.

2. The genetic recombination involves the exchange or sharing of genetic information between multiple parental sequences to create new 'progeny' genes.

3. Genetic recombination occurs when genetic material is exchanged between two different chromosomes or between different regions within the same chromosome.

4. Homologous recombination can occur between homologous chromosomes or sister chromatids in meiotic cell and mitotic cells as well. The homologous recombination is also called general recombination.

5. In the heteroduplex no nucleotide sequences are changed at the site of exchange due to cleavage and rejoining events.

6. The RecBCD proteins act as DNA helicase because these hydrolyze ATP and travel along DNA helix. Thus, the RecBCD proteins result in formation of single stranded whisker at the recognition site which is displaced from the helix. This initiates a base pairing interaction between the two complementary sequences of DNA double helix.

7. Site-specific recombination events are mediated by sequence-specific recombination enzymes often encoded by viruses or transposable elements.

8. The base pairing reaction depends on protein mediated recognition of the two DNA sequences that will combine.

9. Lambda integrase binds to a specific attachment site of DNA sequence on each chromosome. It makes cuts and breaks a short homologous DNA sequences.

10. During steps of TSS recombinational events the integrase makes a cut in one strand at each end of the viral DNA sequences, and exposes the 3'...
OH group that protrudes out. Therefore, each of these 32’-OH ends directly invades a phosphodiester bond on opposite strands of a randomly selected site on a target chromosome. This facilitates to insert the viral DNA sequence into the target chromosome, leaving two short single stranded gaps on each side of recombinational DNA molecule.

5.6 SUMMARY

- Genetic recombination refers to the process of recombining genes to produce new gene combinations that differ from those of either parent.
- Genetic recombination produces genetic variation in organisms that reproduce sexually.
- Genetic recombination happens as a result of the separation of genes that occurs during gamete formation in meiosis, the random uniting of these genes at fertilization, and the transfer of genes that takes place between chromosome pairs in a process known as crossing-over.
- Crossing over allows alleles on DNA molecules to change positions from one homologous chromosome segment to another.
- Genetic recombination is responsible for genetic diversity in a species or population.
- Chromosomes are located within the nucleus of our cells and are formed from chromatin (mass of genetic material consisting of DNA that is tightly coiled around proteins called histones).
- A chromosome is typically single-stranded and consists of a centromere region that connects a long arm region (q arm) with a short arm region (p arm).
- When a cell enters the cell cycle, its chromosomes duplicate via DNA replication in preparation for cell division.
- Each duplicated chromosome is comprised of two identical chromosomes called sister chromatids that are connected to the centromere region.
- During cell division, chromosomes form paired sets consisting of one chromosome from each parent. These chromosomes, known as homologous chromosomes, are similar in length, gene position, and centromere location.
- Genetic recombination that involves crossing over occurs during prophase I of meiosis in sex cell production.
- The duplicated pairs of chromosomes (sister chromatids) donated from each parent line up closely together forming what is called a tetrad. A tetrad is composed of four chromatids.
- As the two sister chromatids are aligned in close proximity to one another, one chromatid from the maternal chromosome can cross positions with a
Recombination and its Types

- Crossing over occurs when the chiasma breaks and the broken chromosome segments get switched onto homologous chromosomes.
- The broken chromosome segment from the maternal chromosome gets joined to its homologous paternal chromosome, and vice versa.
- At the end of meiosis, each resulting haploid cell will contain one of four chromosomes. Two of the four cells will contain one recombinant chromosome.
- In eukaryotic cells (those with a defined nucleus), crossing over can also occur during mitosis.
- Somatic cells (non-sex cells) undergo mitosis to produce two distinct cells with identical genetic material.
- Any crossover that occurs between homologous chromosomes in mitosis does not produce a new combination of genes.
- Crossing over that occurs in non-homologous chromosomes can produce a type of chromosome mutation known as a translocation.
- A translocation happens when a chromosome segment detaches from one chromosome and moves to a new position on another non-homologous chromosome. This type of mutation can be dangerous as it often leads to the development of cancer cells.
- Prokaryotic cells, like bacteria which are unicellular with no nucleus, also undergo genetic recombination.
- Although bacteria most commonly reproduce by binary fission, this mode of reproduction does not produce genetic variation.
- In bacterial recombination, genes from one bacterium are incorporated into the genome of another bacterium through crossing-over.
- Bacterial recombination is accomplished by the processes of conjugation, transformation, or transduction.
- In conjugation, one bacterium connects itself to another through a protein tube structure called a pilus. Genes are transferred from one bacterium to the other through this tube.
- In transformation, bacteria take up DNA from their environment. The DNA remnants in the environment most commonly originate from dead bacterial cells.
- In transduction, bacterial DNA is exchanged through a virus that infects bacteria known as a bacteriophage. Once the foreign DNA is internalized by a bacterium via conjugation, transformation, or transduction, the bacterium can insert segments of the DNA into its own DNA. This DNA transfer is
Recombination and its Types

accomplished via crossing-over and results in the creation of a recombinant bacterial cell.

- DNA of the donor cell recombines with recipient DNA by reciprocal exchange of DNA strands. The recombining DNA molecules have homologous sequences.
- The DNA molecules align or pair with each other and undergo crossing-over and homologous recombination. The recombinant DNA has new genetic constitution.
- Bacteria can acquire DNA of other closely related bacterial species and can become transformed. This is known as transformation.
- Transformation was first demonstrated by Griffith in *Diplococcus pneumoniae* bacteria to confirm DNA as genetic material. Homologous recombination is catalysed by RecA protein.
- Often the transposable elements replicate to generate two copies. One copy remains at the original site while the other jumps to the target site.
- When Phage 1 (lambda) infects *Escherichia coli* bacterium, DNA of the phage λ is inserted into the DNA of the bacterium. In this way phage DNA becomes integrated into host DNA and becomes a part of the host chromosome.
- Bacterium containing a complete set of phage DNA is called lysogen. Phage DNA is inserted at a specific site into the host DNA.
- The attachment site of both DNAs possesses identical 15 bp sequence. A phage encoded protein called integrase catalyses the integration.
- A host protein called Integration Host Factor (IGF) is also involved. Integrase is a topoisomerase enzyme which breaks, rotates and then re-joins the strands of both molecules. In eukaryotes site-specific recombination produces antibody diversity.
- Although common, genetic recombination is a highly complex process. It involves the alignment of two homologous DNA strands (the requirement for homology suggests that this occurs through complementary base pairing, but this has not been definitively shown), precise breakage of each strand, exchange between the strands, and sealing of the resulting recombinant molecules. This process occurs with a high degree of accuracy at high frequency in both Eukaryotic and Prokaryotic cells.
- The basic steps of recombination can occur in two pathways, according to whether the initial break is single or double stranded. In the single-stranded model, following the alignment of homologous chromosomes, a break is introduced into one DNA strand on each chromosome, leaving two free ends.
Each end then crosses over and invades the other chromosome, forming a structure called a Holliday junction. The next step, called branch migration, takes place as the junction travels down the DNA. The junction is then resolved either horizontally, which produces no recombination, or vertically, which results in an exchange of DNA.

In the alternate pathway initiated by double-stranded breaks, the ends at the breakpoints are converted into single strands by the addition of 3' tails. These ends can then perform strand invasion, producing two Holliday junctions. From that point forward, resolution proceeds as in the single-stranded model.

Third model of recombination, Synthesis-Dependent Strand Annealing (SDSA), has also been proposed to account for the lack of crossover typical of recombination in mitotic cells and observed in some meiotic cells to a lesser degree.

The molecular processes they catalyze may rely on very short stretches of homology between the interacting DNAs, or they may be entirely non-homologous.

Recombination that involves very limited or no homology between the interacting DNA sequences is termed illegitimate or non-homologous recombination.

Sometimes a few matched base pairs are seen precisely at illegitimate recombination junctions, and these are called microhomologies.

An event supported by homologies of 100 bp or more would typically be classified as homologous, a match of 10 bp or fewer would be non-homologous, and there is evidently a gray area in between.

### 5.7 KEY WORDS

- **Genetic recombination**: Genetic recombination refers to the process of recombining genes to produce new gene combinations that differ from those of either parent.
- **Non-homologous recombination**: Recombination between DNA molecules of similar sequences. Our cells carry out general recombination during meiosis.
- **Homologous recombination**: Recombination between DNA molecules of similar sequences.
- **Crossing over**: Crossing over is the process of the chiasma breaks and the exchange of broken chromosome segments onto non-sister chromatids of the homologous chromosomes.
Recombination and its Types

NOTES

Conservative recombination: In conservative recombination events, the number of copies of the interacting chromosomes or DNA sequences is maintained throughout the process.

5.8 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What is recombination?
2. Give the formation of heteroduplex joint during homologous recombination with the help of diagram.
3. Explain the two Holliday junctions for homologous recombination.
4. Explain Holliday model for homologous or general recombination with the help of diagram.
5. Give the role of RecA protein in homologous genetic recombination.
6. Distinguish between homologous recombination in Prokaryotes and Eukaryotes.
7. What are the types of site-specific recombination?
8. Give the biological role of site-specific recombination.

Long Answer Questions

1. What is homologous recombination? Discuss the mechanism of homologous recombination with suitable examples.
2. Discuss about the types of recombination in detail.
3. What is genetic recombination? Write in detail the mechanism of recombination based on Holliday’s model.
4. Explain with the help of diagram McClintock and Creighton’s model showing physical evidences of recombination.
5. Give an illustrated account of microbial genetic recombination with necessary step-by-step mechanism.
6. Draw a well labelled diagram to show the mechanism of removal of error in newly made strand by Mismatch Repair system.
7. What is recombination? Discuss the types and significance of recombination.
8. Describe in detail the types and mechanism of site-specific recombination.
5.9 FURTHER READINGS

UNIT 6  CONJUGATION AND F-FACTOR

Structure
6.0 Introduction
6.1 Objectives
6.2 Conjugation
   6.2.1 Conjugation by Escherichia coli F Factor
   6.2.2 Fertility Factor or F Factor
   6.2.3 Hfr Conjugation and Chromosomal Transfer
   6.2.4 The F\(^+\) (F Prime) Factor
   6.2.5 Interrupted Mating and Conjugal Mapping
6.3 Answers to Check Your Progress Questions
6.4 Summary
6.5 Key Words
6.6 Self Assessment Questions and Exercises
6.7 Further Readings

6.0 INTRODUCTION

Conjugation is merely the fusion of two compatible bacterial cells. Bringing two genotypes together and allowing them to conjugate is the equivalent of making a cross in Eukaryotes. The conjugation process is explained on the gut bacterium Escherichia coli or E. coli. Conjugation and gene transfer in Escherichia coli are driven by a circular DNA plasmid called the Fertility Factor or Sex Factor (F), which is found in some but not all cells. Therefore to understand how to make a cross in Escherichia coli, the properties of F have to be understood.

Cells carrying the F plasmid are designated F\(^+\), and those lacking it are F\(^-\). The F plasmid contains approximately 100 genes, which give the plasmid several important properties. The F plasmid can replicate its own DNA, allowing the plasmid to be maintained in a dividing cell population. The rare cells in which the F factor is integrated into the host chromosome can be isolated from the bacterial population to cultivate pure strains derived from these cells. In these strains, every cell donates chromosomal alleles during F transfer, so the frequency of recombinants for these strains is much higher as compared to cells in the original population, where the F factor is not integrated in most cells. Consequently, strains with an integrated F factor are termed High frequency of recombination (Hfr) strains to distinguish them from normal F\(^+\) strains, which contain only a few rare Hfr cells and thus display only a low frequency of recombination for the strain as a whole. Because they transfer chromosomal markers efficiently, hence the Hfr strains are used for genetic mapping.
Fundamentally, the Bacterial conjugation is the mechanism whereby DNA is transferred from a donor to a recipient cell through a complex mechanism that is encoded by a transfer system.

In this unit, you will study about the conjugation, bacterial conjugation by *Escherichia coli*, F factor, structure of F factor, regulation of F factor fertility, establishment of cell contact, DNA mobilization, transfer and separation of mating pair, Hfr conjugation and chromosomal transfer, interrupted mating and conjugational mapping.

### 6.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss what bacterial conjugation is
- Explain conjugation by *Escherichia coli*
- Understand F factor - the structure of F factor and regulation of F factor fertility
- Analyse DNA mobilization, Hfr conjugation and chromosomal transfer

### 6.2 CONJUGATION

Conjugation is the process by which one bacterium transfers genetic material to another through direct contact. During conjugation, one bacterium serves as the donor of the genetic material, and the other serves as the recipient. The donor bacterium carries a DNA sequence called the Fertility factor or F factor. The F factor allows the donor to produce a thin, tube like structure called a pilus, which the donor uses to contact the recipient. The pilus then draws the two bacteria together, at which time the donor bacterium transfers genetic material to the recipient bacterium. Typically, the genetic material is in the form of a plasmid, or a small, circular piece of DNA. The genetic material transferred during conjugation often provides the recipient bacterium with some sort of genetic advantage. For instance, in many cases, conjugation serves to transfer plasmids that carry antibiotic resistance genes.

Bacterial cells may carry besides the main chromosome, one or more small DNA molecules in the cytoplasm called plasmids. Though there are various kinds of plasmids, but only a few are involved in conjugation and are therefore called conjugative plasmids. The Sex element or Fertility factor or F factor is the type of conjugative plasmid.

The presence of F factor in different strains has given rise to two mating types in bacteria namely, the donor which possesses the Fertility factor or F factor and referred to as $F^+$ strain, the second which lacks Fertility factor or F factor is the $F^-$ strain. The F factor is itself the genetic element which is passed from donor
to recipient cells during conjugation. Remember that there is no conjugation between two F+ strains or between two F− strains.

6.2.1 Conjugation by *Escherichia coli* F Factor

Principally, the ‘Conjugation’ is referred as the fusion of two compatible bacterial cells. Bringing two genotypes together and allowing them to conjugate is the equivalent of making a cross in Eukaryotes. The conjugation process is explained on the gut bacterium *Escherichia coli* or *E. coli*. Conjugation and gene transfer in *Escherichia coli* are driven by a circular DNA plasmid called the Fertility factor or Sex factor (F factor), which is found in some but not all cells. Therefore to understand how to make a cross in *Escherichia coli*, the properties of F factor have to be understood.

6.2.2 Fertility Factor or F Factor

The Fertility factor or the F factor was first named ‘F’ by one of its discoverers Esther Lederberg. It is also called the sex factor in *Escherichia coli* or the F sex factor, and also the F plasmid. The F factor permits genes to be transferred from one bacterium carrying the factor to another bacterium lacking the factor by the process termed as **conjugation**. The F plasmid belongs to a class of **conjugative plasmids** that control sexual functions of bacteria with a fertility inhibition, the ‘Fin’ system.

**Structure of F Factor**

The most common functional segments that constitute the F factors include the following:

- **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)**: These are composed of one copy of IS2, two copies of IS3, and one copy of IS1000, the so-called ‘Selfish Genes’, the sequence fragments which can integrate copies of themselves at different locations.
- **traA**: Pilin, the major subunit of the pilus, the F plasmid genes.

**Relation to the Genome**

The episome that harbors the F factor can exist as an independent plasmid or integrate into the bacterial cell’s genome. The following are the names for the possible states:

- **Hfr Bacteria**: It possesses the entire F episome integrated into the bacterial genome.
F+ Bacteria: It possesses F factor as a plasmid independent of the bacterial genome. The F plasmid contains only F factor DNA and no DNA from the bacterial genome.

F’ (F Prime) Bacteria: These are formed by incorrect excision from the chromosome, resulting in F plasmid carrying bacterial sequences that are next to where the F episome has been inserted.

F− Bacteria: These do not contain F factor and acts as the recipients.

F element contains about 2 percent of the cell’s total DNA. It is capable of autonomous replication and is made up of a circular, double stranded DNA molecule having the molecular weight approximately $35 \times 10^6$. It contains about 15 genes, out of which 8 control the formation of F pilis or sex pilis which are hair-like appendages extending from the surface of F+ cells. The function of F pili is conjugation.

Structure of Pili
The F pili (singular, pilus) originate from cell membrane and project outward beyond the cell wall as shown in Figure 6.1. Pili are minute proteinaceous tubules that allow the F+ cells to attach to other cells and to maintain contact with them. The width of pili in different bacteria varies between 4 and 35 nm. The pilus is made up of a Phosphate-Carbohydrate-Protein complex with a single Polypeptide subunit called ‘Pilin’ of 11,000 to 12,000 Daltons.

Each pilin subunit has 2 residues of Phosphate and one of Glucose. The pili have been isolated and analysed by electron microscopy and X-ray diffraction techniques. The F pilus consists of a hollow cylinder 80 Å in diameter. The central hollow core is about 20 Å. The pilin subunits are arranged in the form of four helical chains. Figure 6.1 illustrates the Escherichia coli cell showing Pili on the surface.

![Fig. 6.1 The Escherichia coli Cell Showing Pili on the Surface](image)
F Factor - Mating Process

In a combination of F+ and F- cells, an F+ donor cell can communicate with an F- recipient cell by the F pilus. Therefore, the pilus is essential for recognition of recipient cell with which mating can take place. Figure 6.2 illustrates the transfer of F element factor from F+ to F- cells during the conjugation process.

![Figure 6.2 Transfer of F Element Factor from F+ to F- Cells during Conjugation](image)

After initial contact between the pilus and recipient cell is established, the pilus serves as a protoplasmic connection between the two cells and is called the conjugation tube. A donor cell devoid of pili cannot conjugate. The sex plasmid passes from the donor F+ to the recipient F- cell through the conjugation tube. Transfer takes place when DNA replicates through the rolling circle method.

A nick or incision produced by an endonuclease in one strand of the plasmid DNA duplex produces a free 5' and a 3' end. The strand moves across the cytoplasmic bridge with the 5' end first, into the F- cell. The second inner strand of the plasmid DNA duplex is retained in the F+ cell and synthesizes its complementary strand. The two cells separate after mating and are known as ex-conjugants. Thus the originally mixed population of F- and F+ cells comes to have all F+ cells only.
The transfer of sex element from F+ to F– cell has one more significant feature. Not only can the plasmid exist in the cytoplasm as an autonomous entity, but it can also become incorporated in the main bacterial chromosome in a frequency of about 1 in 10,000 F+ cells.

Integration takes place at a specific site in the host chromosome which has homologous sequences. Such an integrated plasmid is known as episome and promotes the transfer of the main bacterial genophore from donor F+ to recipient F– cells during conjugation, an occurrence followed by recombination.

Cells carrying the F plasmid are termed as F+ while those lacking it are termed as F–. The F plasmid contains approximately 100 genes, which provide several significant properties to plasmid, such as:

1. The F plasmid is capable of replicating its own DNA that allows the plasmid to be maintained in a dividing cell population (Refer Figure 6.3 (a)).

2. Cells that carry the F plasmid stimulate the synthesis of pili on the bacterial cell surface. Pili allow the F+ cells to be attached to other cells and to maintain contact with them, i.e., to conjugate (Refer Figure 6.3 (b)).

3. Fundamentally, the F+ and F– cells can conjugate. When conjugation occurs, the F+ cells can act as F donors. The F plasmid DNA replicates and the newly synthesized copy of the circular F molecule is transferred to the F– recipient (Refer Figure 6.3 (c)). Though, a copy of F permanently remains behind in the donor cell. The recipient cell is converted into F+, since it contains a circular F genome. The transfer of the F plasmid from F+ to F– is fast, therefore the F plasmid is capable of spreading quickly throughout a population from strain to strain.

4. F+ cells are typically inhibited after building contact with other F+ cells, therefore the F plasmid is not transferred from F+ to F+.

**Fig. 6.3 F Plasmid Replicating its Own DNA**
5. Occasionally F within its genome carries one or more IS (Insertion Sequence) elements. An IS element is referred as a mobile segment of DNA that moves from one location to another location within the host chromosome or between the chromosome and the plasmid. The presence of a specific IS element both in the plasmid and in the chromosome can provide a site at which homologous crossing over can occur occasionally. A crossover between the two circular DNAs results in the integration of the plasmid into the bacterial chromosome, as shown in Figure 6.4.

6.2.3 Hfr Conjugation and Chromosomal Transfer

After the discovery of F– strains, a distinctive kind of strain was observed which was several hundred times more fertile in crosses with F+ than any known F+ strain. This strain was isolated by Lederberg et al. (1952) and was called Hfr strain or High frequency recombination. The Hfr strain produces about 1,000 times more prototrophs than in the F+ × F– cross.

In the mating system of Hfr strains the main bacterial chromosome containing an integrated F factor is transferred to F– cells. The Hfr bacteria arise spontaneously from F+ cells at a low frequency by integration of F factor in the main chromosome.

When Hfr cells are mixed with F– cells there is conjugation and a high frequency of transfer of only portions of the main bacterial chromosome (some selected markers) from donor to F– recipient cells. The recipient cell remains F–.

An F– cell is converted to Hfr when F integrates into the main chromosome by reciprocal recombination. The process is reversible so that an Hfr cell becomes F– when another recombinational event causes detachment of the F factor.
Hayes, Wollman and Jacob conducted experiments which demonstrated recombination during mating between Hfr and F−. They took wild type Hfr capable of synthesizing all its organic requirements, which could also utilise the sugars Galactose and Lactose, and was susceptible to being killed by Streptomycin. The second strain they took was of the F− type which could not synthesize some Amino Acids (Leucine and Threonine), nor utilise Galactose and Lactose, and was resistant to Streptomycin.

The Hfr and mutant F− strains were mixed and grown together. For analysing the progeny cells, samples of the cell mixture were grown on minimal medium containing Streptomycin. Recombinants had appeared in the progeny.

**Linear Chromosome Transfer by Hfr Strains**

Wollman and Jacob (1956) studied kinetics of genetic transfer by the interrupted mating technique. After mixing up two parental populations, cell samples are withdrawn at different intervals and then stressed in a mixer so that the mating pairs become separated and conjugation comes to an end. The mixture of cells is diluted and plated on selective medium and the number of recombinants formed during that interval are determined. The appearance of recombinants indicates formation of zygotes.

The different genetic markers appear in the progeny of interrupted matings after different time periods have been allowed before mating is disrupted by stress/pressure. Closely linked markers appear at the same time, whereas distantly placed markers appear at different times.

The markers Threonine and Leucine appear after about 8 minutes whereas Gal appears after 26 minutes. The entire chromosome containing about $5 \times 10^6$ base pairs is transferred in 90 minutes. Therefore, it is possible to map locations of markers on the donor chromosome.

Further investigations showed that the Hfr donor cells transfer only a part of their genome to F− cells. Moreover, there are different Hfr strains which are distinct from each other in transferring a different part of the genome to F− cells. The *Escherichia coli* genome is a closed circular loop.

In the Hfr donor cell the loop is broken at a point characteristic for that strain. The break occurs within the F element so that a part of it is at the leading end and is transferred to the F− cell, the other part of F is at the extreme distal end which trails behind. Transfer takes place by injection of the linear structure into the recipient cell.

The foremost or leading end carries with it gene loci nearest to it (Refer Figure 6.5) until conjugation is interrupted. The transfer of DNA may be broken off at any time due to spontaneous rupture of the connection between conjugating cells.
Fig. 6.5 Integration of F Element into Main Chromosome of F’ to Form Hfr

Figure 6.5 illustrates the integration of F element into main chromosome of F’ to form Hfr. In Figure 6.5, (d) illustrates the break at specific site and transfer from Donor Hfr cell.

After transfer to the F’ cell the donor DNA fragment becomes incorporated into a homologous region of the host cell chromosome. The corresponding segment of the F’ cell DNA is lost. Crossovers occur between the donor Hfr fragment and the F’ host cell chromosome. This integration is essential before donor genetic markers can express themselves.

The presence of F factor in Hfr and F’ cells endows specific surface properties through formation of F pilus, due to which these cells can act as donors. In 1960 Loeb found out that certain bacteriophages could lyse only donor *Escherichia coli* cells but not the recipient cells.

The phages R17 and M12 adsorb to pili present on donors but not on recipient cells and are referred to as male-specific phages. Further, the male-specific RNA phages are observed to adsorb along the length of the sex pilus, whereas male-specific DNA phages adsorb to the tip of the pilus.

The surface of the recipient cell also appears to play an important role in mating. When an F factor is present in a cell, it prevents the cell from acting as a
recipient, so that super-infection does not occur. As this effect is due to a surface component which depends on the F factor, the phenomenon is known as surface exclusion.

Some of the surface proteins coded by the main bacterial chromosome genes are also involved in mating. The con- mutants of *Escherichia coli* are not able to function as recipients and form mating pairs in conjugation. These mutants are found to be deficient in two of the surface proteins.

### 6.2.4 The F’ (F Prime) Factor

The F element can also become integrated into the main bacterial chromosome. Rarely an integrated F can undergo excision and become detached, carrying with it some bacterial genes that remain attached to it. Such an F element is called an F¢ factor.

It behaves like the F factor of F¢ cells and can be transferred to F– cells. Because F¢ carries bacterial genes, it is able to pair with the corresponding region in the bacterial chromosome (Refer Figure 6.6). A bacterium receiving an F¢ factor becomes a partial diploid for the bacterial genes carried by F¢. Figure 6.6 illustrates the genomes of *Escherichia coli* cell with F¢ element.

![Fig. 6.6 Genomes of Escherichia coli Cell with F’ Element](image)

**The Transfer Genes**

There are certain mutant strains in which transfer of F factor cannot take place. The transfer deficient mutants have been useful for identifying the presence of transfer (tra) genes in the F factor. Transfer genes are found to be necessary for conjugation. About 19 tra genes have been identified so far and these are classified into 4 groups as follows:

1. The genes of first group control pilus formation and recognition of recipient cell.
2. The genes of second group are involved in stabilisation of mating pairs.
3. The genes of the third group are essential for some metabolic changes in DNA required for conjugation.
4. The genes of the fourth group (tra J) controls the function of all the other tra genes.
6.2.5 Interrupted Mating and Conjugational Mapping

In conjugation, it was witnessed that the number of genes that are transmitted from donor to recipient was directly proportional to the time interval for which conjugation was permissible. Considering this feature, Francois Jacob and Ellie Wollman developed a specific technique called 'Interrupted Mating Technique' for mapping the bacterial chromosome. In this technique, the donor Hfr and recipient F strains are mixed and permitted to conjugate for a short time period. Then samples are removed at periodic intervals and are violently agitated to break the conjugation tube. The length of transmitted donor chromosome can then be determined and mapped in terms of time units that are required for transfer. It is identified that 8 minutes are required for conjugation to start and then chromosome is transferred slowly in terms of time units, where one time unit is equal to one minute. Therefore the complete chromosome of *Escherichia coli* is transferred in about 89 minutes and therefore the bacterial chromosome is 89 time units in length.

Circular Linkage Map

When linkage maps in bacteria were prepared using several Hfr strains, then these maps typically differ with respect to the first and the last genes of the map. Though, if the first gene transferred is distinguished and also the direction of transfer of subsequent genes, then the order is predetermined. There is a circular map which cleaves at any point due to attachment of F factor to form a linear chromosome. This linear chromosome enters the recipient cell from the end point which is far from the site of attachment of F factor. Different experimental evidences prove that Hfr certainly results due to the insertion of F into the chromosome. Circular nature of linkage group was also established through the evidences for physical circular nature of bacterial chromosome.

Linkage Information from Transformation

Transformation is accomplished through the acceptance of naked DNA that is extracted from one strain of bacteria by another strain of bacteria. While extracting this DNA, there may be some breakage or spiriting into small pieces. When two genes are closer to each other, then there is possibility of being carried on the same piece of DNA, therefore initiating double transformation. Alternatively if the genes are widely separated, then there will be possibility of being carried on separate DNA segments and the frequency of double transformation will be the product of single transformation frequencies. Therefore as per the product rule of probability for the observed frequency of double recombinants prove close linkage.

Recombination after Gene Transfer

When a chromosome segment is transferred from a donor to a recipient strain, or through transduction or transformation, then this transferred segment must be integrated into host genome through an exchange mechanism to produce a stable recombinant. The recipient cell at this stage is termed a merozygote (partial diploid).
which has the complete genome of F- termed endogenote and an incomplete genome termed exogenote derived from F+ or Hfr. An even number of crossovers, instead of a single or odd number of crossovers, permits incorporation of a part of the genome from exogenote into the endogenote, one of the two products of exchange, i.e., a linear fragment is usually lost.

**Conjugation Mapping**

Using different strains of F plasmids and the interrupted mating technique, the order of genes on the chromosome can be determined.

**Step 1.** For a strain of Hfr, use the interrupted mating technique for determining the order of the genes in the region of the chromosome near the plasmid insertion point.

**Step 2.** By obtaining the order of the genes by means of various strains and the interrupted mating technique (as defined in Step 1).

**Step 3.** By means of these orders, a map of the chromosome can be deduced.

### Check Your Progress

1. Explain the term conjugation giving example.
2. What are plasmids and conjugative plasmids?
3. What is F factor?
4. What does F element contain? How it functions?
5. How integration takes place at a specific site in the host chromosome?
6. Explain Hfr strain or High frequency recombination.
7. What is interrupted mating technique?

### 6.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Conjugation is the process by which one bacterium transfers genetic material to another through direct contact. During conjugation, one bacterium serves as the donor of the genetic material, and the other serves as the recipient. The donor bacterium carries a DNA sequence called the Fertility factor or F factor. The F factor allows the donor to produce a thin, tube-like structure called a pilus, which the donor uses to contact the recipient. Principally, the ‘Conjugation’ is referred as the fusion of two compatible bacterial cells. The conjugation process is explained on the gut bacterium *Escherichia coli* or *E. coli*. Conjugation and gene transfer in *Escherichia coli*.
Coli are driven by a circular DNA plasmid called the Fertility factor or Sex factor (F factor), which is found in some but not all cells.

2. Typically, the genetic material is in the form of a plasmid, or a small, circular piece of DNA. Bacterial cells may carry besides the main chromosome, one or more small DNA molecules in the cytoplasm called plasmids. Though there are various kinds of plasmids, but only a few are involved in conjugation and are therefore called conjugative plasmids. The Sex element or Fertility factor or F factor is the type of conjugative plasmid.

3. The Fertility factor or the F factor was first named ‘F’ by one of its discoverers Esther Lederberg. It is also called the sex factor in Escherichia coli or the F sex factor, and also the F plasmid. The F factor permits genes to be transferred from one bacterium carrying the factor to another bacterium lacking the factor by the process termed as conjugation.

4. The F element contains about 2 percent of the cell’s total DNA. It is capable of autonomous replication and is made up of a circular, double stranded DNA molecule having the molecular weight approximately $35 \times 10^6$. It contains about 15 genes, out of which 8 control the formation of F pili or sex pili which are hair-like appendages extending from the surface of F+ cells. The function of F pili is conjugation.

5. Integration takes place at a specific site in the host chromosome which has homologous sequences. Such an integrated plasmid is known as episome and promotes the transfer of the main bacterial genophore from donor F+ to recipient F- cells during conjugation, an occurrence followed by recombination.

6. After the discovery of F+ strains, a distinctive kind of strain was observed which was several hundred times more fertile in crosses with F- than any known F- strain. This strain was isolated by Lederberg et al. (1952) and was called Hfr strain or High frequency recombination. The Hfr strain produces about 1,000 times more prototrophs than in the F+ × F- cross. In the mating system of Hfr strains the main bacterial chromosome containing an integrated F factor is transferred to F- cells. The Hfr bacteria arise spontaneously from F+ cells at a low frequency by integration of F factor in the main chromosome.

7. In conjugation, it was witnessed that the number of genes that are transmitted from donor to recipient was directly proportional to the time interval for which conjugation was permissible. Considering this feature, Francois Jacob and Ellie Wollman developed a specific technique called ‘Interrupted Mating Technique’ for mapping the bacterial chromosome. In this technique, the donor Hfr and recipient F strains are mixed and permitted to conjugate for a short time period.
Conjugation and F-Factor

6.4 SUMMARY

- Conjugation is the process by which one bacterium transfers genetic material to another through direct contact.
- During conjugation, one bacterium serves as the donor of the genetic material, and the other serves as the recipient. The donor bacterium carries a DNA sequence called the Fertility factor or F factor.
- The F factor allows the donor to produce a thin, tube-like structure called a pilus, which the donor uses to contact the recipient. The pilus then draws the two bacteria together, at which time the donor bacterium transfers genetic material to the recipient bacterium.
- Typically, the genetic material is in the form of a plasmid, or a small, circular piece of DNA.
- The genetic material transferred during conjugation often provides the recipient bacterium with some sort of genetic advantage. For instance, in many cases, conjugation serves to transfer plasmids that carry antibiotic resistance genes.
- Bacterial cells may carry besides the main chromosome, one or more small DNA molecules in the cytoplasm called plasmids. Though there are various kinds of plasmids, but only a few are involved in conjugation and are therefore called conjugative plasmids. The Sex element or Fertility factor or F factor is the type of conjugative plasmid.
- Principally, the ‘Conjugation’ is referred as the fusion of two compatible bacterial cells. Bringing two genotypes together and allowing them to conjugate is the equivalent of making a cross in Eukaryotes. The conjugation process is explained on the gut bacterium Escherichia coli or E. coli.
- Conjugation and gene transfer in Escherichia coli are driven by a circular DNA plasmid called the Fertility factor or Sex factor (F factor), which is found in some but not all cells.
- The Fertility factor or the F factor was first named ‘F’ by one of its discoverers Esther Lederberg. It is also called the sex factor in Escherichia coli or the F sex factor, and also the F plasmid.
- The F factor permits genes to be transferred from one bacterium carrying the factor to another bacterium lacking the factor by the process termed as conjugation.
- 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.
Conjugation and F-Factor

NOTES

Self-Instructional Material

- tra-Region (Transfer Genes): Genes coding the F pilus and DNA transfer process.
- The F element contains about 2 percent of the cell’s total DNA. It is capable of autonomous replication and is made up of a circular, double stranded DNA molecule having the molecular weight approximately \(35 \times 10^6\). It contains about 15 genes, out of which 8 control the formation of F pili or sex pili which are hair-like appendages extending from the surface of F+ cells. The function of F pili is conjugation.
- In a combination of F+ and F– cells, an F+ donor cell can communicate with an F– recipient cell by the F pilus. Therefore, the pilus is essential for recognition of recipient cell with which mating can take place.
- Integration takes place at a specific site in the host chromosome which has homologous sequences. Such an integrated plasmid is known as episome and promotes the transfer of the main bacterial genophore from donor F+ to recipient F– cells during conjugation, an occurrence followed by recombination.
- After the discovery of F+ strains, a distinctive kind of strain was observed which was several hundred times more fertile in crosses with F– than any known F+ strain. This strain was isolated by Lederberg et al. (1952) and was called Hfr strain or High frequency recombination.
- The Hfr strain produces about 1,000 times more prototrophs than in the F+ × F– cross.
- In the mating system of Hfr strains the main bacterial chromosome containing an integrated F factor is transferred to F– cells.
- The Hfr bacteria arise spontaneously from F+ cells at a low frequency by integration of F factor in the main chromosome.
- When Hfr cells are mixed with F– cells there is conjugation and a high frequency of transfer of only portions of the main bacterial chromosome (some selected markers) from donor to F– recipient cells. The recipient cell remains F–.
- An F– cell is converted to Hfr when F integrates into the main chromosome by reciprocal recombination. The process is reversible so that an Hfr cell becomes F+ when another recombinational event causes detachment of the F factor.
- In the Hfr donor cell the loop is broken at a point characteristic for that strain. The break occurs within the F element so that a part of it is at the leading end and is transferred to the F– cell, the other part of F is at the extreme distal end which trails behind. Transfer takes place by injection of the linear structure into the recipient cell.
- The F element can also become integrated into the main bacterial chromosome. Rarely an integrated F can undergo excision and become...
detached, carrying with it some bacterial genes that remain attached to it. Such an F element is called an F\textsuperscript{+} factor.

- It behaves like the F factor of F\textsuperscript{+} cells and can be transferred to F\textsuperscript{−} cells. Because F\textsuperscript{+} carries bacterial genes, it is able to pair with the corresponding region in the bacterial chromosome.
- The transfer deficient mutants have been useful for identifying the presence of transfer (tra) genes in the F factor.
- Transfer genes are found to be necessary for conjugation. About 19 tra genes have been identified so far and these are classified into 4 groups.
- In conjugation, it was witnessed that the number of genes that are transmitted from donor to recipient was directly proportional to the time interval for which conjugation was permissible. Considering this feature, Francois Jacob and Ellie Wollman developed a specific technique called ’Interrupted Mating Technique’ for mapping the bacterial chromosome. In this technique, the donor Hfr and recipient F strains are mixed and permitted to conjugate for a short time period.

6.5 **KEY WORDS**

- **Conjugation:** This is the process by which one bacterium transfers genetic material to another through direct contact.
- **Plasmids:** Bacterial cells may carry besides the main chromosome, one or more small DNA molecules in the cytoplasm called plasmids.
- **Conjugative plasmids:** The plasmids that are involved in conjugation are called conjugative plasmids.
- **F factor:** The Fertility factor or the F factor permits genes to be transferred from one bacterium carrying the factor to another bacterium lacking the factor by the process termed as conjugation.
- **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.
- **Hfr bacteria:** It possess the entire F episome integrated into the bacterial genome.
- **F\textsuperscript{+} bacteria:** It possess F factor as a plasmid independent of the bacterial genome. The F plasmid contains only F factor DNA and no DNA from the bacterial genome.
Conjugation and F-Factor

NOTES

F’ (F Prime) bacteria: These are formed by incorrect excision from the chromosome, resulting in F plasmid carrying bacterial sequences that are next to where the F episome has been inserted.

F- bacteria: These do not contain F factor and acts as the recipients.

6.6 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions
1. What is conjugation?
2. What are plasmids?
3. Explain F factor.
4. Define the term mating pair.
5. What is Hfr bacteria?
7. How will you identifying the presence of transfer (tra) genes in the F factor?
8. What is circular linkage map?

Long Answer Questions
1. Briefly discuss about the significance and function of conjugation giving appropriate examples.
2. Discuss conjugation by Escherichia coli F factor.
3. Explain the role of plasmids in bacterial conjugation.
4. What is F factor? Explain the structure of F-factor giving appropriate examples.
5. Explain about the transfer and separation of mating pair.
6. Discuss about the Hfr conjugation and chromosomal transfer process with the help of diagrams.
7. Briefly explain the interrupted mating and conjugational mapping.

6.7 FURTHER READINGS


UNIT 7 TRANSFORMATION

7.0 INTRODUCTION

Transformation is one of the methods of horizontal gene transfer between bacteria. It is a process in which a recipient cell takes up naked DNA from the surrounding medium. This DNA is integrated into the recipient genome by homologous recombination. Basically, the donor cell lyses and fragments of its DNA are released out in the medium. This exogenous, naked DNA is taken up by the recipient cell. Unlike conjugation, no physical contact is required between the donor and recipient cells during transformation.

Difference from transduction is that in this case ssDNA enters the recipient cell cytoplasm whereas in case of transduction dsDNA enters the cell. Both Gram-Positive and Gram-Negative Bacteria can exhibit the ability of transformation in nature mainly under stress conditions. It was discovered in *Streptococcus pneumoniae* and has been extensively studied in various medically Important Gram-Negative Bacteria species, such as *Helicobacter pylori, Haemophilus influenzae* and *Vibrio cholerae*.

In this unit, you will study about concept of transformation, concept of transformation, mechanism of natural and artificial transformation and gene mapping using transformation in detail.

7.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the concept of transformation
- Discuss the types of transformation
7.2 DISCOVERY OF TRANSFORMATION

Transformation was first reported in *Streptococcus pneumoniae* by Frederick Griffith in 1928. DNA as the transforming principle was demonstrated by Avery et al. in 1944. In 1928 Griffith used two strains of *Pneumococcus* (*Streptococcus pneumoniae*) bacteria which infect mice – a Type III-S (Smooth) which was virulent, and a Type II-R (Rough) strain which was non-virulent. In this experiment, bacteria from the III-S strain were killed by heat, and their remains were added to II-R strain bacteria. While neither alone harmed the mice, the combination was able to kill its host. Griffith was also able to isolate both live II-R and live III-S strains of pneumococcus from the blood of these dead mice. Griffith concluded that the Type II-R had been ‘transformed’ into the lethal III-S strain by a ‘Transforming Principle’ that was part of the dead III-S strain bacteria (Refer Figure 7.1).

The ‘transforming principle’ Griffith observed was the DNA of the III-s strain bacteria. While the bacteria had been killed, the DNA had survived the heating process and was taken up by the II-R strain bacteria.

![Fig. 7.1 Griffith’s Experiment Reporting Transformation](image)

The exact nature of the transforming principle (DNA) was verified in the experiments done by Avery, McLeod and McCarty (Refer Figure 7.2). In 1944, Avery, Macleod, McCarty revisited the experiment by Griffith with genetic material transforming in strands of bacteria. Using the same strain of bacteria, S strain,
Avery and his team extracted pure DNA, pure proteins and other materials from the bacterial cells. The materials were mixed with the R strain of bacteria. Rather than using mice, test tubes were used to show the mixing of materials. After examining each mixture, only the R strain mixed with the pure DNA transformed into the S strain bacteria. Since this transformation only occurred within the pure DNA composition, implications that DNA held genetic material were produced. Because of this reason, transformation is sensitive to DNAse (enzymes that cleave DNA).

Fig. 7.2 Avery, Macleod, McCarty Experiment for Finding out the Transforming Principal

Check Your Progress
1. When was transformation reported for the first time?
2. What conclusion did Griffith draw from his experiment?
3. When was the exact nature of the transforming principle (DNA) verified?

7.3 COMPETENCE

Competence is the ability of a cell to take up naked DNA. Competence usually arises at late log phase or stationary phase. It is often triggered by nutritional shortages or stress conditions. Natural competence is genetically encoded. It results from changes in cell wall of bacteria. Receptors on the cell surface are responsible for initial binding of DNA. The number of active receptors varies from 80 for *Pneumococcus* and 4 in *Haemophilus*. 
Transformation

NOTES

Competence initially arises in only small fraction of cells. These cells secrete competence factors that act as signaling molecules or autoinducers or pheromones to make the rest of the cells in the population competent. This is an example of the phenomenon known as quorum sensing. Quorum sensing describes the bacterial communication between cells that enables populations of bacteria to synchronously regulate gene expression and therefore behavior. Competence pheromones are short peptides that are secreted into the culture medium by dividing cells. In late log phase, when density of bacteria is high, sufficient levels of pheromones is achieved and this triggers competence. This mechanism is meant to ensure that any DNA taken up will come from related bacteria as competence is only induced when there are large number of cells of same species.

Types of Competence

There are two types of competence, i.e., as follows:

- Natural Competence
- Artificial Competence

**Naturally Transformable Bacterium** (or naturally competent bacterium) can take up DNA from the environment without requiring special treatment. Examples of naturally competent cells include *Bacillus subtilis, Streptococcus pneumonia, Haemophilus influenza, Neisseria gonorrhoeae* and *Helicobacter pylori*.

Reasons or evolutionary significance of natural competence in bacteria include:

- Selective advantage of genetic diversity.
- DNA uptake as a source of Nucleotides (DNA as food). The recipient cells require nucleotides for their own DNA and RNA synthesis during starvation conditions, so they become naturally component and take up exogenous DNA and utilize their nucleotides for their own synthesis. Some naturally competent bacteria also secrete nucleases into their surroundings, and all bacteria can take up the free nucleotides these nucleases generate from environmental DNA.
- Repair of DNA damage. The selective advantage of a new strand of DNA to promote homologous recombinatorial repair of damaged DNA. The problem of DNA damage is most pronounced during periods of oxidative stress that occur during crowding or starvation conditions.

**Artificial Competence** can also be induced in laboratory making the recipient cells passively permeable to DNA, by exposing it to conditions that do not normally occur in nature.

---

Check Your Progress

1. Define competence.
2. When competence arises?
3. What binds DNA?
4. How many types of competence are there?
7.4 NATURAL AND ARTIFICIAL TRANSFORMATION

Transformation is one of three basic mechanisms for genetic exchange in bacteria. Transformation may be either a natural process that is, one that has evolved in certain bacteria or it may be an artificial process whereby the recipient cells are forced to take up DNA by a physical, chemical, or enzymatic treatment. In both cases, exogenous DNA (DNA that is outside the host cell), is taken into a recipient cell where it is incorporated into the recipient genome, changing the genetic makeup of the bacterium. Natural and artificial transformation is described in detail below.

7.4.1 Natural Transformation

Gram-Positive and Gram-Negative Bacteria differ in the structure of their cell envelope, thus there are some differences in the mechanisms of DNA uptake in these cells.

The basic steps of transformation include the following (Refer Figure 7.3)

- Binding of double-stranded DNA to the outer cell surface of bacterium.
- Movement of DNA across the cell wall and outer membrane (no outer membrane in gram positive bacterium).
- Degradation of one of the DNA strands.
- Translocation of the remaining single strand of DNA into cytoplasm of the cell across inner membrane.
- Once in the cell, the single-stranded transforming DNA might synthesize the complementary strand and re-establish itself as a plasmid, stably integrate into the chromosome, or get degraded.

The uptake of DNA is generally non-sequence specific, although in some species the presence of specific DNA uptake sequences may facilitate efficient DNA uptake.

Fig. 7.3 Overview of DNA Uptake in Bacteria
Mechanism of Natural Transformation in Bacillus subtilis

Various competence factors are involved in uptake of DNA in Gram-Positive Bacterium, Bacillus subtilis. Some of these include Pilin Complex (ComGC), DNA Binding Protein (ComEA), Nuclease (N), Channel Protein (ComEC) and DNA Translocase that moves the DNA into the Cytoplasm (ComFA) (Refer Figure 7.4).

In Gram-Positive Bacteria pseudopilus system is responsible for DNA uptake. It acts as a piston, moving up and down in the thick peptidoglycan cell wall layer thereby binding to extracellular dsDNA and pulling it towards the cell membrane where DNA uptake proteins reside. In Bacillus subtilis, major pseudopaline complex ComGC is able to assemble into a polymeric structure that is essential for transformation. Next, dsDNA bind to ComEA, DNA binding protein which shuffles the DNA to the inner membrane transporter protein, ComEC. Transport across the inner membrane is accomplished by ComEC concomitantly with the degradation of one strand of dsDNA by nuclease producing ssDNA. Next, this ssDNA binds to ComFA that moves it into the cytoplasm. Inside the cytoplasm, this ssDNA is bound by various SSB (Single Strand Binding) Proteins for protection against degradation. Finally, it gets integrated into the recipient genomic DNA by homologous recombination catalysed by RecA Proteins.

Mechanism of Natural Transformation in Streptococcus pneumoniae

Streptococcus pneumoniae is an important human pathogen. Natural transformation was discovered in this species. It involves internalization of exogenous DNA and its incorporation into the genome. Environmental factors trigger the competence. Once competent, transformation is mediated by a series of competence factors or proteins.
Steps in transformation of *Streptococcus* include the following (Refer Figure 7.5):

- Exogenous dsDNA attaches to the Pilius (ComGC).
- It is passed to the DNA Receptor Protein ComEA.
- ComEA directs it to EndA Nuclease which degrades one strand of the dsDNA resulting in ssDNA. This ssDNA is transferred across the membrane via ComEC pore.
- After entry into the cell, ssDNA is coated with SSBs which protect it from nuclease degradation.
- Next, recombinase DprA also binds the ssDNA and recruits recombinase RecA onto ssDNA.
- RecA can then polymerize on the ssDNA, and promotes strand exchange.
7.4.2 Artificial Transformation

Most of the cells are not naturally competent. In addition, natural competence and transformation are efficient for only linear molecules, such as chromosomal DNA but not for circular plasmid molecules. Thus, prokaryotic as well as eukaryotic cells need to be made competent artificially in laboratory. This is an essential step in recombinant DNA technology, mainly to incorporate plasmids carrying gene of interest into the host cells. Unlike natural competence, artificial competence is not encoded by genes of the bacteria.

Methods of Artificial Competence in Bacteria

There are various methods to prepare competent cells in the laboratory. Few of them include:

- Calcium Chloride Mediated Transformation
- Freeze Thaw Method
- Electroporation

Calcium Chloride Method: It is the most common chemical method of making cells competent. Treatment with Calcium ions make some Bacteria competent. The Calcium ions increase the permeability of cell membrane making transient pores which facilitate DNA entry. Basically, role of Calcium ions is to neutralize the negative charges present on the surface of DNA and LipoPolySaccharide (LPS) to avoid electrostatic repulsion between the two. Positively charged Calcium ions (Ca\(^{2+}\)) attract both the negatively charged DNA backbone (Phosphate) and the negatively charged groups in the LPS inner core. Other divalent cations, such as MnCl, and KCl can also increase the efficiency of transformation. DiMethyl Sulfoxide (DMSO) can also increase the efficiency by generating ionic shield.

Once competent cells are prepared, the plasmid DNA is then added to the cells by heat shock method, where chilled cells are heated to a higher temperature of 42°C for short time (generally two minutes). A sudden increase in temperature further creates pores in the plasma membrane of the bacteria and allows for plasmid DNA to enter the bacterial cell. After heat shock, the mixture is immediately placed on ice for 1-2 minutes before adding culture medium for growing. Intact plasmid DNA molecules replicate in bacterial host cells. To help the bacterial cells recover from the heat shock cells are briefly incubated with non-selective growth media. As the cells recover, plasmid genes are expressed. Bacterial colonies selected using antibiotic selection techniques (Refer Figure 7.6).
Transformation

NOTES

Fig. 7.6 Steps of Bacterial Transformation by Calcium Chloride Method

In case of artificial transformation, transformation efficiency can also be calculated. Transformation efficiency is defined as the number of Colony Forming Units (CFUs) which would be produced by transforming 1 μg of plasmid into a given volume of competent cells.

Freeze Thaw Method: In the freeze thaw method, actively growing cells are pelleted down by centrifugation at 5000 rpm at room temperature. Pellet is washed. Cells are then frozen in liquid nitrogen and stored at -70°C. Competent cells are thawed on ice and DNA is added. Cells are then incubated for recovery. Transformed cells are selected by antibiotic selection.

Electroporation: Electroporation is a method of transformation that allows the introduction of foreign DNA into host cells via the application of high-voltage electric pulses.

Cells are placed in buffer and put into electroporator. DNA is added and subjected to a high-voltage electrical pulse of defined magnitude and length. The cells are then allowed to recover and are finally selected by antibiotic selection technique.

Check Your Progress

8. Is transformation natural or artificial?
9. How does uptake of DNA occur in bacteria pseudopilus system?
10. What are the methods of artificial competence in bacteria?
11. Define the term electroporation.
7.5 GENE LINKAGE AND MAPPING BY TRANSFORMATION

Genetic linkage is the tendency of DNA sequences that are close together on a chromosome to be inherited together. Genes that are located on the same chromosome are described as linked genes. If genes are very close together, they are more likely to be transmitted as a block. The distance between genes is designated as map units.

Transformation is used to map genes where mapping by conjugation or transduction is not possible. Transformation experiments can be used to determine the following:

- Whether the genes are linked.
- Order of genes on the genetic map.
- Map distance between the genes.

In case transformation experiments, donor DNA is purified, fragmented and added to competent recipient bacteria. Donor and recipient both have detectable differences in phenotype and therefore genotype. If the DNA fragment undergoes homologous recombination with the recipient’s chromosome, a new phenotype may be produced. Transformants are detected by testing for phenotypic changes (Refer Figure 7.7).

![Fig. 7.7 Gene Mapping Experiments using Transformation](image)

Gene linkage and order can also be determined by cotransformation, i.e., the ability of two genetic markers to be transformed together. If two donor genes are located close together on the chromosome, then there is a greater chance that they will be carried on the same piece of transforming DNA and hence will cause a double transformation or cotransformation. Conversely, if genes are widely separated on the chromosome, then most likely they will be carried on separate transforming segments and the frequency of double transformants will equal the
Transformation

NOTES

product of the single-gene transformation frequencies. Thus, it is possible to test for close linkage. For example, if $a^+ b^+$ donor DNA is used to transform $a^– b^–$ recipient cells, then, if $a$ and $b$ are closely linked, the proportion of $a^+ b^+$ cotransformants, should exceed the product of the proportions of single $a^+$ and $b^+$ transformants. Relative map distances of closely linked genes can be deduced from cotransformation percentages in an approach similar to that of cointernal frequencies.

Check Your Progress

12. What is genetic linkage?
13. What are linked genes?
14. How can the distance between genes be determined?
15. Give the uses of transformation experiments.

7.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Transformation was first reported in *Streptococcus pneumoniae* by Frederick Griffith in 1928.
2. Griffith concluded that the type II-R had been ‘transformed’ into the lethal III-S strain by a ‘Transforming Principle’ that was part of the dead III-S strain bacteria.
3. The exact nature of the transforming principle (DNA) was verified in the experiments done by Avery, McLeod and McCarty.
4. Competence is the ability of cell to take up naked DNA.
5. Competence usually arises at late log phase or stationary phase.
6. Receptors on the cell surface are responsible for initial binding of DNA. The number of active receptors varies from 80 for *Pneumococcus* and 4 in *Haemophilus*.
7. There are two types of competence, i.e., as follows:
   - Natural Competence
   - Artificial Competence
8. Transformation may be either a natural process that is, one that has evolved in certain bacteria or it may be an artificial process whereby the recipient cells are forced to take up DNA by a physical, chemical, or enzymatic treatment.
9. In Gram-Positive Bacteria pseudopilus system is responsible for DNA uptake. It acts as a piston, moving up and down in the thick peptidoglycan cell wall layer thereby binding to extracellular dsDNA and pulling it towards the cell membrane where DNA uptake proteins reside.
There are various methods to prepare competent cells in the laboratory. Few of them include:
- Calcium Chloride Mediated Transformation
- Freeze Thaw Method
- Electroporation

Electroporation is a method of transformation that allows the introduction of foreign DNA into host cells via the application of high-voltage electric pulses.

Genetic linkage is the tendency of DNA sequences that are close together on a chromosome to be inherited together.

Genes that are located on the same chromosome are described as linked genes.

The distance between genes is designated as map units.

Transformation experiments can be used to determine the following:
- Whether the genes are linked.
- Order of genes on the genetic map.
- Map distance between the genes.

7.7 SUMMARY

Transformation was first reported in *Streptococcus pneumoniae* by Frederick Griffith in 1928. DNA as the transforming principle was demonstrated by Avery *et al.*, in 1944.

In 1928 Griffith used two strains of *Pneumococcus (Streptococcus pneumoniae)* Bacteria which infect mice – a Type III-S (Smooth) which was virulent, and a Type II-R (Rough) strain which was non-virulent.

Griffith was also able to isolate both live II-R and live III-S strains of pneumococcus from the blood of these dead mice.

Griffith concluded that the Type II-R had been ‘transformed’ into the lethal III-S strain by a ‘Transforming Principle’ that was part of the dead III-S strain bacteria.

The ‘transforming principle’ Griffith observed was the DNA of the III-s strain bacteria. While the bacteria had been killed, the DNA had survived the heating process and was taken up by the II-R strain bacteria.

The exact nature of the transforming principle (DNA) was verified in the experiments done by Avery, McLeod and McCarty.

In 1944, Avery, Macleod, McCarty revisited the experiment by Griffith with genetic material transforming in strands of bacteria.

Competence is the ability of cell to take up naked DNA. Competence usually arises at late log phase or stationary phase. It is often triggered by nutritional shortages or stress conditions.
Transformation

NOTES

- Natural competence is genetically encoded. It results from changes in cell wall of bacteria.
- Receptors on the cell surface are responsible for initial binding of DNA. The number of active receptors varies from 80 for *Pneumococcus* and 4 in *Haemophilus*.
- Competence initially arises in only small fraction of cells. These cells secrete competence factors that act as signaling molecules or autoinducers or pheromones to make the rest of the cells in the population competent.
- Competence pheromones are short peptides that are secreted into the culture medium by dividing cells. In late log phase, when density of bacteria is high, sufficient levels of pheromones is achieved and this triggers competence.
- Naturally Transformable Bacterium (or naturally competent bacterium) can take up DNA from the environment without requiring special treatment.
- Transformation is one of three basic mechanisms for genetic exchange in bacteria.
- Transformation may be either a natural process that is, one that has evolved in certain bacteria or it may be an artificial process whereby the recipient cells are forced to take up DNA by a physical, chemical, or enzymatic treatment.
- In both cases, exogenous DNA (DNA that is outside the host cell), is taken into a recipient cell where it is incorporated into the recipient genome, changing the genetic makeup of the bacterium.
- Gram-Positive and Gram-Negative Bacteria differ in the structure of their cell envelope, thus there are some differences in the mechanisms of DNA uptake in these cells.
- The uptake of DNA is generally non-sequence specific, although in some species the presence of specific DNA uptake sequences may facilitate efficient DNA uptake.
- Various competence factors are involved in uptake of DNA in Gram Positive Bacterium, *Bacillus subtilis*. Some of these include Pilin Complex (ComGC), DNA Binding Protein (ComEA), Nuclease (N), Channel Protein (ComEC) and DNA Translocase that moves the DNA into the Cytoplasm (ComFA).
- In Gram-Positive Bacteria pseudopilus system is responsible for DNA uptake. It acts as a piston, moving up and down in the thick peptidoglycan cell wall layer thereby binding to extracellular dsDNA and pulling it towards the cell membrane where DNA uptake proteins reside.
- *Streptococcus pneumoniae* is an important human pathogen. Natural transformation was discovered in this species.
- Most of the cells are not naturally competent. In addition, natural competence and transformation are efficient for only linear molecules, such as chromosomal DNA but not for circular plasmid molecules.
In the freeze thaw method, actively growing cells are pelleted down by centrifugation at 5000 rpm at room temperature.

Electroporation is a method of transformation that allows the introduction of foreign DNA into host cells via the application of high-voltage electric pulses.

Genetic linkage is the tendency of DNA sequences that are close together on a chromosome to be inherited together.

Genes that are located on the same chromosome are described as linked genes.

If genes are very close together, they are more likely to be transmitted as a block.

The distance between genes is designated as map units.

Transformation is used to map genes where mapping by conjugation or transduction is not possible.

In case transformation experiments, donor DNA is purified, fragmented and added to competent recipient bacteria.

Donor and recipient both have detectable differences in phenotype and therefore genotype.

If the DNA fragment undergoes homologous recombination with the recipient’s chromosome, a new phenotype may be produced.

Gene linkage and order can also be determined by cotransformation, i.e., the ability of two genetic markers to be transformed together.

7.8 KEY WORDS

**Transformation:** It is type of HGT in bacteria involving uptake of free DNA from the environment.

**Transforming principle:** Component which is able to cause transformation of the recipient cell. DNA is the transforming principle.

**Competence:** It is the ability of cell to take up naked DNA.

**Quorum sensing:** It is the regulation of gene expression in response to fluctuations in cell-population density.

**Cotransformation:** It is the ability of transformation of two closely placed genetic markers together.

**Transformation efficiency:** Transformation efficiency is defined as the number of Colony Forming Units (CFUs) which would be produced by transforming 1 µg of plasmid into a given volume of competent cells.
7.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Define the term transformation.
2. What is competence?
3. Write the differences between natural and artificial competence.
4. Enlist the competence factors involved in natural transformation of *Streptococcus*.
5. What is gene linkage?
6. What is cotransformation and its application?

Long Answer Questions

1. Discuss about transformation in detail. Also write about the discovery of transformation.
2. Write a note on types of transformation.
3. Diagrammatically explain the mechanism of competence in *Bacillus subtilis*.
4. What is natural transformation? Explain the mechanism of natural transformation in *Streptococcus pneumoniae*.
5. Explain artificial transformation and methods of artificial competence.
6. Elaborate a note on gene linkage. How can transformation be used to determine linkage of genes?
7. Elaborate on the various methods of artificial transformation.
8. Describe the experiment of discovery of transformation.
9. A transformation experiment is carried out using a donor strain that has three chromosomal antibiotic resistance mutations, namely strR, rifR, and nalR. These mutations confer resistance to the antibiotics Streptomycin, Rifampicin, and Nalidixic Acid. The recipient strain is sensitive to all 3 antibiotics. The results are:

<table>
<thead>
<tr>
<th>Antibiotics in Agar Plates</th>
<th># Colonies</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>100,000</td>
</tr>
<tr>
<td>str</td>
<td>1,000</td>
</tr>
<tr>
<td>rif</td>
<td>1,000</td>
</tr>
<tr>
<td>nal</td>
<td>100</td>
</tr>
<tr>
<td>str + rif</td>
<td>800</td>
</tr>
<tr>
<td>rif + nal</td>
<td>10</td>
</tr>
<tr>
<td>str + rif + nal</td>
<td>0</td>
</tr>
</tbody>
</table>

What do you conclude about the positions of the genes?
7.10 FURTHER READINGS

Horizontal Gene Transfer (HGT) is the transfer of genetic material from one bacteria to another. It is mediated by three methods, viz., conjugation, transduction and transformation. Conjugation has been discussed in detail in the previous unit. Here, we will study transduction. Transduction is the process by which DNA is transferred from one bacteria to another via a bacteriophage (virus that infects bacteria). Such a virus particle carrying bacterial DNA is called a transducing particle. Unlike conjugation, transduction does not require physical contact between the two cells neither it is inactivated by DNase as in case of transformation.

Transduction is a type of horizontal gene transfer in bacteria leading to transfer of genes from one bacteria to another via bacteriophages. It is of two type, viz., generalized and specialized. In case of generalized transduction, any part of the host bacterial genome can be transferred. It is mediated by lytic phages, such as T4 Phage and results from error in packaging of the genome in phage head. In case of specialized transduction, only specific regions of the host genome, i.e., the genes flanking the site of integration of the phage genome can be transferred. It is mediated by lysogenic phages such as lambda phage and results from error in excision of the phage genome from the host genome. The genes transferred via transduction can recombine in the recipient cell via homologous recombination. Transduction can be used as a means of genetic mapping. Cotransduction frequency of two genetic markers can be calculated. Closer the genes, more is their probability of being transduced together, higher will be their cotransduction frequency. This basis is used to infer the linkage between genes. Three factor crosses can further
be used to decipher the order of the gene. Besides mapping, other applications of transduction include their use as cloning vectors and for hybridization studies.

In this unit, you will study about types of life cycle of bacteriophages, concept of genetic transfer in bacteria by bacteriophages, types of transduction, uses of specialized transduction and gene mapping using bacteriophages in detail.

8.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss the types of life cycle of bacteriophages
- Explain the concept of genetic transfer in bacteria by bacteriophages
- Understand the types of transduction
- Analyse the uses of specialized transduction
- Discuss gene mapping using bacteriophages

8.2 BACTERIOPHAGE AND ITS TYPES

Transduction is virus mediated bacterial DNA transfer. Virus that infect bacteria are called bacteriophages. Although each virus has unique aspects to its life cycle, a general pattern of replication is observable. The typical virus life cycle consists of five steps, namely attachment to the host cell, entry into the host cell, synthesis of viral nucleic acid and proteins within the host cell, self-assembly of virions (progeny virus) within the host cell, and release and maturation of virions from the host cell.

On the basis of their replication cycle bacteriophages are of two types, i.e., as follows (Refer Figure 8.1):

- Lytic Bacteriophages
- Lysogenic Bacteriophages

Lytic Bacteriophages

Lytic bacteriophages are those that break open or lyse the host bacterial cell immediately after replication of the virion. As the bacterial cell is lysed then the phage progeny can find new hosts to infect. The phages that carry out Lytic Cycle are called Virulent or Lytic Phages. An example of a Lytic Bacteriophage is Phage T4, which infects Escherichia coli.

Lysogenic Bacteriophages

Lysogenic phages are the one that do not result in immediate lysis of the host bacterial cell. Instead their viral genome gets stably integrated into the host genome forming a prophase. This prophage DNA which has the viral DNA incorporated
can be passed down to the daughter cells when the bacteria divides. The phage remains dormant thus does not harm the host. However, under unfavorable and stressful conditions, such as depletion of nutrients or UV exposure, the phage genome excises out of the host genome and the phage switches to Lytic cycle. They initiate their genome replication, assemble and result in lysis of host cell to release out. The phage that carry out Lysogenic cycle are called temperate phages. An example of such a bacteriophage is Lambda (λ) phage, which infects Escherichia coli.

Fig. 8.1 Lytic versus Lysogenic Cycle of Bacteriophage

Check Your Progress
1. How many types of bacteriophages are there?
2. Define lytic bacteriophages.
3. What is the difference between virulent and temperate bacteriophage?
4. What are lysogenic phages?

8.3 TRANSDUCTION: DISCOVERY AND TYPES

Transduction was discovered by Norton Zinder and Joshua Lederberg at the University of Wisconsin–Madison in 1952 using Salmonella typhimurium. They used two Salmonella auxotrophic strains LA-22 and LA-2. Auxotrophic strain is a mutant that cannot grow on minimal medium and requires certain supplements whereas prototrophic strains are wild type as it will grow in minimal
medium or medium lacking the supplement. In this experiment, *Salmonella* strain LA-2 was auxotroph for methionine and histidine (met his-) and strain LA-22 was auxotroph for phenylalanine and tryptophan (phe trp-). They mixed the two strains together in the U-tube, and when the mixture was plated on minimal medium, they recovered prototrophic cells (Refer Figure 8.2). This observation showed that there is some Filterable Agent (FA) that transferred the Phenylalanine and Tryptophan marker from strain LA-2 to LA-22 making it prototrophic.

![Fig. 8.2 Lederberg-Zinder Experiment of Bacterial Transduction](image)

Three observations were used to identify the Filterable Agent (FA):

- The FA was produced by the LA-2 cells only when they were grown in association with LA-22 cells.
- The addition of DNase, which enzymatically digests DNA, did not render the FA ineffective. Therefore, the FA is not exogenous DNA, ruling out transformation.
- The FA could not pass across the filter of the Davis U-tube when the pore size was reduced below the size of bacteriophages.

Aided by these observations and aware that temperate phages could lysogenize *Salmonella*, researchers proposed that the genetic recombination event was mediated by bacteriophage P22 and they named this process as transduction.
8.3.1 Types of Transduction

There are two types of transduction depending on which part of bacterial DNA is transferred to the recipient cell. The two forms are:

- **Generalized Transduction:** Any piece of the bacterial genome can be transferred.
- **Specialized Transduction:** Only specific pieces of the bacterial chromosome can be transferred.

8.3.2 Generalized Transduction

Generalized transduction is the process by which any part of bacterial DNA can be transferred from donor to recipient bacteria via a bacteriophage. It is a rare event with 1 in 10,000 phages being transducing. It is a random event as any part of the bacterial genome can be transferred to the recipient cell. Lytic bacteriophages carry out generalized transduction. Generalized transducing phage have a packaging error and thus by mistake pack segment of bacterial DNA in place of viral DNA. Bacteriophage pack their genome by headful packaging mechanism so any part of bacterial host genome which is similar in size to its own genome can be packaged leading to formation of a transducing phage. Thus, generalized transducing phage produce some particles that contain only DNA obtained from the host bacterium, rather than phage DNA. Examples of generalized transducing phage include *Escherichia coli* P1 and *Salmonella typhimurium* P22.

Requirements for generalized transducing phage are:
- Phage must not degrade the host DNA completely.
- Less specificity during packaging.
- Package by the headful rule (based only on size of the DNA).

The transducing phage loaded with bacterial DNA continues to infect another bacterial cell.

Fate of DNA that enters the recipient cell:
- The DNA can be degraded.
- It can circularize to become a plasmid.
- If it matches with a homologous region of the recipient cell's chromosome, it can be exchanged by bacterial recombination.

I. *Escherichia coli* Phage P1 - Generalized Transducing Phage

During infection of *Escherichia coli* by phage P1, a phage-encoded nuclease is made that causes fragmentation of bacterial DNA. A single fragment of bacterial DNA comparable in size to P1 DNA is packaged into a phage particle instead of P1 DNA. The positions of the nuclease cuts in the host chromosome are fairly random, so a transducing particle may contain a fragment derived from any region...
of the host. When a transducing particle adsorbs to a bacterium, the DNA carried within the particle head is injected into the cell and becomes available for double crossing over with the homologous region of the bacterial chromosome. For example, Phage P1 can transfer Leucine marker converting leu− recipient strain of *Escherichia coli* into leu+ (Refer Figure 8.3).

**Fig. 8.3** Generalized Transduction by *Escherichia coli* Phage P1

**II. Abortive Transduction**

- Transduction requires that the recipient cell be Rec+ for recombination to occur. No transductants are observed with RecA− or RecBC recipients. The donor cell can be either Rec+ or Rec−. RecA protein is required for single strand displacement and RecBC protein unwinds the helix.
- Abortive transduction is an event in which transducing DNA fails to be incorporated into the recipient chromosome.
• The bacterial DNA remains in the cytoplasm and does not replicate but is transmitted to one progeny cell following each division.
• Only a single cell, partially diploid for the transduced genes is produced.

8.3.3 Specialized Transduction

In case of specialized transduction, only specific parts of host genome can be transferred to the recipient cell. It takes place via temperate phages, i.e., phages that undergo both lysogenic and lytic cycle. It is produced during induction of Lysogen. As we discussed earlier, integrated phage DNA can excise out from the bacterial genome and follow lytic cycle. Specialized transducing phages are produced when there is error in excision. Host genes that flank the integrated phage DNA are also excised out by mistake leaving some of its own genes. Thus, transducing phage produces particles containing both phage and bacterial genes linked in a single DNA molecule.

Example of a specialized transducing phage is Lambda (λ) phage. It gets integrated between the Galactose (Gal) and Biotin (Bio) Operon. Due to error in excision or abnormal out looping, it can carry Genes of Gal or Bio Operon leaving behind some of its own genes. As some of its own genes have been replaced by bacterial genes, these phages are defective and cannot reproduce. The dgal or gal carries Bacterial Gal Genes and lacks tail fiber genes and thus cannot make Progeny or Plaques (Refer Figure 8.4). Similarly, dbio or bio carries the Biotin Genes and lacks integration, exision genes thus cannot integrate.

![Fig. 8.4: Formation of Specialized Transducing Phage - λdgal](image)

Types of Specialized Transduction

Depending on the frequency of transducing phages obtained after lysis of the host cells, specialized transduction is of two types:

- Low Frequency Transduction (LFT)
- High Frequency Transduction (HFT)
Low Frequency Transducing (LFT) Lysate

- A lysate formed from a single Lysogen in which wrong excision occurs only infrequently is called a LFT lysate.
- LFT Lysates are the Lysates containing mostly normal phages and just a few specialized transducing phages.
- LFT is a result of rare incorrect excision events from a single Lysogen produce transducing particles at about $10^6$ to $10^7$ of the total phage particles in the Lysate.

High Frequency Transducing (HFT) Lysate

- Lysates containing a relatively large number of specialized transducing phages.
- They are created by co-infecting a host cell with a helper phage (normal phage) and a transducing phage, i.e., Dilysogen (Refer Figure 8.5).
- In case of transducing phage certain genes are replaced by bacterial genes so it cannot replicate. However, co-infection with helper phage allows the transducing phage to replicate as it provides functional counterparts of genes that are lacking in transducing phage, thus increasing the number of transducing phages in the Lysate.
- The Dilysogens yield Lysates half of whose phage are transducing particles.

![Fig. 8.5: Dilysogen with $\lambda$gal and $\lambda$ helper Integrated into the Escherichia coli Genome](image)

Both LFT and HFT also contain phage that are not transducing particles for LFT Lysates these will be 99.99% of the total particles in the population and for HFT Lysates these will be 50% of the total particles in the population.

Uses of Specialized Transducing Phages

- Can be used as cloning vehicle as they can transfer specific bacterial genes into the recipient cell.
- The $\lambda$gal can serve as a source of Gal DNA for base-sequencing analysis.
- Specialized transducing phages have also served as hybridization probes for identifying specific mRNA molecule.

8.3.4 Transduction Mapping

- Only a small amount of chromosome, a few genes, can be transferred by transduction. The closer the two genes are to each other, the more likely
they are to be transduced by the same phage. Thus, ‘Cotransduction
Frequency’ is the key parameter used in mapping genes by transduction.
Cotransduction frequency is defined as the ratio of transductants that co-
hherit both markers divided by the total number of transductants.

- Linkages are expressed as cotransduction frequency. Generalized
transduction can be used to derive linkage information about bacterial genes
in cases in which genes are close enough that the phage can pick them up
and transduce them as a single piece of DNA.

- Transduction studies can also determine the precise order of these genes.

- Consider three genes ‘a’, ‘b’ and ‘c’ for which the frequencies of
cotransduction are:
  
  a-b : 90%
  a-c : 33%
  b-c : 32%

  The values indicate that ‘b’ is closer to ‘a’ than to ‘c’, but the order cannot
still be deduced as it can either be ‘a-b-c’ or ‘b-a-c’. Only order ‘a-c-b’ can be
excluded.

Three factor crosses can be used to determine the order of closely linked
markers with respect to a third gene. Principle of this method is that the probability
of appearance of a genotype requiring four exchanges for its formation is much
smaller than the probability of one formed by two exchanges. The results of
reciprocal crosses are compared. A reciprocal cross is cross with the phenotype
each sex reversed as compared with the original cross.

If for Cross 1 the frequency of ‘+++’ is observed to be much less than that
of Cross 2 then, the order must be ‘a-b-c’.

**Another Method for Mapping**

- Transduction between Donor Strain A+B+C- to recipient that is A-B-C+.
 One phenotype A+ is checked by plating on medium lacking A and 100 A+
colonies are tested further for B and C genotypes by replica plating.

<table>
<thead>
<tr>
<th>Genotypes Observed</th>
<th>Number of Colonies</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+B+C+</td>
<td>5</td>
</tr>
<tr>
<td>A+B+C-</td>
<td>19</td>
</tr>
<tr>
<td>A+B-C+</td>
<td>49</td>
</tr>
<tr>
<td>A+B, C-</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>
1. Calculate Cotransduction Frequencies as follows:
   \[ \frac{A+B+ (5 + 19)}{100} = 0.24 \]
   \[ \frac{A+C- (19 + 27)}{100} = 0.46 \]
   Thus, C is nearer to A than B is to A.
   Order can be A-C-B or C-A-B
2. Rule to Analyze the Data: The rarest recombination class is that achieved by the maximum number of exchanges.
3. Rarest class is A+B+C+ and thus Order 1 (A-C-B) is the correct or accurate order.

Check Your Progress
5. What is generalized transduction?
6. Which specific part of genome can be transferred during specialized transduction?
7. Why does HFT Lysate have large number of specialized transducing phages?

8.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS
1. On the basis of their replication cycle bacteriophages are of two types, i.e.,
   - Lytic Phage
   - Lysogenic Phage
2. Lytic bacteriophages are those that break open or lyse the host bacterial cell immediately after replication of the virion.
3. Virulent bacteriophage carries out lytic cycle of replication whereas temperate phage can carry out both lysogenic as well as lytic cycle.
4. Lysogenic phages are the one that do not result in immediate lysis of the host bacterial cell.
5. It is called generalized transduction because any random part of the host bacterial genome can be transferred to the recipient cell.
6. Only the host genes flanking the site of integration of the phage genome can be transferred via specialized transduction.
7. HFT Lysate has large number of specialized transducing phages because the genome of the helper phage gets integrated along with the transducing phage genome forming a Dilysoegen. Helper phage aids in transducing phage replication.
8.5 SUMMARY

- Transduction is virus mediated bacterial DNA transfer. Virus that infect bacteria are called bacteriophages.
- Although each virus has unique aspects to its life cycle, a general pattern of replication is observable.
- The typical virus life cycle consists of five steps, namely attachment to the host cell, entry into the host cell, synthesis of viral nucleic acid and proteins within the host cell, self-assembly of virions (progeny virus) within the host cell, and release and maturation of virions from the host cell.
- Lytic bacteriophages are those that break open or lyse the host bacterial cell immediately after replication of the virion.
- As the bacterial cell is lysed, the phage progeny can find new hosts to infect.
- The phages that carry out lytic cycle are called virulent or lytic phages.
- Lysogenic phages are the one that do not result in immediate lysis of the host bacterial cell.
- The phage that carry out lysogenic cycle are called temperate phages. An example of such a bacteriophage is Lambda (λ) phage, which infects *Escherichia coli*.
- Transduction was discovered by Norton Zinder and Joshua Lederberg at the University of Wisconsin–Madison in 1952 using *Salmonella typhimurium*.
- Auxotrophic strain is a mutant that cannot grow on minimal medium and requires certain supplements whereas prototrophic strains are wild type as it will grow in minimal medium or medium lacking the supplement.
- Generalized transduction is the process by which any part of bacterial DNA can be transferred from donor to recipient bacteria via a bacteriophage.
- Generalized transducing phage have a packaging error and thus by mistake pack segment of bacterial DNA in place of viral DNA.
- Generalized transducing phage produce some particles that contain only DNA obtained from the host bacterium, rather than phage DNA.
- The transducing phage loaded with bacterial DNA continues to infect another bacterial cell.
- During infection of *Escherichia coli* by phage P1, a phage-encoded nuclease is made that causes fragmentation of bacterial DNA.
- A single fragment of bacterial DNA comparable in size to P1 DNA is packaged into a phage particle instead of P1 DNA.
Transduction and its Types

When a transducing particle adsorbs to a bacterium, the DNA carried within the particle head is injected into the cell and becomes available for double crossing over with the homologous region of the bacterial chromosome.

In case of specialized transduction, only specific parts of host genome can be transferred to the recipient cell. It takes place via temperate phages, i.e., phages that undergo both Lysogenic and Lytic cycles.

Specialized transducing phages are produced when there is error in excision.

Integrated phage DNA can excise out from the bacterial genome and follow Lytic cycle.

Host genes that flank the integrated phage DNA are also excised out by mistake leaving some of its own genes.

Transducing phage produces particles containing both phage and bacterial genes linked in a single DNA molecule.

In case of transducing phage certain genes are replaced by bacterial genes so it cannot replicate.

Only a small amount of chromosome, a few genes can be transferred by transduction.

The closer the two genes are to each other, the more likely they are to be transduced by the same phage.

Cotransduction frequency is the key parameter used in mapping genes by transduction.

Cotransduction frequency is defined as the ratio of transductants that co-inherit both markers divided by the total number of transductants.

8.6 KEY WORDS

- **Bacteriophage**: Bacteriophage is a virus that infects the bacteria.
- **Lytic cycle**: Viral life cycle involving entry of viral genome, replication, assembly and release by lysis of the host cell.
- **Lysogenic cycle**: Viral life cycle involving integration of the viral genome into the bacterial host genome, followed by its proliferation.
- **Prophage**: A prophage is a bacteriophage genome integrated into the circular bacterial genome.
- **Generalized transduction**: Transfer of any part of the bacterial genome from donor to recipient cells via a Lytic bacteriophage.
- **Specialized transduction**: Transfer of only specific parts of bacterial genome from donor to recipient via a Lysogenic phage.
Transduction and Its Types

NOTES

8.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What is a bacteriophage?
2. How was transduction discovered?
3. What are virulent and temperate bacteriophages?
4. What is generalized transduction?
5. Write in brief about abortive transduction.
6. What are the uses of specialized transducing phages?
7. What is a Dilysogen? State its importance in HFT.

Long Answer Questions

1. Explain the types of transduction.
2. Tabulate the differences between generalized and specialized transduction.
3. Describe any one generalized transducing phage.
4. Write a note on *Escherichia coli* Phage P1 generalized transducing phage.
5. What is specialized transduction? Explain its types and give its uses.
6. What is the basis of gene mapping using transduction?
7. Differentiate between LFT and HFT Lysate.

8.8 FURTHER READINGS


9.0 INTRODUCTION

A gene is the basic physical and functional unit of heredity. Genes are made up of DNA. Some genes act as instructions to make molecules called proteins. However, many genes do not code for proteins. In humans, genes vary in size from a few hundred DNA bases to more than 2 million bases. Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product. These products are often proteins, but in non-protein coding genes, such as transfer RNA (tRNA) or small nuclear RNA (snRNA) genes, the product is a functional RNA. The DNA of prokaryotes is organized into a circular chromosome, supercoiled within the nucleoid region of the cell cytoplasm. Proteins that are needed for a specific function, or that are involved in the same biochemical pathway, are encoded together in blocks called operons. For example, all of the genes needed to use lactose as an energy source are coded next to each other in the lactose (or lac) operon, and transcribed into a single mRNA.

Regulation of gene expression, or gene regulation, includes a wide range of mechanisms that are used by cells to increase or decrease the production of specific gene products (protein or RNA). Sophisticated programs of gene expression are widely observed in biology, for example to trigger developmental pathways, respond to environmental stimuli, or adapt to new food sources. Virtually any step of gene expression can be modulated, from transcriptional initiation,
to RNA processing, and to the post-translational modification of a protein. Often, one gene regulator controls another, and so on, in a gene regulatory network. Gene regulation is essential for viruses, prokaryotes and eukaryotes as it increases the versatility and adaptability of an organism by allowing the cell to express protein when needed.

In biology, an operon is a functioning unit of DNA containing a cluster of genes under the control of a single promoter. The genes are transcribed together into an mRNA strand and either translated together in the cytoplasm, or undergo splicing to create monocistronic mRNAs that are translated separately, i.e., several strands of mRNA that each encode a single gene product. The result of this is that the genes contained in the operon are either expressed together or not at all. Several genes must be co-transcribed to define an operon.

Originally, operons were thought to exist solely in Prokaryotes, but since the discovery of the first operons in Eukaryotes in the early 1990s, more evidence has arisen to suggest they are more common than previously assumed. In general, expression of Prokaryotic Operons leads to the generation of polycistronic mRNAs, while Eukaryotic Operons lead to monocistronic mRNAs. The Lac Operon (Lactose Operon) is an operon required for the transport and metabolism of lactose in Escherichia coli and many other enteric bacteria. Although glucose is the preferred Carbon source for most bacteria, the lac operon allows for the effective digestion of lactose when glucose is not available through the activity of beta-galactosidase. Gene regulation of the lac operon was the first genetic regulatory mechanism to be understood clearly, so it has become a foremost example of prokaryotic gene regulation.

In this unit, you will study about the gene concept, regulation of bacterial gene expression, lactose system, coordinate regulation, Lac components, positive and negative regulations, and catabolite repression.

9.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Understand the gene concept
- Explain about the regulation of bacterial gene expression
- Discuss the significant features of lactose system
- Elucidate on coordinate regulation and Lac components
- Analyse the positive and negative regulation, and catabolite repression

9.2 GENE CONCEPT

Genes are the small segments present on DNA. They encode for proteins which describes an individual’s phenotype. They are characterized by the presence of
start codon in the beginning and stop codon at the end. There are few genes which do not code for functional proteins and many are hypothetical due to their unknown functions. Interestingly, a gene may code for more than one proteins in eukaryotes by the phenomenon known as alternative splicing. In short, genes are the basic unit of hereditary in all the organisms from simpler like prokaryotes to highly complex like eukaryotes. As we all must know by now that DNA consists of Nucleic Acids, Sugars and Phosphate, i.e., Nucleotides which are of four types: Adenine (A), Guanine (G), Cytosine (C) and Thymine (T). These nucleotide are arranged one after the other in a genome and encode for proteins which perform functions responsible for a phenotype. A codon is made up of three nucleotides which encodes for an amino acid. There are 64 possible Codons and 20 Amino acids, each amino acid is encoded by more than one codon. Also, three start and one stop codon is there which initiate or terminates the expression processes.

The term ‘Gene’ was coined by Johannsen in 1909 and he named hereditary units of Mendel as ‘Genes’. Subsequently many concepts and views emerged on genes calling them as the hereditary component, thread like structures, etc. Basically, a gene encodes for a single polypeptide has a start codon in the beginning and stop codon at the end. Following points are under consideration of modern gene concept:

- Both male and female parents pass on their genes to the offspring, only cytoplasmic inheritance is pass on from mother to offspring.
- DNA is arranged in chromosomes or linkage groups and in diploid organisms, genes are paired. The phenotype is the result of the behaviour resultant whether dominant or recessive, co-dominance, incomplete dominance, etc.
- Genes present on same chromosomes are generally transmitted together, called as linked genes.
- Gene has a position on chromosomes called locus. Due to chromosomal aberrations, locus may change which can affect phenotypic expression.
- Gamete formation leads to segregation of genes and haploid gametes have only one gene of its type.
- Non-linked genes assort differently and independently.

**Gene Expression Control in Prokaryotes**

Prokaryotes are the simplest among the organisms by being single celled and simpler life processes. However, similar to eukaryotes, they perform the essential functions like eating, digestion, respiration, osmoregulation, excretion, reproduction, etc. For this, or any activity, proteins or enzymes are required which are obtained by the process of transcription and translation of the genetic material. These processes do not occur continuously, instead are being under the control of other
genetic factors which make balance by keeping transcription and translation switch on or off. Hence, the mechanism to control the expression of genes can be considered as positive or negative on the basis of whether it turn on and enhance the gene expression or turn off and repress. This is important as in different processes different proteins are required. Some proteins or enzymes are antagonist in behaviour so, the control mechanisms help in deciding which process should take place. We shall now focus on the different types of gene expression control in prokaryotes.

The mechanism of gene expression was first observed in 1900 when the bacteria was grown under the presence of lactose showed a production of an enzyme which was not observed in the absence of lactose. This gives an idea that certain products are utilised by the bacteria when the machinery of its breakdown is turned on. As in presence of lactose it is being utilised but in its absence, bacteria still can obtain carbon source from different sources. These enzymes are called as adaptive or facultative. More appropriately, adaptive enzymes are called as inducible, whose production is induced by the presence of a factor or substrate, commonly called as inducer. This is an example of positive control of gene expression. Similarly, if an enzyme or protein production is repressed by some compound that enzyme is called as repressible and the compound which repressed is called as repressor. Under negative control or repression, the expression of a gene remain continued till the repressor molecule get activated and repressed the expression. There are some genes whose expression is continuous and not dependent upon the circumstances or environment. These are called constitutive enzymes. Let us discuss the positive and negative control in detail below.

Negative Control of Gene Expression and Lac Operon

Research were started in the year 1940s on lactose metabolism and many scientists did significant contribution. These include, Jacques Monad, Joshua Lederberg, Francois Jacob and Andre L’Wolf. It was evident that in the presence of lactose, enzyme which cause lactose breakdown increases in concentration from few molecules to thousand molecules. However, Lac the absence of lactose decreases the production of lactose metabolising enzyme. Thus, this enzyme is inducible and lactose is a substrate as well as an inducer here. It was determined that lactose metabolism is under the control of three structural genes which are arranged together in a cluster in genome of prokaryotes (Refer Figure 9.1). These are:

- **LacZ Gene**: It encodes for beta-galactosidase, an enzyme which convert lactose (disaccharide) to the monomer units, glucose and galactose (monosaccharides).
- **LacY Gene**: This gene encodes for permease which helps in the entry of lactose through the cell wall of bacteria.
- **LacA Gene**: This gene encodes for transacetylase which known to have probable role in exporting out as by product of lactose breakdown.
The above Figure 9.1 shows the lac operon, in which; (A) components of lac operon; (B) transcription is blocked in the absence of lactose as the repressor binds to the operator; (C) transcription of structural genes into mRNA transcripts in the presence of lactose. Lactose binds allosterically to the repressor and inactivate it.

In order to identify their functional role, researchers produced the various mutants of these three genes one by one to observe the effects. They found that lacZ and lacY plays a key role and if any one of them is not functional then an organism cannot metabolise lactose. Also, these genes, if present, are always reside one after the other in a cluster as lacZ-lacY-lacA. This operon is called as lactose operon or lac operon. Along with three structural genes, it also has a promoter, operator and a repressor gene called as lacI. Promoter and operator together form a regulatory region. LacI produces a repressor molecule which is allosteric in nature. Jacob and Monod worked and suggested that the repressor molecule interacts allosterically with operator to control the gene expression. It do so by affecting the binding of RNA polymerase, thus slow or shut down the transcription. But, in the presence of lactose, lactose molecule binds to the repressor leading to the confirmations change allosterically. As a result, the repressor get deactivated and unable to bind to the operator region. Hence, the operon continued transcription and translation. This justifies the role of lactose as an inducer and substrate as its presence initiates its conversion into glucose and galactose. This operon and regulatory mechanisms of prokaryotes are dependent on the presence of lactose and inactivation of repressor molecule hence, this mechanism is an example of negative control of gene expression in prokaryotes.
There are two different types of mutations which can explain these conditions in detail, I- and O-μ. I- is a mutation which inactivates or produces a mutant repressor, non-functional and O-μ is the constitutive mutation in operator region which leads to constitutive expression of structural genes. In both the cases, the lac operon is turned on and continuous expression of genes which facilitate the lactose breakdown takes place. Following are the combinations of gene mutations and lactose present/absent and their corresponding results of switching on/off the operon.

**Tryptophan Operon and Repressible Gene System**

Unlike lac operon where the presence of lactose was required to switch on the transcription, tryptophan operon works differently. In 1953, Monad and colleagues identified the tryptophan operon and its functioning. In prokaryotes, enzymes necessary for the production of amino acids are present, which activate in the absence of external source of amino acids (in a media or environment). When, there is a scarcity of tryptophan, the operon gets activated and continued. The repressor molecule do not bind to the operator during this time. But, when tryptophan is present and can be obtained from the environment, it acts as a co-repressor. Tryptophan binds to the repressor molecule and activate it allosterically. Afterwards, repressor-trp complex binds to the operator and switches off the gene expression of trp operon. Hence, the control of this operon is different from lac operon as trp is acting as a co-repressor while lactose is an inducer.

Structurally, trp operon consists of five genes whose product is important for trp biosynthesis. These are trpA, trpB, trpC, trpD and trpE. TrpP is a promoter and trpO is an operator where RNA polymerase binds and initiates transcription (Refer Figure 9.2). There is 5' leader sequence following trpP which is of 162 nucleotide long which consists of a site called attenuator. Following this, genes are arranged in form of trpE, trpD, trpC, trpB and trpA. Both leader peptide and attenuator have sequences which induce the formation of terminator loops in the presence of trp. The leader sequence consists of two triplet sequence of UUG which encodes for trp amino acid. Upstream to these sequence, AUG codon is present which encode for start codon. Hence transcription is initiated from here. In the presence of trp, tRNA^trp is abundant and thus participate in forming termination loops. In the absence or scarcity of trp, termination loops do not form, instead, anti-termination hairpin is produced together by leader sequence and attenuator which allow RNA polymerase to surpass and continue transcription. This model of trp operon is present in Escherichia coli. In Bacillus subtilis, there is another way to control the expression of trp operon. Trp RNA-binding Attenuation Protein (TRAP) is a molecule present in the cell. When trp is present, it binds to trp and form a complex. TRAP HAS 11 subunits and each subunit binds to the trp. When fully saturated, it binds to 5’ leader peptide which consists of 11 triplet codons of either UAG or GAG, separated by spacer nucleotides. As a result, it disturb the production of anti-termination hairpin, instead terminator loop is formed which
prevent the expression of trp operon. In the absence or less concentration of trp, TRAP remain unsaturated and do not interfere with the production of anti-termination hairpin, thus transcription proceeds.

**NOTES**

In the above Figure 9.2 shows the trp operon, in which (A) the components of trp operon. (B) in the absence of trp, repressor molecule cannot bind to the operator hence, transcription occur. (C) in the presence of trp, it act as a co-repressor, binds to the repressor and blocks the transcription.

**Arabinose (ara) Operon: Positive and Negative Control**

Arabinose operon is a good example of both positive and negative control of gene expression. It performs the metabolic action of arabinose and consists of three structural genes, araB, araA and araD. Their transcription is controlled by a regulatory protein AraC which is encoded by araC gene. This protein interacts with two regions of the operon, aral and araO2. AraI is a inducer region which promotes the transcription of structural genes. AraO2 is the operator region responsible for resisting the expression of ara operon.

The inducer site consists of Catabolite Activating Protein (CAP protein). Functionally, in the presence of arabinose and cAMP, AraC get activated, dimerised and binds to the inducer site. Also, CAP binds to cAMP and form a complex of CAP-cAMP which interact with CAP region. This initiates the transcription of structural genes. In the absence of arabinose, or cAMP, the dimer of AraC binds to the inducer as well as araO2 region which facilitates the looping of the DNA molecule at the region so that the dimers interact. This loop formation prevents the expression of genes. Thus, ara operon and interaction with AraC shows both positive and negative control of gene expression.
9.2.1 Gene Expression in Eukaryotes

Eukaryotes or multicellular organisms are complex and differ in regulating gene expression as compared with prokaryotes. The differences are as follows:

- Eukaryotic DNA is much longer than prokaryotic DNA so, the packaging of Eukaryotic DNA is much more complex by using histones around which DNA is wrapped in a chromatin structure. One of the control of gene expression is the packaging of eukaryotic DNA. The tight or closed packaging is responsible for no expression or DNA is not accessible for transcription while, the open configuration or loose chromatin folding make DNA accessible for gene expression.

- In Eukaryotes, DNA is arranged in many chromosomes rather one or two. Their transcription and translation is controlled differently. These chromosomes reside in the nucleus bounded by double layered nuclear membrane.

- In Eukaryotes, transcription and translation takes place in different compartments. Transcription takes place inside the nucleus whereas translation proceeds in cytoplasm.

- Also, it is required that the mRNA so produced from transcription should undergo three different processes of post transcriptional modification, i.e., 5’ methylguanosine capping, splicing and addition of poly Adenine (A) tail. After this, the mature transcript proceeds for translation.

- The half-life of Eukaryotic mRNA is much more than Prokaryotic as Prokaryotes harbour wide environmental changes and require to produce proteins faster and decay them soon when not needed.

- Each cell of eukaryotes consists of same DNA molecule but they have specialised tissues, organs, etc., for particular function. Hence, few genes are expressed only in one tissue or organ which bear related function. Rest genome stay non-functional in that tissue. Similarly, different genes are activated in different organs.

- mRNA of eukaryotes is more stable than Prokaryotes, hence translation controls are more active and applied.

This suggests that in case of Eukaryotes, the gene expression control is at the stage of:

- Transcription
- Post Transcriptional Modifications
- Transfer of mRNA to Cytoplasm
- Stability of mRNA
- Translation Control by selecting mRNA for Translation
- Post Translational Modifications

Let us discuss the control of gene expression in Eukaryotes in detail.
Transcriptional Control of Gene Expression

There are three regions which play an important role in modulating the transcription levels or switching it ON or OFF, promoters, enhancers and silencers. These are called as cis regulators as they control the gene regulation by being present on the same chromosome. Trans regulators, on the other hand, are the ones which control gene regulation by being present on another chromosome. These are generally transcription factors and modulators. Promoter region is situated upstream to the genes for initiating transcription. These are the sites recognised by transcriptional machinery. They are up to 100 bp long and consists of the repeat of TATA nucleotides, i.e., Thiamine (T) and Adenine (A). Hence, these are also referred as TATA box. It may also contain other boxes like CAAT, GC, etc. TATA box is situated 25-35 bp above the gene and consensus ATAT region of 7-8 bp long. CAAT BOX is situated 70 -80 bp above the gene and GC box with consensus sequence of GGCGGG is situated 110 bp above the gene. Mutations in promoter regions have shown the effect on the rate of transcription.

Enhancers are another type of cis regulators, controls gene expression by residing either upstream or downstream, even thousand bp away from the gene or within the gene. Enhancers interact with promoters, transcriptional factors, regulatory proteins, etc., to enhance the rate of transcription. If enhancers are placed apart from the gene, they interact by bending of the DNA molecule in a way to come closer to the regulatory region. They have a tendency to enhance transcription by many folds. It has been shown that if enhancer is placed at some new location in the genome, it enhances the expression of other genes too. Also, if an unrelated gene is placed near the enhancer or at a place where a gene was already present, it get enhanced in expression by the activated enhancer. In yeast, regions similar to enhancers are found upstream to the genes they regulate called Upstream Activator Sequences (UAS). They are only found upstream and not downstream to a gene they regulate hence, differ from enhancers.

Apart from these, chromatin remodeling is also important way to control gene expression at transcriptional level. This refers to change in organisation of DNA with histone and non-histone protein. The loose interaction of proteins with the DNA make it accessible to transcriptional factors. This condition is called as euchromatin when DNA is transcriptionally active. When DNA is tightly bound with proteins then it stay transcriptionally inaccessible and called as heterochromatin. This inactivation can also achieved by methylation or acetylation of DNA nucleotides.

Chromatin remodeling make promoters and genes free from histone molecules. One example is SWI/SWF, a nucleosome remodeling complex which has 11 subunits. One subunit allow binding with non-specific DNA another is an ATPase. This is directed to specific targets by the presence of leucine zippers, binding to acetylated and methylated DNA. This can make changes like sliding
the nucleosome or twisting the nucleosome to pull out DNA molecule so that make it accessible for transcription.

Another way is histone modifications. This modifications is catalysed by the enzymes called as Histone Acetyl Transferase (HAT). It adds an acetate group to basic amino acid of histone tail. This interaction decrease the attraction between basic amino acids and acidic DNA molecule making it accessible to transcriptional factors. Reversing this reaction, Histone Deacetylase (HDAC) removes acetyl group from histone proteins. Along with acetylation, methylation and phosphorylation are also other mode of modifications. Also, transcriptional factors interact with other factors or co-activators which modulate their activity.

Methylation of DNA is an important way to reduce the gene expression. The methyl group protrudes out from the DNA molecule and affects the binding of RNA polymerase and transcription factors. Methylation occur more often at CG repeats. To determine the presence of methylated residues, a restriction enzyme HpaII can be used which recognises CCGG region. If second C is methylated, the enzyme cannot make a cut in recognition sequence, evident of methylation. Hence, methylation and gene expression are opposite as more the methylation, lesser will be gene expression. One example is X chromosome inactivation in human females for dosage compensation with human males. The inactivated X chromosome show high level of methylation than active X chromosome. Also, methylation can be observed in genes sequences which remain silent in tissues or organs where they do not have any function to perform.

**Regulation at Post-Transcriptional Modification**

As mentioned in above section, we mentioned that there are three types of post transcriptional modifications, 5' -methyl guanosine cap, splicing and 3' poly A tail. Among these, regulation of gene expression can be attained at the level of alternative splicing by which different proteins can be manufactured and on the basis of stability of mRNA.

In alternative splicing, a single transcript can give rise to multiple proteins depends on the requirements of the cell. This is achieved as few exons become introns and introns become exons during splicing of the regions from the primary transcripts. In humans, hundred thousands of proteins are encoded by nearly 30,000 genes because of alternative splicing mechanism. It maintain the regulation and also provide direction to the transcript for production of a particular protein. By this way, gene regulation can also be controlled after the transcription.

mRNA stability is defined as the turnover of mRNA or its half life till it undergone translation. Also, RNA interference is a method identified to control the translation of a mRNA. There are short RNA of nearly 21 nucleotides called as siRNA (small interfering RNA) which participate in cleavage of mRNA. The process starts when a double stranded RNA (nearly 70 bp long) binds to a protein bearing RNAase activity called as dicer. It cleaves the RNA in short RNA called as siRNA. Then, the both helix of double stranded siRNA unwinds and the anti-
Gene Concept and Lactose System

NOTES

Another example is miRNA (micro RNA), which are 70-110 bp long and binds with dicer and target mRNA. The interaction is not exact and even if the 19-24 nucleotides paired, the target mRNA is cleaved and degraded. Thus, its translation is blocked. Phage are those viruses which infect and harbour the DNA of bacteria and archaea and may modulate the eukaryotic genome also. They enter in the cell and through lysogenic cycle they are capable to incorporate their genome into the genome of host organism. RNA interference is initially evolved as a mechanism to protect against the phage or virus attack (it can be lytic also) to protect the species and maintain its survival. To prevent, miRNA and siRNA which consists of foreign sequences or random sequences encounter phage genome and destroy them. In bacteria, there are specific regions called as CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) elements which consists of repeatedly present spacers and repeat regions. When a phage infects a bacteria, it keep the memory of that encounter by keeping a region of short nucleotide stretch (20-24 bp) integrated in its genome surrounded by repeat regions called as spacer. On the repeat attack, the bacteria produces mRNA of the complementary spacer and perform the degradation of phage genome by the process similar to RNA interference.

There is another type of regulation is identified from the RNA molecules called as RNA Directed DNA Methylation (RdDM). These RNA molecules binds to the complementary DNA and initiates its methylation. Thus, regulating their expression by methylating the target DNA.

There are proteins called as RNA-Binding Proteins or RBPs which can bind to the untranslated regions, or UTRs of mRNA. UTRs are the region which cannot be translated and help in mRNA localization, stability, and protein translation. Region present before the protein-coding region is 5' UTR, whereas after the coding region is 3' UTR. Binding of proteins at these regions can greatly affect the stability of mRNA as it may enhance or decrease.

Translation and Post-Translational Control of Expression

Translation is a process by which a mature mRNA is converted into designated protein molecule in eukaryotes. Regulation of this process is achieved by sending a message for proceeding or not for translation. One such example is synthesis of $\alpha$- and $\beta$-tubulin which are subunits of microtubules. Colchicine treatment to the cell can induce the disassembly of microtubules and accumulate the $\alpha$- and $\beta$-tubulin in the cell. This shut down the production of $\alpha$- and $\beta$-tubulin in the cell. Oppositely, treatment with a drug called vinblastine causes microtubule disassembly but initiates the production of $\alpha$- and $\beta$-tubulin in the cell. This is because vinblastine precipitates these subunits and make them unavailable to the cell which in turn
induces their production in the cell. This regulation is takes place at translational level called as autoregulation.

After translational, a protein undergoes proper binding, folding of the motifs in Golgi complex and its transportation. When a cell or extracellularly, a protein is required then its processing takes place without any delay and smoothly by the cell but if not then a well formed protein or enzyme is processed to inactivated by the process of phosphorylation, acetylation, methylation, dephosphorylation, etc. By this way, an enzyme get deactivated but not permanently as reversing the condition make the enzyme available again for the reaction.

Check Your Progress

1. Explain the term gene. How are they characterized?
2. Who coined the term gene? How coding is done by genes?
3. When gene expression was first observed? What are adaptive, facultative and repressible enzymes?
4. How the presence and absence of lactose effects the production of lactose metabolising enzyme? Which three structural genes determine the lactose metabolism?
5. Explain about cis regulators and trans regulators that control the gene regulation.
6. What is translation process by which a mature mRNA is converted into designated protein molecule? How it is regulated?

9.3 REGULATION OF BACTERIAL GENE EXPRESSION

Cell is defined as the basic unit of life. It grows and divides in different phases of cycle. Cell cycle consists of mainly four phases, Growth phase 1 (G1), Synthesis phase (S), Growth phase 2 (G2) and Mitosis (M) or Meiosis. In G1 phase, cell prepares for the next stage, i.e., DNA synthesis or replication in S phase. In S phase, chromosomal DNA is replicated. This is followed by G2 phase where cell continue to grow and increase the amount of organelles, other molecules so that in next stage it can be divided. In M phase, the chromosomes get condensed (prophase) followed by aligning at equatorial pole (metaphase). Then, the sister chromatids get separated to different poles by the action of kinetochores (anaphase). After this, division takes place in (telophase) and cytoplasm is also equally divided (cytokinesis).

The cell cycle progression is continued in this way till the cell is signaled to undergone to the progression arrest from internal factors. In such cases, cell
undergoes programmed cell death by activating different factors in a cascade and triggering their own death called as apoptosis. Also, during tissue injury, cell also suffer a sudden death which is called as necrosis. We shall discuss these in later sections of this unit. It is important to note here that all the cells are programmed to death when the signal arrives and they cannot be immortal. This is achieved by the regulation of cell cycle by the action of internal factors or external trigger. Although, if a cell is failed to regulate the cell cycle, the cell undergo uncontrollable growth and divisions which can lead to lethal conditions like benign tumor or malignant tumor. This raised a good point that cell cycle regulation is very important. In next sections, we shall focus that how regulation is maintained.

**Cell Cycle Regulation**

During cell cycle progression, if a cell signalled to stop dividing and growing, it arrest in stage G1. Then it enters in a G0 stage where no growth or division takes place, instead, a cell stay and perform its functions normally. Most differentiated cells stay in G0 stage indefinitely and do not enters in cell cycle again. However, if signalled from the external environments then they may enter in the cell cycle again. During the loss of this regulation, a cell proceeds with all the four stages of cell cycle one after the other continuously and cannot stop growing and dividing.

In a cell cycle, there are three points where cell monitors and respond to its internal equilibrium before moving to the next stage. These are between G1 and S phase G1/S, G2 and M phase G2/M and M phase. These are also called as checkpoints as they check the progression and can arrest the cell for proceeding. Literally, a checkpoint is a stage in the eukaryotic cell cycle during which cell determines internal and external signals and decides whether or not to move forward to the next phase. Different check points monitors differently and are explained as below:

- **G1/S Check Point:** It checks the size of the cell, growth factors, nutrients and determine any damage to DNA molecule which should be repaired before it proceeds to the next stage. Size of the cell should be enough large which can support for division later. Another notable factor is that whether a cell is reside on proper extracellular matrix which supports its division. The cycle remain halted for longer periods till the cellular machinery corrects the faults and signal for the progression. Initially, if a cell has normal size and no defects in DNA then, it continue to the next phase of cycle. It is also important because at this point cell decides that whether to or not proceed for the division. If this check point is crossed then cell has to cross the entire cycle. If the fault is not corrected then the cell enters in G0 phase and remain non-dividing.
• **G2/M Check Point:** In this check point, physiological conditions of the cell is monitored. The cell is analyzed for the complete DNA replication and whether DNA is damaged. If so, then cell cycle stay arrested at this point and cannot be proceeded to the next phase which is the mitotic division of the cell. Hence, these checkpoints prevent the spread of any damage or mistake in the DNA which may affect the physiology of the organism.

• **M Check Point:** It is third checkpoint and it is monitored during the time of mitosis. It checks the formation of spindle fibers and their proper attachment with the kinetochores which helps in pulling the sister chromatids towards the opposite poles. If spindle fibers are not formed properly or the attachment is not proper then the mitosis stage is arrested.

If the DNA damage is not corrected at the checkpoints or the errors are not possible to remove then cell trigger the programmed cell death to ensure that damaged or faulty DNA or cell should not proliferate.

**Role of Cyclins and Cyclin-Dependent Kinases (CDKs) in Cell Cycle Regulation**

We discussed about the different checkpoints of cell cycle progression. In this section we will focus on the molecules which mainly control the arrest at the checkpoints. The regulation of cell-cycle is performed by two type of proteins:

**Cyclins:** Cyclins are one of the important cell cycle regulators. Four basic types of cyclins are found in humans and other eukaryotes. These are G1, G1/S, S and M cyclins. According to their names, each cyclin is related to particular phase or transition of the cell cycle. The S cyclin promotes the DNA replication during S phase. A cyclin’s concentration also varies throughout the cell cycle, i.e., their concentration stays low throughout the cell cycle but increases during the phase or transition at which they required. The M cyclin is present in highest concentration during M phase. S cyclin is at highest concentration during S and G2 phase also. G1 cyclin is an exception as it stays in a similar concentration throughout the cycle as needed at many points. Hence, cyclins are synthesized during different stages of cell cycle and subsequently destroyed also to maintain the progression. These cyclins are also named as cyclin A, cyclin B, cyclin D and cyclin E.

**Cyclin-Dependent Kinases (CDKs):** To perform functions of activating or deactivating many proteins inside the cell, cyclins bind to a specialized group of proteins called as Cyclin Dependent Kinases (CDKs). Without cyclins they remain inactive but binding with cyclins activates them making it a functional enzyme and allowing it to modify target proteins. As the name suggests, CDKs are the kinases which phosphorylates or attach phosphate groups to target proteins. These target proteins are generally specific to specific kinases. This affect the activity of target protein by making them more or less active. Cyclins not only activate CDKs but
also directs them to a specified set of proteins for phosphorylation so that the cell corrects the errors. Concentration of CDKs remain constant throughout the cell cycle but it is the concentration of cyclins which make them active and inactive. CDKs itself must also be phosphorylate to get activated.

Cyclins are synthesized during different stages of cell cycle and subsequently destroyed also to maintain the progression. Cyclins bind to CDK molecules and form a cyclin/CDK complex which is specific for a particular check point. The main function of this complex is to phosphorylate (due to the presence of kinase) the other proteins and activate the changes necessary for helping the cell for progression. Cyclins and CDKs are evolutionarily conserved and are found in many different types of species, from yeast to humans. The details of the system vary as, yeast has just one CDK molecule, while humans and other mammals have multiple CDKs for different stages of the cell cycle.

**Maturation Promoting Factor (MPF):** In 1980s, MPF, was discovered as a Cdk bound to its M cyclin partner. During M phase, M cyclin accumulates, it binds to MPF which add phosphate group to different proteins in nuclear envelope, resulting its breakdown. It also promote chromosome condensation and other events of M phase. After its work is done, MPF promotes its own destruction by activating a complex called as Anaphase-Promoting Complex/Cyclosome (APC/C). It is a protein complex that degrade M cyclins during in anaphase. Destruction of M cyclins pushes the cell to enter in mitosis. This complex also causes destruction of the proteins which held sister chromatids together thus, allowing them to separate during anaphase and move to opposite poles. APC/C do not phosphorylate its target proteins instead, it adds a small protein tag called as Ubiquitin (Ub). This target can now sent to the proteasome in the cell, which is a recycle bin of the cell, and degraded. During metaphase, this complex also sends signals to destroys cohesin, the protein glue that holds sister chromatids together. This is done by first adding ubiquitin tag to a protein called securing which cause its degradation. Securin binds and inactivates a protein called separase. Hence, separase is now become active, it cleaves the cohesion and allows the sister chromatids to separate.

**CDK and Cyclin Complexes a Different Cell Cycle Stages**

As mentioned above, different CDK and cyclin complexes regulate at different checkpoints (Refer Figure 9.3). These complexes are:

- In G1 phase: CDK4 or CDK6 with cyclin D
- In G1/S phase: CDK2 with cyclin E
- In S phase: CDK2 with cyclin A
- In G2/M phase: CDK1 (CDC2) with cyclin B or A
- In M phase: CDK1 with cyclin B or A
In G1 phase, the complex of CDK4/cyclin D activates expression of those genes whose products required during DNA replication during S phase. It also activates a protein called as pRB which initiates the expression of number of genes that encode for cyclins E and A, Cdk1, and proteins involved in replication. In G1/S phase, cyclin E form a complex with CDK2 and initiates replication. Similarly, as cell cycle progress, cyclin A forms a complex with CDK2 and promotes replication. Then, during transition from G2 to M phase, CDK1/ cyclin A and CDK1/cyclin B complexes phosphorylate variety of proteins which are necessary for mitosis, such as cytoskeletal proteins, histones, and proteins of the nuclear envelope.

9.4 LACTOSE SYSTEM - COORDINATE REGULATION

Operon

- The Basic Concept Structural Gene = Gene that Codes for a Polypeptide.
- A regulated cluster of adjacent structural genes with related functions.
- Has a single promoter region, so an RNA polymerase will transcribe all structural genes on an all-or-none basis.
- Transcription produces a single polycistronic mRNA with coding sequences for all enzymes in a metabolic pathway.

Polycistronic mRNA

- A large mRNA molecule that is a transcript of several genes. Contains stop and start codons for the translation of each polypeptide.
Operator
- A DNA segment between an operon’s promoter and structural genes, which controls access of RNA polymerase to structural genes.
- Acts as an ON/OFF switch for movement of RNA polymerase and transcription of the operon’s structural genes.

Repressor
- Specific protein that binds to an operator and blocks transcription of the operon.
- Blocks attachment of RNA polymerase to the promoter.
- Repressors are encoded by regulatory genes.

Regulatory Genes
- Genes that code for repressor or regulators of other genes.
- Are often located some distance away from the operons they control.
- Are involved in switching on or off the transcription of structural genes.

Corepressor
- A molecule, usually a metabolite that binds to a repressor protein, causing the repressor to change into its active conformation.
- Only the repressor-corepressor complex can attach to the operator and turn off the operon.

Inducer
- An inducer is a molecule that regulates gene expression. An inducer can bind to protein repressors or activators.
- Inducers function by disabling repressors.
- The gene is expressed because an inducer binds to the repressor.
- The binding of the inducer to the repressor prevents the repressor from binding to the operator. RNA polymerase can then begin to transcribe operon genes.

Repressible Versus Inducible Enzymes
Repressible Enzymes: Enzymes which have their synthesis inhibited by a metabolite (for example, tryptophan).
Inducible Enzymes: Enzymes which have their synthesis stimulated or induced by specific metabolites (for example, lac operon).
Positive control of a regulatory system occurs only if an activator molecule interacts directly with the genome to turn on transcription.

**CRP (cAMP Receptor Protein)**

- A protein that binds within an operon’s promoter region and enhances the promoter’s affinity for RNA polymerase.
- When glucose is missing, the cell accumulates cyclic AMP (cAMP), a nucleotide derived from ATP. cAMP activates CAP so that it can bind to the lac promoter.
- When glucose concentration rises, glucose catabolism decreases the intracellular concentration of cAMP. Thus, cAMP releases CRP.

**Catabolite Repression**

- Repression of a variety of unrelated catabolic enzymes when cells are grown in a medium containing glucose.

**Enhancer/Silencer**

- Regulatory regions on eukaryotic DNA that bind activator/repressor proteins controlling single gene expression.

### 9.5 LAC COMPONENTS

An operon is a cluster of functionally-related genes that are controlled by a shared operator. Operons consist of multiple genes grouped together with a promoter and an operator. Operons are present in prokaryotes (bacteria and archaea), but are absent in eukaryotes. In some situations multiple operons are controlled by the same regulatory protein; in these cases the operons form a regulon. Operons were first identified as a mode of gene expression control in 1961 by François Jacob and Jacques Monod.

**Operon Structure**

Operons are regions of DNA that contain clusters of related genes. They are made up of a promoter region, an operator, and multiple related genes. The
operator can be located either within the promoter or between the promoter and the genes. RNA polymerase initiates transcription by binding to the promoter region. The location of the operator is important as its regulation either allows or prevents transcription of the genes into mRNA.

**Operon Function**

An operon is a complete package for gene expression and synthesis of polypeptides. By combining the related genes, all polypeptides required for a specific function are synthesized in response to a single stimulus. For example, the bacterium *Escherichia coli* contains a number of genes clustered into operons and regulons: the Lac operon which is involved in lactose degradation, the Trp operon which is involved in tryptophan biosynthesis, and the His operon which is involved in histidine biosynthesis. These operons are turned on when the gene products are needed.

**Positive and Negative Control**

Operons can be under negative or positive control. Negative control involves turning off the operon in the presence of a repressor; this can be either repressible or inducible. A repressible operon is one that is usually on but which can be repressed in the presence of a repressor molecule. The repressor binds to the operator in such a way that the movement or binding of RNA polymerase is blocked and transcription cannot proceed. An inducible operon is one that is usually off. In the absence of an inducer the operator is blocked by a repressor molecule. When the inducer is present it interacts with the repressor protein, releasing it from the operator and allowing transcription to proceed. Repressible operons are generally involved in anabolic pathways, or the synthesis of an essential component, while inducible operons are generally involved in catabolic pathways, or the breakdown of a nutrient. Positive control of an operon is when gene expression is stimulated by the presence of a regulatory protein.

**Lac Operon**

The Lac operon is the classic operon example, and is responsible for the degradation of the milk protein lactose. The Lac operon is an inducible operon; in the absence of lactose the operon is blocked by a repressor protein. The operon is made up of a promoter with operator, and three genes (lacZ, lacY, and lacA) which encode β-galactosidase, permease, and transacetylase. The three genes are involved in the breakdown of lactose into its metabolites: β-galactosidase breaks lactose down into glucose and galactose, while the other two proteins aid in the metabolic process. The expression of the Lac operon is controlled by the regulatory gene lacI, located immediately adjacent to the promoter region. LacI encodes an allosteric repressor protein that keep the Lac operon 'OFF'.

In order for the Lac operon to be turned on, an inducer molecule must inactivate the repressor protein. The inducer molecule in this system is allolactose,
an isomer of lactose. When lactose and its isomer are present in the cell, allolactose will bind to allosteric sites on the repressor protein, changing its conformation and rendering it inactive. As the repressor protein detaches from the operator, RNA polymerase can bind to the promoter, transcription can occur, and the three lactose degradation genes can be synthesized (Refer Figure 9.4).

![Fig. 9.4 Structure of the Lac Operon and the Adjacent lacR Repressor Gene](image)

The Lac operon is also under positive gene regulation. While the removal of the repressor protein in the presence of lactose is required for synthesis of the lacZ, lacY, and lacA genes, the gene expression will remain low. The level of gene expression is controlled by the amount of the preferred energy source, glucose, in the cell. This control is regulated by an allosteric regulatory protein, Catabolite Activator Protein (CAP). When glucose levels in the cell are low, the organic molecule cyclic AMP is in high concentration. Cyclic AMP activates CAP by binding to the allosteric sites, causing CAP to attach to the Lac operon promoter. Unlike the repressor proteins, binding of CAP to the Lac operon stimulates gene expression. When the cell glucose levels increase, the cyclic AMP levels in the cell decrease, and the activator protein will disassociate from the promoter. Transcription will return to low levels, or will turn off if the repressor protein reattaches (Refer Figure 9.5).

![Fig. 9.5 Lac Operon When Both Glucose and Lactose are Present](image)

The Figure 9.6 depicts the Lac operon and how its gene expression is under both positive and negative control.
The lac operon exhibits both systems. It is a negative control system because expression is typically blocked by an active repressor (the lac repressor) that turns off transcription. The lac repressor binds to the operator region and negatively controls (prevents) transcription.

However, when CAP (Catabolite Activating Protein) binds upstream of this operator region near the promoter and transcription increases, this is an example of a positive control system. We see this positive control of transcription happen when glucose levels decline.

**Positive Regulation of the Lac Operon**

- A bacterium’s environment is too complex for its genes to be controlled by one signal.
- Other factors besides lactose affect the expression of the lac genes, such as the availability of glucose.
- Glucose, in *Escherichia coli* preferred energy source. Other sugars can serve as the main or sole nutrient, but extra steps are required to prepare...
Gene Concept and Lactose System

Expression of the Lac Operon when Both Glucose and Lactose are Present

- A regulatory mechanism known as catabolite repression restricts expression of the genes required for catabolism of lactose, arabinose, and other sugars in the presence of glucose, even when these secondary sugars are also present.
- The effect of glucose is mediated by cAMP, as a co-activator, and an activator protein known as cAMP receptor protein, or CRP (the protein is sometimes called CAP, for catabolite gene activator protein).
- CRP is a homodimer (subunit Mr 22,000) with binding sites for DNA and cAMP.
- Binding is mediated by a helix-turn-helix motif within the protein’s DNA-binding domain. When glucose is absent, CRP-cAMP binds to a site near the lac promoter and stimulates RNA transcription 50-fold.
- CRP-cAMP is therefore a positive regulatory element responsive to glucose levels.
- CRP-cAMP has little effect on the lac operon when the Lac repressor is blocking transcription, and dissociation of the repressor from the lac operator has little effect on transcription of the lac operon unless CRP-cAMP is present to facilitate transcription; when CRP is not bound, the wild-type lac promoter is a relatively weak promoter.
- The open complex of RNA polymerase and the promoter does not form readily unless CRP-cAMP is present.
- CRP interacts directly with RNA polymerase through the polymerase’s α subunit.
- The effect of glucose on CRP is mediated by the cAMP interaction. CRP binds to DNA most avidly when cAMP concentrations are high.
- In the presence of glucose, the synthesis of cAMP is inhibited and efflux of cAMP from the cell is stimulated. As [cAMP] declines, CRP binding to DNA declines, thereby decreasing the expression of the lac operon.
- Strong induction of the lac operon therefore requires both lactose (to inactivate the lac repressor) and a lowered concentration of glucose (to trigger an increase in [cAMP] and increased binding of cAMP to CRP.
- CRP and cAMP are involved in the coordinated regulation of many operons, primarily those that encode enzymes for the metabolism of secondary sugars, such as lactose and arabinose.
- A network of operons with a common regulator is called a regulon.
CAP must complex cAMP before binding to the promoter of the lac operon. The binding of cAMP–CAP to the promoter activates transcription by facilitating the binding of RNA polymerase. Levels of cAMP are inversely related to glucose: low glucose stimulates high cAMP; high glucose stimulates low cAMP.

**Check Your Progress**

7. What is cell cycle progression?
8. Explain the cyclin process.
9. Give the features of operator, repressor and regulatory genes.
10. What is operon?
11. Explain about the term Lac operon.
12. What is catabolite repression?
13. How the effect of glucose is mediated by cAMP?

### 9.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Genes are the small segments present on DNA. They encode for proteins which describe an individual’s phenotype. They are characterized by the presence of start codon in the beginning and stop codon at the end. There are few genes which do not code for functional proteins and many are hypothetical due to their unknown functions.
2. The term Gene was coined by Johannsen in 1909 and he named hereditary units of Mendel as ‘Genes’.

A gene may code for more than one proteins in Eukaryotes by the phenomenon known as alternative splicing. Therefore, the genes are the basic unit of heredity in all the organisms from simpler like Prokaryotes to highly complex like Eukaryotes. As we all must know by now that DNA consists of Nucleic Acids, Sugars and Phosphate, i.e., Nucleotides which are of four types: Adenine (A), Guanine (G), Cytosine (C) and Thymine (T). These nucleotide are arranged one after the other in a genome and encode for proteins which perform functions responsible for a phenotype.

3. The mechanism of gene expression was first observed in 1900 when the bacteria was grown under the presence of lactose showed a production of an enzyme which was not observed in the absence of lactose. Certain products are utilised by the bacteria when the machinery of its breakdown is turned on. In presence of lactose it is being utilised but in its absence, bacteria still can obtain carbon source from different sources. These enzymes are called as adaptive or facultative. More appropriately, adaptive enzymes are called as inducible, whose production is induced by the presence of a factor or substrate, commonly called as inducer. This is an example of positive control of gene expression. Similarly, if an enzyme or protein production is repressed by some compound that enzyme is called as repressible and the compound which repressed is called as repressor. Under negative control or repression, the expression of a gene remain continued till the repressor molecule get activated and repressed the expression.

4. It was evident that in the presence of lactose, enzyme which cause lactose breakdown increases in concentration from few molecules to thousand molecules. However, Lac the absence of lactose decreases the production of lactose metabolising enzyme. Thus, this enzyme is inducible and lactose is a substrate as well as an inducer here. It was determined that lactose metabolism is under the control of three structural genes which are arranged together in a cluster in genome of prokaryotes. These are:

LacZ gene: It encodes for beta-galactosidase, an enzyme which convert lactose (disaccharide) to the monomer units, glucose and galactose (monosaccharides).

LacY gene: This gene encodes for permease which helps in the entry of lactose through the cell wall of bacteria.

LacA gene: This gene encodes for transacetylase which known to have probable role in exporting out as by product of lactose breakdown.

5. There are three regions which plays an important role in modulating the transcription levels or switching it ON or OFF, promoters, enhancer and silencers. These are called as cis regulators as they control the gene regulation by being present on the same chromosome.
Trans regulators, on the other hand, are the ones which control gene regulation by being present on another chromosome. These are generally transcription factors and modulators. Promoter region is situated upstream to the genes for initiating transcription. These are the sites recognised by transcriptional machinery. They are up to 100 bp long and consists of the repeat of TATA nucleotides, i.e., Thiamine (T) and Adenine (A).

6. Translation is a process by which a mature mRNA is converted into designated protein molecule in eukaryotes. Regulation of this process is achieved by sending a message for proceeding or not for translation. One such example is synthesis of $\alpha$- and $\beta$-tubulin which are subunits of microtubules. Colchicine treatment to the cell can induce the disassembly of microtubules and accumulate the $\alpha$- and $\beta$-tubulin in the cell.

7. During cell cycle progression, if a cell signalled to stop dividing and growing, it arrest in stage G1. Then it enters in a G0 stage where no growth or division takes place, instead, a cell stay and perform its functions normally. Most differentiated cells stay in G0 stage indefinitely and do not enters in cell cycle again. However, if signalled from the external environments then they may enter in the cell cycle again. During the loss of this regulation, a cell proceeds with all the four stages of cell cycle one after the other continuously and cannot stop growing and dividing.

In a cell cycle, there are three points where cell monitors and respond to its internal equilibrium before moving to the next stage. These are between G1 and S phase G1/S, G2 and M phase G2/M and M phase. These are also called as checkpoints as they check the progression and can arrest the cell for proceeding.

8. Cyclins are one of the important cell cycle regulators. Four basic types of cyclins are found in humans and other eukaryotes. These are G1, G1/S, S and M cyclins. According to their names, each cyclin is related to particular phase or transition of the cell cycle. The S cyclin promotes the DNA replication during S phase. A cyclin’s concentration also varies throughout the cell cycle, i.e., their concentration stays low throughout the cell cycle but increases during the phase or transition at which they required. The M cyclin is present in highest concentration during M phase. S cyclin is at highest concentration during S and G2 phase also. G1 cyclin is an exception as it stays in a similar concentration throughout the cycle as needed at many points. Hence, cyclins are synthesized during different stages of cell cycle and subsequently destroyed also to maintain the progression. These cyclins are also named as cyclin A, cyclin B, cyclin D and cyclin E.

9. Operator
   - A DNA segment between an operon’s promoter and structural genes, which controls access of RNA polymerase to structural genes.
   - Acts as an on/off switch for movement of RNA polymerase and transcription of the operon’s structural genes.
Repressor
- Specific protein that binds to an operator and blocks transcription of the operon.
- Blocks attachment of RNA polymerase to the promoter.
- Repressors are encoded by regulatory genes.

Regulatory Genes
- Genes that code for repressor or regulators of other genes.
- Are often located some distance away from the operons they control.
- Are involved in switching on or off the transcription of structural genes.

10. An operon is a cluster of functionally-related genes that are controlled by a shared operator. Operons consist of multiple genes grouped together with a promoter and an operator. Operons are present in prokaryotes (bacteria and archaea), but are absent in eukaryotes. In some situations multiple operons are controlled by the same regulatory protein; in these cases the operons form a regulon. Operons were first identified as a mode of gene expression control in 1961 by François Jacob and Jacques Monod.

11. The Lac operon is the classic operon example, and is responsible for the degradation of the milk protein lactose. The Lac operon is an inducible operon; in the absence of lactose the operator is blocked by a repressor protein. The operon is made up of a promoter with operator, and three genes (lacZ, lacY, and lacA) which encode β-galactosidase, permease, and transacetylase.

12. A regulatory mechanism known as catabolite repression restricts expression of the genes required for catabolism of lactose, arabinose, and other sugars in the presence of glucose, even when these secondary sugars are also present.

13. The effect of glucose is mediated by cAMP, as a co-activator, and an activator protein known as cAMP receptor protein, or CRP (the protein is sometimes called CAP, for Catabolite gene Activator Protein).

9.7 SUMMARY
- Genes are the small segments present on DNA. They encode for proteins which describes an individual’s phenotype. They are characterized by the presence of start codon in the beginning and stop codon at the end.
- There are few genes which do not code for functional proteins and many are hypothetical due to their unknown functions.
- A gene may code for more than one proteins in Eukaryotes by the phenomenon known as alternative splicing. In short, genes are the basic
unit of hereditary in all the organisms from simpler like Prokaryotes to highly complex like Eukaryotes. As we all must know by now that DNA consists of Nucleic Acids, Sugars and Phosphate, i.e., Nucleotides which are of four types: Adenine (A), Guanine (G), Cytosine (C) and Thymine (T). These nucleotides are arranged one after the other in a genome and encode for proteins which perform functions responsible for a phenotype.

- A codon is made up of three nucleotides which encodes for an amino acid. There are 64 possible codons and 20 amino acids, each amino acid is encoded by more than one codon. Also, three start and one stop codon is there which initiate or terminates the expression processes.

- The term Gene was coined by Johannsen in 1909 and he named hereditary units of Mendel as ‘Genes’. Basically, a gene encodes for a single polypeptide has a start codon in the beginning and stop codon at the end.

- Gene has a position on chromosomes called locus. Due to chromosomal aberrations, locus may change which can affect phenotypic expression.

- The mechanism of gene expression was first observed in 1900 when the bacteria was grown under the presence of lactose showed a production of an enzyme which was not observed in the absence of lactose. This gives an idea that certain products are utilised by the bacteria when the machinery of its breakdown is turned on.

- In presence of lactose the produced enzyme is being utilised but in its absence, bacteria still can obtain carbon source from different sources. These enzymes are called as adaptive or facultative.

- The adaptive enzymes are called as inducible, whose production is induced by the presence of a factor or substrate, commonly called as inducer. This is an example of positive control of gene expression.

- If an enzyme or protein production is repressed by some compound that enzyme is called as repressible and the compound which repressed is called as repressor. Under negative control or repression, the expression of a gene remain continued till the repressor molecule get activated and repressed the expression.

- There are some genes whose expression is continuous and not dependent upon the circumstances or environment.

- It was evident that in the presence of lactose, enzyme which cause lactose breakdown increases in concentration from few molecules to thousand molecules. However, Lac the absence of lactose decreases the production of lactose metabolising enzyme. Thus, this enzyme is inducible and lactose is a substrate as well as an inducer here.

- It was determined that lactose metabolism is under the control of three structural genes which are arranged together in a cluster in genome of prokaryotes.
- LacZ gene encodes for beta-galactosidase, an enzyme which converts lactose (disaccharide) to the monomer units, glucose and galactose (monosaccharides).
- LacY gene encodes for permease which helps in the entry of lactose through the cell wall of bacteria.
- LacA gene encodes for transacetylase which is known to have a probable role in exporting out as byproduct of lactose breakdown.
- In prokaryotes, enzymes necessary for the production of amino acids are present, which activates in the absence of external source of amino acids (in a media or environment). When, there is scarcity of tryptophan, the operon gets activated and continued. The repressor molecule does not bind to the operator during this time.
- When tryptophan is present and can be obtained from the environment, it acts as a co-repressor. Tryptophan binds to the repressor molecule and activate it allosterically, afterwards, repressor-trp complex binds to the operator and switched off the gene expression of trp operon. Hence, the control of this operon is different from lac operon as trp is acting as a co-repressor while lactose is an inducer.
- Structurally, trp operon consists of five genes whose product is important for trp biosynthesis. These are trpA, trpB, trpC, trpD and trpE. TrpP is a promoter and trpO is an operator where RNA polymerase binds and initiates transcription.
- Trp RNA-binding Attenuation Protein (TRAP) is a molecule present in the cell. When trp is present, it binds to trp and form a complex.
- Arabinose operon is a good example of both positive and negative control of gene expression. It performs the metabolic action of arabinose and consists of three structural genes, araB, araA and araD. Their transcription is controlled by a regulatory protein AraC which is encoded by araC gene.
- In eukaryotes, DNA is arranged in many chromosomes rather than one or two. Their transcription and translation is controlled differently. These chromosomes reside in the nucleus bounded by double layered nuclear membrane.
- In eukaryotes, transcription and translation take place in different compartments. Transcription takes place inside the nucleus whereas translation proceeds in cytoplasm.
- There are three regions which play an important role in modulating the transcription levels or switching it on or off, promoters, enhancer and silencers. These are called cis-regulators as they control the gene regulation by being present on the same chromosome.
- Trans regulators are the ones which control gene regulation by being present on another chromosome. These are generally transcription factors and
modulators. Promoter region is situated upstream to the genes for initiating transcription. These are the sites recognised by transcriptional machinery. They are up to 100 bp long and consists of the repeat of TATA nucleotides, i.e., Thiamine (T) and Adenine (A).

- Methylation of DNA is an important way to reduce the gene expression. The methyl group protrudes out from the DNA molecule and affects the binding of RNA polymerase and transcription factors. Methylation occurs more often at CG repeats.

- In humans, hundred thousands of proteins are encoded by nearly 30,000 genes because of alternative splicing mechanism. It maintains the regulation and also provides direction to the transcript for production of a particular protein. By this way, gene regulation can also be controlled after transcription.

- There is another type of regulation identified from the RNA molecules called as RNA Directed DNA Methylation (RdDM). These RNA molecules bind to the complementary DNA and initiates its methylation. Thus, regulating their expression by methylating the target DNA.

- There are proteins called as RNA-Binding Proteins or RBPs which can bind to the untranslated regions, or UTRs of mRNA. UTRs are the region which cannot be translated and help in mRNA localization, stability, and protein translation. Binding of proteins at these regions can greatly affect the stability of mRNA as it may enhance or decrease.

- Translation is a process by which a mature mRNA is converted into designated protein molecule in eukaryotes. Regulation of this process is achieved by sending a message for proceeding or not for translation. One such example is synthesis of α- and β-tubulin which are subunits of microtubules. Colchicine treatment to the cell can induce the disassembly of microtubules and accumulate the α- and β-tubulin in the cell.

- During cell cycle progression, if a cell signalled to stop dividing and growing, it arrest in stage G1. Then it enters in a G0 stage where no growth or division takes place, instead, a cell stay and perform its functions normally. Most differentiated cells stay in G0 stage indefinitely and do not enter in cell cycle again. However, if signaled from the external environments then they may enter in the cell cycle again. During the loss of this regulation, a cell proceeds with all the four stages of cell cycle one after the other continuously and cannot stop growing and dividing.

- In a cell cycle, there are three points where cell monitors and respond to its internal equilibrium before moving to the next stage. These are between G1 and S phase G1/S, G2 and M phase G2/M and M phase. These are also called as checkpoints as they check the progression and can arrest the cell for proceeding.
- G1/S check point checks the size of the cell, growth factors, nutrients and determine any damage to DNA molecule which should be repaired before it proceeds to the next stage.
- G2/M check point, the physiological conditions of the cell is monitored. The cell is analyzed for the complete DNA replication and whether DNA is damaged.
- Cyclins are one of the important cell cycle regulators. Four basic types of cyclins are found in humans and other eukaryotes. These are G1, G1/S, S and M cyclins.
- According to their names, each cyclin is related to particular phase or transition of the cell cycle. The S cyclin promotes the DNA replication during S phase.
- A cyclin’s concentration also varies throughout the cell cycle, i.e., their concentration stays low throughout the cell cycle but increases during the phase or transition at which they required.
- The M cyclin is present in highest concentration during M phase. S cyclin is at highest concentration during S and G2 phase also.
- G1 cyclin is an exception as it stays in a similar concentration throughout the cycle as needed at many points.
- The cyclins are synthesized during different stages of cell cycle and subsequently destroyed also to maintain the progression. These cyclins are also named as cyclin A, cyclin B, cyclin D and cyclin E.
- To perform functions of activating or deactivating many proteins inside the cell, cyclins bind to a specialized group of proteins called as Cyclin Dependent Kinases (CDKs). Without cyclins they remain inactive but binding with cyclins activates them making it a functional enzyme and allowing it to modify target proteins.
- The CDKs are the kinases which phosphorylates or attach phosphate groups to target proteins. These target proteins are generally specific to specific kinases. This affect the activity of target protein by making them more or less active.
- An inducer is a molecule that regulates gene expression. An inducer can bind to protein repressors or activators.
- Repressible enzymes have their synthesis inhibited by a metabolite (for example, tryptophan).
- Inducible enzymes have their synthesis stimulated or induced by specific metabolites (for example, lac operon).
- An operon is a cluster of functionally-related genes that are controlled by a shared operator. Operons consist of multiple genes grouped together with a promoter and an operator.
Operons are present in prokaryotes (bacteria and archaea), but are absent in eukaryotes. In some situations multiple operons are controlled by the same regulatory protein; in these cases the operons form a regulon.

Operons were first identified as a mode of gene expression control in 1961 by François Jacob and Jacques Monod.

An operon is a complete package for gene expression and synthesis of polypeptides. By combining the related genes, all polypeptides required for a specific function are synthesized in response to a single stimulus.

Operons can be under negative or positive control.

Negative control involves turning off the operon in the presence of a repressor; this can be either repressible or inducible. A repressible operon is one that is usually on but which can be repressed in the presence of a repressor molecule.

Positive control of an operon is when gene expression is stimulated by the presence of a regulatory protein.

The Lac operon is the classic operon example, and is responsible for the degradation of the milk protein lactose. The Lac operon is an inducible operon; in the absence of lactose the operator is blocked by a repressor protein.

The operon is made up of a promoter with operator, and three genes (lacZ, lacY, and lacA) which encode β-galactosidase, permease, and transacetylase.

The expression of the Lac operon is controlled by the regulatory gene lacI, located immediately adjacent to the promoter region. LacI encodes an allosteric repressor protein that keeps the Lac operon "OFF".

A bacterium's environment is too complex for its genes to be controlled by one signal. Other factors besides lactose affect the expression of the lac genes, such as the availability of glucose.

A regulatory mechanism known as catabolite repression restricts expression of the genes required for catabolism of lactose, arabinose, and other sugars in the presence of glucose, even when these secondary sugars are also present.

The effect of glucose is mediated by cAMP, as a co-activator, and an activator protein known as cAMP receptor protein, or CRP (the protein is sometimes called CAP, for Catabolite gene Activator Protein).

CRP-cAMP is therefore a positive regulatory element responsive to glucose levels.
9.8 KEY WORDS

- **Genes**: These are the small segments present on DNA. They encode for proteins which describes an individual’s phenotype. They are characterized by the presence of start codon in the beginning and stop codon at the end.
- **Locus**: Gene has a position on chromosomes called locus. Due to chromosomal aberrations, locus may change which can affect phenotypic expression.
- **LacZ gene**: It encodes for beta-galactosidase, an enzyme which convert lactose (disaccharide) to the monomer units, glucose and galactose (monosaccharides).
- **LacY gene**: This gene encodes for permease which helps in the entry of lactose through the cell wall of bacteria.
- **LacA gene**: This gene encodes for transacetylase which known to have probable role in exporting out as by product of lactose breakdown.
- **G1/S check point**: It checks the size of the cell, growth factors, nutrients and determine any damage to DNA molecule which should be repaired before it proceeds to the next stage.
- **G2/M check point**: In this check point, physiological conditions of the cell is monitored, the cell is analyzed for the complete DNA replication and whether DNA is damaged.
- **Inducer**: An inducer is a molecule that regulates gene expression, it can bind to protein repressors or activators.
- **Repressible enzymes**: Enzymes which have their synthesis inhibited by a metabolite (for example, tryptophan).
- **Inducible enzymes**: Enzymes which have their synthesis stimulated or induced by specific metabolites (for example, lac operon).
- **Operon**: An operon is a cluster of functionally-related genes that are controlled by a shared operator.

9.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What is the significance of gene?
2. How the regulation of bacterial gene expression is done?
3. Define the terms - operon, repressor, inducer, polycistronic mRNA and regulatory genes giving examples.
4. Write a short note on CRP (cAMP receptor protein).
5. What do you understand by positive and negative control of operons?
6. How is cell cycle regulation done? Why it is important?
7. What is the relation between cyclins and CDKs?
8. What is lactose system?
9. Explain the Lac components.
10. Differentiate between positive and negative regulation.
11. What is catabolite repression?

Long Answer Questions

1. Briefly discuss the significance and concept of gene giving appropriate examples.
2. Explain the significant features of regulation of bacterial gene expression.
3. Differentiate between repressible enzymes and inducible enzymes.
4. Describe the structure and functions of operons.
5. What do you understand by the expression of the lac operon when both glucose and lactose are present?
6. Explain the role of cyclins and CDKs in cell cycle regulation. What if only one cyclin or CDK loss their function?
7. Briefly discuss about the lactose system and Lac operon.
8. Explain the following terms giving appropriate examples:
   (a) Inducer
   (b) Enhancer
   (c) Operon Structure
   (d) Lac Components
   (e) Positive and Negative Regulation
   (f) Catabolite Repression

9.10 FURTHER READINGS


UNIT 10 TRYPTOPHAN OPERON
AND ARABINOSE OPERON

Structure
10.0 Introduction
10.1 Objectives
10.2 Tryptophan Operon
10.3 Arabinose Operon and its Regulation
10.4 Answers to Check Your Progress Questions
10.5 Summary
10.6 Key Words
10.7 Self Assessment Questions and Exercises
10.8 Further Readings

10.0 INTRODUCTION

The tryptophan operon is the regulation of transcription of the gene responsible for biosynthesis of tryptophan from chorismate. Tryptophan operon consists of structural gene and regulatory gene. The regulatory gene include Promoter, Repressor, Operator and Leader sequence while the structural gene include TrpE, TrpD, TrpC, TrpB and TrpA. Fundamentally, the Trp operon is an operon in which a group of genes is used or transcribed together that codes for the components for production of tryptophan. The “Trp” operon is present in many bacteria, but was first characterized in Escherichia coli. The operon is regulated so that when tryptophan is present in the environment, the genes for tryptophan synthesis are not expressed. It was an important experimental system for learning about gene regulation, and is commonly used to teach gene regulation. Attenuation is a second mechanism of negative feedback in the Trp operon. The repression system targets the intracellular Trp concentration whereas the attenuation responds to the concentration of charged tRNA<sup>Trp</sup>. Attenuation is possible in prokaryotes (which have no nucleus), the ribosomes begin translating the mRNA while RNA polymerase is still transcribing the DNA sequence. This allows the process of translation to affect transcription of the operon directly.

Arabinose is a 5 Carbon sugar that can be used as an alternative Carbon and energy source by the bacteria Escherichia coli. The enzymes necessary for the metabolism of this sugar are encoded by the araBAD operon. The operon encodes a single polycistronic mRNA with three open reading frames. Each open reading frame encodes one of the enzymes of the catabolic pathway for Arabinose. The regulator protein AraC is sensitive to the level of arabinose and plays a dual role as both an activator in the presence of arabinose and a repressor in the absence of arabinose to regulate the expression of araBAD.
In this unit, you will study about the Tryptophan operon, attenuation mechanism in the Trp operon, Arabinose operon and its regulation mechanism in detail.

### 10.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand what Tryptophan operon is
- Explain the attenuation mechanism
- Discuss about the Arabinose operon and its regulation mechanism

### 10.2 TRYPTOPHAN OPERON

The tryptophan operon is the regulation of transcription of the gene responsible for biosynthesis of tryptophan from chorismate. Tryptophan operon consists of structural gene and regulatory gene. The regulatory gene include Promoter, Repressor, Operator and Leader sequence while the structural gene include TrpE, TrpD, TrpC, TrpB and TrpA. Fundamentally, the Trp operon is an operon in which a group of genes is used or transcribed together that codes for the components for production of tryptophan. The ‘Trp’ operon is present in many bacteria, but was first characterized in *Escherichia coli*. The operon is regulated so that when tryptophan is present in the environment, the genes for tryptophan synthesis are not expressed. It was an important experimental system for learning about gene regulation, and is commonly used to teach gene regulation.

The operon operates by a negative repressible feedback mechanism. The repressor for the trp operon is produced upstream by the trpR gene, which is constitutively expressed at a low level. Synthesized trpR monomers associate into dimers. When tryptophan is present, these tryptophan repressor dimers bind to tryptophan, causing a change in the repressor conformation, allowing the repressor to bind to the operator. This prevents RNA polymerase from binding to and transcribing the operon, so tryptophan is not produced from its precursor. When tryptophan is not present, the repressor is in its inactive conformation and cannot bind the operator region, so transcription is not inhibited by the repressor.

The tryptophan operon is the regulation of transcription of the gene responsible for biosynthesis of tryptophan from chorismate.
Tryptophan Operon consists of structural gene and regulatory gene.

I. **Regulatory gene** are Promoter, Repressor, Operator, and Leader sequence.

II. **Structural gene** are TrpE, TrpD, TrpC, TrpB and TrpA

   1. trpE: It Encodes the Enzyme Anthranilate Synthase I
   2. trpD: It Encodes the Enzyme Anthranilate Synthase II
   3. trpC: It Encodes the Enzyme N-5’-Phosphoribosyl Anthranilate Isomerase and Indole-3-Glycerolphosphate Synthase
   4. trpB: It Encodes the Enzyme Tryptophan Synthase-B sub unit
   5. trpA: It Encode the Enzyme Tryptophan Synthase-A sub unit

Tryptophan operon is regulated by following mechanism:

1. **Repression**

   (i) **When Tryptophan is Absent in Cell**

   Repressor gene (trpR) encodes the repressor protein which is originally inactive. In the absence of tryptophan, transcription of structural gene occur for the biosynthesis of tryptophan from chorismate (Refer Figure 10.2).
(ii) When Tryptophan is High in Cell

When tryptophan is high in cell then it binds with repressor protein and change its confirmation so that it become active and bind to the operator near promoter.

Binding of repressor protein to operator overlaps the promoter, so RNA polymerase cannot bind to the promoter. Hence transcription is halted.

Since tryptophan is already high in cell, no transcription of structural gene is required for biosynthesis of tryptophan. This is also known as negative regulation.

2. Attenuation

Attenuation is a second mechanism of negative feedback in the Trp operon. The repression system targets the intracellular Trp concentration whereas the attenuation responds to the concentration of charged tRNA\(^{\text{trp}}\). Attenuation is possible in prokaryotes (which have no nucleus), the ribosomes begin translating the mRNA while RNA polymerase is still transcribing the DNA sequence. This allows the process of translation to affect transcription of the operon directly.

In bacteria, transcription and translation occurs simultaneously. The translation starts before transcription completes. In this attenuation mechanism, rate of translation determines whether transcription continues or terminates. Therefore the attenuation mechanism is only found in bacteria but not in eukaryotic cell.
Leader sequence (trpL) play important role in attenuation. Leader sequence contains such a nucleotide sequence that mRNA transcribed from it contains four specific region. Region 1, Region 2, Region 3 and Region 4. Region 3 is complementary to both Region 2 and Region 4.

If region 3 and region 4 base pair with each other, they form a loop like structure called attenuator and it function as transcriptional termination. If pairing occur between Region 3 and Region 2, then no such attenuator form so that transcription continues.

Region 1 is the most important region that determines whether to form loop between Region 2-Region 3 or Region 3-Region 4. The Region 1 consists of sequence of 14 Codons, out of which two codons are Tryptophan Codon (Codon 10 and 11) (Refer Figure 10.3).

Fig. 10.3 Attenuation

When tryptophan is high in cell then tRNA carrying tryptophan encodes Codon 10 and 11, such that ribosome encloses the Region 2 which is near to the tryptophan codon. Hence region 3 base pair with Region 4 to form attenuator as Region 2 is not available for pairing. Consequently, transcription is halted.

When tryptophan is low or absent in cell, then translation stops at the position of Tryptophan Codon. Such that loop between Region 2 and Region 3 forms. Transcription continues.
Feedback Mechanism of Trp Operon

1. When tryptophan is high in cell then transcription of structural gene does not occur.
2. When tryptophan is absent or very low then transcription continues.
3. If chorismate is high in cell then it favors the transcription of structural gene.
4. If chorismate is low or absent in cell then it inhibit transcription.

10.3 ARABINOSE OPERON AND ITS REGULATION

Arabinose is a 5 Carbon sugar that can be used as an alternative Carbon and energy source by the bacteria *Escherichia coli*. The enzymes that are essential for the metabolism of this sugar are encoded by the araBAD operon. The operon encodes a single polycistronic mRNA with three open reading frames. Each open reading frame encodes one of the enzymes of the Catabolic Pathway for Arabinose. Expression of the araBAD operon is highly regulated. Key to its regulation is the composition of sugars in the environment, for example, the operon shows a classic pattern of substrate level regulation. When arabinose is absent from the environment then this operon is not expressed, therefore it is essential that to express this arabinose operon it must be present in the environment. This is an adaptive mechanism that ensures the enzymes required to catabolize arabinose and are only produced when arabinose is present in the environment.

The araBAD operon also exhibits the phenomena of catabolite repression. High levels of glucose in the environment prevent the expression of the araBAD operon even if arabinose is present. This is believed to be adaptive because bacteria can extract energy more efficiently from glucose than from arabinose. Therefore if glucose is present in the environment, the enzymes for metabolism of arabinose are not produced.

The L-Arabinose operon, also called the ara or araBAD operon, is an operon required for the breakdown of the five Carbon sugar the L-arabinose, in *Escherichia coli*. The L-arabinose operon contains three structural genes - araB, araA, araD, collectively known as araBAD, which encode for three metabolic enzymes that are required for the metabolism of L-arabinose. AraB (Ribulokinase), AraA (an Isomerase), AraD (an Epimerase) produced by these genes catalyse conversion of L-arabinose to an intermediate of the pentose phosphate pathway, D-Xylulose-5-Phosphate.

Figure 10.4 illustrates the structure of the araBAD operon. The three open reading frames, araB, araA and araD, are denoted by the boxes. The cis elements regulate transcription. In addition to the promoter (P_{bad}), there are three critical regulatory elements araI_{1}, araI_{2}, and araO_{2} that are involved in substrate level regulation.
AraC is the regulatory protein that mediates substrate level regulation of transcription. It is a sequence specific binding protein that binds to the sequences at araI1, araI2, and araO2. Because AraC functions as a homodimer, it always binds two of these regulatory elements at the same time. AraC also has an Arabinose Binding Pocket. The binding of Arabinose to AraC alters its conformation, making the AraC protein more flexible. This increase flexibility is significant to the regulation of the operon. The AraC also has a domain that interacts with RNA polymerase, helping to recruit RNA polymerase to PBAD and thereby promoting transcription of araBAD.

In the presence of arabinose, the AraC homodimer has a more flexible structure which allows the AraC homodimer to simultaneously bind the nearby adjacent sites araI1 and araI2, as shown in Figure 10.5. Binding of AraC to araI1/araI2 allows it to actively recruit RNA polymerase to PBAD and thereby promotes transcription.

In the absence of Arabinose, the AraC homodimer has a ridged structure. This ridged structure interferes in binding to the closely adjacent sites araI1 and araI2. As an alternative the ridged AraC binds to araO2. For a single homodimer of AraC to bind to these two distant DNA sites requires DNA looping. The binding of AraC to these two sites prevents transcription of the araBAD operon in two different ways. First, since AraC is bound to araO2 instead of araI2, it is not able to promote recruitment of RNA polymerase. Second, the loop of DNA hinders the finding of the RNA polymerase to PBAD, as well as the binding of CAP to the cap binding site.

Figure 10.6 illustrates the map of the Ara region, in which the B, A, and D genes together with the I and O sites constitute the Ara operon.
The specific structural genes, araB, araA and araD encode the metabolic enzymes for breaking down Arabinose are transcribed as a multigenic mRNA. Transcription is activated at araI, termed as the initiator region, which contains both an operator site and a promoter. When the araC gene, when bound to arabinose, encodes an activator protein that activates transcription of the Ara operon, possibly by facilitating RNA polymerase bind to the promoter that is located within the araI region. An additional activation event is typically mediated by the similar CAP–cAMP catabolite repression system that is responsible for the regulation of lac operon expression.

When the arabinose is present then both the CAP–cAMP complex and the AraC arabinose complex essentially bind to the initiator region so that the RNA polymerase binds to the promoter and transcribe the Ara operon (Refer Figure 10.7). When the arabinose is not present then the AraC protein assumes a dissimilar or different conformation and represses the Ara operon by binding both to araI and to a second operator region AraO, thus forming a loop in order to prevent transcription. Consequently, the AraC protein has two conformations, first that acts as an activator and the second that acts as a repressor. Figure 10.7 illustrates the control mechanism of the Ara operon.

Check Your Progress

1. Explain the term tryptophan operon giving its types.
2. What happens when tryptophan is high in cell?
3. What is attenuation?
4. Explain about Arabinose operon giving example.
5. What happens when arabinose is absent from the environment?
6. How araBAD operon exhibits the phenomena of catabolite repression?
7. What is L-Arabinose operon?
1. The tryptophan operon is the regulation of transcription of the gene responsible for biosynthesis of tryptophan from chorismate. Tryptophan operon consists of structural gene and regulatory gene. The regulatory gene include Promoter, Repressor, Operater and Leader sequence while the structural gene include TrpE, TrpD, TrpC, TrpB and TrpA. Fundamentally, the Trp operon is an operon in which a group of genes is used or transcribed together that codes for the components for production of tryptophan. The 'Trp' operon is present in many bacteria, but was first characterized in Escherichia coli. The operon is regulated so that when tryptophan is present in the environment, the genes for tryptophan synthesis are not expressed.

2. When tryptophan is high in cell then it binds with repressor protein and change its confirmation so that it become active and bind to the operator near promoter. Binding of repressor protein to operator overlaps the promoter, so RNA polymerase cannot bind to the promoter. Hence transcription is halted. Since tryptophan is already high in cell, no transcription of structural gene is required for biosynthesis of tryptophan. This is also known as negative regulation.

3. Attenuation is a second mechanism of negative feedback in the Trp operon. The repression system targets the intracellular Trp concentration whereas the attenuation responds to the concentration of charged tRNA\(^\text{Trp}\). Attenuation is possible in prokaryotes (which have no nucleus), the ribosomes begin translating the mRNA while RNA polymerase is still transcribing the DNA sequence. This allows the process of translation to affect transcription of the operon directly.

4. Arabinose is a 5 Carbon sugar that can be used as an alternative Carbon and energy source by the bacteria Escherichia coli. The enzymes that are essential for the metabolism of this sugar are encoded by the araBAD operon. The operon encodes a single polycistronic mRNA with three open reading frames. Each open reading frame encodes one of the enzymes of the Catabolic Pathway for Arabinose. Expression of the araBAD operon is highly regulated. Key to its regulation is the composition of sugars in the environment, for example, the operon shows a classic pattern of substrate level regulation.

5. When arabinose is absent from the environment then this operon is not expressed, therefore it is essential that to express this arabinose operon it must be present in the environment. This is an adaptive mechanism that ensures the enzymes required to catabolize arabinose and are only produced when arabinose is present in the environment.
6. The araBAD operon exhibits the phenomena of catabolite repression. High levels of glucose in the environment prevent the expression of the araBAD operon even if arabinose is present. This is believed to be adaptive because bacteria can extract energy more efficiently from glucose than from arabinose. Therefore if glucose is present in the environment, the enzymes for metabolism of arabinose are not produced.

7. The L-Arabinose operon, also called the ara or araBAD operon, is an operon required for the breakdown of the five Carbon sugar the L-arabinose, in Escherichia coli. The L-arabinose operon contains three structural genes - araB, araA, araD, collectively known as araBAD, which encode for three metabolic enzymes that are required for the metabolism of L-arabinose. AraB (Ribulokinase), AraA (an Isomerase), AraD (an Epimerase) produced by these genes catalyse conversion of L-arabinose to an intermediate of the pentose phosphate pathway, D-Xylulose-5-Phosphate.

10.5 SUMMARY

- The tryptophan operon is the regulation of transcription of the gene responsible for biosynthesis of tryptophan from chorismate.
- Tryptophan operon consists of structural gene and regulatory gene. The regulatory gene include Promoter, Repressor, Operator and Leader sequence while the structural gene include TrpE, TrpD, TrpC, TrpB and TrpA.
- Fundamentally, the Trp operon is an operon in which a group of genes is used or transcribed together that codes for the components for production of tryptophan.
- The 'Trp' operon is present in many bacteria, but was first characterized in Escherichia coli. The operon is regulated so that when tryptophan is present in the environment, the genes for tryptophan synthesis are not expressed.
- The tryptophan operon is the regulation of transcription of the gene responsible for biosynthesis of tryptophan from chorismate.
- The operon operates by a negative repressible feedback mechanism. The repressor for the trp operon is produced upstream by the trpR gene, which is constitutively expressed at a low level.
- Synthesized trpR monomers associate into dimers. When tryptophan is present, these tryptophan repressor dimers bind to tryptophan, causing a change in the repressor conformation, allowing the repressor to bind to the operator.
- Repressor gene (trpR) encodes the repressor protein which is originally inactive. In the absence of tryptophan, transcription of structural gene occur for the biosynthesis of tryptophan from chorismate.
When tryptophan is high in cell then it binds with repressor protein and change its confirmation so that it become active and bind to the operator near promoter.

- Binding of repressor protein to operator overlaps the promoter, so RNA polymerase cannot bind to the promoter. Hence transcription is halted.
- Since tryptophan is already high in cell, no transcription of structural gene is required for biosynthesis of tryptophan. This is also known as negative regulation.
- Attenuation is a second mechanism of negative feedback in the Trp operon. The repression system targets the intracellular Trp concentration whereas the attenuation responds to the concentration of charged tRNA<sup>Trp</sup>.
- Attenuation is possible in prokaryotes (which have no nucleus), the ribosomes begin translating the mRNA while RNA polymerase is still transcribing the DNA sequence. This allows the process of translation to affect transcription of the operon directly.
- In bacteria, transcription and translation occurs simultaneously. The translation starts before transcription completes. In this attenuation mechanism, rate of translation determines whether transcription continues or terminates. Therefore the attenuation mechanism is only found in bacteria but not in eukaryotic cell.

When tryptophan is high in cell then tRNA carrying tryptophan encodes Codon 10 and 11 such that ribosome encloses the Region 2 which is near to the Tryptophan Codon. Hence Region 3 base pair with Region 4 to form attenuator as Region 2 is not available for pairing. Consequently, transcription is halted.

- Arabinose is a 5 Carbon sugar that can be used as an alternative Carbon and energy source by the bacteria *Escherichia coli*.
- The enzymes that are essential for the metabolism of this sugar are encoded by the araBAD operon. The operon encodes a single polycistronic mRNA with three open reading frames. Each open reading frame encodes one of the enzymes of the Catabolic Pathway for Arabinose.
- Expression of the araBAD operon is highly regulated. Key to its regulation is the composition of sugars in the environment, for example, the operon shows a classic pattern of substrate level regulation.
- When arabinose is absent from the environment then this operon is not expressed, therefore it is essential that to express this arabinose operon it must be present in the environment. This is an adaptive mechanism that ensures the enzymes required to catabolize arabinose and are only produced when arabinose is present in the environment.
The araBAD operon also exhibits the phenomena of catabolite repression. High levels of glucose in the environment prevent the expression of the araBAD operon even if arabinose is present. This is believed to be adaptive because bacteria can extract energy more efficiently from glucose than from arabinose.

Therefore if glucose is present in the environment, the enzymes for metabolism of arabinose are not produced.

The L-Arabinose operon, also called the ara or araBAD operon, is an operon required for the breakdown of the five Carbon sugar the L-Arabinose, in *Escherichia coli*.

The L-arabinose operon contains three structural genes - araB, araA, araD, collectively known as araBAD, which encode for three metabolic enzymes that are required for the metabolism of L-arabinose. AraB (Ribulokinase), AraA (an Isomerase), AraD (an Epimerase) produced by these genes catalyse conversion of L-arabinose to an intermediate of the pentose phosphate pathway, D-Xylulose-5-Phosphate.

AraC is the regulatory protein that mediates substrate level regulation of transcription. It is a sequence specific binding protein that binds to the sequences at araI\(_1\), araI\(_2\) and araO\(_2\). Because AraC functions as a homodimer, it always binds two of these regulatory elements at the same time.

AraC has an Arabinose Binding Pocket. The binding of Arabinose to AraC alters its allosteric conformation, making the AraC protein more flexible. This increase flexibility is significant to the regulation of the operon.

The AraC has a domain that interacts with RNA polymerase, helping to recruit RNA polymerase to PBAD and thereby promoting transcription of araBAD.

AraC is the regulatory protein that mediates substrate level regulation of transcription.

In the presence of arabinose, the AraC homodimer has a more flexible structure which allows the AraC homodimer to simultaneously bind the nearby adjacent sites araI\(_1\) and araI\(_2\).

In the absence of Arabinose, the AraC homodimer has a ridged structure. This ridged structure interferes in binding to the closely adjacent sites araI\(_1\) and araI\(_2\). As an alternative the ridged AraC binds to araI\(_1\) and araO\(_2\).

When the arabinose is present then both the CAP-cAMP complex and the AraC arabinose complex essentially bind to the initiator region so that the RNA polymerase binds to the promoter and transcribe the Ara operon.
10.6 KEY WORDS

- **Tryptophan operon**: The tryptophan operon is the regulation of transcription of the gene responsible for biosynthesis of tryptophan from chorismate, the tryptophan operon consists of structural gene and regulatory gene.
- **Attenuation**: It is a second mechanism of negative feedback in the Trp operon, the repression system targets the intracellular Trp concentration whereas the attenuation responds to the concentration of charged tRNA<sub>Trp</sub>.
- **Arabinose**: Arabinose is a 5 Carbon sugar that can be used as an alternative Carbon and energy source by the bacteria *Escherichia coli*.
- **AraC**: It is the regulatory protein that mediates substrate level regulation of transcription.

10.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

**Short Answer Questions**

1. What is tryptophan operon?
2. Differentiate between regulatory and structural genes.
3. What happens when tryptophan is absent in cells?
4. Explain the term attenuation giving example.
5. What is Arabinose operon?
6. How the Arabinose operon is regulated?
7. What is araBAD?

**Long Answer Questions**

1. Briefly discuss about the tryptophan operon and its types giving examples.
2. Explain the mechanism required for the regulation of tryptophan operon.
3. What is attenuation? Explain giving examples.
4. Discuss the significant features of Arabinose operon.
5. Explain in detail how the regulation of Arabinose operon is done with the help of diagrams.
6. “AraC is the regulatory protein that mediates substrate level regulation of transcription”. Discuss.
10.8 FURTHER READINGS


NOTES

Self-Instructional Material 219

Tryptophan Operon and Arabinose Operon
UNIT 11 PLASMIDS: TYPES AND ITS PROPERTIES

11.0 INTRODUCTION

Plasmid, in microbiology, an extrachromosomal genetic element that occurs in many bacterial strains. Plasmids are circular DeoxyriboNucleic Acid (DNA) molecules that replicate independently of the bacterial chromosome. They are not essential for the bacterium but may confer a selective advantage. 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. Some Col factors and R factors can transfer themselves from one cell to another and thus are capable of spreading rapidly through a bacterial population. A plasmid that is attached to the cell membrane or integrated into the bacterial chromosome is called an episome.

Plasmids are extremely valuable tools in the fields of molecular biology and genetics, specifically in the area of genetic engineering. They play a critical role in such procedures as gene cloning, recombinant protein production, and gene therapy research. In such procedures, a plasmid is cut at a specific site using enzymes called restriction endonucleases. A foreign DNA element is then spliced into the plasmid. The resulting circular structure, a recombinant DNA molecule, is then introduced into bacterial cells. The autonomous replication of the plasmid within the bacterial cells makes it possible to produce large numbers of copies of the recombinant DNA molecule for experimental manipulation or commercial purposes. Plasmids are well suited to genetic engineering in other ways. Their antibiotic resistance genes, for example, prove useful in identifying those bacterial cells that have taken up the recombinant DNA molecule in a high background of untransformed cells.
In this unit, you will study about plasmids, its different types - F, R and Col plasmids, properties of plasmids - sex factors, drug resistant, Colicinogenic, Agrobacterium Ti and broad host range plasmid in detail.

11.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand about plasmids
- Discuss the types of plasmids
- Explain the properties of plasmids

11.2 PLASMIDS: TYPES AND PROPERTIES

A plasmid is a small 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, such as by providing antibiotic resistance. While the chromosomes are big and contain all the essential genetic information for living under normal conditions, plasmids usually are 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. Figure 11.1 shows the image of a plasmid.

Plasmids were discovered by William Hayes and Joshua Lederberg (1952) in bacterial cells. They are small, double stranded, closed circular, symbiotic DNA molecules that occur naturally in bacteria, outside the bacterial chromosome. They are regarded as extra chromosomal DNA containing extrachromosomal genome. A bacterial cell may possess one or many copies of one or more plasmids. They are inherited from parent bacterial cell to daughter cell and have capability to self-replication in the cytoplasm of bacterial cell. The circular molecule of DNA can be broken to yield a linear molecule which passes from one bacterial cell to another.
There are many kinds of plasmids which differ from one another in size and in composition of genes. They vary in size from 2 kilobars to more than 400 kilobars. The major components of a plasmid comprise of an origin site (the replication origin) or a specific portion of their genome that serves as start signal for self-replication. Under natural conditions, each plasmid replicates to produce 20-30 copies per cell. This number can be artificially increased. In presence of certain antibiotics the number can be increased to about 1000 copies.

Another valuable feature of plasmids is the presence of specific restriction sites where the enzyme restriction endonuclease makes a cut so that a foreign DNA segment may be jointed to the plasmid. This property helps plasmids to act as trusted cloning vehicles during gene transfer in genetic engineering. Plasmids are considered replicons, units of DNA capable of replicating autonomously within a suitable host. However, plasmids, like viruses, are not generally classified as life. 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. The relationship between microbes and plasmid DNA is neither parasitic nor mutualistic, because each implies the presence of an independent species living in a detrimental or commensal state with the host organism. Rather, plasmids provide a mechanism for horizontal gene transfer within a population of microbes and typically provide a selective advantage under a given environmental state. Plasmids may carry genes that provide resistance to naturally occurring antibiotics in a competitive environmental niche, or the proteins produced may act as toxins under similar circumstances, or allow the organism to utilize particular organic compounds that would be advantageous when nutrients are scarce.

**Vectors**

In molecular cloning, a vector is a DNA molecule used as a vehicle to artificially carry foreign genetic material into another cell, where it can be replicated and/or expressed (for example, plasmid, cosmid, Lambda Phages). A vector containing foreign DNA is termed recombinant DNA. The four major types of vectors are plasmids, viral vectors, cosmids, and artificial chromosomes. Of these, the most commonly used vectors are plasmids. Common to all engineered vectors are an origin of replication, a multi-cloning site, and a selectable marker.
**Viral Vectors:** Viral vectors are generally genetically engineered viruses carrying modified viral DNA or RNA that has been rendered noninfectious, but still contain viral promoters and also the transgene, thus allowing for translation of the transgene through a viral promoter. However, because viral vectors frequently are lacking infectious sequences, they require helper viruses or packaging lines for large-scale transfection.

Viral vectors are often designed for permanent incorporation of the insert into the host genome, and thus leave distinct genetic markers in the host genome after incorporating the transgene. For example, retroviruses leave a characteristic retroviral integration pattern after insertion that is detectable and indicates that the viral vector has incorporated into the host genome.

**Artificial Chromosomes:** Artificial chromosomes are manufactured chromosomes in the context of Yeast Artificial Chromosomes (YACs), Bacterial Artificial Chromosomes (BACs), or Human Artificial Chromosomes (HACs). An artificial chromosome can carry a much larger DNA fragment than other vectors. YACs and BACs can carry a DNA fragment up to 300,000 nucleotides long. Three structural necessities of an artificial chromosome include an origin of replication, a centromere, and telomeric end sequences. Artificially constructed plasmids may be used as vectors in genetic engineering. These plasmids serve as an important tool in genetics and biotechnology labs, where they are commonly used to clone and amplify (make many copies of) or express particular genes.

A wide variety of plasmids are commercially available for such uses. The gene to be replicated is normally inserted into a plasmid that typically contains a number of features for their use. These include a gene that confers resistance to particular antibiotics (Ampicillin is most frequently used for bacterial strains), an origin of replication to allow the bacterial cells to replicate the plasmid DNA, and a suitable site for cloning (referred to as a multiple cloning site). In molecular cloning, a **vector** is a DNA molecule used as a vehicle to artificially carry foreign genetic material into another cell, where it can be replicated and/or expressed (for example, plasmid, cosmid, Lambda Phages). A vector containing foreign DNA is termed recombinant DNA. The four major types of vectors are plasmids, viral vectors, cosmids, and artificial chromosomes. Of these, the most commonly used vectors are plasmids. Common to all engineered vectors are an origin of replication, a multi cloning site, and a selectable marker. The vector itself is generally a DNA sequence that consists of an insert (transgene) and a larger sequence that serves as the ‘backbone’ of the vector. The purpose of a vector which transfers genetic information to another cell is typically to isolate, multiply, or express the insert in the target cell.

All vectors may be used for cloning and are therefore cloning vectors, but there are also vectors designed especially for cloning, while others may be designed specifically for other purposes, such as transcription and protein expression. Vectors designed specifically for the expression of the transgene in the target cell are...
Plasmids: Types and its Properties

NOTES

Called expression vectors, and generally have a promoter sequence that drives expression of the transgene. Simpler vectors called transcription vectors are only capable of being transcribed but not translated - they can be replicated in a target cell but not expressed, unlike expression vectors. Transcription vectors are used to amplify their insert. The manipulation of DNA is normally conducted on *Escherichia coli* vectors, which contain elements necessary for their maintenance in *Escherichia coli*.

However, vectors may also have elements that allow them to be maintained in another organism such as yeast, plant or mammalian cells, and these vectors are called shuttle vectors. Such vectors have bacterial or viral elements which may be transferred to the non-bacterial host organism, however other vectors termed intragenic vectors have also been developed to avoid the transfer of any genetic material from an alien species.

Generally, the bacterial plasmids are 1 to 5% of the chromosomal DNA in size. Plasmids vary widely in size. The smaller plasmids have molecular weights ranging from 4 - 5 × 10⁶ Daltons, while the larger ones have molecular weights of 25 to 95 × 10⁶ Daltons. Plasmids not only vary in size, but also in copy number which denotes the number of copies of a specific plasmid in a cell. Copy number is discontinuously variable, i.e., some plasmids generally the smaller ones - have a high copy number, while the larger plasmids have characteristically a low copy number. Those having a high copy number are known as relaxed plasmids and those having low copy number are called stringent plasmids. Whatever may be the copy number, plasmids are generally distributed equally in the daughter cell during cell division. Rarely, a plasmid-free cell may arise spontaneously at a frequency of about 1 in 10⁴ cells. Plasmid free cells may also be produced artificially by the use of mutagens. The process is commonly called curing of plasmids. Usually, the low-copy number large plasmids have one or two copies per cell and are easier to be cured. The smaller plasmids, in contrast, may have 10 to 100 copies per cell. So far as the biological functions of plasmids are concerned, they are not indispensable constituents of the bacteria. This is proved by the fact that the bacteria cured of plasmids can grow normally without any difficulty. However, the genes carried on the plasmid DNA confer special properties to the host bacteria and such properties may become advantageous under special environmental conditions.

For example, bacteria carrying the R-plasmids (resistance plasmids) can survive when the environment contains inhibitory concentrations of one or more antibiotics. Obviously, the R-plasmid-less bacteria are destroyed under such conditions. Another example is provided by the plasmids of some species of *Pseudomonas* which carry genes for production of enzymes catalyzing degradation of complex hydrocarbons.

Bacteria carrying such plasmids are capable of using such unusual substrates for growth and, obviously, enjoy special advantage over others lacking them. The F plasmid gives the power to carry out a type of sexual reproduction to bacteria making it possible to exchange genetic materials leading to genetic recombination.
Again, some bacteria, like *Escherichia coli*, *Pseudomonas* and *Lactobacillus*, etc., produce special type of proteins, called bacteriocins which are coded by plasmid genes. These proteins are able to kill other closely related bacteria and, thereby, they can eliminate competition for food and space. Thus it is seen that even though plasmids are not absolutely essential for the life of bacteria under normal conditions of growth, their presence may become valuable and advantageous for the host under special conditions, or may even prove critical for survival as in case of the R-plasmids.

The R-plasmids with the help of the resistance genes produce proteins which can inactivate or destroy specific antibiotics. Besides the advantageous properties attributable to the plasmids, these extra-chromosomal genetic elements have played an important role in the development of recombinant DNA technology. In this technology, the plasmids are used as vectors for transferring a gene of interest from one organism to another organism. Such transfer of a gene is possible, not only from one bacterium to another, but also from eukaryotic organism to a bacterium, or vice versa. A segment of DNA containing the specific gene is isolated from a suitable donor and inserted by recombinant DNA technology into a plasmid. The recombinant plasmid is next introduced into a suitable host cell where the gene is expressed producing the gene product. In this way, several human genes producing therapeutically important proteins have been introduced into bacteria. Also, some bacterial genes have been transferred to eukaryotic hosts, like plants, and some viral genes have been transferred to yeasts. In most of such gene transfers, plasmids play a key role as vectors or carriers.

### 11.2.1 Types of Plasmids

Following are the types of plasmids:

**F Plasmid**

The F plasmid, also known as the fertility factor or sex factor, determines the sex of *Escherichia coli* bacteria. 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.

The process of transfer takes place by conjugation of the F+ cell with the F- cell. The F plasmid is a conjugative plasmid.

We know that a characteristic feature of the F plasmid is that it can either remain as an independent entity replicating separately along with the chromosomal DNA, or it can be inserted into the chromosome as its integral part. When an F plasmid is integrated into the *Escherichia coli* chromosome, the bacterial cell changes from F to an Hfr strain (High frequency of recombination).

There are many sites on the *Escherichia coli* chromosome where the F plasmid can be integrated. Depending on the site, each integration gives rise to a
Plasmids: Types and its Properties

NOTES

Self-Instructional Material

226

different Hfr-strain. In F+ × F− conjugation, the plasmid alone is transmitted, but in Hfr × F− conjugation, chromosomal genes are transmitted and rarely also the F plasmid.

The F plasmid is a large self-transmissible plasmid having a double-stranded circular DNA molecule. Its molecular weight is 63 × 10^6 Daltons and it contains about genes controlling the transfer of the plasmid from the donor to the recipient. A mutation in any of the essential genes results in the loss of transmissibility of the plasmid.

Just as an F plasmid can be integrated into the chromosome of the host cell, so it can though on rare occasions, be separated or excised from the *Escherichia coli* chromosome in the free state to form the circular plasmid. It has been observed that the excision process is sometimes imperfect in the sense that some parts of the *Escherichia coli* chromosome adjoining the linearly inserted F plasmid are included in the excised F DNA and at the same time, parts of the plasmid DNA are retained in the *Escherichia coli* chromosome.

The F plasmids containing parts of the chromosomal DNA are designated as F1 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 sexduction (Refer Figure 11.2).

![Fig. 11.2 F Plasmid](image-url)
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 *Shigella dysenteriae*. Since then these plasmids have been found in *Escherichia 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 $30 \times 10^6$ 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 $6 \times 10^6$ Daltons are non-transmissible. Most of the self-transmissible large plasmids like R100 of *Shigella* conferring multiple drug resistance are co-integrates of 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.

The resistance genes located in the other segment elaborate enzymes for destruction of the antibacterial drugs, like *Penicillins*, *Streptomycin*, *Chloramphenicol*, *Tetracyclines*, *Kanamycin*, *Sulfonamides*, etc.

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*.

Col-Plasmids

The Col-plasmids are present in different strains of *Escherichia 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 (Cor).
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 $60 \times 10^6$ 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 $5 \times 10^6$ 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 pyocin produced by *Pseudomonas aeruginosa*, megasine elaborated 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, protein synthesis 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.

**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 break-down 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 Ananda Mohan Chakraborty 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.

**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. 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 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 transposase to 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.

A gross structure of the Ti plasmid and generation of the T complex have been diagrammatically represented (Refer Figure 11.3).

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 are productions 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 synthesizes an insecticidal protein. Another interesting achievement though not of practical significance was production of bioluminescent tomato plants by introducing the gene controlling bioluminescence in firefly.
Eukaryotic Plasmids

Plasmids occur rarely in Eukaryotic cells. Some plasmids have been found in yeast (*Saccharomyces cerevisiae*) and in several plants. The only RNA plasmid discovered till now has been found in yeast. It is a double-stranded RNA having a molecular weight of $15 \times 10^6$ Daltons. It contains 10 genes including one coding for a bacteriocin-like protein. The protein can kill other yeast cells lacking the plasmid. This yeast plasmid has been designated as killer particle.

Yeast also contains small DNA plasmids with high copy number. They are located in the nucleus and like the chromosomal DNA are associated with basic proteins — histones. Some yeast DNA plasmids have been genetically engineered in such a way that they are capable of multiplication in both *Escherichia coli* and yeast.

One such engineered yeast plasmid is Yep which can function as a shuttle vector. This plasmid has been used in transfer of useful genes from other organisms into yeast cells via *Escherichia coli* for production of valuable therapeutically important proteins. A successful application of the Yep plasmid is the transfer of the gene coding the coat glycoprotein of Hepatitis B Virus to yeast.

The transgenic yeast can express the gene successfully with production of the viral glycoprotein. The glycoprotein has been used for preparation of Hepatitis B Vaccine for human application. Shuttle vectors are especially useful in transferring Eukaryotic genes, because such genes are often not successfully expressed in bacterial hosts.

Besides yeast, DNA plasmids have been discovered in several plants, like maize and sorghum, as also in several fungi. These plasmids are made of usually linear double stranded DNA molecules whereas all bacterial plasmids are circular.

Replication of Plasmids

In the non-dividing plasmids, the double-stranded DNA exists as a right-handed super-helical coil having 400-600 base pairs per turn of the coil. During replication, the plasmids can multiply autonomously, although replication requires the host cell enzymes. That is why plasmids can multiply only within host cells.

Each plasmid has its own origin of replication. Some plasmids also have genes which code for proteins necessary for their own multiplication. This is proved by the fact that a temperature sensitive mutant of F plasmid ($F^\text{ts}$, i.e., temperature sensitive replicon) is unable to replicate at 42°C, although it can function normally at 37°C.

Different Aspects of Plasmid Replication are briefly discussed below:

**Non-Transmissible Plasmids:** Replication of plasmid DNA starts at the site of origin and may proceed either bi-directionally as in case of bacterial chromosome, or may proceed unidirectionally depending on the nature of plasmid. In bidirectional replication, replication terminates when the two replication forks meet.
meet each other. In unidirectional replication, termination occurs when the replication fork reaches the site of origin. In both cases, the circularity of the plasmid DNA is maintained throughout the process.

**Self-Transmissible Plasmids:** In case of conjugative plasmids, like F plasmid or R-plasmid, replication occurs by the rolling-circle model. The supercoiled DNA undergoes a nick in one of the strand resulting in relaxation of the supercoiled state to form an open circle. The enzyme catalyzing the nick remains attached to the 5'-P end of the relaxed molecules. Such a single-stranded nick becomes necessary for transfer of a copy of the plasmid during conjugation to the mating partner.

By rolling circle replication, the donor cell retains its double-stranded plasmid, while a single-stranded copy is transferred through the mating bridge to the recipient cell, where a complimentary strand is synthesized and ligated to form a double-stranded copy of the plasmid.

**Control of Copy Number:** The large plasmids are characterized by low copy number (one to few) and small plasmids by high copy number (10 to 100). The copy number is controlled by an inhibitor coded by the plasmid DNA itself. The inhibitor concentration in the bacterial cell determines the rate of initiation of plasmid replication.

When a cell containing two large plasmids divides to produce two daughter cells, each having one plasmid, the inhibitor concentration in these cells is the same as that of the mother cell. Now, the daughter cells grow in size to attain maturity resulting in lowering of the inhibitor concentration in the cytoplasm.

As a consequence, DNA synthesis is initiated in the plasmid leading to its replication producing two copies. As each plasmid copy possesses an inhibitor gene, the production of inhibitor doubles and the inhibitor concentration becomes high enough to stop plasmid DNA synthesis and further replication. Thus, the copy number is restricted to two per cell.

A similar mechanism of control of copy member is believed to operate in case of high copy number plasmids also. However, in this case, the inhibitor concentration must reach a higher threshold level to stop initiation of plasmid DNA synthesis in comparison to that of low copy number plasmids.

**Plasmid Amplification:** Another important point of plasmid replication is that chromosomal DNA synthesis and plasmid DNA synthesis are independent of each other, though, in both, DNA synthesis is followed by replication. Thus it is possible to stop chromosomal DNA synthesis and replication without affecting plasmid DNA synthesis and replication.

Such situation can be practically created by adding chloramphenicol to a bacterial culture. This antibiotic specifically inhibits prokaryotic protein synthesis. When it is added to a growing bacterial culture, chromosomal DNA synthesis is inhibited, but plasmid DNA synthesis and replication continue at the cost of the available replication proteins which are not used for chromosomal DNA synthesis.
The net result is that each bacterial cell contains a large number of plasmid copies. This is known as plasmid amplification. When a specific gene which has been transferred (cloned) to a plasmid requires to be isolated, plasmid amplification becomes a useful tool, because of high plasmid DNA concentration in the total cellular DNA.

**Transfer of Non-Self Transmissible Plasmids:** There are some plasmids which do not possess genes for self-transmission, but can be transferred to other cells with the help of a self-transmissible plasmid when both plasmids occur in the same cell. They are known as mobilizable plasmids.

Such plasmids possess genes for proteins needed for nicking its own DNA at the site of origin of replication, but lack in genes needed for building the conjugation tube. When they coexist with a self-transmissible plasmid, like F or R, the latter can build the mating bridge through which a copy of the mobilizable plasmid produced by rolling-circle replication is transferred to a recipient cell.

A different type of mobilization occurs when a donor cell having a self-transmissible plasmid conjugates with a recipient having a mobilizable plasmid. In this type of conjugation, both the donor and the recipient acquire a copy of both types of plasmids by a process of retro transfer. First, the self-transmissible plasmid replicates by rolling-circle model and a single-stranded copy is transferred through the mating bridge to the recipient, where it forms a complimentary strand leading to the formation of a copy of the self-transmissible plasmid in the usual way.

The mobilizable plasmid in the recipient cell then replicates and a single-stranded copy is transferred to the other cell which now acts as the recipient of the mobilizable plasmid. Finally, the two cells separate and each has a copy of the self-transmissible plasmid and a copy of the mobilizable plasmid.

**Incompatibility of Plasmids**

Generally, two closely related plasmids cannot coexist in a bacterial cell. In the population of progeny cells derived from a cell containing two such plasmids, the proportion of cells having only one of the two plasmids increases with every cell division. This is known as plasmid incompatibility.

On the other hand, two different unrelated plasmids, for example F plasmid and ColEl can exist together without any difficulty, because these plasmids belong to two different incompatibility groups. Whereas, two F plasmids cannot coexist in the same cell.

One mechanism by which a plasmid already resident in a cell prevents the entry of a second similar plasmid into the same cell is by surface exclusion. For example, an F plasmid of *Escherichia coli* does not allow entry of another F plasmid by inhibiting it from leaving the cell where it is already located. The effect is mediated at the surface of the cell whereby the F DNA cannot come out of the cell.
A different mechanism operates when a cell already has two closely related plasmids, say X and Y. We know that the copy number of plasmids is controlled by specific inhibitors coded by the plasmid itself.

As X and Y are two closely related plasmids, it would be expected that their inhibitors would also be closely similar and that replication of both the plasmids would be regulated by the inhibitor produced either by X or Y.

During replication, X and Y may be selected at random, so that, during first replication of the plasmid, a cell initially containing one copy of each plasmid may produce two copies of either X or Y, so that the cell has now two copies of either X or Y and one copy of the un-replicated plasmid, i.e., \( 2X + Y \) or \( X + 2Y \). In the second round of plasmid replication, each cell will contain 4 plasmids, but depending on which plasmid is replicated, the combination may be \( X + 3Y \), \( 2X + 2Y \) or \( 3X + Y \).

Now the cell divides to produce two daughter cells, each with 2 plasmids and the plasmid combinations of the daughter cells may be \( X + X \), \( X + Y \) or \( Y + Y \). Thus the probability of progeny cells having either two X plasmids or two Y plasmids is equal to those having two different plasmids, i.e., \( X + Y \). In other words, the probability of elimination of one plasmid is 50%. Such probability increases with more cell generations.

**Plasmid Library**

A plasmid library is a gene library which contains a collection of bacterial cultures, each of which contains a plasmid, but plasmid of one culture differs from that of another in having a separate DNA fragment of a genome of an organism. The total genome isolated from an organism is fragmented and the fragments are inserted (cloned) separately into individual plasmids.

These recombinant plasmids are then introduced into suitable host bacteria. Thus each bacterial culture contains a plasmid with a fragment. The total collection of cultures would be expected to contain the entire genome of an organism and would constitute a gene library of the particular organism.

*Agrobacterium tumefaciens* is a plant pathogen with the capacity to deliver a segment of oncogenic DNA carried on a large plasmid called the tumor-inducing or Ti plasmid to susceptible plant cells. *Agrobacterium tumefaciens* belongs to the class Alphaproteobacteria, whose members include other plant pathogens (*Agrobacterium rhizogenes*), plant and insect symbionts (*Rhizobium* spp. and *Wolbachia* spp., respectively), human pathogens (*Brucella* spp., *Bartonella* spp., *Rickettsia* spp.), and non-pathogens (*Caulobacter crescentus, Rhodobacter sphaeroides*). Many species of Alphaproteobacteria carry large plasmids ranging in size from ~100 kilobases (kbs) to nearly 2 megabases (Mbs). These large replicons typically code for functions essential for cell physiology, pathogenesis or symbiosis. Most of these elements rely on a conserved gene cassette termed RepABC for replication and partitioning, and
Plasmids: Types and its Properties

11.2.2 Properties and Characteristics

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
Plasmids: Types and its Properties

NOTES

Self-Instructional Material

metal, while others may produce virulence factors that enable a bacterium to
colonize a host and overcome its defences, 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 megaplasmids of several megabase pairs (Mbp). At the upper end,
little can differentiate 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.

Check Your Progress

1. What are plasmids?
2. Where are artificial plasmids used?
3. Write in short about vector.
4. What are F plasmid?
5. Define R-plasmids.
6. Write a short note on non-transmissible plasmids.

11.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. A plasmid is a small 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, such as by providing antibiotic resistance.

2. Artificial plasmids are widely used as vectors in molecular cloning, serving to drive the replication of recombinant DNA sequences within host organisms.

3. In molecular cloning, a vector is a DNA molecule used as a vehicle to artificially carry foreign genetic material into another cell, where it can be replicated and/or expressed (for example, plasmid, cosmid, Lambda Phages). A vector containing foreign DNA is termed recombinant DNA. The four major types of vectors are plasmids, viral vectors, cosmids, and artificial chromosomes. Of these, the most commonly used vectors are plasmids. Common to all engineered vectors are an origin of replication, a multi cloning site, and a selectable marker.

4. The F plasmid, also known as the fertility factor or sex factor, determines the sex of *Escherichia coli* bacteria. 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.

5. 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 *Shigella dysenteriae*. Since then these plasmids have been found in *Escherichia coli* and other enteric bacteria. Such plasmids have proved a great threat to the medical science.

6. Replication of plasmid DNA starts at the site of origin and may proceed either bi-directionally as in case of bacterial chromosome, or may proceed unidirectionally depending on the nature of plasmid. In bidirectional replication, replication terminates when the two replication forks meet each other. In unidirectional replication, termination occurs when the replication fork reaches the site of origin. In both cases, the circularity of the plasmid DNA is maintained throughout the process.

### 11.4 SUMMARY

- A plasmid is a small DNA molecule within a cell that is physically separated from chromosomal DNA and can replicate independently.
- In nature, plasmids often carry genes that benefit the survival of the organism, such as by providing antibiotic resistance.
- Artificial plasmids are widely used as vectors in molecular cloning, serving to drive the replication of recombinant DNA sequences within host organisms.
A bacterial cell may possess one or many copies of one or more plasmids. They are inherited from parent bacterial cell to daughter cell and have capability to self-replication in the cytoplasm of bacterial cell.

The circular molecule of DNA can be broken to yield a linear molecule which passes from one bacterial cell to another. There are many kinds of plasmids which differ from one another in size and in composition of genes.

Another valuable feature of plasmids is the presence of specific restriction sites where the enzyme restriction endonuclease makes a cut so that a foreign DNA segment may be jointed to the plasmid.

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.

Plasmids provide a mechanism for horizontal gene transfer within a population of microbes and typically provide a selective advantage under a given environmental state.

In molecular cloning, a vector is a DNA molecule used as a vehicle to artificially carry foreign genetic material into another cell, where it can be replicated and/or expressed (for example, plasmid, cosmid, Lambda phages).

A vector containing foreign DNA is termed recombinant DNA. The four major types of vectors are plasmids, viral vectors, cosmids, and artificial chromosomes.

Viral vectors are generally genetically engineered viruses carrying modified viral DNA or RNA that has been rendered noninfectious, but still contain viral promoters and also the transgene, thus allowing for translation of the transgene through a viral promoter.

Artificial chromosomes are manufactured chromosomes in the context of Yeast Artificial Chromosomes (YACs), Bacterial Artificial Chromosomes (BACs), or Human Artificial Chromosomes (HACs).

The four major types of vectors are plasmids, viral vectors, cosmids, and artificial chromosomes. Of these, the most commonly used vectors are plasmids

The purpose of a vector which transfers genetic information to another cell is typically to isolate, multiply, or express the insert in the target cell.

Vectors designed specifically for the expression of the transgene in the target cell are called expression vectors, and generally have a promoter sequence that drives expression of the transgene.

Transcription vectors are used to amplify their insert. The manipulation of DNA is normally conducted on Escherichia coli vectors, which contain elements necessary for their maintenance in Escherichia coli.
Plasmids vary widely in size. The smaller plasmids have molecular weights ranging from 4 - 5 x 10^6 Daltons, while the larger ones have molecular weights of 25 to 95 x 10^6 Daltons.

- Copy number is discontinuously variable, i.e., some plasmids generally the smaller ones — have a high copy number, while the larger plasmids have characteristically a low copy number.
- A plasmid free cell may arise spontaneously at a frequency of about 1 in 10^4 cells. Plasmid free cells may also be produced artificially by the use of mutagens. The process is commonly called curing of plasmids.
- Bacteria carrying such plasmids are capable of using such unusual substrates for growth and, obviously, enjoy special advantage over others lacking them.
- The F plasmid gives the power to carry out a type of sexual reproduction to bacteria making it possible to exchange genetic materials leading to genetic recombination.
- The R-plasmids with the help of the resistance genes produce proteins which can inactivate or destroy specific antibiotics.
- A segment of DNA containing the specific gene is isolated from a suitable donor and inserted by recombinant DNA technology into a plasmid.
- The recombinant plasmid is next introduced into a suitable host cell where the gene is expressed producing the gene product.
- The F plasmid, also known as the fertility factor or sex factor, determines the sex of *Escherichia coli* bacteria.
- 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.
- 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 *Shigella dysenteriae*.
- 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.
- The Col-plasmids are present in different strains of *Escherichia 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 (Cor).
- Degradation or dissimilation of organic compounds in course of mineralization is often controlled by plasmid borne genes in many microorganisms.
Plasmids: Types and its Properties

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.

- 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.
- Plasmids occur rarely in Eukaryotic cells. Some plasmids have been found in yeast (Saccharomyces cerevisiae) and in several plants.
- The protein can kill other yeast cells lacking the plasmid. This yeast plasmid has been designated as killer particle.
- In the non-dividing plasmids, the double-stranded DNA exists as a right-handed super-helical coil having 400-600 base pairs per turn of the coil.
- Each plasmid has its own origin of replication. Some plasmids also have genes which code for proteins necessary for their own multiplication.
- The total genome isolated from an organism is fragmented and the fragments are inserted (cloned) separately into individual plasmids.
- 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, plasmid 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.
- 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.
- 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.

11.5 KEY WORDS

- **Plasmid**: A plasmid is a small DNA molecule within a cell that is physically separated from chromosomal DNA and can replicate independently.
- **Relaxed plasmids**: Plasmids having a high copy number are known as relaxed plasmids.
- **Stringent plasmids**: Plasmids having low copy number are called stringent plasmids.
11.6 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What are plasmids?
2. Name the types of plasmids.
3. Draw a well-labelled diagram of a plasmid.
4. What is a viral vector?
5. What is F plasmid?
6. Draw a well-labelled diagram of Ti plasmid.
7. What are the different aspects of plasmid replication?
8. Give some general properties of plasmids.

Long Answer Questions

1. What are plasmids? Discuss the types of plasmids giving examples.
2. Elaborate a note on artificial chromosomes.
3. Draw a well-labelled diagram to show F plasmid.
4. Write a detailed note on degradative plasmids.
5. Discuss about Ti plasmid of Agrobacterium.
6. Explain the replication of plasmids.
7. Write a note on incompatibility of plasmids.
8. Draw a well-labelled diagram of types of plasmid integration into a host bacteria.
9. Discuss about the properties of plasmids.

11.7 FURTHER READINGS

NOTES


UNIT 12 DETECTION, PURIFICATION AND TRANSFER OF PLASMID DNA, AND PLASMID REPLICATION

Structure
12.0 Introduction
12.1 Objectives
12.2 Plasmid DNA
12.3 Detection and Purification of Plasmid DNA
12.3.1 Isolation and Purification of Plasmid DNA
12.4 Transfer of Plasmid DNA
12.5 Replication of Plasmid
12.6 Plasmid Copy Number
12.7 Answers to Check Your Progress Questions
12.8 Summary
12.9 Key Words
12.10 Self Assessment Questions and Exercises
12.11 Further Readings

12.0 INTRODUCTION

A plasmid is a small 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, such as by providing antibiotic resistance. While the chromosomes are big and contain all the essential genetic information for living under normal conditions, plasmids usually are very small and contain only additional genes that may be useful in certain situations or conditions.

The term ‘plasmid’ was coined by Joshua Lederberg in 1592. Originally evolved from bacteria, plasmids are extra-chromosomal genetic elements that are present in most species of Archae, Eukarya and Eubacteria that have ability to replicate independently. Plasmids are circular double stranded DNA molecule that...
are distinct from the cells chromosomal DNA. The structure and function of a bacterial cell is directed by the genetic material contained within the chromosomal DNA. In some cases plasmids are generally not essential for the survival of the host bacterium. Plasmids specify traits that allow the host to persist in environments that would otherwise be either lethal or restrictive for growth. For example, antibiotic resistance and protein expression. Antibiotic resistance genes are typically encoded by the plasmid, which allows the bacteria to survive in an antibiotic containing environment, thereby providing the bacterium with a competitive advantage over antibiotic-sensitive species. In addition, the plasmids as a tool can be modified to express the protein of interest, for example production of human insulin using recombinant DNA technology.

Every gene manipulation procedure requires genetic material like DNA and RNA. Fundamentally, the nucleic acids occur naturally in association with proteins and lipoprotein organelles. The dissociation of a nucleoprotein into nucleic acid and protein moieties and their subsequent separation, are the essential steps in the isolation of all species of nucleic acids. Isolation of nucleic acids is followed by quantitation of nucleic acids which is normally done by either spectrophotometric or by using fluorescent dyes to determine the average concentrations and purity of DNA or RNA present in a mixture. Isolation of the genetic material (DNA) from cells (bacterial, viral, plant or animal) involves three basic steps, namely the rupturing of cell membrane to release the cellular components and DNA, separation of the nucleic acids from other cellular components and purification of nucleic acids.

In this unit, you will study about the detection and purification of plasmid DNA, transfer of plasmid DNA, replication of plasmid, control of copy number, plasmid amplification, curing and incompatibility.

12.1 OBJECTIVES

After going through this unit, you will be able to:
- Discuss what plasmid DNA is
- Explain the detection and purification of plasmid DNA
- Analyse the transfer of plasmid DNA
- Define replication of plasmid
- Understand control of copy number and plasmid amplification
- Elaborate on curing and incompatibility

12.2 PLASMID DNA

A plasmid is a small 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, such as by providing antibiotic resistance. While the chromosomes are big and contain all the essential genetic information for living under normal conditions, plasmids usually are 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.

Plasmids are considered replicons, units of DNA capable of replicating autonomously within a suitable host. However, plasmids, like viruses, are not generally classified as life. 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. 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 essential 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 addition, the plasmids as a tool can be modified to express the protein of interest, for example production of human insulin using recombinant DNA technology.

Characteristics of Plasmids

Plasmids present in the bacterium differ in their physical properties, such as in size (kbp), geometry and copy number.

**Plasmid Size:** Plasmids range in size from 1 kbp (kilo base pair) to 1000 (kilo base pair) mega-plasmids that are many hundred base pairs in size.

**Plasmid Geometry:** Even though most plasmids possess a circular geometry, there are now many examples of plasmids that are linear in most of the bacteria. Plasmid DNA may appear in one of the five conformations nicked open circular DNA which has one strand cut. The undisturbed circular DNA is completely intact with both strands uncut, but has been enzymatically relaxed. The linear DNA has free ends, while the supercoiled DNA is fully intact with both strands uncut.

**Plasmid Copy Numbers:** Copy number is the significant feature and refers to the average or expected number of copies per host cell. Plasmids have either low, medium or high copy number. After knowing that to which category plasmid belongs to is very significant to start an experiment. When working with a low copy number plasmid which is associated with a low yield may therefore require more cultures to be done. Alternatively, if the yield obtained from a high copy plasmid is poor, then troubleshooting is essential. In bacterium with high copy number plasmids,
during cell division the plasmids get segregate randomly in the daughter cells, whereas bacterium with low copy numbers, during cell division and partition the plasmids divided equally in the daughter cells. An advantage of high copy number is the greater stability of the plasmid when random partitioning, i.e., the partitioning of plasmids into daughter cells occurs at cell division.

Every gene manipulation procedure requires genetic material like DNA and RNA. Fundamentally, the nucleic acids occur naturally in association with proteins and lipoprotein organelles. The dissociation of a nucleoprotein into nucleic acid and protein moieties and their subsequent separation, are the essential steps in the isolation of all species of nucleic acids. Isolation of nucleic acids is followed by quantitation of nucleic acids which is normally done by either spectrophotometric or by using fluorescent dyes to determine the average concentrations and purity of DNA or RNA present in a mixture. Isolation of the genetic material (DNA) from cells (bacterial, viral, plant or animal) involves three basic steps, namely the rupturing of cell membrane to release the cellular components and DNA, separation of the nucleic acids from other cellular components and purification of nucleic acids.

### 12.3 DETECTION AND PURIFICATION OF PLASMID DNA

Plasmids are important vehicles for rapid adaptation of bacterial populations to changing environmental conditions. Among the various methods available for the preparation of plasmid DNA for rapid screening, a protocol involving the use of an alkaline solution to lyse the cells, salt precipitation to remove cell debris and chromosomal DNA and application to bind DNA column to eliminate proteins and other contaminants has been widely used. Gel electrophoresis is considered as the popular method that can be easily performed resolution technique in nucleic acid research. Gel Electrophoresis using Agarose, is considered as highly purified linear polysaccharide derived from agar and is widely used in the detection and characterization of plasmids, and also the linear DNA fragments. Plasmids of sizes ranging from less than one kilo base (kb) to over a few hundred kb can be easily resolved by conventional Agarose Gel Electrophoresis.

Plasmid purification is a technique used to isolate and purify plasmid DNA from genomic DNA, proteins, ribosomes, and the bacterial cell wall. A plasmid is a small, circular, double-stranded DNA that is used as a carrier of specific DNA molecules. When introduced into a host organism via transformation, a plasmid will be replicated, creating numerous copies of the DNA fragment under study.

A plasmid preparation is a method of DNA extraction and purification for plasmid DNA. Various methods have been developed to purify plasmid DNA from bacteria. All the purification methods involve following three steps:

Step 1: Growth of the Bacterial Culture.
Step 2: Harvesting and Lysis of the Bacteria.

Step 3: Purification of Plasmid DNA.

The purification of plasmid DNA from bacterial cells is a significant step specifically in the cloning process. During plasmid purification, the bacterial cells are lysed, freeing DNA and other cellular components from the cell wall. Cellular components are then removed, and the DNA containing lysate is processed to further remove contaminants to separate the plasmid DNA from the genomic DNA.

12.3.1 Isolation and Purification of Plasmid DNA

The isolation of plasmid DNA from bacteria is a critical technique in molecular biology and is an essential step in many procedures, such as cloning, DNA sequencing, transfection, and gene therapy. These manipulations require the isolation of high purity plasmid DNA. The purified plasmid DNA can be used for immediate use in all molecular biology procedures, such as digestion with restriction enzymes, cloning, PCR, transfection, in vitro translation, blotting and sequencing.

Alkaline lysis is a method used in molecular biology, to isolate plasmid DNA or other cell components, such as proteins by breaking the cells. Bacteria containing the plasmid of interest is first grown, and then permitted to lyse with an alkaline lysis buffer consisting of a detergent Sodium Dodecyl Sulfate (SDS) and a strong base Sodium Hydroxide (NaOH). The detergent splits the phospholipid bilayer of membrane and the alkali denatures the proteins which are involved in maintaining the structure of the cell membrane. This involves series of steps, such as agitation, precipitation, centrifugation, and the removal of supernatant. The cellular debris is removed and the plasmid is isolated and purified.

Principle

Purification of plasmid DNA from bacterial DNA is based on the differential denaturation of chromosomal and plasmid DNA that uses the alkaline lysis for separating the two. The basic steps of plasmid isolation are disruption of the cellular structure to create a lysate, separation of the plasmid from the chromosomal DNA, cell debris and other insoluble material.

Bacteria are lysed with a lysis buffer solution containing Sodium Dodecyl Sulfate (SDS) and Sodium Hydroxide (NaOH). Throughout this step, the disruption of most cells is done, i.e., the chromosomal as well as plasmid DNA are denatured and the resulting Lysate is cleared by means of centrifugation, filtration or magnetic clearing. Subsequent neutralization with Potassium Acetate helps only the covalently closed plasmid DNA to reanneal and to remain solubilized. Most of the chromosomal DNA and proteins precipitate in a complex are formed with Potassium and SDS, which is removed by centrifugation.

The bacteria is resuspended in a resuspension buffer (50mM Tris-Cl, 10 mM EDTA, 100 μg/ml RNase A, pH 8.0) and then treated by 1% SDS (w/v) / Alkaline Lysis buffer (200mM NaOH) to liberate the plasmid DNA from the
Detection, Purification and Transfer of Plasmid DNA, and Plasmid Replication

NOTES

Self-Instructional Material

Escherichia coli host cells. Neutralization buffer (3.0 M Potassium Acetate, pH 5.0) neutralizes the resulting Lysate for creating appropriate or suitable conditions to bind the plasmid DNA to the Silica membrane column. Subsequently, the precipitated Protein, Genomic DNA, and Cell Debris are pelleted through a centrifugation step and the supernatant is loaded onto a column. Contamination, such as the salts, metabolites, and soluble macromolecular cellular components are removed using the simple washing with Ethanol Wash buffer (1.0 M NaCl, 50mM MOPS, pH 7.0, Isopropanol (v/v) 15 %). Finally, the ‘Pure Plasmid DNA’ is eluted under low ionic strength conditions with slightly Alkaline buffer (5 mM Tris / HCl, pH 8.5).

Culture Media

Yield and quality of plasmid DNA extremely depends on the type of culture media being used. Most of the plasmid purification are optimized with cultures grown in standard Luria Bertani (LB) medium.

To prepare the LB medium dissolve 10 g Tryptone, 5 g Yeast Extract, and 10 g NaCl in 800 ml Distilled Water. Adjust the pH to 7.0 with 1 N NaOH. Adjust the volume to 1 liter by adding distilled water and sterilize by autoclaving. The cell culture should be incubated at 37°C with constant shaking (200–250 rpm) preferably 12–16 hours overnight. Alternatively, rich media like 2 x YT (Yeast / Tryptone), TB (Terrific Broth), or CircleGrow can be used.

To find the optimal culture conditions, the culture medium and incubation times have to be optimized individually for each host strain / plasmid construct combination.

Lysate and Neutralization

Lysis formulas can differ depending on which extract DNA/RNA/Plasmid is required. All methods of lysing bacteria will yield plasmid solutions that is contaminated with chromosomal DNA and RNA. Centrifugation removes the majority of chromosomal DNA that forms a pellet, while the plasmid DNA remains soluble. The treatment with RNase will eliminate contaminating RNA.

In general, lysis buffers contain a high concentration of chaotropic salts. Chaotropes have two important roles in nucleic acid extraction. First is they destabilize Hydrogen bonds, van der Waals forces and hydrophobic interactions, leading to destabilization of proteins, including nucleases. Second is they disrupt the association of nucleic acids with water, hence providing optimal conditions for their transfer to Silica.

Separation and removal of the plasmids from the bacterial cell is performed by resuspension of 1-5 mL of culture in a resuspension buffer (50mM Tris-Cl, 10 mM EDTA, 100 µg/ ml RNase A, pH 8.0) and pellet cells in a microcentrifuge at 11000 x g for 30 seconds.
Lysate is achieved by adding 250 μL of lysis buffer with neutralization buffer, as it helps in complete precipitation of SDS, Protein, and Genomic DNA. Incomplete neutralization will lead to reduced yield. Do not shake the released plasmid DNA too much or too strongly because it will damage the DNA.

**Binding and Washing in Silica Membrane**

After centrifuging the Lysate through Silica membrane, the desired nucleic acids are bound to the columned and impurities, such as Protein and Polysaccharides should be removed. Principally, the plant samples contain Polysaccharides and Pigments, while for blood samples contain the membrane which may be slightly brown or yellow in colour. The washing steps will remove such impurities. Typically, there are two washing steps, though it varies depending on the sample type. The first washing step includes a low concentration of Chaotropic Salts to remove residual Proteins and Pigments and is always followed with an Ethanol Wash to remove the Salts.

Columns containing the Silica resin selectively binds to DNA/RNA. The DNA of interest can be isolated due to its ability to bind Silica in the presence of high concentrations of Chaotropic Salts. These salts are then removed with an Alcohol Based Wash and the DNA is eluted using a low-ionic-strength solution, such as TE buffer or Water. The binding of DNA to Silica appears to be determined by dehydration and Hydrogen bond formation, which competes against weak electrostatic repulsion. Hence, a high concentration of Salt will help in the determination of DNA adsorption onto Silica, and a low concentration will release the DNA.

**Elution**

The elution buffer volume and method is specifically adapted to the subsequent application for achieving higher yield and/or concentration that uses the standard method. Elution buffer is used to wash the Unbound Proteins at first and at a greater concentration it releases the Desired Protein from the ligand. It is significant that the elution buffer mechanism works fast without changing the function or activity of the Desired Protein. For maximal DNA elution, the buffer is allowed to stand in the membrane for a few minutes before centrifugation. Elution Buffer AE (5 mM Tris/HCl, pH 8.5) can also be replaced by TE buffer or Water. Using a weakly buffered slightly alkaline buffer containing no EDTA is preferred especially if the plasmid DNA is intended for sequencing reactions.

**Analytical Gel Analysis**

It is recommended to remove and save the aliquots during the purification procedure. If the plasmid DNA is of low yield or low in quality, then the samples can be analyzed by Agarose Gel Electrophoresis to determine at what stage of the purification procedure the problem occurred.
Detection, Purification and Transfer of Plasmid DNA, and Plasmid Replication

NOTES

Self-Instructional Material

Procedure: It includes the following steps.

Step 1: Harvesting Bacterial and Resuspended Cells

1. Choose a single colony from a freshly streaked selective plate and inoculate a starter culture of 2–5 ml LB medium containing the appropriate selective antibiotic. Incubate for approximately 8 hours at 37°C with vigorous shaking (approximately 300 rpm).

2. Dilute the starter culture 1/500 to 1/1000 into 3 ml selective LB medium. Grow at 37°C for 12–16 hours with vigorous shaking (approximately 300 rpm).

3. Harvest the bacterial cells by centrifugation at 6000 x g for 15 minutes and remove as much of the supernatant as possible. Resuspend the bacterial pellet in 0.1-0.5 ml of resuspension buffer (50 mM Tris-Cl, 10 mM EDTA, 100 μg/mL RNase A, pH 8.0). The bacteria should be resuspended completely by vortexing or pipetting up and down until no cell clumps remain.

Step 2: Cell Lysis

4. Add 0.25 ml of lysis buffer, mix thoroughly by vigorously inverting the sealed tube 4–6 times, and incubate at room temperature (15–25°C) for 5 minutes. Do not vortex, as this will result in shearing of genomic DNA. The lysate should appear viscous. Do not allow the lysis reaction to proceed for more than 5 minutes.

Step 3: Neutralization

5. Add 0.3 mL of neutralization buffer, mix immediately and thoroughly by vigorously inverting 4–6 times, and incubate on ice for 5 minutes. Precipitation is enhanced by using chilled neutralization buffer and incubating on ice. After addition of neutralization buffer, a fluffy white material forms and the Lysate becomes less viscous. The precipitated material contains Genomic DNA, Proteins, Cell Debris, and KDS. The Lysate should be mixed thoroughly to ensure even Potassium Dodecyl Sulphate precipitation. If the mixture still appears viscous, more mixing is required to completely neutralize the solution. A homogeneous colorless suspension indicates that the SDS has been effectively precipitated.

Step 4: Loading Lysate on Column

6. Before loading the column, carefully remove the supernatant and then transfer it to a collection tube containing the column and centrifuge at 13,000 rpm for 1 minute.

7. Discard the flow-through liquid and remove supernatant containing plasmid DNA promptly. After centrifugation, the supernatants should be clear.

8. If the supernatant is not clear, a second, shorter centrifugation should be carried out to avoid applying any suspended or particulate material to the column. Suspended material, which can cause the sample to appear turbid, will clog the column and reduce or eliminate flow.
Step 5: Binding and Washing

9. Add 0.7 ml of wash buffer to the column placed in the collection tube and centrifuge for 10 minutes at 13000 rpm for 1 minute. Equilibrate by applying 1 mL equilibration buffer (750 mM NaCl, 50 Mm MOPS, pH 7.0, 15 % Isopropanol) and allow the column to empty by gravity flow. Flow of buffer will begin automatically by reduction in surface tension due to the presence of detergent in the equilibration buffer.

10. Apply the supernatant from Step 6 to the column and allow it to enter the resin by gravity flow.

Step 6: Plasmid Elution

11. Elute DNA with 0.8 mL elution buffer (1.23 M NaCl, 50 mm Tris-Cl, pH 8.5, 15 % Isopropanol). Collect the elute in a 1.5 mL or 2 mL microcentrifuge tube.

12. Precipitate DNA by adding 0.7 volumes (0.56 mL per 0.8 mL of elution volume) of room temperature Isopropanol to the eluted DNA. Mix and centrifuge immediately at ≥10,000 rpm for 30 minutes in a microcentrifuge. Cautiously decant the supernatant. All solutions should be at room temperature with the objective of minimize salt precipitation.

13. Wash DNA pellet with 1 mL of 70% Ethanol and centrifuge at 10,000 rpm for 10 minutes.

14. Carefully decant the supernatant without disturbing the pellet.

15. The 70% Ethanol removes precipitated salt and replaces Isopropanol with the more volatile Ethanol, making the DNA easier to re-dissolve.

16. Air-dry the pellet for 5–10 minutes, and re-dissolve the DNA in a suitable volume of buffer (e.g., TE buffer, pH 8.0, or 10mM Tris-Cl, pH 8.5). Re-dissolve the DNA pellet by rinsing the walls for recovering all the DNA.

Determination of Yield

For the determination of yield, DNA concentration must be determined by means of UV spectrophotometry at 260 nm and also by quantitative analysis on an Agarose Gel. To quantitate the nucleic acid concentration, dilute the plasmid DNA 1 : 100 or 1 : 50, this typically depends on the plasmid copy number, in TE buffer and measure the absorbance (optical density) at 260 nm (A260) and 280 nm (A280). The TE buffer must be used as the blank. This measurement can be done using the direct calculation of the nucleic acid concentration using the formula:

\[
[\text{DNA}] (\mu g/mL) = A_{260} \times \text{Dilution Factor} \times 50
\]

Where, 50 is the Extinction Coefficient of DNA. The ratio A260/A280 provides a reasonable or realistic estimate of the purity of the preparation.
Plasmids are referred as extra chromosomal circular DNA molecules that exist in most bacterial species and in some species of Eukaryotes. Under normal conditions, a specific plasmid is dispensable to its host cell. Many plasmids contain plasmid genes that are essential in certain specified environments, for example 'R' plasmids carry genes that provide resistance to numerous antibiotics such that the cell containing this plasmid can resist antibiotic action. Figure 12.1 illustrates the process of horizontal gene transfer.

Fundamentally, all known plasmids are supercoiled circular DNA except the one exceptional, i.e., the killer plasmid of Yeast, a RNA molecule. The molecular weight ranges from $10^6 – 10^8$. The number of copies of a Plasmid Per Cell ranges from 1–2 for low copy number plasmids and to 10–60 for high copy number plasmids.

Plasmid DNA can be isolated from bacteria using the culturing plasmid containing bacteria. Detergent is added to it to get Lysed bacteria. Then the Lysate is centrifuged. The smaller plasmid DNA remains in the supernatant. Cesium Chloride (CsCl) and Ethidium Bromide is added to the supernatant. The CsCl forms a density gradient. The supercoiled DNA with high density can be identified by Ethidium Bromide and easily removed.

Transfer of Plasmid DNA

The bacterium *Escherichia coli* possesses 2 mating types, i.e., donors or male and recipients or female. The determinant of male feature is the ‘F’ factor or sex.
plasmid or donor, therefore a male cell is designated as F+. Female lacks ‘F’ plasmid and designated as F−. When the culture of male and female is mixed, the conjugation takes place between the male and female pairs forming a conjugation bridge. Pairing induces looped rolling circle replication of ‘F’, the recipient cell contains the ‘F’ plasmid. During transfer both ‘Donor’ and ‘Recipient’ cells show DNA synthesis. In ‘Donor’, the synthesis replaces single strand to be transferred, the synthesis in the recipient converts the transferred single strand to double stranded DNA.

‘F’ plasmid has an ability to integrate into the bacterial chromosome. Principally, the ‘F’ fuses with chromosome increasing its size. The cell is now called Hfr cell or High frequency recombination cell. When a culture of Hfr is mixed with F− culture, the conjugation occurs. Replication starts in the Hfr cell and is transferred to F− cell. The direction of replication is specific, i.e., small portion of ‘F’ is transferred first and then the major part is transferred. Female receives both small and large fragments of the male chromosome. The Hfr is produced when ‘F’ integrates firmly into the chromosome. The plasmid containing both ‘F’ genes and chromosomal genes is called as ‘F’ plasmid.

12.5 REPLICATION OF PLASMID

An essential feature of bacterial plasmids is their ability to replicate as autonomous genetic elements in a controlled way within the host. Therefore, they are used for exploring the mechanisms involved in the DNA replication and for analyzing the different strategies that couple DNA replication to other critical procedures in the cell cycle.

Plasmid replication process is divided into three stages, namely initiation, elongation, and termination. The inability of DNA polymerases to initiate de novo replication for creating the essential independent generation of a primer.

This is explained, in circular plasmids, through following two main strategies:
- Opening of the strands followed by RNA priming (theta and strand displacement replication).
- Cleavage of one of the DNA strands to generate a 3'-OH end (rolling-circle replication).

Initiation is catalyzed regularly through one or a few plasmid-encoded initiation Proteins that are capable of recognizing the plasmid-specific DNA sequences and determining the point from which replication starts, i.e., the origin of replication. In some cases, these Proteins also participate directly in the generation of the primer. These initiators are also involved in the role of Pilot Proteins that guide the assembly of the host Reposome at the Plasmid Origin.

Elongation of plasmid replication is carried out basically through DNA Polymerase III Holoenzyme and in some cases using DNA Polymerase I at an
early stage, with the participation of other Host Proteins that form the Replisome. Termination of replication includes the precise requirements and implications for reinitiation.

The initiation stage performs an additional role, as at this stage the mechanisms controlling replication function or operate. The key objective of this control is to maintain a fixed concentration of plasmid molecules in a growing bacterial population, i.e., duplication of the plasmid pool paced with duplication of the bacterial population. The molecules that are directly involved in this control can be,

- RNA (Antisense RNA)
- DNA Sequences (Iterons)
- Antisense RNA and Proteins involved in the Mechanism.

Characteristically, the control elements maintain an average frequency of one plasmid replication through ‘per plasmid copy per cell cycle’ and to ‘sense’ and correct deviations from this average. The plasmid replication and its control is based on the results of analyses performed using pure cultures under steady-state growth conditions. Thus setting significant parameters required to understand the maintenance of these genetic elements in mixed populations and under environmental conditions.

As already discussed the plasmids are extra chromosomal DNA elements with characteristic copy numbers within the host. These replicons occur in species from the three representatives of the living world, namely, the domains Archaea, Bacteria, and Eukarya. Plasmids may constitute a substantial amount of the total genetic content of an organism, representing more than 25% of the genetic material of the cell in some members of the Archaea. They can, therefore, incorporate and deliver genes by recombination or transposition, thus supporting genetic exchanges in bacterial populations. Since plasmids can be hosted or introduced into new hosts through various mechanisms, hence they are considered as a group of extra chromosomal DNA which is shared between populations. The genetic information are carried by plasmids, and their effect in the microbial communities, and the potential of these elements perform as natural cloning vectors that have stimulated exploration into plasmids not only from the fundamental but also from the clinical, biotechnological, and environmental aspects.

Following are the three key factors that contribute to the development of plasmid investigation or exploration:

- The genetic organization of these elements is apparently simple.
- They can be easily isolated and manipulated in vitro.
- Because plasmids are dispensable, hence their manipulation does not appear, in principle, to have adverse consequences to the hosts.

The characteristic feature that expresses plasmids is that they can replicate in an autonomous and self-controlled method. The plasmid replication analysis and its control, such as the existence of antisense RNAs, have contributed to the
unraveling of mechanisms of DNA replication, macromolecular interactions, and control of gene expression. Some plasmids have the ability to pass across the so-called genetic barriers among different living organisms and thus rendering questions about general mechanisms that govern the replication and the communication between plasmid replication components and the host machinery involved in DNA replication.

Despite their autonomous replication, the plasmids extensively use the replication machinery of the host, and therefore plasmid replication facilitates the exploration of the mechanisms involved in chromosome replication.

Origins of Replication

Plasmid origins of replication can be defined as follows:

- The minimal cis acting region that can support autonomous replication of the plasmid.
- The region where DNA strands are melted to initiate the replication process.
- The base(s) at which leading-strand synthesis starts.

Replication origins contain sites that are required for interactions of plasmid encoded proteins and/or host encoded proteins.

General Features: With some exceptions, initiation of plasmid DNA replication requires a specific plasmid-encoded Rep initiator protein. This is reflected by the presence, at the origin of replication, of specific sequences with which the Rep protein interacts. Additional features found in many origins of theta-replicating plasmids include an adjacent AT-rich region containing sequence repeats, where opening of the strands and assembly of host initiation factors occur, and one or more sites where the host DNA A initiator protein binds. Multiple methylation sequences that are present in the origin of replication of the *Escherichia coli* chromosome, oriC, can also be found at the origin of replication of plasmids, such as P1 and pSC101. Methylation is not essential for replication, because its primary role is in post-replication.

Iteron Containing Origins: In many situations, the origin of replication contains directly repeated sequences, termed iterons, which are the binding sites for the plasmid-encoded Rep Proteins having control properties. The iterons are not only essential for replication but are also the key elements for the control of plasmid replication. Among plasmids which restrict their establishment to a single or a few species of *Enterobacteria*, iterons have been described for several replicons like P1, F, pSC101, etc. Iterons can be adjacent or separated by intervening sequences.

12.6 PLASMID 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 per cell) from which several very high copy number cloning vectors (~1000 copies per cell) have been derived.

**Regulation**

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, namely Antisense RNA or Iteron Binding Groups. Low copy number plasmids, also called stringent plasmids, require tighter control of replication.

**ColE1 Derived Plasmids: Antisense RNA**

In ColE1 derived plasmids, replication is primarily regulated through a small plasmid-encoded RNA called RNA I. A single promoter initiates replication in ColE1: the RNA II promoter. The RNA II transcript forms a stable RNA-DNA hybrid with the DNA template strand near the origin of replication, where it is then processed by RNaseH to produce the 3' OH primer that DNA Polymerase I uses to initiate leading strand DNA synthesis. RNA I serves as a major plasmid-encoded inhibitor of this process whose concentration is proportional to plasmid copy number. RNA I and RNA II first form a weak interaction which is stabilized by a Protein called Rop (Repressor of primer) and a double-stranded RNA-I/RNA-II, RNA duplex is formed. This altered shape prevents RNA II from hybridizing to the DNA and being processed from RNaseH to produce the primer necessary for initiation of plasmid replication. More RNA I is produced when the concentration of the plasmid is high, and high concentration of RNA I inhibits replication, resulting in regulation of copy number.

**R1 and ColIb-p9 Plasmids: Antisense RNA**

Most plasmids require a plasmid-encoded protein, usually called Rep, to separate the strands of DNA at the origin of replication (oriV) to initiate DNA replication. Rep binds to specific DNA sequences in oriV which are unique to a plasmid type. The synthesis of Rep protein is controlled in order to limit plasmid replication and therefore regulate copy number. In R1 plasmids RepA can be transcribed from two different promoters. It is made from the first promoter until the plasmid reaches its copy number, upon which the protein CopB represses this primary promoter. RepA expression is also regulated post-transcriptionally from the secondary promoter by an Antisense RNA called CopA. CopA interacts with its RNA target in the RepA mRNA and forms a RNA-RNA duplex. The resultant double stranded RNA is cleaved by RNase III, preventing synthesis of RepA. The higher
concentration of the plasmid, the more CopA RNA is produced and the less RepA protein can be synthesized, increasing inhibition of plasmid replication.

**Col1b-P9: Antisense RNA**

Replication of the low-copy-number Col1b-P9 depends upon Rep, which is produced by expression of the repZ gene. The repZ expression requires formation of a pseudoknot in the mRNA. Basically, repZ is repressed by a small antisense Inc RNA, which binds to repZ mRNA, forms an Inc RNA-mRNA duplex, and prevents formation of the pseudoknot to inhibit repZ translation into Rep.

**pSC101: Iteron Plasmid**

Iteron plasmids, including F and RK2-related plasmids, have oriV regions containing multiple (~3-7) repeats of 17-22 bp iteron sequences. The pSC101 represents a simple model of an iteron plasmid. Iteron plasmids control copy number through two combined methods, suitable for low-copy number stringent plasmids. One method is control of RepA synthesis. RepA is the only plasmid-encoded protein required for replication in pSC101. RepA protein represses its own synthesis by binding to its own promoter region and blocking transcription of itself (Transcriptional Autoregulation). Thus, the more RepA is made, the more its synthesis is repressed, and subsequently limiting plasmid replication. The coupling hypothesis proposes that the second method is coupling of plasmids through the Rep Protein and Iteron Sequences. When the plasmid concentration is high, RepA plasmids bound to iterons form dimers in between two plasmids, ‘handcuffing’ them at the origin of replication and inhibiting replication.

**Incompatibility**

Plasmids can be incompatible if they share the same replication control mechanism. Under these circumstances, both plasmids contribute to the total copy number and are regulated together. They are not recognized as distinct plasmids. As such, it becomes much more likely that one of the plasmids may be out-copied by the other and lost during cell division, the cell is ‘cured’ of the plasmid. This is particularly likely with low copy number plasmids. Plasmids can also be incompatible due to shared partitioning systems.

**Incompatibility Groups**

1. Not all plasmids can live together.
2. Plasmids that are able to coexist in the same cell do not interfere with each other’s replication.
3. A single cell can have as many Inc group plasmids as it can tolerate and replicate.

Detection and purification of plasmid DNA. Transfer of plasmid DNA. Replication of plasmid. Control of copy number, plasmid amplification, curing and incompatibility.
Check Your Progress
1. Explain the term plasmid.
2. Why plasmids are considered replicons?
3. How are plasmids transmitted?
4. What is plasmid copy numbers? Explain with example.
5. What is plasmid purification? How it is done?
6. How the isolation of plasmid DNA is done from bacteria?
7. How is plasmid DNA transferred?
8. Explain the term plasmid replication.
9. How the control elements maintain an average frequency of one plasmid replication?
10. When plasmids can be incompatible?

12.7 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS
1. A plasmid is a small 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, such as by providing antibiotic resistance. Plasmids present in the bacterium differ in their physical properties, such as in size (kbp), geometry and copy number.

2. Plasmids are considered replicons, units of DNA capable of replicating autonomously within a suitable host.

3. 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. 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 essential for their own transfer.

4. Copy number is the significant feature and refers to the average or expected number of copies per host cell. Plasmids have either low, medium or high copy number. After knowing that to which category plasmid belongs to is very significant to start an experiment. When working with a low copy number plasmid which is associated with a low yield may therefore require...
more cultures to be done. Alternatively, if the yield obtained from a high copy plasmid is poor, then troubleshooting is essential. In bacterium with high copy number plasmids, during cell division the plasmids get segregate randomly in the daughter cells, whereas bacterium with low copy numbers, during cell division and partition the plasmids divided equally in the daughter cells. An advantage of high copy number is the greater stability of the plasmid when random partitioning, i.e., the partitioning of plasmids into daughter cells occurs at cell division.

5. Plasmid purification is a technique used to isolate and purify plasmid DNA from genomic DNA, proteins, ribosomes, and the bacterial cell wall. A plasmid is a small, circular, double-stranded DNA that is used as a carrier of specific DNA molecules. When introduced into a host organism via transformation, a plasmid will be replicated, creating numerous copies of the DNA fragment under study.

A plasmid preparation is a method of DNA extraction and purification for plasmid DNA. Various methods have been developed to purify plasmid DNA from bacteria. All the purification methods involve following three steps:

Step 1: Growth of the Bacterial Culture.
Step 2: Harvesting and Lysis of the Bacteria.
Step 3: Purification of Plasmid DNA.

During plasmid purification, the bacterial cells are lysed, freeing DNA and other cellular components from the cell wall. Cellular components are then removed, and the DNA containing lysate is processed to further remove contaminants to separate the plasmid DNA from the genomic DNA.

6. The isolation of plasmid DNA from bacteria is a critical technique in molecular biology and is an essential step in many procedures, such as cloning, DNA sequencing, transfection, and gene therapy. These manipulations require the isolation of high purity plasmid DNA. Alkaline lysis is a method used in molecular biology, to isolate plasmid DNA or other cell components, such as proteins by breaking the cells. Bacteria containing the plasmid of interest is first grown, and then permitted to lyse with an alkaline lysis buffer consisting of a detergent Sodium Dodecyl Sulfate (SDS) and a strong base Sodium Hydroxide (NaOH). The detergent splits the phospholipid bilayer of membrane and the alkali denatures the proteins which are involved in maintaining the structure of the cell membrane. This involves series of steps, such as agitation, precipitation, centrifugation, and the removal of supernatant. The cellular debris is removed and the plasmid is isolated and purified.

7. ‘F’ plasmid has an ability to integrate into the bacterial chromosome. Principally, the ‘F’ fuses with chromosome increasing its size. The cell is now called Hfr cell or High frequency recombination cell. When a culture of
Hfr is mixed with F– culture, the conjugation occurs. Replication starts in
the Hfr cell and is transferred to F– cell. The direction of replication is specific,
i.e., small portion of ‘F’ is transferred first and then the major part is
transferred. Female receives both small and large fragments of the male
chromosome. The Hfr is produced when ‘F’ integrates firmly into the
chromosome. The plasmid containing both ‘F’ genes and chromosomal
genes is called as ‘F’ plasmid.

8. An essential feature of bacterial plasmids is their ability to replicate as
autonomous genetic elements in a controlled way within the host. Therefore,
they are used for exploring the mechanisms involved in the DNA replication
and for analyzing the different strategies that couple DNA replication to
other critical procedures in the cell cycle. Plasmid replication process is
divided into three stages, namely initiation, elongation, and termination. The
inability of DNA polymerases to initiate de novo replication for creating the
essential independent generation of a primer.

9. Characteristically, the control elements maintain an average frequency of
one plasmid replication through ‘per plasmid copy per cell cycle’ and to
‘sense’ and correct deviations from this average. The plasmid replication
and its control is based on the results of analyses performed using pure
cultures under steady-state growth conditions. Thus setting significant
parameters required to understand the maintenance of these genetic elements
in mixed populations and under environmental conditions.

10. Plasmids can be incompatible if they share the same replication control
mechanism. Under these circumstances, plasmids contribute to the total
copy number and are regulated together. They are not recognized as distinct
plasmids.

One of the plasmids may be out-copied by the other and lost during cell
division, the cell is ‘cured’ of the plasmid. This is particularly likely with low
copy number plasmids. Plasmids can also be incompatible due to shared
partitioning systems.

12.8 SUMMARY

- A plasmid is a small 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,
such as by providing antibiotic resistance.

- While the chromosomes are big and contain all the essential genetic
information for living under normal conditions, plasmids usually are 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.

- Plasmids are considered replicons, units of DNA capable of replicating autonomously within a suitable host. However, plasmids, like viruses, are not generally classified as life.

- 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.

- 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.

- 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.

- Plasmids present in the bacterium differ in their physical properties, such as in size (kbp), geometry and copy number.

- Plasmids range in size from 1 kbp (kilo base pair) to 1000 (kilo base pair) mega-plasmids that are many hundred base pairs in size.

- Plasmid DNA may appear in one of the five conformations nicked open circular DNA which has one strand cut. The undisturbed circular DNA is completely intact with both strands uncleared, but has been enzymatically relaxed.

- Copy number is the significant feature and refers to the average or expected number of copies per host cell. Plasmids have either low, medium or high copy number.

- Among the various methods available for the preparation of plasmid DNA for rapid screening, a protocol involving the use of an alkaline solution to lyse the cells, salt precipitation to remove cell debris and chromosomal DNA and application to bind DNA column to eliminate proteins and other contaminants has been widely used.

- Gel electrophoresis is considered as the popular method that can be easily performed resolution technique in nucleic acid research.

- Gel Electrophoresis using Agarose, is considered as highly purified linear polysaccharide derived from agar and is widely used in the detection and characterization of plasmids, and also the linear DNA fragments.

- Plasmids of sizes ranging from less than one kilo base (kb) to over a few hundred kb can be easily resolved by conventional Agarose Gel Electrophoresis.
Plasmid purification is a technique used to isolate and purify plasmid DNA from genomic DNA, proteins, ribosomes, and the bacterial cell wall.

A plasmid is a small, circular, double-stranded DNA that is used as a carrier of specific DNA molecules. When introduced into a host organism via transformation, a plasmid will be replicated, creating numerous copies of the DNA fragment under study.

A plasmid preparation is a method of DNA extraction and purification for plasmid DNA. Various methods have been developed to purify plasmid DNA from bacteria.

The isolation of plasmid DNA from bacteria is a critical technique in molecular biology and is an essential step in many procedures, such as cloning, DNA sequencing, transfection, and gene therapy. These manipulations require the isolation of high purity plasmid DNA.

The purified plasmid DNA can be used for immediate use in all molecular biology procedures, such as digestion with restriction enzymes, cloning, PCR, transfection, in vitro translation, blotting and sequencing.

Alkaline lysis is a method used in molecular biology, to isolate plasmid DNA or other cell components, such as proteins by breaking the cells. Bacteria containing the plasmid of interest is first grown, and then permitted to lyse with an alkaline lysis buffer consisting of a detergent Sodium Dodecyl Sulfate (SDS) and a strong base Sodium Hydroxide (NaOH).

The detergent splits the phospholipid bilayer of membrane and the alkali denatures the proteins which are involved in maintaining the structure of the cell membrane. This involves series of steps, such as agitation, precipitation, centrifugation, and the removal of supernatant. The cellular debris is removed and the plasmid is isolated and purified.

Yield and quality of plasmid DNA extremely depends on the type of culture media being used. Most of the plasmid purification are optimized with cultures grown in standard Luria Bertani (LB) medium.

The elution buffer volume and method is specifically adapted to the subsequent application for achieving higher yield and/or concentration that uses the standard method.

Fundamentally, all known plasmids are supercoiled circular DNA except the one exceptional, i.e., the killer plasmid of Yeast, a RNA molecule. The molecular weight ranges from $10^6$ – $10^8$. The number of copies of a Plasmid Per Cell ranges from 1 – 2 for low copy number plasmids and to 10 – 60 for high copy number plasmids.

‘F’ plasmid has an ability to integrate into the bacterial chromosome. Principally, the ‘F’ fuses with chromosome increasing its size. The cell is now called Hfr cell or High frequency recombination cell. When a culture of Hfr is mixed with F’ culture, the conjugation occurs.
Replication starts in the Hfr cell and is transferred to F- cell. The direction of replication is specific, i.e., small portion of ‘F’ is transferred first and then the major part is transferred. Female receives both small and large fragments of the male chromosome.

The Hfr is produced when ‘F’ integrates firmly into the chromosome. The plasmid containing both ‘F’ genes and chromosomal genes is called as ‘F’ plasmid.

An essential feature of bacterial plasmids is their ability to replicate as autonomous genetic elements in a controlled way within the host. Therefore, they are used for exploring the mechanisms involved in the DNA replication and for analyzing the different strategies that couple DNA replication to other critical procedures in the cell cycle.

Plasmid replication process is divided into three stages, namely initiation, elongation, and termination. The inability of DNA polymerases to initiate de novo replication for creating the essential independent generation of a primer.

Characteristically, the control elements maintain an average frequency of one plasmid replication through ‘per plasmid copy per cell cycle’ and to ‘sense’ and correct deviations from this average.

The plasmid replication and its control is based on the results of analyses performed using pure cultures under steady-state growth conditions. Thus setting significant parameters required to understand the maintenance of these genetic elements in mixed populations and under environmental conditions.

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 per cell) from which several very high copy number cloning vectors (~1000 copies per cell) have been derived.

Plasmids can be incompatible if they share the same replication control mechanism. Under these circumstances, plasmids contribute to the total copy number and are regulated together. They are not recognized as distinct plasmids.

One of the plasmids may be out-copied by the other and lost during cell division, the cell is ‘cured’ of the plasmid. This is particularly likely with low copy number plasmids. Plasmids can also be incompatible due to shared partitioning systems.
12.9 KEY WORDS

- **Plasmid**: A plasmid is a small DNA molecule within a cell that is physically separated from chromosomal DNA and can replicate independently.
- **Plasmid size**: Plasmids range in size from 1 kbp (kilo base pair) to 1000 (kilo base pair) mega-plasmids that are many hundred base pairs in size.
- **Plasmid copy numbers**: Copy number is the significant feature and refers to the average or expected number of copies per host cell. Plasmids have either low, medium or high copy number.
- ‘F’ plasmid: It has an ability to integrate into the bacterial chromosome.
- **Incompatibility**: Plasmids can be incompatible if they share the same replication control mechanism, under these circumstances, plasmids contribute to the total copy number and are regulated together.

12.10 SELF ASSESSMENT QUESTIONS AND EXERCISES

**Short Answer Questions**
1. What is plasmid DNA?
2. What is plasmid copy number?
3. Explain the significance of purification of plasmid DNA.
4. How the transfer of plasmid DNA is done?
5. What is replication of plasmid?
6. How the copy number is controlled?
7. What is plasmid incompatibility?

**Long Answer Questions**
1. Briefly discuss about the significance and function plasmid DNA.
2. Explain the detection and purification process of plasmid DNA giving appropriate examples.
3. Discuss the transfer mechanism of plasmid DNA giving appropriate examples.
4. What is replication of plasmid? Explain giving significance and examples.
5. Briefly discuss the concept control of copy number.
6. Explain plasmid curing and incompatibility process.
12.11 FURTHER READINGS


NOTES

Detection, Purification and Transfer of Plasmid DNA, and Plasmid Replication
UNIT 13 TRANSPOSABLE ELEMENTS

13.0 INTRODUCTION

A Transposable Element (TE) or Transposon or Jumping Gene is a DNA sequence that can change its position within a genome, sometimes creating or reversing mutations and altering the cell’s genetic identity and genome size. Transposition often results in duplication of the same genetic material. Transposable Elements or TEs make up a large fraction of the genome and are responsible for much of the mass of DNA in a Eukaryotic cell. Although TEs are selfish genetic elements, many are significant in genome function and evolution. Transposons are also very useful to researchers as a means to alter DNA inside a living organism.

There are at least two classes of TEs, namely Class I TEs or Retrotransposons generally function via Reverse Transcription, while Class II TEs or DNA Transposons encode the Protein Transposase, which they require for insertion and excision, and some of these TEs also encode other Proteins. TEs, both active and inactive, occupy approximately half the human genome and a substantially greater fraction of some plant genomes. These movable elements are ubiquitous in the biosphere, and are highly successful in propagating themselves. Some TEs are also viruses, for instance, some retroviruses can integrate into a host genome to form endogenous retroviruses. Certainly, some viruses may be derived from natural transposable elements and vice versa. Since viruses move between individuals, at least some transposable elements can move between genomes (between individuals) as well as within an individual’s genome.
Transposition is related to replication, recombination and repair. The process of moving from one place to another involves a type of recombination, insertions of transposable elements can cause mutations, and some transpositions are replicative, generating a new copy while leaving the old copy intact. However, this ability to move is a unique property of transposable elements.

Bacteriophage Mu, also known as mu phage or mu bacteriophage, is a mulikevirus (the first of its kind to be identified) of the family Myoviridae which has been shown to cause genetic transposition. It is of specific importance as its discovery in *Escherichia coli* by Larry Taylor was among the first observations of insertions in a genome. Transposons are mobile genetic elements which can move anywhere in a genome by two mechanisms ‘copy and paste (Retrotransposons)’ and ‘cut and paste (DNA Transposons)’ and code for many necessary enzymes. Integrons are genetic elements able to acquire and rearrange Open Reading Frames (ORFs) embedded in gene cassette units and convert them to functional genes by ensuring their correct expression.

In this unit, you will study about the transposable elements, types of transposable elements, genetic organization and mechanism of transposition of Tn5, Tn3 and related transposons, bacteriophage Mu, Tn7 and IS911, integrons and retrotransposons.

13.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand what transposable elements are
- Explain the types of transposable elements
- Discuss the structure, genetic organization and mechanism of transposition of Tn5, Tn3 and related transposons
- Analyse the bacteriophage Mu, Tn7 and IS911
- Define integrons and retrotransposons

13.2 TRANSPOSABLE ELEMENTS

A Transposable Element (TE) or Transposon or Jumping Gene is a DNA sequence that can change its position within a genome, sometimes creating or reversing mutations and altering the cell’s genetic identity and genome size. Transposition often results in duplication of the same genetic material. Transposable Elements or TEs make up a large fraction of the genome and are responsible for much of the mass of DNA in a Eukaryotic cell. Although TEs are selfish genetic elements, many are important in genome function and evolution. Transposons are also very useful to researchers as a means to alter DNA inside a living organism.
Transposable Elements

There are at least two classes of TEs, namely Class I TEs or Retrotransposons generally function via Reverse Transcription, while Class II TEs or DNA Transposons encode the Protein Transposase, which they require for insertion and excision, and some of these TEs also encode other Proteins. Figure 13.1 illustrates a bacterial DNA transposon.

Fig. 13.1 A Bacterial DNA Transposon

TEs, both active and inactive, occupy approximately half the human genome and a substantially greater fraction of some plant genomes. These movable elements are ubiquitous in the biosphere, and are highly successful in propagating themselves. Some TEs are also viruses, for instance, some retroviruses can integrate into a host genome to form endogenous retroviruses. Certainly, some viruses may be derived from natural transposable elements and vice versa. Since viruses move between individuals, at least some transposable elements can move between genomes (between individuals) as well as within an individual’s genome.

Transposition is related to replication, recombination and repair. The process of moving from one place to another involves a type of recombination, insertions of transposable elements can cause mutations, and some transpositions are replicative, generating a new copy while leaving the old copy intact. However, this ability to move is a unique property of transposable elements.

Properties and Effects of Transposable Elements or TEs

The defining property of transposable elements is their mobility, i.e., they are genetic elements that can move from one position to another in the genome. In addition to the common property of mobility, transposable elements show considerable diversity. Some move by DNA intermediates, and others move by RNA intermediates. Much of the mechanism of transposition is distinctive for these two classes, but all transposable elements effectively insert at staggered breaks in chromosomes. Some transposable elements move in a replicative manner, whereas others are non-replicative, i.e., they move without making a copy of themselves.

Transposable elements are major forces in the evolution and rearrangement of genomes (Refer Figure 13.2). Some transposition events inactivate genes, since the coding potential or expression of a gene is disrupted by insertion of the
A classic example is the ‘r’ allele (Rugosus) of the gene encoding a starch branching enzyme in peas is nonfunctional due to the insertion of a transposable element. This allele causes the wrinkled pea phenotype in homozygotes originally studied by Mendel. In other cases, transposition can activate nearby genes by bringing an enhancer of transcription (within the transposable element) adequately close to a gene to stimulate its expression. If the target gene is not usually expressed in a certain cell type, then this activation can lead to pathology, such as activation of a proto-oncogene causing a cell to become cancerous. In other cases, no obvious phenotype results from the transposition. A particular type of transposable element can activate, inactivate or have no effect on nearby genes, depending on exactly where it inserts, its orientation and other factors.

In Figure 13.2, the transposable element is shown as rectangle and the Target Gene (X) is composed of multiple exons. The angled arrow indicates the start site for transcription.

Transposable elements can cause deletions or inversions of DNA. When transposition generates two copies of the same sequence in the same orientation, recombination can delete the DNA between them. If the two copies are in the opposite orientations, recombination will invert the DNA between them.

As part of the mechanism of transposition, additional DNA sequences can be mobilized. DNA located between two copies of a transposable element can be moved together with them when they move. In this manner, transposition can move DNA sequences that are not normally part of a transposable element to new locations. Certainly, the ‘host’ sequences can be acquired by viruses and propagated by infection of other individuals. This may be a natural means for evolving new strains of viruses. One of the most striking examples is the acquisition and modification of a proto-oncogene, such as cellular c-src, by a retrovirus to generate...
a modified, transforming form of the gene, called v-src. These and related observations provided insights into the progression of events that turn a normal cell into a cancerous one. They also point to the continual acquisition and possibly deletion of information from host genomes as a natural part of the evolution of viruses.

13.2.1 Types of Transposable Elements

Transposable Elements or TEs represent one of several types of mobile genetic elements. TEs are assigned to one of two classes according to their mechanism of transposition, which can be described as either copy and paste ‘Class I TEs’ or cut and paste ‘Class II TEs’.

**Retrotransposon**

Class I TEs are copied in two stages. First, they are transcribed from DNA to RNA, and the RNA produced is then reverse transcribed to DNA. This copied DNA is then inserted back into the genome at a new position. The reverse transcription step is catalyzed by a reverse transcriptase, which is often encoded by the TE itself. The characteristics of retrotransposons are similar to retroviruses, such as HIV (Human Immunodeficiency Virus).

Retrotransposons are commonly grouped into three main orders as follows:

- **Retrotransposons, with Long Terminal Repeats (LTRs)**, which encode reverse transcriptase, similar to retroviruses.
- **Retroposons, Long Interspersed Nuclear Elements (LINEs, LINE-1s, or L1s)**, which encode reverse transcriptase but lack LTRs, and are transcribed by RNA Polymerase II.
- **Short Interspersed Nuclear Elements (SINEs)** do not encode reverse transcriptase and are transcribed by RNA Polymerase III.

Retroviruses can also be considered TEs. For example, after conversion of Retroviral RNA into DNA inside a host cell, the newly produced Retroviral DNA is integrated into the genome of the host cell. These integrated DNAs are termed proviruses. The provirus is a specialized form of Eukaryotic Retrotransposon, which can produce RNA intermediates that may leave the host cell and infect other cells. The transposition cycle of retroviruses has similarities to that of prokaryotic TEs, suggesting a distant relationship between the two.

**DNA Transposons**

The ‘Cut and Paste’ transposition mechanism of Class II TEs does not involve an RNA intermediate. The transpositions are catalyzed by several transposase enzymes. Some transposases non-specifically bind to any target site in DNA, whereas others bind to specific target sequences. The transposase makes a staggered cut at the target site producing sticky ends, cuts out the DNA transposon and ligates it into the target site. A DNA polymerase fills in the resulting gaps from
the sticky ends and DNA ligase closes the Sugar-Phosphate backbone. This results in target site duplication and the insertion sites of DNA transposons may be identified by short direct repeats (a staggered cut in the target DNA filled by DNA polymerase) followed by inverted repeats (which are important for the TE excision by transposase).

The ‘Cut and Paste TEs’ may be duplicated if their transposition takes place during S phase of the cell cycle, when a donor site has already been replicated but a target site has not yet been replicated. Such duplications at the target site can result in gene duplication, which plays an important role in genomic evolution.

Not all DNA transposons transpose through the cut and paste mechanism. In some situations, a replicative transposition is observed in which a transposon replicates itself to a new target site, for example Helitron.

Class II TEs comprise less than 2% of the human genome, making the rest Class I.

Figure 13.3 illustrates the structure of DNA transposons and mechanism of transposition. Figure 13.3 (A) illustrates the structure of DNA transposons (Mariner type), the two Tandem Inverted Repeats (TIR) flank the transposase gene. Two short Tandem Site Duplications (TSD) are present on both sides of the insert, while the Figure 13.3 (B) illustrates the mechanism of transposition in which two transposases recognize and bind to TIR sequences, join together and promote DNA double-strand cleavage. The DNA-transposase complex then inserts its DNA cargo at specific DNA motifs elsewhere in the genome, creating short TSDs upon integration.

![Fig. 13.3](A) Structure of DNA Transposons (Mariner Type), (B) Mechanism of Transposition
Transposable Elements

NOTES

Autonomous and Non-Autonomous

Transposition can be classified as either ‘Autonomous’ or ‘Non-Autonomous’ in both Class I and Class II TEs. Autonomous TEs can move by themselves, whereas non-autonomous TEs require the presence of another TE to move. This is often because dependent TEs lack transposase for Class II or reverse transcriptase for Class I.

Activator (Ac) element is an example of an autonomous TE, and Dissociation (Ds) elements is an example of a non-autonomous TE. Without Ac, Ds is not able to transpose.

Instances of TEs

• The first TEs were discovered in maize (Zea mays) by Barbara McClintock in 1948, for which she was later awarded a Nobel Prize. She noticed chromosomal insertions, deletions, and translocations caused by these elements. These changes in the genome could, for example, lead to a change in the color of corn kernels. About 85% of the maize genome consists of TEs. The Ac/Ds system described by McClintock are Class II TEs. Transposition of Ac in Tobacco has been demonstrated by B. Baker.

• In the pond microorganisms, Oxytricha, TEs play a critical role that when removed, the organism fails to develop.

• One family of TEs in the fruit fly Drosophila melanogaster are called P elements. They seem to have first appeared in the species only in the middle of the twentieth century; within the last 50 years, they spread through every population of the species.

• Transposons in bacteria usually carry an additional gene for functions other than transposition, often for antibiotic resistance. In bacteria, transposons can jump from chromosomal DNA to plasmid DNA and back, allowing for the transfer and permanent addition of genes such as those encoding antibiotic resistance. Bacterial transposons of this type belong to the ‘Tn’ family. When the transposable elements lack additional genes, they are known as insertion sequences.

• The most common transposable element in humans is the ‘Alu’ sequence. It is approximately 300 bases long and can be found between 300,000 and one million times in the human genome. Alu alone is estimated to make up 15–17% of the human genome.

• Mariner like elements are another prominent class of transposons found in multiple species, including humans. The Mariner transposon was first discovered by Jacobson and Hartl in Drosophila melanogaster. This Class II TE is known for its uncanny ability to be transmitted horizontally in many species. There are an estimated 14,000 copies of Mariner in the human genome comprising 2.6 million base pairs. The first Mariner element transposons outside of animals were found in Trichomonas vaginalis.
Transposable Elements

• Mu phage transposition is the best-known example of replicative transposition.

• Yeast, *Saccharomyces cerevisiae*, genomes contain five distinct retrotransposon families, namely Ty1, Ty2, Ty3, Ty4 and Ty5.

• A Helitron is a TE found in Eukaryotes that is thought to replicate by a rolling-circle mechanism.

• In human embryos, two types of transposons combined to form noncoding RNA that catalyzes the development of stem cells. During the early stages of a fetus’s growth, the embryo’s inner cell mass expands as these stem cells enumerate. The increase of this type of cells is crucial, since stem cells later change form and give rise to all the cells in the body.

• In peppered moths, a transposon in a gene called cortex caused the moths’ wings to turn completely black. This change in coloration helped moths to blend in with ash and soot-covered areas during the Industrial Revolution.

Genetic Diseases and TEs

TEs are mutagens and their movements are often the causes of genetic diseases. They can damage the genome of their host cell in different ways, such as:

• A transposon or a retrotransposon that inserts itself into a functional gene will most likely disable that gene.

• After a DNA transposon leaves a gene, the resulting gap will probably not be repaired correctly.

• Multiple copies of the same sequence, such as Alu sequences, can hinder precise chromosomal pairing during mitosis and meiosis, resulting in unequal crossovers, one of the main reasons for chromosome duplication.

Diseases often caused by TEs include Hemophilia A and B, severe combined immunodeficiency, Porphyria, predisposition to Cancer, and Duchenne muscular dystrophy. LINE1 (L1) TEs that reside on the human Factor VIII have been shown to cause Haemophilia and insertion of L1 into the APC gene causes Colon Cancer, confirming that TEs play an important role in disease development. Transposable element dysregulation can cause neuronal death in Alzheimer’s disease and similar tauopathies. Additionally, many TEs contain promoters which drive transcription of their own transposase. These promoters can cause aberrant expression of linked genes, causing disease or mutant phenotypes.

Rate of Transposition, Induction and Defense

One study estimated the rate of transposition of a particular retrotransposon, the Ty1 element in *Saccharomyces cerevisiae*. Using several assumptions, the rate of successful transposition event per single Ty1 element came out to be about once every few months to once every few years. Some TEs contain heat-shock like promoters and their rate of transposition increases if the cell is subjected to stress.
thus increasing the mutation rate under these conditions, which might be beneficial to the cell.

Cells defend against the proliferation of TEs in a number of ways. These include piRNAs and siRNAs, which silence TEs after they have been transcribed.

If organisms are mostly composed of TEs, one might assume that disease caused by misplaced TEs is very common, but in most cases TEs are silenced through epigenetic mechanisms like DNA methylation, chromatin remodeling and piRNA, such that little to no phenotypic effects nor movements of TEs occur as in some wild-type plant TEs. Certain mutated plants have been found to have defects in methylation-related enzymes, Methyl Transferase, which cause the transcription of TEs, thus affecting the phenotype.

One hypothesis suggests that only approximately 100 LINE1 related sequences are active, despite their sequences making up 17% of the human genome. In human cells, silencing of LINE1 sequences is triggered by an RNA interference (RNAi) mechanism. Surprisingly, the RNAi sequences are derived from the 5’ UnTranslated Region (UTR) of the LINE1, a long terminal which repeats itself. Supposedly, the 5’ LINE1 UTR that codes for the sense promoter for LINE1 transcription also encodes the antisense promoter for the miRNA that becomes the substrate for siRNA production. Inhibition of the RNAi silencing mechanism in this region showed an increase in LINE1 transcription.

Evolution of TEs

TEs are found in almost all life forms, and the scientific community is still exploring their evolution and their effect on genome evolution. It is unclear whether TEs originated in the last universal common ancestor, arose independently multiple times, or arose once and then spread to other kingdoms by horizontal gene transfer. While some TEs confer benefits on their hosts, most are regarded as selfish DNA parasites. In this way, they are similar to viruses. Various viruses and TEs also share features in their genome structures and biochemical abilities, leading to speculation that they share a common ancestor.

Because excessive TE activity can damage exons, many organisms have acquired mechanisms to inhibit their activity. Bacteria may undergo high rates of gene deletion as part of a mechanism to remove TEs and viruses from their genomes, while Eukaryotic organisms typically use RNA interference to inhibit TE activity. Nevertheless, some TEs generate large families often associated with speciation events. Evolution often deactivates DNA transposons, leaving them as introns (Inactive Gene Sequences). In vertebrate animal cells, nearly all 100,000 + DNA Transposons Per Genome have genes that encode Inactive Transposase Polypeptides.

The first synthetic transposon designed for use in vertebrate (including human) cells, the ‘Sleeping Beauty Transposon System’, is a ‘Tc1/Mariner like Transposon’. Its dead or ‘fossil’ versions are spread widely in the Salmonid genome and a functional version was engineered by comparing these versions. The Sleeping
Beauty Transposon System has been used extensively as an insertional tag for identifying Cancer Genes. Human Tc1-like transposons are divided into ‘Hsmar1’ and ‘Hsmar2’ subfamilies. Although both types are inactive, one copy of Hsmar1 found in the SETMAR gene is under selection as it provides DNA-binding for the histone-modifying protein. Many other human genes are similarly derived from transposons. Hsmar2 has been reconstructed multiple times from the fossil sequences.

Interspersed repeats within genomes are created by transposition events accumulating over evolutionary time. Because interspersed repeats block gene conversion, they protect novel gene sequences from being overwritten by similar gene sequences and thereby facilitate the development of new genes. TEs may also have been co-opted by the vertebrate immune system as a means of producing antibody diversity. The V(D)J recombination system operates by a mechanism similar to that of some TEs.

TEs can contain many types of genes, including those conferring antibiotic resistance and ability to transpose to conjugative plasmids. Some TEs also contain integrons, genetic elements that can capture and express genes from other sources. These contain integrase, which can integrate gene cassettes. There are over 40 antibiotic resistance genes identified on cassettes, as well as virulence genes.

Transposons do not always excise their elements precisely, sometimes removing the adjacent base pairs, this phenomenon is called ‘exon shuffling’. Shuffling two unrelated exons can create a novel gene product or, more likely, an intron.

### 13.3 GENETIC ORGANIZATION AND MECHANISM OF TRANPOSITION

The phenomenon of moving genetic segments from one location to the other in a genome is known as transposition. There are two types of transposition, replicative and conservative 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.

In addition, the conservative transposition does not involve replication. Simply the elements move to a new site. When the target site is present within a gene both types of transposition takes place. The frequency of transposition varies among different elements.

The overall rate of transposition is $10^4$ - $10^5$ per element per generation. Different IS elements contain different number of bases.

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 recognizes and cleaves at either 5' or 3' of both ends of transponson, and catalyses at either 5' or 3' of both ends of transponson and catalyses a staggered cut at the target site. Depending on transponson, 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 transponson sequence.

**Types of Transposable Elements**

The maize Ac transposable element is only one of several types found in nature. Transposable elements can be divided into two major classes based on method of transposition:

**Retrotransposons (Class 1)**
- Use reverse transposase to make RNA intermediate for transposition.
- Encode an integrase and reverse transcriptase for transposition.
- Found in viruses.

**Transposons (Class 2)**

DNA fragments transpose directly from DNA segment to DNA segment as follows:
- Producing a DNA copy that transposes (Replicative Transposition).
- Cut/paste into a new locus (Conservative Transposition).

The features include:
- Encode a transposase for transposition.
- Can carry additional genes.
- Found in Eukaryotes and Prokaryotes.

Transposase and integrase proteins carry a ribonuclease-like catalytic domain and can use the same target site to catalyse both DNA cleavage and DNA strand transfer. However, transposases and integrases are only active when assembled into a synaptic complex (transpososome) on the DNA. The transpososome provides a scaffold to support the transposition reactions, changing its conformation to accommodate the different steps in transposition.

**Transposase Families**

Five families have been classified till now, although this number will most likely grow as new transposases are characterised. These families use distinct catalytic mechanisms for break/rejoining of DNA. For example, some transposases cut/transfer/paste original DNA, while others copy DNA into the target site.
These families include the following:

**DDE Transposases:** These transposases carry a triad of conserved amino acids, namely aspartate (D) and glutamate (E), which are required for the coordination of a metal ion required for catalysis, although the DDE chemistry can be integrated into the transposition cycle in differing ways. These employ a cut and paste mechanism of the original transposon. This family includes the maize Ac transposon, as well as the *Drosophila P* element, Bacteriophage Mu, Tn5 and Tn10, Mariner, IS10, and IS50.

**Tyrosine (Y) Transposases:** These also use a cut and paste mechanism of transposition, but employ a site-specific tyrosine residue. The transposon is excised from its original site (which is repaired); the transposon then forms a closed circle of DNA, which is integrated into a new site by a reversal of the original excision step. These transposons are usually found only in Bacteria, and include Kangaroo, Tn916, and DIRS1.

**Serine (S) Transposases:** These transposases use a cut and paste (cut out/paste in) mechanism of transposition involving a circular DNA intermediate, which is similar to that of tyrosine transposases, only they employ a site-specific serine residue. These transposons are usually found only in Bacteria, and include Tn5397 and IS607.

**Rolling Circle (RC) or Y2 Transposases:** These employ either a copy in mechanism, where they copy a single strand directly into the target site by DNA replication, so that the old (template) and new (copied) transposons both have one newly synthesized strand. These transposons usually employ host DNA replication enzymes. Examples include IS91 and Helitrons.

**Reverse Transcriptases/Endonucleases (RT/En):** Retrotransposons can vary in their mechanism of transposition. Some use the RT/En method, employing an endonuclease to nick the target site DNA, the nick serving as a primer for reverse transcription of an RNA copy by the reverse transcriptase enzyme. Examples include LINE-1 and TP-Retrotransposons.

### 13.3.1 Tn3 Transposon

The Tn3 transposon is a 4957 base pair mobile genetic element, found in prokaryotes. It encodes following three proteins:

- **β-Lactamase**, an enzyme that confers resistance to β-Lactam Antibiotics and is encoded by the gene *bla*.
- **Tn3 Transposase** encoded by Gene *tnpA*.
- **Tn3 Resolvase** encoded by Gene *tnpR*.

Initially discovered as a repressor of transposase, resolvase also plays a significant role in facilitating Tn3 replication (Sherratt 1989). The transposon is flanked by a pair of 38 bp inverted repeats.
Mechanism of Replication

The mechanism of replication involves the following steps.

**Step 1 – Replicative Integration**

This first stage is catalysed by transposase.

The plasmid containing the transposon (the donor plasmid) fuses with a host plasmid (the target plasmid). In the process, the transposon and a short section of host DNA are replicated. The end product is a ‘cointegrate’ plasmid containing two copies of the transposon.

Shapiro (1978) proposed the following mechanism for this process:

1. Four single-strand cleavages occur – one on each strand of the donor plasmid and one on each strand of the target plasmid.
2. The donor and target plasmids are ligated together, but there are two single-stranded regions, due to the positioning of the original cleavages.
3. DNA replication makes the single-stranded regions double stranded, using the existing strand as a template. It is in this stage that the transposon is replicated.

Figure 13.4 illustrates the replicative integration process.

**Step 2 – Resolution**

To separate the host and target molecules Tn3 resolvase executes site-specific recombination between the old and new copy of transposon at a specific site called ‘Res’, which is present in each copy of the transposon. Res is 114 bp long and it consists of 3 sub-sites, namely Site I, Site II and Site III. Each of these sites is of different lengths (28, 34 and 25 bp, respectively) and they are unevenly spaced with 22 bp separating Site I and Site II and only 5 bp between Site II and Site III. The sites consist of 6 bp inverted repeat motifs flanking a central sequence of variable length. These motifs act as binding sites for resolvase, so that each site binds a resolvase dimer but with varying affinity and probably a slightly different protein-DNA complex architecture. All three sub-sites are essential for recombination.
At recombination, two directly repeated res sites with resolvase dimers bound to each sub-site, come together to form a large complex structure called the Synaptosome. Resolvase bound to Site II and Site III initiates the assembly of this complex. In this structure, exact architecture of which is still unclear, two Res sites are intertwined in such a way as to juxtapose two copies of Site I, allowing resolvase dimers bound to each site to form a tetramer. Again, it is the interaction between the resolvase dimers bound at accessory sites (Site II and Site III) and resolvase at Site I that causes the two dimers to synapse and form a tetramer. After the tetramer is formed it becomes activated and the top and bottom DNA strands are simultaneously cleaved in the middle of the Site I with a 2 bp overhang. The strand exchange ensues by as yet unknown mechanism with a resulting net rotation of 180°.

The strand exchange is then followed by the religation (Stark et al., 1992). Recombination between two directly repeated Res sites separates, or resolves, the ‘cointegrate’ into two original molecules, each one now containing a copy of the ‘Tn3 transposon’.

After resolution these two molecules remain linked as a simple two-node catenane which can be easily separated in vivo by a Type II topoisomerase (Grindley 2002). Wild type resolvase system absolutely requires a supercoiled substrate and that the recombination sites are oriented in a direct repeat on the same DNA molecule. However, a number of ‘deregulated’ or ‘hyperactive’ mutants that have lost the requirement for the accessory sites have been isolated. These mutants are capable of catalysing recombination between two copies of Site I only, which basically reduces the recombination site size from 114 bp to only 28 bp. Furthermore, these mutants have no supercoiling or connectivity requirements (Arnold et al., 1999) and have been shown to work in mammalian cells. Hyperactive resolvase mutants have so far proven useful in creating resolvases with altered sequence specificity but also in structural work.

The entire resolvase recombination reaction can be reproduced in vitro, requiring only resolvase, a substrate DNA and multivalent cations, using either wild type protein or hyperactive mutants. Hyperactive resolvase mutants, if further developed, could become an alternative to Cre and FLP, the most commonly used recombination systems in molecular biology to date.

13.3.2 Tn5 Transposon

Tn5 was one of the first transposons to be identified. As a result of Tn5’s early discovery and its simple macromolecular requirements for transposition, the Tn5 system has been a very productive tool for studying the molecular mechanism of DNA transposition. These studies are of broad value because they offer insights into DNA transposition in general, because DNA transposition is a useful model with which to understand other types of Protein-DNA interactions, such as retroviral DNA integration and the DNA cleavage events involved in immunoglobin gene
Transposable Elements

formation, and because Tn5 derived tools are useful adjuncts in genetic experimentation.

Transposons are a class of genetic elements that can move from one site in a cell’s genome to another independently of the cell’s general recombination system. Little is known about the mechanism of transposition of compound transposons such as Tn5, but it is thought that a transposon-encoded protein (a transposase) must recognize the outer ends of the element and, together with host factors, catalyse the transfer of the internal DNA into a new site in a manner that may involve replication. It has previously been shown that the synthesis of an IS50R-encoded Protein (Protein 1) is an essential requirement for Tn5 transposition. Here we demonstrate that a structure containing only the outer 186 base pairs (bp) of both inverted repeats is capable of being efficiently complemented to transpose in *Escherichia coli*, provided IS50R is located close by on the same replicon. In addition, Bal31-generated deletions indicate that 16-18 bp of the outer end of IS50L are required for transposition. This 16-18 bp sequence contains the 8-9 bp small inverted repeat present at each end of IS50 plus a 9 bp sequence which is homologous to an interrelated sequence present in four copies in the chromosomal origin of replication in a variety of Gram-Negative Bacteria. This sequence organization suggests that the ends of Tn5 may function to provide a recognition site for the Tn5 transposase adjacent to a sequence recognized by the host replication system.

The bacterial transposon Tn5 encodes two proteins, the transposase and a related protein, the transposition inhibitor, whose relative abundance determines, in part, the frequency of Tn5 transposition. The synthesis of these proteins is programmed by a complex set of genetic regulatory elements. The host DNA methylation function, dam, inhibits transposase promoter recognition and indirectly enhances the transposition inhibitor promoter. The inhibitor lacks the N-terminal 55 amino acids of the transposase, suggesting that this sequence plays a key role in the transposition process. An intact N-terminal sequence is required for the transposase’s recognition of the 19 bp end DNA sequences. This is the first critical step in the transposition process. Transposase-end DNA interaction is itself regulated by an intricate series of reactions involving several host proteins, such as DnaA, Dam, and Fis. The transposase is a unique protein in that it acts primarily in cis and inhibits its own activity in trans. Models to explain these properties are described. Finally circumstantial evidence suggests that transposition occurs preferentially from newly replicated DNA that has yet to be partitioned to progeny cells.

13.3.3 Bacteriophage Mu

Transposable phage Mu has played a historic role in the development of the mobile DNA element field.

Bacteriophage Mu, also known as mu phage or mu bacteriophage, is a mulikevirus (the first of its kind to be identified) of the family Myoviridae which
Transposable Elements

13.3.4 Insertion Sequence–Tn7 and IS911

Insertion element, also known as an IS, an Insertion Sequence element, or an IS element, is a short DNA sequence that acts as a simple Transposable Element (TE). Insertion sequences have two major characteristics: they are small relative to other transposable elements, generally around 700 to 2500 bp in length, and only code for proteins implicated in the transposition activity. They are thus different from other transposons, which also carry accessory genes, such as antibiotic resistance genes. These proteins are usually the transposase which catalyses the enzymatic reaction allowing the IS to move, and also one regulatory protein which either stimulates or inhibits the transposition activity. The coding region in an insertion
sequence is usually flanked by inverted repeats. For example, the well-known IS911 (1250 bp) is flanked by two 36 bp inverted repeat extremities and the coding region has two genes partially overlapping orfA and orfAB, coding the transposase (OrfAB) and a regulatory protein (OrfA). A particular insertion sequence may be named according to the form ISn, where n is a number, for example IS1, IS2, IS3, IS10, IS50, IS911, IS26, etc. Although insertion sequences are usually considered in the context of Prokaryotic genomes, but certain Eukaryotic DNA sequences belonging to the family of Tc1/Mariner transposable elements may be considered to be the insertion sequences.

In addition to occurring autonomously, insertion sequences may also occur as parts of composite transposons, in a composite transposon, two insertion sequences flank one or more accessory genes, such as an antibiotic resistance gene, such as Tn10 and Tn5. Nevertheless, there exist another sort of transposons, called unit transposons that do not carry insertion sequences at their extremities, such as Tn7.

A complex transposon does not rely on flanking insertion sequences for resolvase. The resolvase is part of the “Tn” genome and cuts at flanking inverted repeats.

Transposition frequency of IS elements is dependent of multiple parameters, including culture growth phase, medium composition, oxygen tension, growth scale, and structural conformation of target sites, for example curvature, presence of certain motifs, DNA composition.

13.3.5 Integrons and Retrotransposons
Integrons are genetic mechanisms that allow bacteria to adapt and evolve rapidly through the stockpiling and expression of new genes. These genes are embedded in a specific genetic structure called gene cassette (a term that is lately changing to integron cassette) that generally carries one promoterless ORF together with a recombination site (attC). Integron cassettes are incorporated to the attI site of the integron platform by site-specific recombination reactions mediated by the integrase.

Integrons were initially discovered on conjugative plasmids through their role in antibiotic resistance. Indeed, these mobile integrons, as they are now known, can carry a variety of cassettes containing genes that are almost exclusively related to antibiotic resistance. Further studies have come to the conclusion that integrons are chromosomal elements, and that their mobilisation onto plasmids has been fostered by transposons and selected by the intensive use of antibiotics. The function of the majority of cassettes found in chromosomal integrons remains unknown.

An integron is minimally composed of:

Gene Encoding for a Site-Specific Recombinase: intI, belonging to the integrase family.
**Proximal Recombination Site:** attI, which is recognized by the integrase and at which gene cassettes may be inserted.

**Promoter:** Pc, which directs transcription of cassette-encoded genes.

**Gene Cassettes:** Additionally, an integron will usually contain one or more gene cassettes that have been incorporated into it. The gene cassettes may encode genes for antibiotic resistance, although most genes in integrons are uncharacterized. An attC sequence (also called 59-be) is a repeat that flanks cassettes and enables cassettes to be integrated at the attI site, excised and undergo horizontal gene transfer.

**Occurrence:** Integrons may be found as part of mobile genetic elements, such as plasmids and transposons. Integrons can also be found in chromosomes.

**Retrotransposons**

Retrotransposons, also called Class I transposable elements or transposons via RNA intermediates, are genetic elements that can amplify themselves in a genome and are ubiquitous components of the DNA of many Eukaryotic organisms. These DNA sequences use a ‘copy and paste’ mechanism, whereby they are first transcribed into RNA, then converted back into identical DNA sequences using reverse transcription, and these sequences are then inserted into the genome at target sites.

Retrotransposons form one of the two subclasses of transposons, where the others are DNA transposons, which does not involve an RNA intermediate.

Retrotransposons are particularly abundant in plants, where they are often a principal component of nuclear DNA. In maize, 49–78% of the genome is made up of retrotransposons. In wheat, about 90% of the genome consists of repeated sequences and 68% of transposable elements. In mammals, almost half the genome (45% to 48%) is transposons or remnants of transposons. Around 42% of the human genome is made up of retrotransposons, while DNA transposons account for about 2–3%.

**Types:** Retrotransposons, also known as Class I transposable elements, consist of two subclasses, the Long Terminal Repeat (LTR-retrotransposons) and the non-LTR retrotransposons. Classification into these subclasses is based on the phylogeny of the reverse transcriptase, which goes in line with structural differences, such as presence/absence of long terminal repeats as well as number and types of open reading frames, encoding domains and target site duplication lengths. LTR Retrotransposons have direct LTRs that range from ~100 bp to over 5 kb in size. In plant genomes, LTR retrotransposons are the major repetitive sequence class, such as able to constitute more than 75% of the maize genome. The Non-LTR Retrotransposons consist of two sub-types, long interspersed elements (LINEs) and short interspersed elements (SINEs). They can also be found in high copy numbers, as shown in the plant species. Non-Long Terminal Repeat (LTR) retroposons are widespread in Eukaryotic genomes. LINEs possess...
two ORFs, which encode all the functions needed for retrotransposition. These functions include reverse transcriptase and endonuclease activities, in addition to a nucleic acid-binding property needed to form a ribonucleoprotein particle. SINEs, on the other hand, co-opt the LINE machinery and function as nonautonomous retroelements. While historically viewed as ‘junk DNA’.

Check Your Progress

1. Explain the term transposable element.
2. What are the classes of transposable element?
3. Explain the defining property of transposable elements.
4. What is transposition? Explain giving example.
5. What is Tn3 transposon? Which three proteins it encodes?
6. Explain insertion sequence element giving examples.

13.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. A Transposable Element (TE) or Transposon or Jumping Gene is a DNA sequence that can change its position within a genome, sometimes creating or reversing mutations and altering the cell’s genetic identity and genome size. Transposition often results in duplication of the same genetic material. Transposable Elements or TEs make up a large fraction of the genome and are responsible for much of the mass of DNA in a Eukaryotic cell. Although TEs are selfish genetic elements, many are important in genome function and evolution. Transposons are also very useful to researchers as a means to alter DNA inside a living organism.

2. There are at least two classes of TEs, namely Class I TEs or Retrotransposons generally function via Reverse Transcription, while Class II TEs or DNA Transposons encode the Protein Transposase, which they require for insertion and excision, and some of these TEs also encode other Proteins. Transposable Elements or TEs represent one of several types of mobile genetic elements. TEs are assigned to one of two classes according to their mechanism of transposition, which can be described as either copy and paste ‘Class I TEs’ or cut and paste ‘Class II TEs’.

3. The defining property of transposable elements is their mobility, i.e., they are genetic elements that can move from one position to another in the genome. In addition to the common property of mobility, transposable elements show considerable diversity. Some move by DNA intermediates, and others move by RNA intermediates.
4. The phenomenon of moving genetic segments from one location to the other in a genome is known as transposition. There are two types of transposition, replicative and conservative 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. In addition, the conservative transposition does not involve replication. Simply the elements move to a new site. When the target site is present within a gene both types of transposition takes place. The frequency of transposition varies among different elements. The overall rate of transposition is $10^5 - 10^6$ per element per generation. Different IS elements contain different number of bases. 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.

5. The Tn3 transposon is a 4957 base pair mobile genetic element, found in prokaryotes. It encodes following three proteins:
   - $\beta$-Lactamase, an enzyme that confers resistance to $\beta$-Lactam Antibiotics and is encoded by the gene Bla.
   - Tn3 Transposase encoded by Gene tnpA.
   - Tn3 Resolvase encoded by Gene tnpR.

6. Insertion element, also known as an IS, an Insertion Sequence element, or an IS element, is a short DNA sequence that acts as a simple Transposable Element (TE). Insertion sequences have two major characteristics: they are small relative to other transposable elements, generally around 700 to 2500 bp in length, and only code for proteins implicated in the transposition activity. A particular insertion sequence may be named according to the form ISn, where n is a number, for example IS1, IS2, IS3, IS10, IS50, IS911, IS26, etc. There exist another sort of transposons, called unit transposons that do not carry insertion sequences at their extremities, such as Tn7.

13.5 SUMMARY

- A Transposable Element (TE) or Transposon or Jumping Gene is a DNA sequence that can change its position within a genome, sometimes creating or reversing mutations and altering the cell’s genetic identity and genome size.
- Transposition often results in duplication of the same genetic material. Transposable Elements or TEs make up a large fraction of the genome and are responsible for much of the mass of DNA in a Eukaryotic cell.
- Although TEs are selfish genetic elements, many are important in genome function and evolution. Transposons are also very useful to researchers as a means to alter DNA inside a living organism.
Transposable Elements

- There are at least two classes of TEs, namely Class I TEs or Retrotransposons generally function via Reverse Transcription, while Class II TEs or DNA Transposons encode the Protein Transposase, which they require for insertion and excision, and some of these TEs also encode other Proteins.

- Transposition is related to replication, recombination and repair. The process of moving from one place to another involves a type of recombination, insertions of transposable elements can cause mutations, and some transpositions are replicative, generating a new copy while leaving the old copy intact. However, this ability to move is a unique property of transposable elements.

- The defining property of transposable elements is their mobility, i.e., they are genetic elements that can move from one position to another in the genome.

- In addition to the common property of mobility, transposable elements show considerable diversity. Some move by DNA intermediates, and others move by RNA intermediates.

- Transposable elements can cause deletions or inversions of DNA. When transposition generates two copies of the same sequence in the same orientation, recombination can delete the DNA between them. If the two copies are in the opposite orientations, recombination will invert the DNA between them.

- Transposable Elements or TEs represent one of several types of mobile genetic elements. TEs are assigned to one of two classes according to their mechanism of transposition, which can be described as either copy and paste ‘Class I TEs’ or cut and paste ‘Class II TEs’.

- Transposition can be classified as either ‘Autonomous’ or ‘Non-Autonomous’ in both Class I and Class II TEs. Autonomous TEs can move by themselves, whereas non-autonomous TEs require the presence of another TE to move.

- The first synthetic transposon designed for use in vertebrate (including human) cells, the ‘Sleeping Beauty Transposon System’, is a ‘Tc1/Mariner like Transposon’. Its dead or ‘fossil’ versions are spread widely in the Salmoid genome and a functional version was engineered by comparing those versions.

- The Sleeping Beauty Transposon System has been used extensively as an insertional tag for identifying Cancer Genes.

- The phenomenon of moving genetic segments from one location to the other in a genome is known as transposition. There are two types of transposition, replicative and conservative transposition.
Transposable Elements

- 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.
- The conservative transposition does not involve replication. Simply the elements move to a new site. When the target site is present within a gene both types of transposition takes place. The frequency of transposition varies among different elements.
- The overall rate of transposition is \(10^7 - 10^9\) per element per generation. Different IS elements contain different number of bases.
- 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.
- DDE transposases carry a triad of conserved amino acids, namely aspartate (D) and glutamate (E), which are required for the coordination of a metal ion required for catalysis, although the DDE chemistry can be integrated into the transposition cycle in differing ways. This family includes the maize Ac transposon, as well as the _Drosophila_ P element, Bacteriophage Mu, Tn5 and Tn10, Mariner, IS10, and IS50.
- Tyrosine (Y) transposases use a cut and paste mechanism of transposition, but employ a site-specific tyrosine residue. The transposon is excised from its original site (which is repaired); the transposon then forms a closed circle of DNA, which is integrated into a new site by a reversal of the original excision step. These transposons are usually found only in Bacteria, and include Kangaroo, Tn916, and DIRS1.
- Serine (S) transposases use a cut and paste (cut out/paste in) mechanism of transposition involving a circular DNA intermediate, which is similar to that of tyrosine transposases, only they employ a site-specific serine residue. These transposons are usually found only in Bacteria, and include Tn5397 and IS607.
- Rolling Circle (RC) or Y2 transposases employ either a copy in mechanism, where they copy a single strand directly into the target site by DNA replication, so that the old (template) and new (copied) transposons both have one newly synthesized strand. These transposons usually employ host DNA replication enzymes. Examples include IS91 and Helitrons.
- The Tn3 transposon is a 4957 base pair mobile genetic element, found in prokaryotes.
- Tn5 was one of the first transposons to be identified. As a result of Tn5’s early discovery and its simple macromolecular requirements for transposition, the Tn5 system has been a very productive tool for studying the molecular mechanism of DNA transposition.
Bacteriophage Mu, also known as mu phage or mu bacteriophage, is a
mulikevirus (the first of its kind to be identified) of the family Myoviridae
which has been shown to cause genetic transposition.

Bacteriophage Mu represents a hybrid gene creature that exists as both a
transposable element and a bacteriophage, a virus that infects bacteria. Mu
integrates into its Escherichia coli host by transposition.

Insertion element, also known as an IS, an Insertion Sequence element, or
an IS element, is a short DNA sequence that acts as a simple Transposable
Element (TE). Insertion sequences have two major characteristics: they are
small relative to other transposable elements, generally around 700 to 2500
bp in length, and only code for proteins implicated in the transposition activity.

A particular insertion sequence may be named according to the form ISn,
where n is a number, for example IS1, IS2, IS3, IS10, IS50, IS911, IS26,
etc.

There exist another sort of transposons, called unit transposons that do not
carry insertion sequences at their extremities, such as Tn7.

Integrons are genetic mechanisms that allow bacteria to adapt and evolve
rapidly through the stockpiling and expression of new genes.

Retrotransposons, also called Class I transposable elements or transposons
via RNA intermediates, are genetic elements that can amplify themselves in
a genome and are ubiquitous components of the DNA of many Eukaryotic
organisms. These DNA sequences use a ‘copy and paste’ mechanism,
whereby they are first transcribed into RNA, then converted back into
identical DNA sequences using reverse transcription, and these sequences
are then inserted into the genome at target sites.

13.6 KEY WORDS

- **Transposable Element (TE):** Also termed as Transposon or Jumping
  Gene is a DNA sequence that can change its position within a genome,
sometimes creating or reversing mutations and altering the cell’s genetic
identity and genome size.
- **DDE transposases:** These transposases carry a triad of conserved amino
  acids, namely aspartate (D) and glutamate (E), which are required for the
  coordination of a metal ion required for catalysis, although the DDE chemistry
  can be integrated into the transposition cycle in differing ways.
- **Tyrosine (Y) transposases:** These also use a cut and paste mechanism of
  transposition, but employ a site-specific tyrosine residue.
- **Serine (S) transposases:** These transposases use a cut and paste (cut
  out/paste in) mechanism of transposition involving a circular DNA
  intermediate, which is similar to that of tyrosine transposases, only they
  employ a site-specific serine residue.
- **Tn3 transposon**: The Tn3 transposon is a 4957 base pair mobile genetic element, found in prokaryotes.

- **Bacteriophage Mu**: Also known as mu phage or mu bacteriophage, is a myrivirus (the first of its kind to be identified) of the family Myoviridae which has been shown to cause genetic transposition.

- **Integrons**: These are genetic mechanisms that allow bacteria to adapt and evolve rapidly through the stockpiling and expression of new genes.

### 13.7 Self Assessment Questions and Exercises

#### Short Answer Questions

1. What is transposable elements?
2. Explain the types of transposable elements.
3. What is genetic organization?
4. When the mechanism of transposition of Tn5 and Tn3 transposons are used?
5. Define Bacteriophage Mu.
6. Explain insertion sequence Tn7 and IS911.
7. Differentiate between integrons and retrotransposons.

#### Long Answer Questions

1. Briefly discuss about the significance and functions of transposable elements giving diagrams and examples.
2. Explain the types of transposable elements.
3. Discuss the structure, genetic organization and mechanism of transposition of Tn5 and Tn3 transposons.
4. Explain the significant features of Bacteriophage Mu.
5. Explain briefly insertion sequence with reference to Tn7 and IS911.
6. Discuss the significance and features of integrons and retrotransposons.

### 13.8 Further Readings


UNIT 14 EPIGENETICS

14.0 INTRODUCTION

Epigenetics is the study of heritable phenotype changes that do not involve alterations in the DNA sequence. Derived from the Greek prefix ‘epi-’ means ‘over, outside of, around’, i.e., in epigenetics implies features that are ‘on top of’ or ‘in addition to’ the traditional genetic basis for inheritance. Principally, the word ‘epigenetic’ literally means ‘in addition to changes in genetic sequence’. The term has evolved to include any process that alters gene activity without changing the DNA sequence, and leads to modifications that can be transmitted to daughter cells, although experiments show that some epigenetic changes can be reversed. Many types of epigenetic processes have been identified—they include methylation, acetylation, phosphorylation, ubiquitylation, and sumoylation. Epigenetic processes are natural and essential to many organism functions, but if they occur improperly, there can be major adverse health and behavioral effects.

Epigenetics most often denotes changes that affect gene activity and expression, but can also be used to describe any heritable phenotypic change. Such effects on cellular and physiological phenotypic traits may result from external or environmental factors, or be part of normal development. The standard definition of epigenetics requires these alterations to be heritable, in the progeny of either cells or organisms. The term also refers to the changes themselves, functionally relevant changes to the genome that do not involve a change in the nucleotide sequence. Examples of mechanisms that produce such changes are DNA methylation and histone modification, each of which alters how genes are expressed without altering the underlying DNA sequence. Gene expression can be controlled...
Epigenetics

Epigenetics

NOTES

through the action of repressor proteins that attach to silencer regions of the DNA. These epigenetic changes may last through cell divisions for the duration of the cell’s life, and may also last for multiple generations even though they do not involve changes in the underlying DNA sequence of the organism, as an alternative, non-genetic factors cause the organism’s genes to behave or ‘express themselves’ differently.

Epigenetic changes modify the activation of certain genes, but not the genetic code sequence of DNA. The microstructure (not code) of DNA itself or the associated chromatin proteins may be modified, causing activation or silencing. This mechanism enables differentiated cells in a multicellular organism to express only the genes that are necessary for their own activity. Epigenetic changes are preserved when cells divide. Most epigenetic changes only occur within the course of one individual organism’s lifetime, however, these epigenetic changes can be transmitted to the organism’s offspring through a process called transgenerational epigenetic inheritance. Moreover, if gene inactivation occurs in a sperm or egg cell that results in fertilization, this epigenetic modification may also be transferred to the next generation.

In this unit, you will study about the epigenetics, its definition, molecular basis, mechanisms, functions and epigenetics in bacteria.

14.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss the significance of epigenetics
- Explain the definition, molecular basis, mechanisms and functions of epigenetics
- Describe the epigenetics in bacteria

14.2 EPIGENETICS

Epigenetics is the study of heritable phenotype changes that do not involve alterations in the DNA sequence. Derived from the Greek prefix ‘epi-’ means ‘over, outside of, around’, i.e., in epigenetics implies features that are ‘on top of’ or ‘in addition to’ the traditional genetic basis for inheritance.

Principally, the word ‘epigenetic’ literally means ‘in addition to changes in genetic sequence’. The term has evolved to include any process that alters gene activity without changing the DNA sequence, and leads to modifications that can be transmitted to daughter cells, although experiments show that some epigenetic changes can be reversed. Many types of epigenetic processes have been
identified—they include methylation, acetylation, phosphorylation, ubiquitylation, and sumoylation. Epigenetic processes are natural and essential to many organism functions, but if they occur improperly, there can be major adverse health and behavioral effects.

Epigenetics most often denotes changes that affect gene activity and expression, but can also be used to describe any heritable phenotypic change. Such effects on cellular and physiological phenotypic traits may result from external or environmental factors, or be part of normal development. The standard definition of epigenetics requires these alterations to be heritable, in the progeny of either cells or organisms. The term also refers to the changes themselves, functionally relevant changes to the genome that do not involve a change in the nucleotide sequence. Examples of mechanisms that produce such changes are DNA methylation and histone modification, each of which alters how genes are expressed without altering the underlying DNA sequence. Gene expression can be controlled through the action of repressor proteins that attach to silencer regions of the DNA. These epigenetic changes may last through cell divisions for the duration of the cell’s life, and may also last for multiple generations even though they do not involve changes in the underlying DNA sequence of the organism, as an alternative, non-genetic factors cause the organism’s genes to behave or ‘express themselves’ differently.

One example of an epigenetic change in eukaryotic biology is the process of cellular differentiation. During morphogenesis, totipotent stem cells become the various pluripotent cell lines of the embryo, which in turn become fully differentiated cells. In other words, as a single fertilized egg cell – the zygote – continues to divide, the resulting daughter cells change into all the different cell types in an organism, including neurons, muscle cells, epithelium, endothelium of blood vessels, etc., by activating some genes while inhibiting the expression of others.

Historically, some phenomena not necessarily heritable have also been described as epigenetic. For example, the term epigenetic has been used to describe any modification of chromosomal regions, especially histone modifications, whether or not these changes are heritable or associated with a phenotype. The consensus definition now requires a trait to be heritable for it to be considered epigenetic. Figure 14.1 illustrates the epigenetic mechanisms.
14.2.1 Definitions of Epigenetics

The term epigenetics in its contemporary usage emerged in the 1990s, but for some years has been used with somewhat variable meanings. A consensus definition of the concept of epigenetic trait as a, ‘Stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence’ was formulated at a Cold Spring Harbor meeting in 2008, although alternate definitions that include non-heritable traits are still being used. The term epigenesis has a generic meaning of ‘extra growth’, and has been used in English since the 17th century.

Waddington’s Definition on Canalisation

From the generic meaning, and the associated adjective epigenetic, British embryologist C. H. Waddington coined the term ‘Epigenetics’ in 1942 as pertaining to epigenesis, in parallel to Valentin Haecker’s ‘phenogenetics’. Epigenesis in the context of the biology of that period referred to the differentiation of cells from their initial totipotent state during embryonic development.

When Waddington coined the term ‘Epigenetics’, the physical nature of genes and their role in heredity was not known. He used it instead as a conceptual model of how genetic components might interact with their surroundings to produce a phenotype, he used the phrase ‘epigenetic landscape’ as a metaphor for biological development. Waddington held that cell fates were established during development in a process he called canalisation much as a marble rolls down to the point of lowest local elevation. Waddington suggested visualising increasing irreversibility of cell type differentiation as ridges rising between the valleys where the marbles (analogous to cells) are travelling.
In recent times, Waddington’s notion of the epigenetic landscape has been rigorously formalized in the context of the systems dynamics state approach to the study of ‘Cell Fate’. Cell Fate determination is predicted to exhibit certain dynamics, such as attractor-convergence, where the attractor can be an equilibrium point, limit cycle or strange attractor or oscillatory.

Contemporary Definitions

Robin Holliday defined epigenetics as, ‘The study of the mechanisms of temporal and spatial control of gene activity during the development of complex organisms’. Thus, in its broadest sense, epigenetic can be used to describe anything other than DNA sequence that influences the development of an organism.

More recent usage of the word in biology follows stricter definitions. It is, as defined by Arthur Riggs and colleagues as, ‘The study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence’.

The term has also been used, however, to describe processes which have not been demonstrated to be heritable, such as some forms of histone modification, there are therefore attempts to redefine the term ‘Epigenetics’ in broader terms that would avoid the constraints of requiring heritability. For example, Adrian Bird defined epigenetics as, ‘The structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states’.

This definition would be inclusive of transient modifications associated with DNA repair or cell cycle phases as well as stable changes maintained across multiple cell generations, but exclude others, such as templating of membrane architecture and prions unless they impinge on chromosome function. Such redefinitions however are not universally accepted and are still subject to debate.

The NIH ‘Roadmap Epigenomics Project’, ongoing as of 2016, uses the following definition:

‘For purposes of this program, epigenetics refers to both heritable changes in gene activity and expression, in the progeny of cells or of individuals, and also stable, long term alterations in the transcriptional potential of a cell that are not necessarily heritable’.

In 2008, a consensus definition of the epigenetic trait, a ‘Stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence’, was made at a Cold Spring Harbor meeting.

The similarity of the word to ‘Genetics’ has generated many parallel usages. The term ‘Epigenome’ is a parallel to the word ‘Genome’, referring to the overall epigenetic state of a cell, and epigenomics refers to global analyses of epigenetic changes across the entire genome. The phrase ‘Genetic Code’ has also been adapted – the ‘Epigenetic Code’ has been used to describe the set of epigenetic features that create different phenotypes in different cells from the same underlying DNA sequence.
Additionally, the term ‘Epigenetic Code’ could represent the total state of the cell, with the position of each molecule accounted for in an epigenomic map, a diagrammatic representation of the gene expression, DNA methylation and histone modification status of a particular genomic region. More typically, the term is used in reference to systematic efforts to measure specific, relevant forms of epigenetic information such as the histone code or DNA methylation patterns.

**14.2.2 Molecular Basis of Epigenetic**

Epigenetic changes modify the activation of certain genes, but not the genetic code sequence of DNA. The microstructure (not code) of DNA itself or the associated chromatin proteins may be modified, causing activation or silencing. This mechanism enables differentiated cells in a multicellular organism to express only the genes that are necessary for their own activity. Epigenetic changes are preserved when cells divide. Most epigenetic changes only occur within the course of one individual organism’s lifetime; however, these epigenetic changes can be transmitted to the organism’s offspring through a process called transgenerational epigenetic inheritance. Moreover, if gene inactivation occurs in a sperm or egg cell that results in fertilization, this epigenetic modification may also be transferred to the next generation.

Specific ‘Epigenetic’ processes include Paramutation, Bookmarking, Imprinting, Gene Silencing, X Chromosome inactivation, Position effect, DNA Methylation reprogramming, Transvection, Maternal Effects, the progress of Carcinogenesis, effects of Teratogens, Regulation of Histone modifications and Heterochromatin, and technical limitations affecting Parthenogenesis and Cloning.

**DNA Damage**

DNA damage can also cause epigenetic changes. The DNA damage is very frequent, occurring on average about 60,000 times a day per cell of the human body. These damages are largely repaired, but at the site of a DNA repair, epigenetic changes can remain. In particular, a double strand break in DNA can initiate unprogrammed epigenetic gene silencing both by causing DNA methylation as well as by promoting silencing types of histone modifications.

In addition, the enzyme Parp1 (poly(ADP)-ribose polymerase) and its product poly(ADP)-ribose (PAR) accumulate at sites of DNA damage as part of a repair process. This accumulation, in turn, directs recruitment and activation of the chromatin remodelling protein ALC1 that can cause nucleosome remodelling. Nucleosome remodelling has been found to cause, for instance, epigenetic silencing of DNA repair gene MLH1. DNA damaging chemicals, such as Benzene, Hydroquinone, Styrene, Carbon Tetrachloride and Trichloroethylene, cause considerable Hypomethylation of DNA, some through the activation of oxidative stress pathways.
Foods are known to alter the epigenetics of rats on different diets. Some food components epigenetically increase the levels of DNA repair enzymes, such as MGMT and MLH1 and p53. Other food components can reduce DNA damage, such as Soy Isoflavones. In one study, markers for oxidative stress, such as modified nucleotides that can result from DNA damage, were decreased by a 3-week diet supplemented with Soy. A decrease in oxidative DNA damage was also observed after the consumption of Anthocyanin-Rich Bilberry, *Vaccinium myrtillus* L.

**Techniques to Study Epigenetics**

Epigenetic research uses a wide range of molecular biological techniques to further understanding of epigenetic phenomena, including chromatin immunoprecipitation, together with its large-scale variants ChIP-on-chip and ChIP-Seq, fluorescent in situ hybridization, methylation sensitive restriction enzymes, DNA adenine methyltransferase identification (DamID) and bisulfite sequencing. Furthermore, the use of bioinformatics methods has a role in computational epigenetics.

**14.2.3 Epigenetics Mechanisms**

Several types of epigenetic inheritance systems have a significant role as cell memory, however not all of these are universally accepted to be examples of epigenetics. Following are some universally accepted mechanisms of epigenetics.

**Covalent Modifications**

Covalent modifications of DNA, for example Cytosine Methylation and Hydroxymethylation or of Histone Proteins, such as Lysine Acetylation, Lysine and Arginine Methylation, Serine and Threonine Phosphorylation, and Lysine Ubiquitination and Sumoylation, play central roles in many types of epigenetic inheritances. Therefore, the word ‘Epigenetics’ is sometimes used as a synonym for these processes. However, this can be misleading. Chromatin remodelling is not always inherited, and not all epigenetic inheritance involves chromatin remodelling. DNA associates with histone proteins to form chromatin.

Because the phenotype of a cell or individual is affected by which of its genes are transcribed, heritable transcription states can give rise to epigenetic effects. There are several layers of regulation of gene expression. One way that genes are regulated is through the remodelling of chromatin. Chromatin is the complex of DNA and the histone proteins with which it associates. If the way that DNA is wrapped around the histones changes, gene expression can change as well. Chromatin remodelling is accomplished through the following two main mechanisms:

**Mechanism 1:** The first method is post translational modification of the amino acids that make up histone proteins. Histone proteins are made up of long chains of amino acids. If the amino acids that are in the chain are changed, the shape of the histone might be modified. DNA is not completely unwound during replication.
It is possible, then, that the modified histones may be carried into each copy of the DNA. Once there, these histones may act as templates, initiating the surrounding new histones to be shaped in the new manner. By altering the shape of the histones around them, these modified histones would ensure that a lineage-specific transcription program is maintained after cell division.

Mechanism 2: The second method is the addition of methyl groups to the DNA, mostly at CpG sites, to convert Cytosine (C) into 5-Methylcytosine (5-MeC). 5-Methylcytosine performs much like a regular Cytosine (C), pairing with a guanine in double-stranded DNA. However, some areas of the genome are methylated more heavily than others, and highly methylated areas tend to be less transcriptionally active than others. Methylation of Cytosine's can also persist from the germ line of one of the parents into the zygote, marking the chromosome as being inherited from one parent or the other (genetic imprinting).

Mechanisms of heritability of histone state are not well understood, however, much is known about the mechanism of heritability of DNA methylation state during cell division and differentiation. Heritability of methylation state is dependent on certain enzymes, such as DNMT1, which have a higher affinity for 5-Methylcytosine than for Cytosine (C). If this enzyme reaches a hemimethylated portion of DNA, where 5-Methylcytosine is in only one of the two DNA strands, the enzyme will methylate the other half.

Although histone modifications occur throughout the entire sequence, the unstructured N-termini of histones (called histone tails) are particularly modified. These modifications include Acetylation, Methylation, Ubiquitylation, Phosphorylation, Sumoylation, and Citrullination. Acetylation is the most highly studied of these modifications. For example, Acetylation of the K14 and K9 Lysines of the tail of histone H3 by Histone AcetylTransferase (HAT) can create a binding site for Chromatin-Modifying Enzymes or transcription machinery. This chromatin remodeler can then cause changes in the state of the chromatin. Certain transcription factors, such as Acetyl-CoA, are found in many enzymes that help activate transcription.
Epigenetics

including the SWI/SNF complex. It may be that acetylation acts in this and the previous way to aid in transcriptional activation.

The idea that modifications act as docking modules for related factors is borne out by histone methylation as well. Methylation of Lysine 9 of Histone H3 has long been associated with constitutively transcriptionally silent chromatin (constitutive heterochromatin). It has been determined that a chromodomain (a domain that specifically binds methyl-lysine) in the transcriptionally repressive protein HP1 recruits HP1 to K9 methylated regions. One example that seems to refute this biophysical model for methylation is that tri-methylation of histone H3 at lysine 4 is strongly associated with transcriptional activation. Tri-methylation in this case would introduce a fixed positive charge on the tail.

It has been shown that the histone lysine methyltransferase (KMT) is responsible for this methylation activity in the pattern of histones H3 and H4. This enzyme utilizes a catalytically active site called the SET domain. The SET domain is a 130-amino acid sequence involved in modulating gene activities. This domain has been demonstrated to bind to the histone tail and causes the methylation of the histone.

Differing histone modifications are likely to function in differing ways; acetylation at one position is likely to function differently from acetylation at another position. Also, multiple modifications may occur at the same time, and these modifications may work together to change the behavior of the nucleosome. The idea that multiple dynamic modifications regulate gene transcription in a systematic and reproducible way is called the histone code, although the idea that histone state can be read linearly as a digital information carrier has been largely debunked. One of the best-understood systems that orchestrates chromatin-based silencing is the SIR protein based silencing of the yeast hidden mating type loci HML and HMR.

DNA methylation frequently occurs in repeated sequences, and helps to suppress the expression and mobility of ‘transposable elements’. Because 5-methylcytosine can be spontaneously deaminated (replacing nitrogen by oxygen) to thymidine, CpG sites are frequently mutated and become rare in the genome, except at CpG islands where they remain unmethylated. Epigenetic changes of this type thus have the potential to direct increased frequencies of permanent genetic mutation. DNA methylation patterns are known to be established and modified in response to environmental factors by a complex interplay of at least three independent DNA methyltransferases, DNMT1, DNMT3A, and DNMT3B, the loss of any of which is lethal in mice. DNMT1 is the most abundant methyltransferase in somatic cells, localizes to replication foci, has a 10–40 fold preference for hemimethylated DNA and interacts with the Proliferating Cell Nuclear Antigen (PCNA).

By preferentially modifying hemimethylated DNA, DNMT1 transfers patterns of methylation to a newly synthesized strand after DNA replication, and
therefore is often referred to as the ‘maintenance’ methyltransferase. DNMT1 is essential for proper embryonic development, imprinting and X-inactivation. To emphasize the difference of this molecular mechanism of inheritance from the canonical Watson–Crick base-pairing mechanism of transmission of genetic information, the term ‘Epigenetic Templating’ was introduced. Furthermore, in addition to the maintenance and transmission of methylated DNA states, the same principle could work in the maintenance and transmission of histone modifications and even cytoplasmic (structural) heritable states.

Histones H3 and H4 can also be manipulated through demethylation using histone lysine demethylase (KDM). This recently identified enzyme has a catalytically active site called the Jumonji domain (JmjC). The demethylation occurs when JmjC utilizes multiple cofactors to hydroxylate the methyl group, thereby removing it. JmjC is capable of demethylating mono-, di-, and tri-methylated substrates.

Chromosomal regions can adopt stable and heritable alternative states resulting in bistable gene expression without changes to the DNA sequence. Epigenetic control is often associated with alternative covalent modifications of histones. The stability and heritability of states of larger chromosomal regions are suggested to involve positive feedback where modified nucleosomes recruit enzymes that similarly modify nearby nucleosomes. A simplified stochastic model for this type of epigenetics is found here.

It has been suggested that chromatin-based transcriptional regulation could be mediated by the effect of small RNAs. Small interfering RNAs can modulate transcriptional gene expression via epigenetic modulation of targeted promoters.

RNA Transcripts

Sometimes a gene, after being turned on, transcribes a product that (directly or indirectly) maintains the activity of that gene. For example, Hnf4 and MyoD enhance the transcription of many liver specific genes and muscle specific genes, respectively, including their own, through the transcription factor activity of the proteins they encode. RNA signalling includes differential recruitment of a hierarchy of generic chromatin modifying complexes and DNA methyltransferases to specific loci by RNAs during differentiation and development. Other epigenetic changes are mediated by the production of different splice forms of RNA, or by formation of double-stranded RNA (RNAi). Descendants of the cell in which the gene was turned on will inherit this activity, even if the original stimulus for gene-activation is no longer present. These genes are often turned on or off by signal transduction, although in some systems where syncytia or gap junctions are important, RNA may spread directly to other cells or nuclei by diffusion. A large amount of RNA and protein is contributed to the zygote by the mother during oogenesis or via nurse cells, resulting in maternal effect phenotypes. A smaller quantity of sperm RNA is transmitted from the father, but there is recent evidence that this epigenetic information can lead to visible changes in several generations of offspring.
MicroRNAs

MicroRNAs (miRNAs) are members of non-coding RNAs that range in size from 17 to 25 nucleotides. The miRNAs regulate a large variety of biological functions in plants and animals. So far, in 2013, about 2000 miRNAs have been discovered in humans and these can be found online in a miRNA database. Each miRNA expressed in a cell may target about 100 to 200 messenger RNAs (mRNAs) that it downregulates. Most of the downregulation of mRNAs occurs by causing the decay of the targeted mRNA, while some downregulation occurs at the level of translation into protein.

It appears that about 60% of human protein coding genes are regulated by miRNAs. Many miRNAs are epigenetically regulated. About 50% of miRNA genes are associated with CpG islands that may be repressed by epigenetic methylation. Transcription from methylated CpG islands is strongly and heritably repressed. Other miRNAs are epigenetically regulated by either histone modifications or by combined DNA methylation and histone modification.

mRNA

In 2011, it was demonstrated that the methylation of mRNA plays a critical role in human energy homeostasis. The obesity-associated FTO gene is shown to be able to demethylate N6-Methyladenosine in RNA.

sRNAs

sRNAs are small (50–250 nucleotides), highly structured, non-coding RNA fragments found in bacteria. They control gene expression including virulence fragments in pathogens and are viewed as new targets in the fight against drug-resistant bacteria. They play an important role in many biological processes, binding to mRNA and protein targets in prokaryotes. Their phylogenetic analyses, for example through sRNA–mRNA target interactions or protein binding properties, are used to build comprehensive databases. sRNA-gene maps based on their targets in microbial genomes are also constructed.

Prions

Prions are infectious forms of proteins. In general, proteins fold into discrete units that perform distinct cellular functions, but some proteins are also capable of forming an infectious conformational state known as a prion. Although often viewed in the context of infectious disease, prions are more loosely defined by their ability to catalytically convert other native state versions of the same protein to an infectious conformational state. It is in this latter sense that they can be viewed as epigenetic agents capable of inducing a phenotypic change without a modification of the genome.

Fungal prions are considered by some to be epigenetic because the infectious phenotype caused by the prion can be inherited without modification of the genome.
Epigenetics

NOTES

PSI+ and URE3, discovered in yeast in 1965 and 1971, are the two best studied of this type of prion. Prions can have a phenotypic effect through the sequestration of protein in aggregates, thereby reducing that protein’s activity. In PSI+ cells, the loss of the Sup35 protein (which is involved in termination of translation) causes ribosomes to have a higher rate of read-through of stop codons, an effect that results in suppression of nonsense mutations in other genes. The ability of Sup35 to form prions may be a conserved trait. It could confer an adaptive advantage by giving cells the ability to switch into a PSI+ state and express dormant genetic features normally terminated by stop codon mutations.

Structural Inheritance

In ciliates, such as *Tetrahymena* and *Paramecium*, genetically identical cells show heritable differences in the patterns of ciliary rows on their cell surface. Experimentally altered patterns can be transmitted to daughter cells. It seems existing structures act as templates for new structures. The mechanisms of such inheritance are unclear, but reasons exist to assume that multicellular organisms also use existing cell structures to assemble new ones.

Nucleosome Positioning

Eukaryotic genomes have numerous nucleosomes. Nucleosome position is not random, and determine the accessibility of DNA to regulatory proteins. This determines differences in gene expression and cell differentiation. It has been shown that at least some nucleosomes are retained in sperm cells where most but not all histones are replaced by protamines. Thus nucleosome positioning is to some degree inheritable. Recent studies have uncovered connections between nucleosome positioning and other epigenetic factors, such as DNA methylation and hydroxymethylation.

14.2.4 Functions and Consequences

Developmental epigenetics can be divided into predetermined and probabilistic epigenesis. Predetermined epigenesis is a unidirectional movement from structural development in DNA to the functional maturation of the protein. ‘Predetermined’ here means that development is scripted and predictable. Probabilistic epigenesis on the other hand is a bidirectional structure-function development with experiences and external molding development.

Somatic epigenetic inheritance, particularly through DNA and histone covalent modifications and nucleosome repositioning, is very important in the development of multicellular eukaryotic organisms. The genome sequence is static (with some notable exceptions), but cells differentiate into many different types, which perform different functions, and respond differently to the environment and intercellular signalling. Thus, as individuals develop, morphogens activate or silence genes in an epigenetically heritable fashion, giving cells a memory. In mammals, most cells terminally differentiate, with only stem cells retaining the ability to
differentiate into several cell types, ‘totipotency’ and ‘multipotency’. In mammals, some stem cells continue producing new differentiated cells throughout life, such as in neurogenesis, but mammals are not able to respond to loss of some tissues, for example, the inability to regenerate limbs, which some other animals are capable of. Epigenetic modifications regulate the transition from neural stem cells to glial progenitor cells, for example, differentiation into oligodendrocytes is regulated by the deacetylation and methylation of histones. Unlike animals, plant cells do not terminally differentiate, remaining totipotent with the ability to give rise to a new individual plant. While plants do utilise many of the same epigenetic mechanisms as animals, such as chromatin remodelling, it has been hypothesised that some kinds of plant cells do not use or require ‘cellular memories’, resetting their gene expression patterns using positional information from the environment and surrounding cells to determine their fate.

Epigenetic changes can occur in response to environmental exposure – for example, mice given some dietary supplements have epigenetic changes affecting expression of the agouti gene, which affects their fur color, weight, and propensity to develop cancer.

Transgenerational

Epigenetic mechanisms were a necessary part of the evolutionary origin of cell differentiation. Although epigenetics in multicellular organisms is generally thought to be a mechanism involved in differentiation, with epigenetic patterns ‘reset’ when organisms reproduce, there have been some observations of transgenerational epigenetic inheritance, for example the phenomenon of paramutation observed in maize. Although most of these multigenerational epigenetic traits are gradually lost over several generations, the possibility remains that multigenerational epigenetics could be another aspect to evolution and adaptation. As mentioned above, some define epigenetics as heritable.

Two important ways in which epigenetic inheritance can be different from traditional genetic inheritance, with important consequences for evolution, are that rates of epimutation can be much faster than rates of mutation and the epimutations are more easily reversible. In plants, heritable DNA methylation mutations are 100,000 times more likely to occur compared to DNA mutations. An epigenetically inherited element, such as the PSI+ system can act as a ‘stop-gap’, good enough for short-term adaptation that allows the lineage to survive for long enough for mutation and/or recombination to genetically assimilate the adaptive phenotypic change. The existence of this possibility increases the evolvability of a species.

More than 100 cases of transgenerational epigenetic inheritance phenomena have been reported in a wide range of organisms, including prokaryotes, plants, and animals. For instance, mourning cloak butterflies will change color through hormone changes in response to experimentation of varying temperatures.
The filamentous fungus *Neurospora crassa* is a prominent model system for understanding the control and function of cytosine methylation. In this organism, DNA methylation is associated with relics of a genome defense system called RIP (Repeat-Induced Point) mutation and silences gene expression by inhibiting transcription elongation.

The yeast prion PSI is generated by a conformational change of a translation termination factor, which is then inherited by daughter cells. This can provide a survival advantage under adverse conditions. This is an example of epigenetic regulation enabling unicellular organisms to respond rapidly to environmental stress. Prions can be viewed as epigenetic agents capable of inducing a phenotypic change without modification of the genome.

Direct detection of epigenetic marks in microorganisms is possible with single molecule real time sequencing, in which polymerase sensitivity allows for measuring methylation and other modifications as a DNA molecule is being sequenced.

### 14.2.5 Epigenetics in Bacteria

While epigenetics is of fundamental importance in eukaryotes, especially metazoans, it plays a different role in bacteria. Most importantly, eukaryotes use epigenetic mechanisms primarily to regulate gene expression which bacteria rarely do. However, bacteria make widespread use of post-replicative DNA methylation for the epigenetic control of DNA-protein interactions. Bacteria also use DNA adenine methylation (rather than DNA cytosine methylation) as an epigenetic signal.

DNA adenine methylation is important in bacteria virulence in organisms, such as *Escherichia coli*, *Salmonella*, *Vibrio*, *Yersinia*, *Haemophilus*, and *Brucella*. In *Alphaproteobacteria*, methylation of Adenine (A) regulates the cell cycle and couples gene transcription to DNA replication. In *Gammaproteobacteria*, Adenine (A) methylation provides signals for DNA replication, chromosome segregation, mismatch repair, packaging of bacteriophage, transposase activity and regulation of gene expression.

There exists a genetic switch controlling *Streptococcus pneumoniae* (the pneumococcus) that allows the bacterium to randomly change its characteristics into six alternative states that could pave the way to improved vaccines. Each form is randomly generated by a phase variable methylation system. The ability of the pneumococcus to cause deadly infections is different in each of these six states. Similar systems exist in other bacterial genera.
1. Explain the term epigenetics. Which processes it includes?
2. Give the C. H. Waddington and Robin Holliday definitions for epigenetics.
3. How DNA damage can cause epigenetic changes?
4. What are prions?
5. Give examples of transgenerational epigenetic inheritance phenomena.

14.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Epigenetics is the study of heritable phenotype changes that do not involve alterations in the DNA sequence. Derived from the Greek prefix ‘epi-’ means ‘over, outside of, around’, i.e., in epigenetics implies features that are ‘on top of’ or ‘in addition to’ the traditional genetic basis for inheritance. Principally, the word ‘epigenetic’ literally means ‘in addition to changes in genetic sequence’. The term has evolved to include any process that alters gene activity without changing the DNA sequence, and leads to modifications that can be transmitted to daughter cells, although experiments show that some epigenetic changes can be reversed. Many types of epigenetic processes have been identified—they include methylation, acetylation, phosphorylation, ubiquitylation, and sumoylation. Epigenetic processes are natural and essential to many organism functions, but if they occur improperly, there can be major adverse health and behavioral effects.

2. British embryologist C. H. Waddington coined the term ‘Epigenetics’ in 1942 as pertaining to epigenesis, in parallel to Valentin Haecker’s ‘phenogenetics’ and defined ‘Epigenesis’ in the context of the biology of that period referred to the differentiation of cells from their initial totipotent state during embryonic development.

Robin Holliday defined epigenetics as, ‘The study of the mechanisms of temporal and spatial control of gene activity during the development of complex organisms’. Thus, in its broadest sense, epigenetic can be used to describe anything other than DNA sequence that influences the development of an organism.

3. DNA damage can also cause epigenetic changes. The DNA damage is very frequent, occurring on average about 60,000 times a day per cell of the human body. These damages are largely repaired, but at the site of a DNA repair, epigenetic changes can remain. In particular, a double strand...
break in DNA can initiate unprogrammed epigenetic gene silencing both by causing DNA methylation as well as by promoting silencing types of histone modifications.

4. Prions are infectious forms of proteins. In general, proteins fold into discrete units that perform distinct cellular functions, but some proteins are also capable of forming an infectious conformational state known as a prion. Although often viewed in the context of infectious disease, prions are more loosely defined by their ability to catalytically convert other native state versions of the same protein to an infectious conformational state. It is in this latter sense that they can be viewed as epigenetic agents capable of inducing a phenotypic change without a modification of the genome.

5. More than 100 cases of transgenerational epigenetic inheritance phenomena have been reported in a wide range of organisms, including prokaryotes, plants, and animals. For instance, mourning cloak butterflies will change color through hormone changes in response to experimentation of varying temperatures.

The filamentous fungus *Neurospora crassa* is a prominent model system for understanding the control and function of cytosine methylation. In this organism, DNA methylation is associated with relics of a genome defense system called RIP (Repeat-Induced Point) mutation and silences gene expression by inhibiting transcription elongation.

The yeast prion PSI is generated by a conformational change of a translation termination factor, which is then inherited by daughter cells. This can provide a survival advantage under adverse conditions. This is an example of epigenetic regulation enabling unicellular organisms to respond rapidly to environmental stress. Prions can be viewed as epigenetic agents capable of inducing a phenotypic change without modification of the genome.

14.4 SUMMARY

- Epigenetics is the study of heritable phenotype changes that do not involve alterations in the DNA sequence.
- Derived from the Greek prefix ‘epi-’ means ‘over, outside of, around’, i.e., in epigenetics implies features that are ‘on top of’ or ‘in addition to’ the traditional genetic basis for inheritance.
- Principally, the word ‘epigenetic’ literally means ‘in addition to changes in genetic sequence’. The term has evolved to include any process that alters gene activity without changing the DNA sequence, and leads to modifications that can be transmitted to daughter cells, although experiments show that some epigenetic changes can be reversed.
Many types of epigenetic processes have been identified—they include methylation, acetylation, phosphorylation, ubiquitylation, and sumoylation. Epigenetic processes are natural and essential to many organism functions, but if they occur improperly, there can be major adverse health and behavioral effects.

A consensus definition of the concept of epigenetic trait as a, ‘Stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence’ was formulated at a Cold Spring Harbor meeting in 2008, although alternate definitions that include non-heritable traits are still being used.

The term epigenesis has a generic meaning of ‘extra growth’, and has been used in English since the 17th century.

British embryologist C. H. Waddington coined the term ‘Epigenetics’ in 1942 as pertaining to epigenesis, in parallel to Valentin Haecker’s ‘phenogenetics’ and defined ‘Epigenesis’ in the context of the biology of that period referred to the differentiation of cells from their initial totipotent state during embryonic development.

Robin Holliday defined epigenetics as, ‘The study of the mechanisms of temporal and spatial control of gene activity during the development of complex organisms’. Thus, in its broadest sense, epigenetic can be used to describe anything other than DNA sequence that influences the development of an organism.

The term ‘Epigencode’ is a parallel to the word ‘Genome’, referring to the overall epigencode state of a cell, and epigenomics refers to global analyses of epigenetic changes across the entire genome. The phrase ‘Genetic Code’ has also been adapted—the ‘Epigenetic Code’ has been used to describe the set of epigenetic features that create different phenotypes in different cells from the same underlying DNA sequence.

Epigenetic changes modify the activation of certain genes, but not the genetic code sequence of DNA. The microstructure (not code) of DNA itself or the associated chromatin proteins may be modified, causing activation or silencing. This mechanism enables differentiated cells in a multicellular organism to express only the genes that are necessary for their own activity. Epigenetic changes are preserved when cells divide.

DNA damage can also cause epigenetic changes. The DNA damage is very frequent, occurring on average about 60,000 times a day per cell of the human body.

Covalent modifications of DNA, for example Cytosine Methylation and Hydroxymethylation or of Histone Proteins, such as Lysine Acetylation, Lysine and Arginine Methylation, Serine and Threonine Phosphorylation,
Epigenetics

DNA methylation frequently occurs in repeated sequences, and helps to suppress the expression and mobility of ‘transposable elements’.

sRNAs are small (50–250 nucleotides), highly structured, non-coding RNA fragments found in bacteria. They control gene expression including virulence genes in pathogens and are viewed as new targets in the fight against drug-resistant bacteria.

Prions are infectious forms of proteins. In general, proteins fold into discrete units that perform distinct cellular functions, but some proteins are also capable of forming an infectious conformational state known as a prion.

Somatic epigenetic inheritance, particularly through DNA and histone covalent modifications and nucleosome repositioning, is very important in the development of multicellular eukaryotic organisms.

Epigenetic modifications regulate the transition from neural stem cells to glial progenitor cells, for example, differentiation into oligodendrocytes is regulated by the deacetylation and methylation of histones.

More than 100 cases of transgenerational epigenetic inheritance phenomena have been reported in a wide range of organisms, including prokaryotes, plants, and animals. For instance, mourning cloak butterflies will change color through hormone changes in response to experimentation of varying temperatures.

The filamentous fungus *Neurospora crassa* is a prominent model system for understanding the control and function of cytosine methylation. In this organism, DNA methylation is associated with relics of a genome defense system called RIP (Repeat-Induced Point) mutation and silences gene expression by inhibiting transcription elongation.

Eukaryotes use epigenetic mechanisms primarily to regulate gene expression which bacteria rarely do. However, bacteria make widespread use of post-replicative DNA methylation for the epigenetic control of DNA-protein interactions.

Bacteria also use DNA adenine methylation (rather than DNA cytosine methylation) as an epigenetic signal. DNA adenine methylation is important in bacteria virulence in organisms, such as *Escherichia coli*, *Salmonella*, *Vibrio*, *Yersinia*, *Haemophilus*, and *Brucella*.

14.5 KEY WORDS

- **Epigenetics**: It is the study of heritable phenotype changes that do not involve alterations in the DNA sequence.
MicroRNAs (miRNAs): These are members of non-coding RNAs that range in size from 17 to 25 nucleotides, the miRNAs regulate a large variety of biological functions in plants and animals.

Prions: These are infectious forms of proteins, basically, proteins fold into discrete units that perform distinct cellular functions, but some proteins are also capable of forming an infectious conformational state known as a prion.

14.6 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions
1. What is epigenetics?
2. Give definitions of epigenetics.
3. What are the molecular basis of epigenetics?
4. Explain the mechanisms and functions of epigenetics.
5. What is epigenetics in bacteria?

Long Answer Questions
1. Briefly discuss about the significance and functions of epigenetics giving appropriate examples.
2. Explain the various definitions of epigenetics.
3. Discuss the role and significance of the molecular basis, mechanisms and functions of epigenetics.
4. Explain the epigenetics process in bacteria.

14.7 FURTHER READINGS


M.Sc. [Microbiology]  
364 21  
MICROBIAL GENETICS  
II - Semester  

ALAGAPPA UNIVERSITY  
[Accredited with ‘A+’ Grade by NAAC (CGPA:3.64) in the Third Cycle  
and Graded as Category-I University by MHRD-UGC]  
KARAIKUDI – 630 003  
DIRECTORATE OF DISTANCE EDUCATION