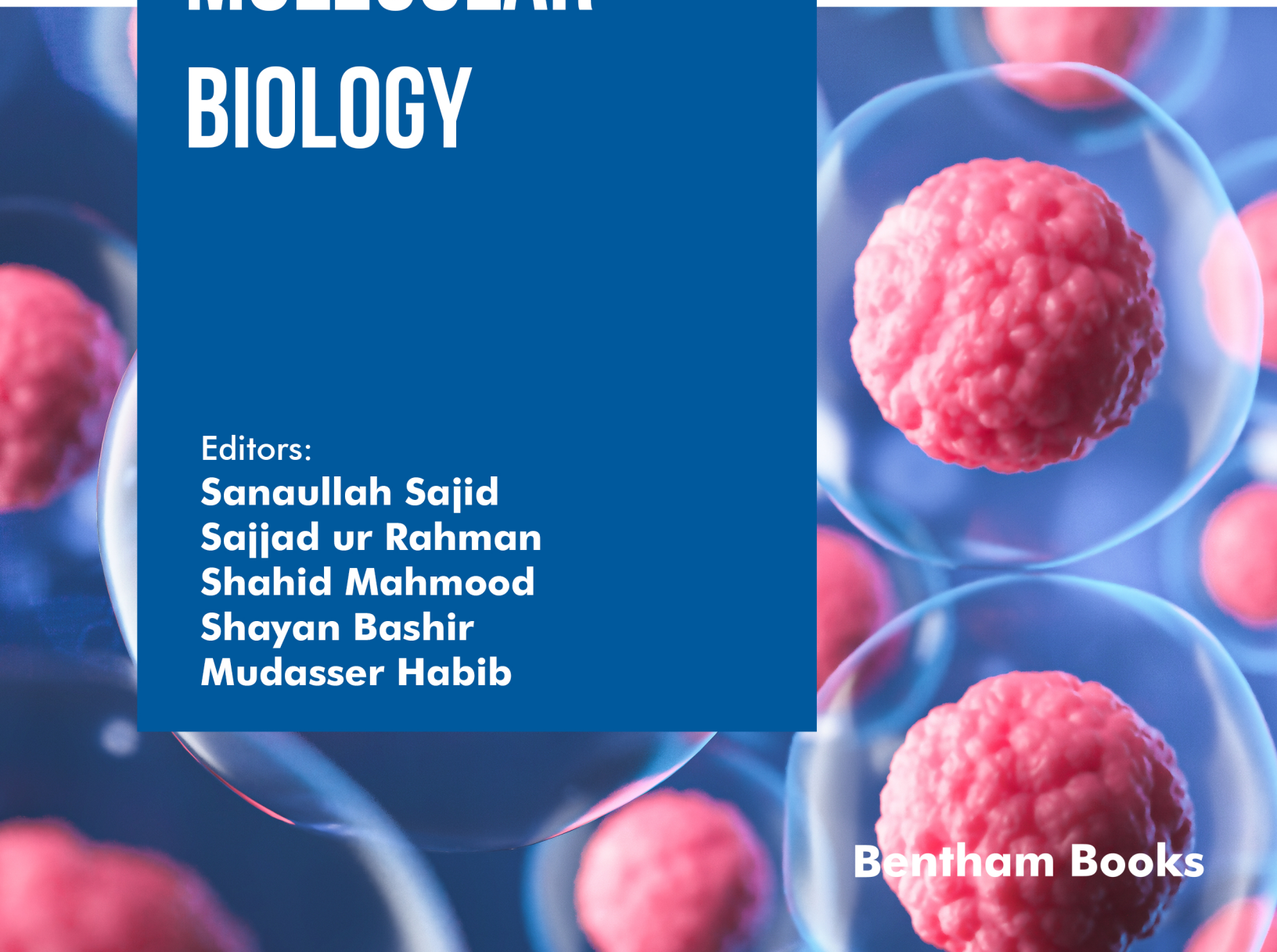
FUNDAMENTALS OF CELLULAR AND MOLECULAR BIOLOGY

Editors:
Sanaullah Sajid
Sajjad ur Rahman
Shahid Mahmood
Shayan Bashir
Mudasser Habib



Bentham Books

Fundamentals of Cellular and Molecular Biology

Edited By

Sanaullah Sajid
Institute of Microbiology
University of Agriculture
Faisalabad, Pakistan

Sajjad ur Rahman
Institute of Microbiology
University of Agriculture
Faisalabad, Pakistan

Shahid Mahmood
Institute of Microbiology
University of Agriculture
Faisalabad, Pakistan

Shayan Bashir
Premier Research, Ontario
Toronto, Canada

&

Mudasser Habib
Nuclear Institute for Agriculture and Biology
Faisalabad, Pakistan

Fundamentals of Cellular and Molecular Biology

Editors: Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood,

Shayan Bashir and Mudasser Habib

ISBN (Online): 978-981-5238-03-7

ISBN (Print): 978-981-5238-04-4

ISBN (Paperback): 978-981-5238-05-1

First published in 2024.

need for a court order if at any point you breach any terms of this License Agreement. In no event will any delay or failure by Bentham Science Publishers in enforcing your compliance with this License Agreement constitute a waiver of any of its rights.

3. You acknowledge that you have read this License Agreement, and agree to be bound by its terms and conditions. To the extent that any other terms and conditions presented on any website of Bentham Science Publishers conflict with, or are inconsistent with, the terms and conditions set out in this License Agreement, you acknowledge that the terms and conditions set out in this License Agreement shall prevail.

Bentham Science Publishers Pte. Ltd.
80 Robinson Road #02-00
Singapore 068898
Singapore
Email: subscriptions@benthamscience.net

CONTENTS

PREFACE

Cellular and molecular biology is a fascinating and rapidly evolving field that has revolutionized our understanding of the living world. From the intricacies of the genetic code to the complex interactions between cells, this field provides insights into the fundamental processes that make life possible.

The aim of this book is to provide a comprehensive introduction to the fundamentals of cellular and molecular biology, covering topics such as cell structure and function, DNA replication and repair, gene expression and regulation, protein synthesis and folding, and cell signaling. Through clear explanations, detailed diagrams, and engaging examples, readers will gain a deep understanding of the underlying principles and mechanisms that govern the behavior of cells and molecules.

Whether you are a student just starting in the field, or a researcher looking to review the basics, this book is an indispensable resource. With its focus on fundamental concepts and its emphasis on clarity and accessibility, it is the perfect guide to the exciting world of cellular and molecular biology.

We hope that this book will inspire readers to delve deeper into the mysteries of life and to pursue their investigations into the intricate workings of the cellular and molecular world.

Sanaullah Sajid
Institute of Microbiology
University of Agriculture
Faisalabad, Pakistan

Sajjad ur Rahman
Institute of Microbiology
University of Agriculture
Faisalabad, Pakistan

Shahid Mahmood
Institute of Microbiology
University of Agriculture
Faisalabad, Pakistan

Shayan Bashir
Premier Research, Ontario
Toronto, Canada

&

Mudasser Habib
Nuclear Institute for Agriculture and Biology
Faisalabad, Pakistan

List of Contributors

Atta Ur Rahman	Institute of Microbiology, Faculty of Veterinary and Animal Sciences, Gomal University, Dera Ismail Khan, KPK, Pakistan
Farwa Mehmood	Department of Pre-clinical and Clinical Operations, Trials 360 (CRO), Lahore, Pakistan
Hafiz Zain Ul Abideen	College of Basic and Clinical Medicine, Dali University, Dali, Yunnan, China
Haseeb Ullah Sajid	Institute of Microbiology, University of Agriculture, Faisalabad, Pakistan
Hira Asghar	Department of Science of Dental Material, Azra Naheed Dental College, Superior University, Lahore, Pakistan
Hayat Ullah	Department of Chemistry, University of Okara, Okara, Punjab, 56300, Pakistan
Itrat Fatima Toor	National Institute of Food Science and Technology, University of Agriculture, Faisalabad, Pakistan
Imran Abdullah	Institute of Nuclear Medicine and Oncology, Lahore, Pakistan
Kainat Gul	Department of Microbiology, Abbottabad University of Science and Technology, Havelian, Abbottabad, Pakistan
Munazzah Sajid	Department of Chemistry, University of Agriculture, Faisalabad, Pakistan
Muneeb Ur Rehman	Medical College of Wuhan, University of Science and Technology, Hubei, Wuhan, China
Maliha Sarfraz	Department of Zoology, Wildlife and Fisheries, University of Agriculture, Faisalabad, Sub-Campus, Toba Tek Singh, Pakistan
Maheen Shafiq	School of Engineering and Built Environment, Griffith University, Nathan, Queensland, Australia
Rai Shafqat Ali	Nuclear Institute for Agriculture and Biology, Faisalabad, Pakistan
Samiullah Sajid	Dali University, Yunnan, China
Sanaullah Sajid	Institute of Microbiology, University of Agriculture, Faisalabad, Pakistan
Shumaila Yousaf	Nuclear Institute for Agriculture and Biology, Faisalabad, Pakistan
Saba Nasir	Nuclear Institute for Agriculture and Biology, Faisalabad, Pakistan
Saima Pervaiz	Department of Immunology, University of Health Sciences, Lahore, Pakistan
Sara Masood Cheema	Department of Pathology, Azra Naheed Medical College, Superior University, Lahore, Pakistan
Shehryar Awan	Department of Pathology, Azra Naheed Dental College, Superior University, Lahore, Pakistan
Zain Ul Abedien	Institute of Microbiology, University of Agriculture, Faisalabad, Pakistan

CHAPTER 1

The Cell and its Molecular Constituents

Samiullah Sajid[1,*] and **Imran Abdullah**[2]

[1] *Dali University, Yunnan, China*

[2] *Institute of Nuclear Medicine and Oncology, Lahore, Pakistan*

Abstract: The cell is the basic unit of life and the fundamental building block of all living organisms. It is a complex and dynamic structure composed of various molecular constituents. These constituents include the plasma membrane, cytoplasm, organelles, and genetic material. The plasma membrane is a thin, flexible layer that separates the cell from its surroundings and regulates the movement of substances into and out of the cell. The cytoplasm is a gel-like substance that contains various organelles, such as mitochondria, ribosomes, and the endoplasmic reticulum, which are involved in multiple cellular functions. The genetic material, composed of DNA and RNA, contains instructions for synthesizing proteins, the building blocks of life. The cell and its molecular constituents play vital roles in maintaining the organism's integrity, responding to environmental cues, and carrying out essential physiological functions. Understanding the organization and function of these molecular constituents is crucial for advancing our knowledge of biology and developing new therapies for various diseases.

Keywords: Cell, Cytoplasm, DNA, Genetic material, Organelles, Proteins, Plasma membrane, Physiological functions, RNA.

INTRODUCTION

All living organisms are made up of small structures called cells. These structures, representing the smallest unit of life, are quite complex and diverse and contain the morphological and physiological characteristics of living organisms. The properties of a given organism depend on its cells, whose continuity occurs through their genetic material. Simple life occurs in isolated cells, which propagate by cell division. Higher organisms, on the other hand, like man himself, are made up of aggregates of cells that perform specialized functions. The cells of different organisms are very similar in structure and molecular constituents, despite fundamental organizational differences. When analyzing the molecular constituents, it is important to consider not only the individual properties of the

* **Corresponding author Samiullah Sajid:** Dali University, Yunnan, China; E-mail: hafizsamiullahsajid@gmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

molecules but also the interactions between them and their location within the cell. That analysis is even more necessary when considering a multicellular organism and the events that occur inside to produce the differentiation and development of that organism. This chapter briefly reviews the cellular structure, its molecular constituents, and the interactions carried out by these compounds [1].

CELLULAR STRUCTURES

The first observations and the very name of cells for structural units, in which all functions necessary for the maintenance and preservation of life are performed, were made by Robert Hooke in 1665. Regardless of the organism's complexity, all cells have the same structure formed by the plasma membrane, which surrounds the cell contents and separates it from the extracellular environment by the cytosol and the nucleus (or nucleoid). The cytosol in the whole cell is composed of a complex aqueous solution with various dispersed particles and molecules. Cell size and shape vary greatly and have no relation to the organism's size. Some cells live in isolation, like unicellular organisms, but those of multicellular organisms generally relate to each other. The plasma membrane, which surrounds all cells, is formed by a double layer of lipids from the phospholipid class and, in variable amounts, if associated with protein molecules [2].

The plasma membrane defines the boundary between the intracellular and extracellular environment. To enter or leave a cell, a substance must cross the cell membrane, which will depend on the membrane's permeability. This lipid bilayer is permeable to certain gases, such as O_2 and CO_2, and impermeable to many substances, such as sugar, amino acids, and inorganic ions (K^+ and Cl^-); water can freely diffuse through the cell. Many proteins are bound to the plasma membrane (permeases or carriers), forming channels in the lipid bilayer and facilitating the passage of certain substances. Thus, all cells of organisms have the characteristic structural architecture of their membranes and many metabolic processes, even DNA replication, synthesis of protein, and chemical energy production.

PROKARYOTIC AND EUKARYOTIC CELLS

Despite the similarity between the cells that make living beings, organisms maintain fundamental differences at the cellular level and can be classified into two large groups: prokaryotes and eukaryotes [3]. Prokaryotic organisms are single-celled and simpler in an organization, although they occur in associated groups, forming colonies with some differentiation of functions. The prokaryotes include bacteria and archaea (bacteria that survive in unusual environments, such as saline lakes, swimming pools, thermals, and swamps). Eukaryotic organisms

are more complex and include not only multicellular plants, animals, fungi, protozoa, and some single-celled organisms such as yeast and green algae (Fig. **1**).

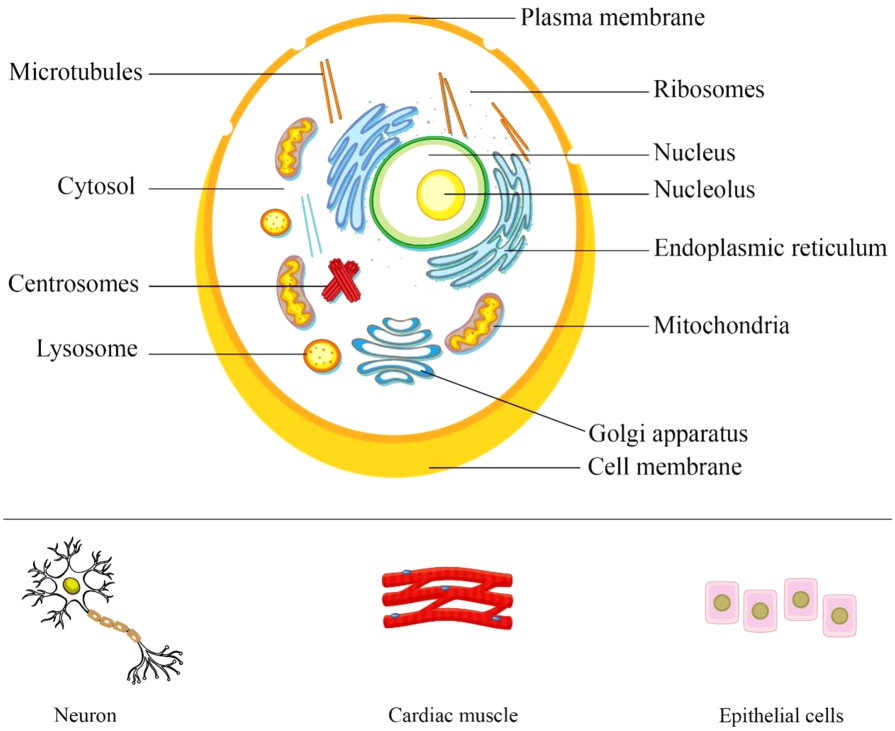


Fig. (1). Schematic representation of an animal cell (eukaryotic cell) with its main structures and cellular organelles. The main features are the presence of a well-defined nucleus containing the genome and the cell organelles, which compartmentalize certain functions.

The main difference between prokaryotes and eukaryotes is that, in eukaryotes, organelles are found, mainly the nucleus, which contains the genome. Organelles are regions bounded by internal membranes, forming specialized functions compartments. In prokaryotes, the absence of an envelope nucleus puts the genome in direct contact with the rest of the cytoplasm in a space within the cell called the nucleoid, staying next to ribosomes, other particles, and a wide variety of dissolved molecules. Prokaryotic cells usually have a cell wall, whose function is to provide greater rigidity and mechanical protection. This membrane is also present in eukaryotic plant cells. The chemical composition of the cell wall of prokaryotes is quite complex, containing molecules of polysaccharides, lipids, and proteins.

CHAPTER 2

Nucleic Acid Structure

Atta Ur Rahman[1,*]

[1] *Institute of Microbiology, Faculty of Veterinary and Animal Sciences, Gomal University, Dera Ismail Khan, KPK, Pakistan*

Abstract: Nucleic acids are essential biomolecules that store and transmit genetic information in all living organisms. The structure of nucleic acids, specifically DNA (deoxyribonucleic acid) and RNA (ribonucleic acid), is crucial for their function. The backbone of nucleic acids is composed of alternating sugar and phosphate groups, while the nucleotide bases (adenine, guanine, cytosine, thymine and uracil) protrude from the sugar-phosphate backbone. The bases form specific hydrogen bonds with complementary bases on the opposite strand, resulting in the double-helical structure of DNA. RNA is usually single-stranded but can form secondary structures through base pairing. The three-dimensional structure of nucleic acids is also important for their function, as it determines the interactions between nucleic acids and other molecules, such as proteins. Understanding nucleic acid structure is fundamental for many areas of biology, including genetics, molecular biology and biotechnology.

Keywords: Cellular process, Gene expression, Genetic information, Nucleotide sequence, Polynucleotide synthesis, Ribonucleic acid, Ribosomal RNA, Transfer RNA.

INTRODUCTION

DNA and RNA are crucial nucleic acids that store and transfer genetic information in living organisms. DNA, a double-stranded molecule, acts as the blueprint for directing cellular functions. Cells transcribe the instructions encoded in DNA into the nucleotide sequence of single-stranded RNA molecules. RNA molecules have various tasks, including the creation of polypeptides, regulation of gene expression, and defense against invading nucleic acids, such as those encountered in viral infections. The study of nucleic acid structure and function, which began over 60 years ago, has provided an unprecedented understanding of biological processes and has transformed several professions, including illness detection, treatment, and forensic inquiry [1].

[*] **Corresponding author Atta Ur Rahman:** Institute of Microbiology, Faculty of Veterinary and Animal Sciences, Gomal University, Dera Ismail Khan, KPK, Pakistan; E-mail: arehman@gu.edu.pk

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

Throughout history, humans have observed patterns of heredity without grasping the underlying mechanisms responsible for conveying physical features and developmental processes from parents to offspring. However, these discoveries have been leveraged by diverse societies to strengthen their economic conditions through selective breeding and crop production. The scientific area of genetics began in the nineteenth century, dedicated to solving the secrets of heredity. At the start of the twentieth century, scientists largely recognized that physical attributes are inherited through discrete units known as genes, which sit on chromosomes within the nucleus of cells. After decades of tedious investigation, the chemical makeup of chromosomes was revealed, leading to the identification of DNA as the bearer of genetic information. The crucial discovery of the DNA structure by James Watson and Francis Crick in 1953 marked a turning point and gave rise to the science of molecular biology.

Molecular biology is a specialist field that focuses on researching the structure of genes and the complicated mechanisms involved in DNA manipulation. Researchers in life sciences, armed with cutting-edge technology developed by molecular biologists and biochemists, have been actively examining the manner in which living creatures organize and process genetic information. This huge corpus of efforts has uncovered numerous essential principles:

1. DNA serves as the central director of cellular operations and is inherited by offspring. It consists of two polynucleotide strands that interlock to form a double helix. The information stored in DNA is delivered by the unique sequence of purine and pyrimidine nucleotides. A gene is a particular sequence of DNA that comprises the required information, including both regulatory sequences that govern gene product synthesis and the instructions for encoding a gene product, such as a polypeptide or various types of protein molecules. DNA replication, known as the synthesis of new DNA strands, includes the precise matching of purine and pyrimidine bases between the original DNA strand and the newly produced strand. The functioning of DNA, both physiologically and genetically, relies on the synthesis of accurate copies of the original and pristine DNA strands. To maintain the fidelity of genetic information, various DNA repair mechanisms are employed by most organisms.
2. The process by which genetic information is decoded and utilized to orchestrate cellular processes commences with the synthesis of another class of nucleic acid called ribonucleic acid (RNA). RNA synthesis, known as transcription, involves the complementary pairing of ribonucleotides' bases with the bases of a DNA molecule. Each newly synthesized RNA molecule is termed a transcript, and the entirety of RNA molecules transcribed from an organism's genome is referred to as the transcriptome.

3. Several types of RNA directly participate in enzyme synthesis, while other proteins are indispensable for the regulated production of all other biomolecules essential for an organism's function. Messenger RNA (mRNA) contains base sequences that specify the primary structure of a particular polypeptide. Ribosomal RNA (rRNA) molecules serve as crucial components of ribosomes, cellular structures responsible for protein synthesis. Each transfer RNA (tRNA) molecule is covalently attached to a specific amino acid and transports it to the ribosome for incorporation into a growing polypeptide chain. The process of protein synthesis, termed translation, transpires within ribosomes, which are molecular machines composed of ribonucleoproteins that interpret the base sequences of mRNA molecules and translate them into amino acid sequences of polypeptides. The complete set of proteins synthesized by a cell is known as the proteome.
4. Gene expression encompasses the mechanisms through which cells temporally regulate gene synthesis in response to environmental or developmental signals. A vast array of proteins known as transcription factors and RNA molecules called non-coding RNA (ncRNA) molecules modulate gene expression by binding to specific DNA sequences. The metabolome refers to the comprehensive collection of low-molecular-weight metabolite molecules generated by a cell as a consequence of its gene expression pattern.
5. The flow of genetic information can be summarized by the central dogma, which illustrates the sequence of molecular events. DNA is transcribed into RNA, and the RNA molecules are translated into proteins. This fundamental principle initially proposed that genetic information flows in one direction only: from DNA to RNA to proteins. However, a significant exception to the central dogma was discovered in recent years. Certain viruses possess RNA genomes and possess an enzyme called reverse transcriptase. Upon infecting a host cell, reverse transcriptase copies the viral RNA to generate a DNA copy, which is then inserted into the host chromosome. An example of such a virus is HIV, as described in the perspective on “HIV Infection”.

The present chapter is dedicated to exploring the structure of nucleic acids. It commences with a detailed description of DNA structure and the scientific investigations that led to its discovery. Subsequently, it provides an overview of the current understanding regarding the structure of the genome and chromosomes, as well as the structure and functions of different forms of RNA. The chapter concludes by elucidating the macromolecular complex formed by nucleic acids and proteins, which are cellular parasites.

The study of nucleic acids has yielded profound insights into the intricate mechanisms of life. It has unraveled the secrets of genetic information storage and transmission, fueling practical applications across diverse fields. The

CHAPTER 3

DNA Replication

Kainat Gul[1,*]

[1] *Department of Microbiology, Abbottabad University of Science and Technology, Havelian, Abbottabad, Pakistan*

Abstract: DNA replication is the process by which a cell makes identical copies of its genetic material. This process is essential for transmitting genetic information from one generation of cells to the next during cell division. DNA replication is a complex process involving unwinding the double helix, separating the two strands, and synthesizing new complementary strands. The process is catalyzed by a large complex of proteins called the replisome, which includes DNA polymerases, helicases, and other enzymes. The accuracy of DNA replication is critical to the maintenance of genetic information, and errors in replication can lead to mutations and genetic diseases. Understanding the molecular mechanisms of DNA replication is fundamental to many fields of biology, including genetics, biotechnology, and medicine.

Keywords: Biotechnology, Cell division, Complementary strands, DNA polymerase, Genetic information, Helicase, Mutations, Molecular mechanisms, Replisome.

INTRODUCTION

Cell division occurs in all organisms and is usually described as a sequence of events that result in genome duplication, called DNA replication. After cell division, the same genetic material must be present in all the cells in both eukaryotic and prokaryotic organisms. For cells to maintain the same genetic material, the entire genome must be precisely replicated before each cell division event. DNA replication requires several enzymatic activities that must act in an integrated and sequential manner in the initiation, chain elongation (or growth), and termination stages. The set of proteins that function in replication is called a replisome. As a fundamental principle of repetition, it can be noted that it will only occur if there is a DNA molecule to be copied (template strand), with a region paired containing a free 3'-OH end where synthesis of the new strand of DNA will begin. The DNA molecule to be replicated can be either double-

* **Corresponding author Kainat Gul:** Department of Microbiology, Abbottabad University of Science and Technology, Havelian, Abbottabad, Pakistan; E-mail: kainatgul195@gmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

stranded or single-stranded. In instances where specific viral genotypes require the replication of DNA with a single strand, it is imperative to have a double-stranded intermediate before initiating the replication process. In addition to these characteristics, some basic rules for the replication process can be applied to the genomes of all prokaryotic and eukaryotic organisms. The addition of nucleotides is always in the direction of 5'-P to 3'-OH (5'→3'); replication always starts at specific origins.

Creating new DNA strands complementary to the original strands occurs during the cell cycle's S phase. Before cell division, a cell repeats its genetic material through DNA replication. This procedure is crucial for retaining the genome's genetic data and ensuring it is appropriately transmitted to daughter cells. Several proteins and enzymes work in unison to replicate DNA, which is a tightly controlled process. The DNA double helix is unwound during initiation when a collection of proteins binds to particular locations on the DNA, known as replication origins. After the DNA has been unraveled, DNA polymerases can start the process of elongation, which involves the synthesis of new DNA strands. Finalization marks the completion of DNA replication, after which the freshly created DNA strands are examined for mistakes and bundled into chromatin in preparation for cell division. All living things must accurately and effectively replicate their genetic material to survive and function. Maintaining the integrity of the genome and preventing the onset of illnesses like cancer require the management of DNA replication. The correct and effective replication of genetic material is regulated by replication origin firing, replication fork stability, replication termination, replication stress response, replication fidelity, replication licensing, and replication-coupled chromatin assembly. The cell can ensure that accurate genetic information is accurately transmitted from one generation to the next and that the genome remains stable and functional by strictly controlling DNA replication.

In addition to regulating DNA replication, the replication process can be affected by various factors, including DNA damage, mutations in replication-related genes, and disruptions to the DNA replication machinery. When DNA damage occurs, cells activate a complex network of DNA damage response pathways that can repair or trigger cell death if the damage is too severe. Mutations in replication-related genes, such as DNA polymerases or helicases, can lead to errors in DNA replication and result in genetic instability or diseases such as cancer.

In addition, it is worth noting that disturbances in the DNA replication apparatus, such as the stalling or collapse of replication forks, can give rise to genomic instability and pathological conditions. The phenomenon of replication fork

stalling arises when the progression of the replication fork is impeded by an obstruction, such as a DNA lesion or a protein complex, which hinders the synthesis of DNA. If the replication fork cannot be restarted, it can collapse, leading to the loss of genetic material and chromosome instability. Understanding the regulation and factors affecting DNA replication is essential for developing new treatments and therapies for diseases associated with genetic instability, such as cancer [1, 2].

UNIDIRECTIONAL AND BIDIRECTIONAL REPLICATION

The replication starts at the origin and proceeds sequentially along the DNA strand in one or both directions to termination, forming a replication bubble. In this replication bubble lies the region of DNA that has already been replicated and flanked by regions where new strands have not yet been synthesized. A replication fork (replisome) starts from the origin in unidirectional replication and continues replicating DNA in one direction, whereas in bidirectional, two replication forks leave the head in opposite directions [3, 4].

In 1958, Meselson and Stahl's experiments demonstrated that DNA replication occurs semiconservative, with each strand of DNA copied into the direction 5'→3', originating new strands. In this replication, the two strands of parental DNA must be separated, and each one will serve as a template for synthesizing new DNA strands. This process is considered semiconservative, as the DNA molecules generated after replication contain only one newly synthesized strand, and the other corresponds to the parental DNA strand. One strand is oriented in the direction 5'→3', and the other in the 3'→5' order (anti-parallelism). Each newly synthesized strand is complementary (AT pairing and GC) to the template strand and presents antiparallel conformation.

Initiation

The first stage of DNA replication is the initiation, in which the replication machinery is assembled at specific sites on the DNA molecule called replication origins. The initiation of DNA replication is a complex process that involves regulatory mechanisms and several proteins. It begins with recognizing the image's origin by a protein complex called the origin recognition complex (ORC) in eukaryotes and the DNAA protein in bacteria. ORC and DnaA bind to specific DNA sequences within the origin, which are rich in adenine-thymine (AT) base pairs and are referred to as AT-rich regions. These regions are easy to denature, which is required for DNA replication. After binding to the origin, ORC and DNA recruit other proteins to form a pre-replication complex (pre-RC) in eukaryotes and a pre-initiation complex (pre-IC) in bacteria. These complexes comprise several proteins, including helicases and single-stranded DNA-binding proteins.

CHAPTER 4

Mutation and DNA Repairs

Shumaila Yousaf[1,*] and **Saba Nasir**[1]

[1] *Nuclear Institute for Agriculture and Biology, Faisalabad, Pakistan*

Abstract: Mutations are changes in the DNA sequence that can occur spontaneously or due to exposure to mutagenic agents such as chemicals, radiation, or viruses. These changes can have a wide range of effects on the organism, from no effect at all to causing genetic disorders or cancer. DNA repair mechanisms exist to correct these mutations, ensuring the integrity of the genetic material. There are several types of DNA repair mechanisms, including base excision repair, nucleotide excision repair, and mismatch repair, each designed to correct different types of DNA damage. The repair mechanisms are highly regulated and involve a complex network of proteins that detect, remove, and replace damaged DNA. Defects in DNA repair mechanisms can lead to an accumulation of mutations, increasing the risk of cancer and other diseases. For example, individuals with inherited mutations in DNA repair genes have a higher risk of developing certain types of cancer, such as breast and ovarian cancer.

Keywords: Chromosomal aberration, *E. coli*, Gene pool, Genetic material, Genetic variability, Nucleotide.

INTRODUCTION

The molecules in an organism's DNA are not static. Often, the nitrogenous bases of their nucleotides are exposed to agents, natural or artificial, that cause modification in their structure or chemical position. Every organism that exhibits a form different from its ancestor's results from a mutant mutation. In a broad sense, mutation refers to any sudden and hereditary modification in an organism's gene pool, not explicable by recombining the pre-existing genetic variability. These changes include changes in the number of chromosomes (euploidy and aneuploidy), the chromosomal structure (chromosomal aberrations), and changes in individual genes. Within the confines of this particular chapter, the term “mutation” shall be employed in a more limited manner, specifically denoting alterations that are identified solely at the individual gene level. The preservation of genetic material across successive generations relies on the imperative of minimizing mutation rates. High mutation rates in the germline would lead to the

[*] **Corresponding author Shumaila Yousaf:** Nuclear Institute for Agriculture and Biology, Faisalabad, Pakistan; E-mail: shumailayousuf_alvi@yahoo.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

extinction of the species, while increased mutation rates in somatic lineages would result in the individual's death. In this way, cells depend on the correct functioning of thousands of genes, each of which can be damaged by a mutation in any of the sites of the coding sequence of the protein or functional RNA or regulatory regions that promote their expression. In general, mutation-carrying organisms in a particular gene have problems in their survival and are eliminated by natural selection. However, not every mutation results in a deleterious consequence for its carrier. All genetic variety comes from mutations, which fuel evolution. This variation is rearranged into new combinations through recombination, and natural or artificial selection protects the combinations best suited to the current environment. Mutation allows organisms to evolve and adapt to their environment [1].

NATURE OF MUTATIONS

All living things mutate due to normal cellular functions or random environmental interactions. These mutations are called spontaneous, and their frequency of occurrence is characteristic for each species, constituting the so-called basal level (background). The occurrence of mutations can be increased by treatment with specific compounds. Those compounds are called mutagens, and the modifications they cause induced mutations. Many mutagens act directly on DNA due to their ability to work as a nitrogenous base or to be incorporated into the polynucleotide chain. The effect of a mutagen is measured by the degree to which it increases the mutation frequency above the basal level. Any nitrogenous base of DNA can be mutated. A point mutation involves modification in a single base pair (substitution) or a few pairs of bases (addition/insertion or deletion). Point mutation may result from a malfunction of the cellular system that replicates or repairs DNA by inserting an incorrect base in the polynucleotide chain, which is being synthesized, or from a chemical interference on one of the bases of DNA [2].

CLASSIFICATION OF MUTATIONS

Mutations must cause some detectable modification for their presence to be recognized. Changes can be minimal, identified only by special genetic and biochemical techniques; they can affect the cell's morphology or be lethal. The extent to which a change alters an organism's characteristics depends on where this mutation occurred and the degree to which it affected the functioning of the gene product. Geneticists classify mutations in various ways because of the wide variety of types and effects. These classifications are not mutually exclusive. The same mutation can be classified in more ways than one.

Classification by Location

Mutations can be classified according to the type of cell or the chromosomal location in which they occur; they can arise in any cell cycle stage. Somatic mutations occur in any cell of the body except the germ cells. They will be perpetuated only in the cells descended from the original cell where the mutation occurred and may not affect the entire organism. Germline mutations occur in the gametes and will be transmitted to the offspring. Autosomal mutations are those that occur in genes located on autosomal chromosomes, and X-linked mutations are those that affect the genes located on the X chromosome. If a somatic cell of a diploid organism experiences an autosomal recessive mutation, the probability of manifesting an observable phenotype is low. The phenotypic manifestation of most of these mutations is likely to be suppressed by the wild-type allele within the cell. The phenotypic consequences of somatic mutations are contingent upon their dominance or, in males, their linkage to the X chromosome. Similarly, somatic or X-linked mutations' impact becomes more evident when they manifest early in the organism's developmental process, during which a limited number of undifferentiated cells undergo proliferation to generate diverse tissues and organs. Germline mutations can be phenotypically manifested in all cells of an offspring. A phonetically detectable autosomal dominant mutation will be displayed in the first generation due to inheritance. Mutations occurring in a gamete of a homogametic female, namely those associated with X-linked recessive traits, have the potential to manifest in the male progeny who inherits the afflicted X chromosome in a hemizygous state. The presence of heterozygosity can lead to the transmission of an autosomal recessive mutation in gametes, regardless of gender, even if it entails a deadly allele. This phenomenon may remain undetected through multiple generations until the mutant allele has successfully disseminated within the population. The manifestation of the novel allele will only occur when chance crossings result in pairing two identical copies of the allele in the homozygous state.

Classification by Type of Molecular Alterations

Mutations are often classified in terms of the nucleotide changes that constitute the mutation. As already mentioned, the substitution of a base pair, on the other hand, in a DNA molecule is called a point mutation or base substitution. The exchange of a naked nucleotide within the coding region of a gene can result in the creation of a new codon, which will correspond to a different amino acid in the protein product; this mutation is called a mutation with a missense mutation. One second possible result is that the original codon is replaced by a stop codon, resulting in premature termination of the translated protein called Nonsense Mutation. If the point mutation changes the codon, but this change does not result

CHAPTER 5

Restriction Enzymes

Zain Ul Abedien[1,*]

[1] *Institute of Microbiology, University of Agriculture, Faisalabad, Pakistan*

Abstract: Restriction enzymes are bacterial enzymes that cleave DNA at specific recognition sequences, usually consisting of four to eight base pairs. These enzymes have become invaluable tools in molecular biology, enabling scientists to manipulate and analyze DNA in various ways. Restriction enzymes are used in various applications, including gene cloning, DNA fingerprinting, and genome mapping. By cleaving DNA at specific sites, restriction enzymes can generate DNA fragments with defined ends, which can then be ligated into vectors for cloning or PCR amplification. Using restriction enzymes in conjunction with gel electrophoresis allows for the separation and analysis of DNA fragments based on their size. There are over 3,000 known restriction enzymes, each with its unique recognition sequence. Many of these enzymes have been isolated from bacteria and are named after the bacterial species from which they were derived. Some restriction enzymes have also been engineered to recognize new recognition sequences, expanding their usefulness in molecular biology. The discovery and development of restriction enzymes have revolutionized molecular biology, allowing scientists to manipulate and analyze DNA in previously impossible ways. As our understanding of the molecular mechanisms of these enzymes continues to grow, they will likely play a critical role in genetics and biotechnology.

Keywords: Biotechnology, Bacterial enzymes, DNA cleavage, DNA fingerprinting, Gene cloning, Genome mapping, Molecular biology, Restriction enzymes, Recognition sequences, Vectors.

INTRODUCTION

The discovery, characterization, and isolation of different types of enzymes have facilitated the research of genomes and their expression, their regulation, and their subsequent correlation with the development of various diseases. Today, the application of molecular biology techniques in multiple areas, particularly in the health sciences, has revolutionized analytical methods, significantly impacting basic and clinical research, especially in diagnosing diseases [1].

[*] **Corresponding author Zain Ul Abedien:** Institute of Microbiology, University of Agriculture, Faisalabad, Pakistan; E-mail: zainuaf82@gmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

Restriction enzymes are a primary tool for the execution of various assays useful in clinical research, from the determination of restriction fragment length polymorphisms (RFLP), as well as in the construction of vectors with recombinant DNA and sequence cloning of DNA from humans, viruses, bacteria, and other microorganisms of interest for the generation of knowledge or application in the medical field. Among enzymes that modify nucleic acids include nucleases, *i.e.*, those capable of cleaving the phosphodiester bonds that link the nucleotides to a chain of nucleic acids. Nucleases are called endonucleases if they can cleave the phosphodiester bond at nucleotides within the nucleic acid chain. They are called exonucleases if they cleave the phosphodiester bonds of nucleotides at the ends of chains. Thus, some 3'exonucleases cleave bonds at the 3' end of the chain, and 5' exonucleases cleave phosphodiester bonds at the 5' end of the chain.

Restriction and other enzymes that modify nucleic acids, such as ligase, allowed the development of recombinant DNA technology. Restriction enzymes exhibit endonuclease activity, are of bacterial origin, and cut phosphodiester DNA bonds at a specific target sequence. Using these enzymes as a biotechnological tool arose from the need to cut the long genomic DNA strands to manipulate and analyse them into smaller fragments.

In 1960, Stewart Linny and Werner Arber isolated enzymes in *E. coli* that methylated DNA molecules and cleaved a phosphodiester bond from the unmethylated DNA. Later, in 1970, the first restriction nuclease characterized by Daniel Nathans and Hamilton Smith, isolated from Hemophilus influenzae, called Hind I. At present, it is known that all species of bacteria synthesize one or more types of endonucleases capable of making cuts in specific nucleotide sequences of an exogenous DNA double-strand; generally, this “foreign” DNA is from bacteriophages, viruses with the ability to infect such bacteria.

These endonucleases are called restriction enzymes because they restrict the permanence of an exogenous DNA in the bacterial cell that produces such an enzyme. The restriction enzymes most frequently used in recombinant DNA technology usually recognize a single sequence of 4 to 8 nucleotides, where they perform the cut.

ORIGIN OF RESTRICTION ENZYMES

Restriction endonucleases are part of the bacterial machinery they rely on to defend themselves against viral infections [2]. The DNA of the bacterium that produces the restriction enzyme is unaffected by its enzyme, as bacteria methylate specific sequences to their DNA to differentiate it from viral DNA utilizing a

methyltransferase enzyme at particular bases, which usually are those recognized by their restriction enzymes (Table **1**).

Table 1. Characteristics of restriction enzymes.

Enzyme Type	Energy Expenditure	DNA Cleavage Site	Methylase Activity
I	ATP	Random, outside the target sequence.	Yes
II	-	Specifically in the target sequence.	No
III	ATP	Random, 25 to 27 base pairs downstream from the target sequence.	Yes

Nomenclature

By international agreement, the nomenclature of restriction enzymes is assigned according to their bacterial origin and based on the following rules:

1. Three letters in italics correspond to the bacterium's scientific name from which it was attracted, *e.g*., *Escherichia coli* (Eco), Hemophilus influenzas (Hin), *etc*. The first letter, in capital letters, corresponds to the genus; the other two to the species.
2. The strain or lineage, if any, *e.g*., EcoRI, isolated from the"RY13" strain of *E. coli*.
3. In Roman numerals, a number is distinguished if more than one endonuclease is isolated from the same species, *e.g*., EcoRI, EcoRV.
4. All should indicate an "R" for restriction or "M" for methylase, depending on the enzyme's function, but it is usually omitted.

For example:

Hind III:

H = genus Hemophilus.

In = species influenza.

d = strain DSM 11121.

III = third endonuclease isolated from this organism.

EcoRI:

E = genus Escherichia.

CHAPTER 6

Cloning and Expression Vectors

Zain Ul Abedien[1,*] and **Rai Shafqat Ali**[2]

[1] *Institute of Microbiology, University of Agriculture, Faisalabad, Pakistan*

[2] *Nuclear Institute for Agriculture and Biology, Faisalabad, Pakistan*

Abstract: Cloning and expression vectors are essential tools in molecular biology that enable researchers to manipulate and study genes and proteins. Cloning vectors are DNA molecules that can carry foreign DNA fragments and introduce them into host cells. Expression vectors are specialized cloning vectors designed to drive the expression of foreign genes in a host cell, allowing researchers to produce large quantities of recombinant proteins. The most commonly used cloning vectors are plasmids; circular DNA molecules that can replicate independently of the host chromosome. Plasmids can be engineered to contain various features such as selectable markers, multiple cloning sites, and regulatory sequences. These features make plasmids versatile application tools, including gene cloning, mutagenesis, and genetic engineering. Expression vectors are typically based on plasmids and contain additional elements that enable the efficient expression of the foreign gene. These elements include a strong promoter, a ribosome binding site, and a transcription terminator, which work together to ensure the production of the recombinant protein. The choice of expression vector depends on the desired protein expression level, the host cell type, and downstream applications. The use of cloning and expression vectors has revolutionized the field of molecular biology, enabling the production of recombinant proteins, genetic engineering of organisms, and gene therapy. However, using these tools requires careful consideration of potential risks, such as unintended genetic modifications or the spread of genetically modified organisms.

Keywords: Complementary DNA, Expression vector, Molecular cloning, Oligonucleotide sequence, Restriction enzyme.

INTRODUCTION

Until the 1970s, DNA molecules were extremely difficult to analyze since it was too long and consisted of only four monomers. At that time, the nucleotide sequence of DNA could only be studied indirectly through the sequence of amino acids in proteins or through their expression in RNA. Therefore, the discovery of restriction enzymes facilitated the study of DNA since it was possible to section

* **Corresponding author Zain Ul Abedien:** Institute of Microbiology, University of Agriculture, Faisalabad, Pakistan; E-mail: zainuaf82@gmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

large DNA molecules and separate them into fragments. Each of these fragments was analyzed separately, and it was possible to multiply them through the polymerase chain reaction technique and even determine the sequence of its nucleotides by sequencing. Restriction enzymes can cut off the recognized sequences with specificity. Identifying, isolating, and cloning DNA fragments from different organisms has made it possible to produce recombinant DNA molecules. The term cloning indicates the act of producing many identical copies of a DNA molecule; currently, cloning is used for the production of recombinant molecules and to determine their functions in the organism [1].

CLONING

DNA cloning, or molecular cloning, introduces a DNA fragment known as insert into a DNA molecule called a vector, which can replicate autonomously and independently of the host cell genome. The result is obtaining millions of copies of a recombinant molecule or molecular clone composed of DNA from the insert and the vector. The insert can be DNA obtained from any organism. It can remain as genomic DNA (gDNA), complementary DNA (cDNA), a product of retro transcription of RNA, a PCR product, or a derived RNA by *in vitro* transcription [2].

Cloning Vectors

A vector is a double-stranded DNA molecule that can harbor an exogenous DNA fragment (of another origin).

Classification of Vectors

According to their use, vectors are classified as cloning vectors and expression vectors.

Cloning Vectors: Their purpose is to store sequences and obtain large quantities of the inserted DNA or recombinant molecule. Cloning vectors are usually plasmids, phages, cosmids, and artificial bacterial or yeast chromosomes.

Expression Vectors: The objective is to produce a transcript (RNA) or transcribe the protein product. Expression vectors can be plasmids or phages.

Elements that Form a Cloning Vector: All cloning vectors have, as part of their structure, the following common essential elements:

a) Origin of replication: it is a sequence in the DNA of the vector that provides a single site of recognition to proteins that identify the origin of the replication site

(ORI). Plasmids are characterized by having a single ORI site in their genome and performing unidirectional replication.

b) Selection marker: a gene that generally confers resistance to an antibiotic or generates a particular phenotype with which the cell can be selected, incorporating the vector. Some vector plasmids have a second selectable marker, the lacZ gene, found in the *E. coli* Lac operon and encodes the enzyme β-galactosidase. Thus, when the vector replicates within the host cell, the β-galactosidase can be expressed and generate a functional enzyme. If the cell containing the vector is put into contact with a substrate of the enzyme, such as X-gal (5-bromo-4-chl-ro-3-indolyl β-D-galactopyranoside), is degraded by the enzyme, which produces a blue color that allows identifying the cells that introduced the vector.

c) Multiple cloning site: it is a DNA fragment containing several unique restriction enzyme recognition sites close to each other, with a wide range of possibilities to insert any piece of DNA.

Plasmids

Plasmids are small circular DNA molecules that replicate and are transmitted independently of the bacterial chromosome. They naturally constitute mobile genetic material of bacteria, transporting genes to other bacteria, such as antibiotic resistance. The plasmids remain episomal in the bacterium, that is, without being incorporated into its genome, and there may be multiple copies of the same plasmid in it. When a bacterium reproduces, these Plasmids are transmitted to progeny cells by equal distribution of the bacterial cytoplasm during cellular division. However, they can also be transmitted through the pili between the bacteria. Plasmid replication and transcription depend on the enzymatic machinery of the host cell. Plasmids have been genetically manipulated in the laboratory to preserve only the desirable characteristics and eliminate unnecessary DNA. The plasmids used as a vector, in general, are formed by 2 to 5 kb of DNA, which facilitate the analysis of the inserts incorporated into it. Hundreds of commercially available plasmids offer various possibilities regarding the restriction enzymes used for the cloning, the length of the product to be inserted, and the selection markers of the transformed clones. The most commonly used plasmids are pBluescript, pUC19, pBR322, and pGLO.

Bacteriophages

Bacteriophages are viruses that naturally infect bacteria and behave as obligate intracellular parasites that multiply using the biosynthetic machinery of bacteria. Like plasmids, bacteriophages replicate autonomously, carry genetic information, and can give bacteria new properties or develop pathological processes. To

CHAPTER 7

Recombinant DNA Technology

Farwa Mehmood[1,*]

[1] *Department of Pre-clinical and Clinical Operations, Trials 360 (CRO), Lahore, Pakistan*

Abstract: Recombinant DNA technology, also known as genetic engineering, has revolutionized the field of molecular biology by allowing researchers to manipulate and transfer DNA sequences between different organisms. This technique involves the use of restriction enzymes to cut DNA molecules, which can then be spliced together to create novel sequences. Recombinant DNA technology has numerous applications in medicine, agriculture, and biotechnology. For example, it has been used to produce human insulin, growth hormone, and other therapeutic proteins. In agriculture, it has been used to create crops with improved yield and resistance to pests and diseases. However, the use of recombinant DNA technology also raises ethical and safety concerns, and its regulation is subject to the ongoing debate.

Keywords: Agriculture, Biotechnology, DNA splicing, Genetic engineering, Regulation, Recombinant DNA.

INTRODUCTION

Recombinant DNA techniques comprise a large number and variety of technological developments, which have constantly been advancing over the last 30 years and have given rise to biotechnology, one of the most active and fastest-growing fields of scientific research in the last two decades. These techniques allow researchers to isolate genes from many different organisms, manipulate the purified DNA in the laboratory, and then transfer these genes to other organisms [1].

Gene cloning experiments have survived the hysteria of the 1970s, during which the fears of creating pathogenic organisms causing uncontrollable plagues threatened to suspend this line of experimentation from its origins. These fears have proved to be unfounded since the Asilomar Conference and the development of procedures by the National Institute of Health (NIH) to conduct recombinant

[*] **Corresponding author Farwa Mehmood:** Department of Pre-clinical and Clinical Operations, Trials 360 (CRO), Lahore, Pakistan; E-mail: farwamehmood4@gmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

DNA experiments. On the other hand, as experimental evidence has accumulated the beneficial potential of these new technologies for both basic and applied research, they are now considered one of the most important scientific developments in history. This chapter mainly deals with the basic concepts related to recombinant DNA technology, as well as briefly a description of some related techniques, the most recent advances in this field, and the achievements and future goals of this technology.

Historical background to biotechnology:

- 1909-British physicist Archibald Garrod is the first to propose the relationship between genes and proteins. He suggested that genes may be involved in the formation of the proteins that carry out the reactions of metabolism.
- 1928-Franklin Griffith discovered that increasing temperature could transfer genetic information from dead to living bacteria. This phenomenon was called transformation.
- 1930-Through experiments performed with mutant strains of Neurospora yeasts, George Beadle and Edward Tatum provide evidence supporting Garrod's hypothesis. This evidence gives rise to the hypothesis: “one gene, one protein,” *i.e*., that each protein in the cell results from gene expression.
- 1944-Oswald Avery, Maclyn McCarty, and Colin MacLeod identify DNA as Griffith's “transforming agent”.
- 1957-During a dysentery epidemic in Japan, biologists discovered that certain strains of bacteria were resistant to antibiotics. It was later found that this resistance was due to plasmids.
- 1961-Sidney Brenner and Francis Crick propose the genetic code, establishing that groups of three nucleotides or codons encode the amino acids of proteins in the DNA.
- 1966- Nirenberg: The genetic code is deciphered by discovering, through biochemical analysis, the amino acid that determines each codon.
- 1970-Hamilton Smith, at the Johns Hopkins School of Medicine, isolates a restriction enzyme for the first time.
- 1972 -Stanley Cohen and Herbert Boyer combine their efforts to generate recombinant DNA technology, giving rise to the biotechnology industry.
- 1976-Herbert Boyer co-founded Genentech, the first company to apply recombinant DNA technology.
- 1978-The first human protein is synthesized using recombinant DNA technology, somatostatin, which regulates growth hormones.

BASICS OF NUCLEIC ACID

The genetic information needed to synthesize the proteins of all cells is contained in DNA (deoxyribonucleic acid). DNA, in the nucleus of eukaryotic, is associated with histone proteins that make up chromatin, a structure with a varying degree of packing that will give rise to chromosomes. In prokaryotes, mitochondria and chloroplasts contain genetic information in a circular DNA molecule. Both cases have two antiparallel strands, which give rise to a double-helix structure. The complementary strands are polymers of nucleotides (nitrogenous base + sugar + phosphate). The genetic information is encoded in the sequence of purines (adenine and guanine) and pyrimidines (cytosine and thymine), which, grouped in threes, make up a codon, or basic unit of genetic information.

Each codon translates into an amino acid and establishes its position in the sequence of a protein. The position occupied by adenine in one strand is paired with thymine in the complementary strand by two hydrogen bonds. Guanine and cytosine pair with each other by three hydrogen bonds. Before cell division, two copies of the DNA must be generated so that each daughter cell retains complete genetic information. This process of producing two identical copies of DNA is called replication or duplication. The information stored in the DNA is transferred to the messenger RNA (mRNA) through transcription. The mRNA is complementary to the DNA strand from which it was copied. The nitrogenous bases that compose it are the same as those of DNA, except thymine, which is replaced by uracil. Therefore, the information in the mRNA will be decoded for the synthesis of proteins through the translation process. This process is carried out in the cell cytoplasm by ribosomes, supra-macromolecular complexes responsible for reading and translating the genetic message into proteins.

The coding region of the DNA contains the codons that will be translated into protein through the mRNA. In prokaryotes, the coding region of DNA is usually uninterrupted and directly produces the mRNA, which practically does not undergo maturation processes and is translated immediately. However, this is not the case in eukaryotes, but the regions of DNA containing the coding sequences (exons) are separated by noncoding sequences (introns). Thus, the initial transcription of the DNA generates a large RNA molecule called heterogeneous nuclear RNA (hnRNA), which will undergo a maturation process inside the nucleus called the adjustment (splicing or editing process), in which the introns will be cut and removed. The exons are spliced together to give rise to the mature mRNA, which will exit the cytoplasm, where it will serve as a template for protein synthesis. The precise function of introns is not yet fully known; nevertheless, it was evident that mRNA production in eukaryotes is much more complex than in prokaryotes due to the negative charges of the phosphate groups;

CHAPTER 8

DNA Polymorphisms and Genetic Fingerprint

Itrat Fatima Toor[1,*]

[1] *National Institute of Food Science and Technology, University of Agriculture, Faisalabad, Pakistan*

Abstract: DNA polymorphisms are variations in the genetic sequence that occur within a population. These polymorphisms can be used as genetic markers to identify individuals, determine familial relationships, and study population genetics. Genetic fingerprinting is a widely used method for identifying individuals based on DNA polymorphisms. Genetic fingerprinting involves the analysis of DNA polymorphisms at multiple loci to generate a unique genetic profile for an individual. The most common types of DNA polymorphisms used for genetic fingerprinting are short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs). STRs are short, repeated sequences of DNA that vary in length between individuals. They are highly polymorphic and generate DNA profiles by analyzing the number of repeats at each locus. SNP markers, on the other hand, are single nucleotide variations that occur at specific positions within the genome. They are less polymorphic than STRs, but genetic mapping and association studies are widely used. Genetic fingerprinting has many applications, including forensic science, paternity testing, and conservation biology. However, the use of genetic data also raises ethical concerns regarding privacy and discrimination.

Keywords: Conservation biology, DNA profiles, DNA polymorphisms, Genetic fingerprinting, Paternity testing, Short tandem repeats (STRs), Single nucleotide polymorphisms (SNPs).

INTRODUCTION

The DNA of all known species of organisms has the same chemical structure; however, each organism is entirely different from another; the difference is due to the order of the nitrogenous bases in the DNA molecule. Organisms of the same species share sequences in their DNA molecule, but within the same species, there are variations between individuals. Organisms of the same species will share up to 99% of their sequences, which gives them similar traits. Close relatives will also

* **Corresponding author Itrat Fatima Toor:** National Institute of Food Science and Technology, University of Agriculture, Faisalabad, Pakistan; E-mail: itratfatima71@gmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

have more similar sequences, but they will never be identical. This is what is meant by intraspecies and interspecies genetic variability.

In human DNA, the gene sequences are not very variable within the species, but the rest of the sequence is very variable. Due to many millions of base pairs per DNA molecule and a high percentage of it does not code for a protein, each person has a unique DNA sequence that can be identified only by its base pair order, called polymorphism. There are two types of genetic polymorphs: those that show a single nucleotide change by substitution of bases and those involving changes in the size of the sequence; this may be due to insertions or deletions of DNA sequences or to base repetition (or combination of bases) continuously in a segment of the DNA.

This variability is a typical biological process since 30% of the DNA sequence is highly repetitive, which makes it possible to establish specific genetic patterns to identify individuals, *i.e*., to establish a genetic footprint or DNA fingerprinting. Therefore, scientists have developed strategies to identify individuals by analyzing repetitive patterns of DNA, which allow relating samples from the exact origin or relative people (common ancestors). These sequences were chosen because they are known to vary between unrelated individuals, so the analysis of a group of them to find concordances allows one to infer associations with high probabilities of having the exact origin or being from the same person. A gene is a chromosomal region identical to a DNA segment with a predetermined sequence and contains the instructions needed to produce a particular protein. An organism's genes contain accurate information about every part and process of a person's training and development. However, genes cannot control behavior, so it is important to remember that the environment plays a significant role in how these genes are expressed throughout an organism's life. A gene presents in a double strand, one that provides the father and the other the mother; these copies are called alleles, so there are two alleles for each gene, one of maternal origin and another paternal. These allelic forms of the same gene exist because the DNA sequence is subject to changes, sometimes due to mutations that happen randomly but that give rise to alternate, stable, and inherited forms of a gene with a different function than the original allele or wild. The mutation is the basis of the gradual variation of the genetic structure of populations; that is, the base of evolution. The physical position where a gene is located in the genome is called a locus (plural loci); therefore, the wild-type or modified alleles of the same gene reside in the same locus. When an individual has different allelic forms at the locus is said to be heterozygous since each allele's sequence can generate variants of the same protein. The protein product will be the same regardless of whether the allele is wild-type or modified. On the other hand, an individual is to be homozygous

when they display the same allelic form (of either their paternal or maternal origin) at the locus in question [1].

POLYMORPHISMS

Organisms have many variable sequences, and population genetic analyses have made finding two or more alleles for each one possible. The more different the alleles are, the more polymorphic they are. It means it is less likely that two people will have the same polymorphic regions, making it easier to tell them apart. It considers that there is a genetic polymorphism when there are multiple alleles of a gene in a defined population or, more specifically, when the base sequence neogenesis of the DNA molecule of a particular locus is variable among the organisms of a population. The word polymorphism is composed of poly (many), morpho (form), and is (process or state), *i.e.*, "in many ways". A polymorphism can be observed in a complete individual, in varying forms of proteins or blood groups (biochemical polymorphism), in the morphological characteristics of chromosomes (chromosomal polymorphism), or in the DNA, by differences in the nucleotide sequence (DNA polymorphism). Nucleic acids and proteins have the property of appearing in different molecular forms or in multiple alleles, which may have implications in molecular pathologies.

The allelic variants of the coding sequences of genes in the DNA, due to normal processes of the cell function, promote the production of polymorphic codons and, with this, in alternate forms of proteins, generally without altering the function of the synthesized product, which allows differentiating a polymorphism from a mutation, which usually occurs at random. The polymorphism may happen in the promotor region of the gene, which influences the messenger RNA gene expression and, therefore, in the protein that it encodes (different phenotypes) or even identifies itself in untranslated regions (introns) so that there is no interpretation of its function, at least known so far. However, they are still sequences that allow differentiating individuals and species. Polymorphisms, like mutations, can be categorized by how they change the genetic code. Silent polymorphisms do not change the translation of the protein product when the sequence changes. As the nucleotide sequence changes, the codon that codes for the original amino acid changes to another codon that codes for the same amino acid or one with similar chemical properties. In addition, it considers that the neutral polymorphs, which are the ones that vary in their sequence in non-coding regions of the DNA, are also silent. Polymorphism is considered neutral if the presence or absence of the allele does not confer any advantage or disadvantage to the individual. In addition, a polymorphism may represent evolutionary advantages for a given population, such as conferring resistance to environmental conditions, according to the geographical area which they inhabit. The term

CHAPTER 9

Application of Molecular Biology in Biotechnology

Hira Asghar[1,*]

[1] *Department of Science of Dental Material, Azra Naheed Dental College, Superior University, Lahore, Pakistan*

Abstract: Molecular biology is the study of molecular interactions and structures that govern cellular processes. Biotechnology is an interdisciplinary field that utilizes biological systems to create new products and technologies. The application of molecular biology in biotechnology has revolutionized the way we understand and manipulate biological systems. This technology has enabled the development of new diagnostic tools, therapies, and drugs. The ability to manipulate DNA and RNA sequences has also allowed the creation of genetically modified organisms with desirable traits, including crops and animals. In addition, molecular biology has facilitated the development of gene therapy, where defective genes are replaced or repaired, and vaccines, where specific antigens are identified and synthesized. Integrating molecular biology and biotechnology has opened up vast opportunities for research and innovation in many fields, including medicine, agriculture, and environmental science. As a result, the application of molecular biology in biotechnology holds great promise for the future of science and technology.

Keywords: Biotechnology, DNA sequencing, Genetic engineering, Gene therapy, Genetically modified organisms (GMOs), Molecular biology, Recombinant DNA technology, Synthetic biology.

INTRODUCTION

Since time immemorial, humankind has used living beings for its benefit. In different ways, sometimes by accident and sometimes with great ingenuity, procedures have been discovered to improve animal and plant varieties. The agricultural revolution, animal domestication, and wine, beer, and yoghurt production are but various manifestations of biotechnology. Biotechnology has been defined as the application of biological organisms, systems, and processes in manufacturing and service industries. However, modern biotechnology uses DNA techniques to improve living things with a view to their utilization. Biotechnology does not exist as a scientific discipline but rather as a wide variety of subject

[*] **Corresponding author Hira Asghar:** Department of Science of Dental Material, Azra Naheed Dental College, Superior University, Lahore, Pakistan; E-mail: hir_08@hotmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

areas. Its fruits have grown from a tree whose roots are the biological sciences, particularly microbiology, genetics, molecular biology, and biochemistry.

This word was first used in the early 1920s by the Leeds City Council in the United Kingdom when its members founded an Institute of Biotechnology. However, biotechnological processes date back some 5,000 years ago, when the production of alcoholic beverages by fermentation was discovered. The ancient Egyptians went so far back in time that they even used mouldy bread as a poultice for healing infected wounds and introduced a service to detect pregnancy based on the effect of urine on the germination rate of wheat. The term biotechnology has only recently been introduced into popular vocabulary due to the great potential resulting from the application of molecular biology techniques. The advent of PCR has revolutionized all areas of scientific research to the point of making a multitude of processes that previously required manual and imprecise procedures. With it, the promise that biotechnology could become a commercial success has become a reality in just a few years. In addition, the relatively new *in vitro* genetic manipulation has expanded the number of products generated by microorganisms. It has provided new methods to increase the yield in obtaining existing ones.

Recombinant DNA technology has also significantly impacted the industry due to the unlimited possibility of separating a gene or DNA sequence from any source and transferring it to another organism so that it can replicate and be passed on to offspring. This sequence can be a gene that codes for a commercially helpful product, transferred to species in which it can be synthesized efficiently and economically, or a new trait for the host microorganism. The essential elements of the recombinant DNA technique are: 1) Obtaining specific DNA fragments (restriction enzymes). Restriction endonucleases are invaluable for fragmenting DNA at specific sites, as they recognize a particular sequence, which differs for different species and strains. 2) To reassociate two DNA molecules, covalent ligation or reassociation is necessary to reconstitute a phosphodiester bond; an enzyme called DNA ligase comes into action. 3) Mechanism to introduce recombinant DNA into the organism (transformation). Mechanism to ensure replication and identification of the recombinant molecule within the cell. If the transferred DNA is to be able to replicate, it must be bound to a molecule capable of multiplication in the host cell. A small molecule is required, which also possesses the replication capacity (*i.e.*, independent of the chromosome); these molecules are called vectors. In nature, there are two kinds of vectors: plasmids (extrachromosomal DNA molecules that are found as multiple copies in many bacterial species and are generally not necessary for cell viability, although they can carry some genes, such as those that confer antibiotic resistance), and bacterial viruses or phages, which can replicate inside a bacterium and give rise to many identical molecules. The bound molecules are introduced into host cells and

selected by marker genes in the cloning vector. Within their sequence, cloning vectors contain signals that induce DNA replication.

Currently, with the introduction of PCR by Kary Mullis in 1983, restriction enzyme and sequencing, and hybridization techniques, the sequences of a single gene can be localized among an abundance of other sequences, cut at specific sites, and amplified exponentially to yield an analyzable quantity. PCR has increased the possibilities offered by DNA technology, as it allows the amplification of specific DNA sequences without the need for cloning in cells. This chapter aims to show how molecular biology has provided the primary impetus in the development of biotechnology and to describe how the potential of these techniques can be utilized in obtaining products or processes useful for man in some of the areas in which molecular biology has had the most significant impact, such as the pharmaceutical and agri-food industries [1].

HEALTH AND PHARMACEUTICAL INDUSTRIES

The emergence of genetic engineering and its impressive development to date has made it possible to initiate an in-depth analysis of the structure and mechanism of action of the gene program of a living being. The applications of this new knowledge have consolidated the origin of a new medicine that seeks to understand the molecular basis of health and disease. Today's therapy successfully uses genetic technology that allows the production of recombinant proteins necessary for diagnosing and treating genetic and infectious diseases and developing new vaccines. These methodologies have also affected the diagnosis and treatment of hereditary disorders. It will soon offer a new therapy to introduce healthy genes into cells with defective genes within our body.

The commercial application of recombinant DNA technology began in the late 1970s with the production of proteins as therapeutic agents. Several new therapeutic agents are based on genetically engineered antibodies, antisense technology, and other gene products and therapies through molecular biology techniques. However, therapeutic proteins have had the most significant impact on their use. In addition, their synthesis has been achieved in the quantities required for medicinal use. Typical examples are the transfer *to Escherichia coli* of the human insulin gene, used in the treatment of diabetes; the interferon genes for viral diseases and in combination with other agents in the treatment of different types of cancer; the human growth hormone messenger in the treatment of pituitary dwarfism; erythropoietin, which stimulates erythrocyte production and is used in the treatment of acute anaemia or related to treatments, and myeloid colony-stimulating factor, to complement anti-cancer therapy and in bone marrow transplants. As molecular biologists began the need for authentic and pure

CHAPTER 10

Application of Molecular Biology in Gene Therapy

Maheen Shafiq[1,*]

[1] *School of Engineering and Built Environment, Griffith University, Nathan, Queensland, Australia*

Abstract: Gene therapy treats genetic and acquired diseases by introducing functional genes into cells to replace or correct defective genes. The field of molecular biology has played a significant role in the development of gene therapy, providing tools and techniques to manipulate and analyze genes and their expression. One of the main challenges in gene therapy is the efficient delivery of therapeutic genes to the target cells. Molecular biology has provided a range of vectors, such as viruses and plasmids, that can be used to deliver genes to cells and methods to modify these vectors to improve their efficacy and safety. Molecular biology has also contributed to the development of gene editing technologies, such as CRISPR-Cas9, which can be used to correct or modify genes at the genomic level. This approach can potentially treat genetic disorders by targeting the underlying genetic mutations. In addition, molecular biology has facilitated the development of methods to regulate gene expressions, such as gene silencing and RNA interference, which can be used to turn off genes that cause disease. Despite the progress made in gene therapy, many challenges remain to be addressed, including ensuring the safety and efficacy of gene therapy approaches. Continued advancements in molecular biology are critical for developing safe and effective gene therapies for treating genetic and acquired diseases.

Keywords: Functional genes, Genetic disorders, Gene editing, Gene expression, Defective genes, Gene silencing, Gene therapy, Molecular biology, Plasmids, Vectors, Viruses.

INTRODUCTION

Gene therapy is an area of genomic medicine that uses the arsenal available in molecular biology to introduce healthy copies of defective genes into specific cells of the organism and modify the course of the disease.

How does gene therapy work?

Researchers have developed several strategies to correct defective genes:

[*] **Corresponding author Maheen Shafiq:** School of Engineering and Built Environment, Griffith University, Nathan, Queensland, Australia; E-mail: maheen.shafiq@griffithuni.edu.au

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

- Insert complementary DNA from a normal gene into the genome to replace the non-functional gene.
- Exchanging an abnormal gene for a normal one by homologous recombination.
- Repairing the abnormal gene by selective reverse mutation returns the gene to its normal function.
- The regulation of a gene can be altered by switching it on or off.

What is gene therapy?

Gene therapy studies the insertion of a “normal” gene to replace an “abnormal” gene (the gene that causes the disease). The carrier molecule is a vector to deliver the therapeutic gene to the patient's target cells.

What is used as a vector in gene therapy?

The most commonly used vector is a virus genetically modified to carry normal human DNA and cannot replicate autonomously. Viruses have naturally evolved a way to encapsulate and deliver their genes to human cells but in a pathogenic form. Scientists have utilized this ability and, after manipulating the virus genome, have inserted therapeutic genes to correct the disease and control the viral vector's replication.

What is the current status of gene therapy research?

The U.S. Food and Drug Administration (FDA) is responsible for approving any products that go on sale and are used for human gene therapy. Much of the gene therapy is in the experimental phase, and has not demonstrated much success in clinical studies. The first clinical study of gene therapy was conducted in 1990. The importance of gene therapy lies mainly in making these advances known to students of medicine and healthcare so they learn about this new methodology.

Gene therapy is a process that involves the introduction of nucleic acid into eukaryotic cells, excluding germ cells, with the aim of modifying the progression of a medical condition or rectifying a metabolic or genetic abnormality. In its early stages, gene therapy was primarily developed to address the specific intention of addressing monogenic illnesses, which are characterized by a mutation or deficiency in a single gene. However, it is worth noting that a solitary gene has the potential to be regarded as a novel “pharmacological agent” for the treatment of numerous illnesses. In the last decade, the concept of gene therapy has transitioned from a theoretical notion to a practical application, as evidenced by the implementation of 835 clinical procedures globally, involving a total of 5,625 patients. Notably, most of these procedures are being conducted within the United States. The early success of this relatively new sector from clinical and

preclinical methodologies has been achieved, despite expectations being exceeded. Despite being a very nascent field, it merits attention due to its significant progress facilitated by recent technological advancements, enabling its application in treating attention due to its significant progress facilitated by recent technology advancements, enabling its application in the treatment of many ailments. The above observation has been demonstrated through the encouraging outcomes observed in individuals afflicted with severe X-linked combined immunodeficiency syndrome (SCID) and hemophilia B.

Nevertheless, it is crucial to acknowledge that several obstacles exist that impede the efficacy of the therapeutic intervention. The primary challenge that has proven to be particularly tough is the lack of efficient means to transport the "therapeutic gene" to a specific cell or tissue, resulting in the inadequate production of the desired gene product, typically a protein, which is necessary to correct the condition. This phenomenon can be attributed, in part, to the development of intricate mechanisms within cells and organs that effectively hinder the internal accumulation of the outside genetic material. The predominant emphasis in the advancement of gene therapy systems for human application has been directed towards somatic cells, mostly due to the substantial apprehension surrounding the potential introduction of foreign genetic material into germ cells. Nevertheless, recent research has investigated the potential of gene transfer to the animal uterus as a means to rectify defects in the developing embryo. The aforementioned facts necessitate additional analysis, although undeniably expand the scope of available treatment interventions. The methods employed for the therapeutic gene transfer of genes into human cells can be categorized into three overarching strategies: *ex vivo*, *in vivo*, and *situ*. One example of *ex vivo* techniques involves the extraction of leukocytes primarily from the patient, followed by their cultivation and stimulation in a controlled environment. Subsequently, genetic material is inserted into the cells, which are then sustained in culture. Once the desired gene is properly expressed, these cells are introduced into the diseased host through injection. In contrast to *in vivo* methodologies, the aforementioned multi-step procedure is circumvented with the direct administration of the therapeutic gene into the bloodstream of the afflicted host, hence facilitating targeted delivery to the specific organ. The process of in *situ* gene delivery involves the direct injection of a specific gene into a suitable vector, such as in solid tumors, such as prostate cancer [1].

THE LIVER AS A TARGET ORGAN

There are a number of qualities present in the liver that make it an appealing target for gene therapy. Because it is an organ in which critical metabolic pathways are carried out, the liver is susceptible to various metabolic conditions

CHAPTER 11

Molecular Biology of Sports

Sanaullah Sajid[1,*]

[1] *Institute of Microbiology, University of Agriculture, Faisalabad, Pakistan*

Abstract: Molecular biology of sports is a rapidly evolving field that investigates the relationship between genetic and molecular factors and athletic performance. Research in this area aims to identify genes and molecules influencing physical traits such as strength, endurance, and speed. By understanding the underlying biology of athletic performance, scientists can develop new approaches to improve athletic training, injury prevention, and rehabilitation. The study of molecular biology in sports also provides insights into the relationship between genetics and lifestyle factors, such as diet and exercise, that affect overall health and well-being. This abstract will provide an overview of the current state of research in molecular biology of sports, including recent advances in genomics, transcriptomics, proteomics, and metabolomics, and how these techniques are applied to sports science.

Keywords: Athletic performance, Endurance, Health, Metabolomics, Proteomics, Physical traits, Strength, Transcriptomics, Well-being.

INTRODUCTION

Physical activity is considered any bodily movement carried out by human beings during a specific period, either as part of their work activity or in their leisure moments; this movement considerably increases energy consumption and resting metabolic levels; that is, physical activity induces calorie consumption. Daily physical activity benefits the health, cardiovascular, musculoskeletal, and immunological systems, *etc*. Physical activities integrated into everyday life, such as walking, bicycling, climbing stairs, doing housework, or simply making purchases, are beneficial, especially if they are frequently made. However, it is best practice to program physical exercise or sports at moderate intensity. Sport is any form of physical activity that aims to express or improve psychophysical attitudes, develop social relationships, or achieve sporting results at any level. The correct performance of a sport requires training. Training systematically plans and implements the necessary measures (contents and methods) to obtain specific persistent and long-lasting physical or psychological effects through physical

[*] **Corresponding author Sanaullah Sajid:** Institute of Microbiology, University of Agriculture, Faisalabad, Pakistan; E-mail: sanaullahsajid@gmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

activity. Exercise requires body movements, which are achieved through skeletal muscle movements. The muscle cell can only obtain chemical energy through adenosine triphosphate (ATP). Therefore macronutrients must give up the energy of their chemical bonds through various biochemical processes. For this to happen, different metabolic pathways provide energy [1].

ENERGY PATHWAYS DURING PHYSICAL ACTIVITY

The primary energy for the contraction of muscle fibers is obtained from the breakdown of ATP into adenosine diphosphate (ADP) and adenosine monophosphate (AMP). This reaction is carried out by the enzyme myosin ATPase; However, the amount of ATP available in the muscle only provides power for 1 to 2 seconds. If muscle contractions are prolonged for longer, ATP has to be supplied through secondary energy sources, divided into anaerobic and aerobic energy pathways. Anaerobic efforts have a maximum duration of two minutes, while endurance efforts last more than 10 minutes, and 80% of the energy source comes from oxidative pathways, called the aerobic pathway, and its primary energy source is carbohydrates [2].

Anaerobic A-lactic Pathway: Phosphocreatine

The first mechanism of secondary energy pathways is the breakdown of the creatine phosphate (PCr) molecule into creatine: ATP + PCr → ATP + Cr. In muscle, PCr is reserved in small amounts; in untrained people, the energy from this metabolic pathway lasts up to six seconds; in high-performance athletes, it lasts 12 to 20 seconds. This energy pathway does not produce lactate, called the anaerobic a-lactic pathway. This pathway is activated during athletics races of 100m, the start of swimming, the long and triple jump, or javelin and discus throws, *etc*.

Anaerobic Pathway: Anaerobic Glycolysis

If the energy demand exceeds six to 20 seconds, the PCr runs out; from that point, energy is obtained from muscle glycogen through anaerobic glycolysis. During this metabolic pathway, the energy of the glucose molecule as a metabolite is generated by two pyruvate molecules. Then, in aerobic glycolysis, pyruvate can be transformed into acetyl coenzyme (CoA) by oxidative decarboxylation through a series of reactions catalyzed by pyruvate dehydrogenase, located in the mitochondrial matrix, and Acetyl-CoA enters the Krebs cycle to generate more energy. When the energy demand exceeds oxygen availability in the blood, pyruvate dehydrogenase fails to convert pyruvate to acetyl-CoA. Fast enough, it builds up, inhibiting glycolysis and reducing ATP production. Under these condi-

tions, the enzyme lactate dehydrogenase (LDH) reduces pyruvate to lactate by the following reaction:

Pyruvate + NADH + H → lactate + NAD

The production of lactate oxidizes the NADH + H to regenerate nicotinamide adenine dinucleotide (NAD), which is necessary for glycolysis and to continue the production of ATP. In the anaerobic breakdown of glycogen, lactate is produced in concentrations ranging from 4 to 20 mmol/L of blood; accumulations greater than 6 to 8 mmol/L prevent efficient sports performance. Anaerobic glycolysis reaches its maximum level after about 45 seconds, and it lasts about two minutes; therefore, the duration of lactic anaerobic efforts is limited by muscle acidification (due to the depletion of the shock absorbing systems of the muscle that try to reduce the concentration of hydrogen cations released in the hydrolysis of ATP) and due to the lack of muscle glycogen. These efforts can be appreciated in the 400 m race and the 100 m swimming. On the other hand, the produced lactate leaves the cell muscle and reaches the liver *via* the bloodstream, where it is transformed back into glucose by gluconeogenesis, known as the Cori cycle.

Aerobic Pathway

This pathway can be divided into long-term physical exercise duration and short and medium-duration physical exercises. In long-term physical activities, energy comes mainly from lipids. Here the speed of energy production is low. Since fat is an abundant energy reserve, it allows for maintaining physical efforts for long periods, as in the case of marathons, triathlons, marches, and bicycle races. On the other hand, the energy necessary for short or medium-form aerobic exercises is obtained mainly from carbohydrates, whose combustion is faster than that of fats; at a particular time, the reservations of carbohydrates are depleted. In this exercise, lactate concentrations may appear in 2 to 4 mmol/L blood. This energy source is used in 400 m swimming events and 500 and 1000 m track races.

Aerobic Pathway: Aerobic Glycolysis

After approximately two minutes, the energy is obtained by the aerobic oxidation of the rest of the glycogen stores; acetyl−CoA is converted, together with oxalic acid (oxaloacetate), in citric acid, in each Krebs cycle which will generate one molecule of oxaloacetate to two CO_2, guanosine triphosphate (GTP) and high electrons potential in the molecules of NADH and $FADH_2$. The net balance of the cycle is:

Acetyl−CoA + 3NAD+ + FAD + GDP + Pi + $2H_2O$ → CoA−SH + 3(NADH + H+) + $FADH_2$ + GTP + $2CO_2$

CHAPTER 12

Molecular Basis of Cancer

Maliha Sarfraz[1,*] and **Hayat Ullah**[2]

[1] *Department of Zoology, Wildlife and Fisheries, University of Agriculture Faisalabad, Sub-Campus, Toba Tek Singh, Pakistan*

[2] *Department of Chemistry, University of Okara, Okara, Punjab, 56300, Pakistan*

Abstract: Cancer is a disease that arises from the uncontrolled growth of cells due to genetic mutations and epigenetic changes. Molecular biology has provided valuable insights into cancer development and progression mechanisms. Cancer cells have alterations in the genes that regulate cell growth, division, and death, leading to the accumulation of mutations that confer a survival advantage. Oncogenes promote cell growth and division, while tumour suppressor genes inhibit cell proliferation and induce cell death. Alterations in these genes and changes in DNA methylation and histone modifications lead to the dysregulation of cell signalling pathways, which contribute to cancer development. In addition, the tumour microenvironment plays a critical role in cancer progression by providing growth factors, cytokines, and extracellular matrix components that promote tumour growth and invasion. Molecular biology techniques such as DNA sequencing, gene expression profiling, and epigenetic analysis have facilitated the identification of driver mutations and key molecular pathways involved in cancer development, leading to targeted therapies that exploit these vulnerabilities. Understanding the molecular basis of cancer can revolutionize cancer diagnosis, treatment, and prevention.

Keywords: Biomarkers, Cell growth, Cell signalling pathways, DNA methylation, Epigenetics, Genetic mutations, Gene expression profiling, Histone modifications, Oncogenes, Tumor suppressor genes, Targeted therapies.

INTRODUCTION

Cancer is characterized as a multifactorial disease that affects the average growth and proliferation of cells and produces alterations in the cell differentiation process, which determines the formation of a tumour in a specific tissue. Thus, the term cancer implies, in addition to uncontrolled cell growth and proliferation, a malignant transformation, *i.e.*, a loss of typical characteristics and functions of the cells in a tissue. These processes are influenced by genetic or epigenetic altera-

* **Corresponding author Maliha Sarfraz:** Department of Zoology, Wildlife and Fisheries, University of Agriculture Faisalabad, Sub-Campus, Toba Tek Singh, Pakistan; E-mail: maliha.sarfraz@uaf.edu.pk

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

tions of numerous genes encoding proteins that regulate this process, contributing to a malignant phenotype. In addition, environmental factors are involved in tumorigenesis with a determining influence on some types of cancer. For a better understanding of cell cycle control, it can be schematized analogically as a biological clock, which operates from the nucleus of cells where various stimuli exist and/or converge that inhibit or induce cell cycle progression.

The cell cycle comprises eukaryotic cells' G1, S, G2, and M phases. Its regulation is a priority and is controlled by the participation of different proteins called cyclins and cyclin-dependent kinases (cyclin-dependent kinases, CDKs). Both act cooperatively, where CDKs phosphorylate certain cyclins at a specific time, allowing the cell to continue the cell cycle. In addition, it is important to consider that between the phases mentioned above, there are checkpoints that involve complex protein machinery that verifies whether the cell is given the right conditions or not to continue with the next cell cycle. These checkpoints control the progression between G1/S, G2/M, and M phases.

Cell cycle regulation involves cyclins and cyclin-dependent kinases as well as other proteins, which are the product of proto-oncogenes and tumour suppressor genes, which likewise participate in an organized manner in this control whose functions will be reviewed throughout the chapter. The complexity and importance of controlling the cell cycle, and their multiple participation ultimately complicate the re-establishment of an expected phenotype in malignant cells [1].

Proto-oncogenes, Oncogenes, and Tumor Suppressor Genes

Usually, cells respond to internal and external signals that stimulate cell proliferation, requiring regulation involving different cell cycle-stimulating proteins. Among the numerous genes involved in malignant cell transformation of cells include two large groups, classified as oncogenes and tumour suppressor genes. Both of them present a common denominator for their involvement in cell proliferation.

ONCOGENES

In 1910, Peyton Rous of the Rockefeller Institute and Nobel Prize winner experimented with chicken sarcoma. Injecting cell-free sarcoma homogenate into healthy chickens induced the same kind of tumour, raising the question of which agent, molecule, or factor was responsible for the tumorigenesis. The etiological agent of this condition was discovered in the following decades: a retrovirus called Rouss sarcoma virus (RSV). In 1970, one of the Rous sarcoma virus genes, the src gene, was described as having the ability to transform normal cells into malignant phenotypes. The study of this gene *via in situ* hybridization

experiments performed by J. Michael Bishop and Harold E. Varmus at the University of California, San Francisco (1976) showed that normal chicken cells contained some elements of the src gene, which gave rise to the following controversy: how is it possible that genes found in the oncogenic virus (RSV) can also be found in normal eukaryotic cells? Subsequent studies showed that similar versions of oncogenic viral genes are found in normal cells and involved in cell proliferation. The src gene belongs to a tyrosine kinase family that regulates embryonic development and cell growth. These genes were named proto-oncogenes, and with this historical background, proto-oncogenes can be defined as genes that encode proteins that regulate regularly and physiologically to regulate the cascade of events that serve to maintain the control of cell cycle progression and the normal state of cell differentiation.

They are prefixed with the prefix "c" for "cellular" to distinguish them from the viral versions, which are prefixed with the prefix "v" (*e.g.*, c-src -cellular- and v-src-viral-). The altered versions of these proto-oncogenes by mutations give rise to oncogenes. As a result, the functions of proto-oncogene and oncogene gene products differ; in oncogenes, these products are oncoproteins. These proteins act pleiotropically by causing cellular and molecular changes, depending on the specific molecular level at which they act. Like oncogenes, they are determinants in the evolution of cancer. Tumour suppressor genes or anti-oncogenes play a crucial role in tumorigenesis since they restrict or suppress cell proliferation under certain conditions.

These genes are extremely important, as their inactivation or loss is necessary for malignant tumour formation, which confers their recessive nature. In 1986, the first retinoblastoma tumour suppressor gene (Rb) was cloned; two years later, the p53 gene emerged, a molecule that was known as "the molecule of the year" labelled by the scientific journal Science because of its importance in the physiological function. Advances in the study of oncogenes and tumour-suppressor genes have placed p53 at the centre of attention in studying carcinogenic processes. Thus, it is possible to speak of genes related to susceptibility to breast cancer, of a model that has established an association between histopathological lesions and the different mutations in proto-oncogenes and tumour suppressor genes, such as in colon cancer (multistep theory) and how the human papillomavirus is one of the main risk factors for the development of cervical cancer [2].

CHAPTER 13

Molecular Mechanisms of Diabetes Mellitus

Itrat Fatima Toor[1,*], Munazzah Sajid[2] and **Haseeb Ullah Sajid[3]**

[1] *National Institute of Food Science and Technology, University of Agriculture, Faisalabad, Pakistan*

[2] *Department of Chemistry, University of Agriculture, Faisalabad, Pakistan*

[3] *Institute of Microbiology, University of Agriculture, Faisalabad, Pakistan*

Abstract: Diabetes Mellitus (DM) is a complex metabolic disorder characterized by high blood glucose levels resulting from defects in insulin secretion, insulin action, or both. The disease affects millions worldwide and is a leading cause of morbidity and mortality. Molecular mechanisms underlying the pathogenesis of diabetes mellitus are complex and involve multiple cellular and molecular processes. In this review, we discuss the current understanding of the molecular mechanisms involved in the development and progression of diabetes mellitus. Specifically, we focus on the role of pancreatic beta-cell dysfunction, insulin resistance, and abnormalities in glucose metabolism, lipids, and proteins. We also examine the contribution of genetic and environmental factors to developing diabetes mellitus. Additionally, we highlight the importance of targeting these molecular mechanisms for developing new and effective therapies for managing diabetes mellitus. A better understanding of the molecular mechanisms involved in diabetes mellitus can lead to more effective treatments and better disease management.

Keywords: Beta-cell dysfunction, Diabetes mellitus, Genetic factors, Glucose metabolism, Insulin resistance, Molecular mechanisms, Therapeutic targets.

INTRODUCTION

Type 1 diabetes is characterized by the destruction of (beta) cells of the islets of Langerhans due to an autoimmune mechanism or an unknown cause, while the phenomenon of insulin resistance characterizes type 2 diabetes. Both types of diabetes mellitus (DM) are endocrine, metabolic diseases characterized by chronic hyperglycemia and disturbances in carbohydrate, fat, and protein metabolism (DM type 2). In the first century, the term diabetes was used to refer to passing large amounts of urine (polyuria). The term was coined by the Greek philosopher Arateus; he called it diabetes mellitus or diabetes saccharin due to the sweet smell

[*] **Corresponding author Itrat Fatima Toor:** National Institute of Food Science and Technology, University of Agriculture, Faisalabad, Pakistan; E-mail: itratfatima71@gmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

of urine; hence it is called in Latin: diabetes mellitus. In recent years, the terms insulin-dependent diabetes mellitus and non-insulin-dependent diabetes mellitus were previously applied to diabetes types 1 and 2, respectively, as both classes may require insulin administration for treatment [1].

EPIDEMIOLOGY

DM is currently recognized as a true epidemic, and it is estimated that in the world, about 200 million people have diabetes, it is predicted that this figure will rise to 300 million by 2025. The Latin American Diabetes Association (ALAD) indicates that the prevalence of diabetes mellitus in Latin America is from 4 to 16%, and an increase is expected in the next 25 years from 25 to 50%. In the United States, in young patients with obesity, the prevalence is 6.6%; 4% for individuals aged 20 to 60 years, 12% for those aged 60 to 64, and 17% for the elderly. An increase in the prevalence of both types of diabetes is expected worldwide. However, the incidence of DM2 may increase more rapidly due to obesity, reduced physical activity, aging of the population, and urbanization of societies. In Mexico, DM2 is the leading cause of premature death and permanent disability, with a prevalence of 10.9%; in addition, one in four people is unaware that they are suffering from the disease, and even with treatment, about 60% do not show optimal blood glucose control.

CLASSIFICATION OF DIABETES

The most commonly used etiological classification was recommended by the American Diabetes Association (ADA) in 1997 and by the Committee of Experts for the Classification and Diagnosis of Diabetes (World Health Organization, 1998) [2]. Both the organizations classified DM in:

DM type 1: It can be of 2 kinds: autoimmune (DM1a) and idiopathic (DM1b). It is also known as juvenile diabetes, as it often occurs during childhood, although it can also occur in adults.

DM type 2: It can vary between insulin-resistant, with a relative deficiency of insulin, to a defect, preferably secretor with or without RI corresponds to the majority (90%) of cases of diabetes. However, in most patients with DM2, multiple genetic and environmental factors contribute to the disease's origin, progression, and late complications.

MODY (Maturity-Onset Diabetes of the Young): It implies genetic defects in pancreatic B-cell function beginning in youth or maturity.

Gestational DM: This condition develops at any time during pregnancy in a woman without diabetes. It is the one that is presented in the course of pregnancy. It consists of the presence of elevated levels of glucose in the blood.

Childhood DM: It can be grouped into several main subtypes:

- Transient or permanent neonatal diabetes (TNDM and PNDM, respectively).
- DM1: DM1a, autoimmune, DM1b, idiopathic.
- DM2: This type of DM is reported in obese children.
- Mitochondrial diabetes (MDM).
- Cystic fibrosis-related diabetes (CFRD).
- Diabetes associated with syndromes such as Down and Turner.

Type 1 Diabetes Mellitus

This disease is characterized by the absolute deficiency of insulin, caused by an immune attack against the beta cells of the pancreas. The Islets of Langerhans are infiltrated with activated T-lymphocytes and cause insulitis. This autoimmune attack leads to the gradual destruction of pancreatic beta cells. Symptoms appear suddenly when 80 to 90% of the beta cells of the pancreas are damaged. The pancreas stop responding appropriately to the stimulation of glucose to secrete insulin, so it must be administered exogenously to restore metabolic control and prevent the ketoacidosis that can put the patient's life at risk. For this autoimmune attack to occur, it is necessary to present a stimulus in the environment (in some cases, it is due to a viral infection), in addition to a determinant genetic disorder that leads the immune system not to recognize the beta cells as their own. People with DM1 should administer exogenous insulin to live.

Type 2 Diabetes Mellitus

Hypertension with hyperinsulinism is already present in the prediabetic normoglycemic stage. It is a chronic disease with complications that involve a significant cause of mortality and are associated with the damage or failure of various organs. Genetic susceptibility factors are polygenic; however, environmental factors exacerbate these abnormalities. A reduced mass of beta cells in the pancreas and increased cellular apoptosis characterize individuals with DM2. Hyperglycemia in fasting results from hepatic resistance to insulin action. Postprandial hyperglycemia results from abnormal insulin secretion by the pancreas's beta cells by consuming food and the inadequate intake of glucose by the insulin-sensitive peripheral tissues, particularly in the skeletal muscle.

The lack of response of the receptor tissues to insulin maintains the stimulus of elevated blood glucose, which increases the production and release of insulin by

CHAPTER 14

Molecular Basis of Obesity

Itrat Fatima Toor[1,*]

[1] *National Institute of Food Science and Technology, University of Agriculture, Faisalabad, Pakistan*

Abstract: Based on the classification by the World Organization of Health (WHO), it considers that a BMI equal to or greater than 30 kg/m^2 corresponds to obesity. Likewise, a BMI value equal to or greater than 25 kg/m^2 increases the chances of developing diseases associated with obesity. It is estimated that heredity in the variation of the BMI is in the range of 0.4 to 0.7; that is, the probability of inheriting obesity is very low and is more associated with exogenous factors. Obesity comorbidities are a risk factor for developing insulin resistance (IR), DM2, CVD, stroke, osteoarthritis, endometrial, breast, and colon cancer, among other chronic noncommunicable conditions. In addition, obesity is also linked to various digestive diseases, including gastroesophageal reflux disease, esophagitis, colorectal polyps, and non-alcoholic steatohepatitis. Obesity and overweight are associated with 44% of DM2 cases, 23% of ischemic heart disease cases, and 7 to 41% of cancer cases.

Keywords: Agouti-related protein, BMI, Bardet-biedl syndrome, Diabetes mellitus, Insulin resistance, Leptin, Neuropeptide Y, Prader-Willi syndrome.

INTRODUCTION

During the last decades, obesity has become a disease with a high prevalence worldwide. As of 2012, Mexico ranks first in adult obesity and second in childhood obesity. Obesity increases the risk of suffering six times from noncommunicable chronic degenerative diseases, such as diabetes mellitus type 2 (DM2), cardiovascular disease (CVD), and cerebrovascular disease. These three diseases constitute the leading causes of mortality in Mexico. To date, scientific and technological advances have provided an incomplete answer to the etiology of obesity. It is evident that genetic and environmental factors, such as an improper diet and physical inactivity, participate in developing this disease. It is essential to understand the molecular mechanisms that regulate hunger and induce food

* **Corresponding author Itrat Fatima Toor:** National Institute of Food Science and Technology, University of Agriculture, Faisalabad, Pakistan; E-mail: itratfatima71@gmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

intake, as well as satiety signaling mechanisms, to find alternatives in preventing and treating obesity [1].

OBESITY

Obesity is a multifactorial chronic degenerative disease characterized by a disproportionate increase of adipose tissue, or adiposity, in the body, which is associated with poor health. This increase in adipose tissue is due to the positive energy balance resulting from an imbalance between caloric intake and energy expenditure [2]. The body mass index (BMI) is an indirect estimate to diagnose obesity and was devised by the statistician L. A. J. Quetelet. The BMI is the ratio resulting from dividing body weight in kilograms (kg) by square meters (m^2) height. The BMI is also known as the Quetelet index.

BMI = weight in kg/height in m^2

Based on the classification of the World Organization of Health (WHO), it is considered that a BMI equal to or greater than 30 kg/m^2 corresponds to obesity. Likewise, a BMI value equal to or greater than 25 kg/m^2 increases the chances of developing diseases associated with obesity. It is estimated that heredity in the variation of the BMI is in the range of 0.4 to 0.7; that is, the probability of inheriting obesity is very low and is more associated with exogenous factors. Obesity and related comorbidities are risk factors for developing insulin resistance (IR), DM2, CVD, stroke, osteoarthritis, endometrial, breast, and colon cancer, among other chronic noncommunicable conditions. In addition, obesity is also linked to various digestive diseases, including gastroesophageal reflux disease, esophagitis, colorectal polyps, and non-alcoholic steatohepatitis. Obesity and overweight are associated with 44% of DM2 cases, 23% of ischemic heart disease cases, and 7 to 41% of cancer cases.

ADIPOSE TISSUE

Adipose or fatty tissue is of mesenchymal origin and represents a connective tissue formed by adipocytes that accumulate triglycerides in their cytoplasm. These adipocytes have two primary functions: reserve of energy and thermogenesis. However, fatty tissues are currently recognized as endocrine organs because they respond to traditional hormones and the central nervous system (CNS) systems and express and secrete various proteins with endocrine functions. These include adiponectin, leptin, resistin, multiple cytokines, complement components, plasminogen activator inhibitory factor-1, and proteins of the renin-angiotensin system. Therefore, the knowledge of these molecules' function will contribute to understanding the complex development of obesity.

In the adults, the positive energy balance causes an increase in the size of the adipocyte, a well-known phenomenon as hypertrophy. In contrast, the rise in the number of adipocytes predominates in children and adolescents, known as hyperplasia. The hypertrophy of the adipocyte causes modifications in its gene expression, increasing the synthesis of tumor necrosis factor-alpha (TNFα), a proinflammatory cytokine synthesized mainly by macrophages. TNFα, in turn, increases the secretion of the proinflammatory cytokine interleukin 6 (IL6). Thus, TNFα and the protein binding to serum retinol and C-reactive protein are considered markers of inflammation in obesity. Another change in gene expression is the alteration in the synthesis of proteins such as adiponectin, leptin, and resistin.

Adiponectin

The adiponectin gene, or ADIPOQ, also known as ACDC, Acrp30, apM1, or GBP28, is located in chromosome 3q27 and is expressed mainly in white adipose tissue. Adiponectin is a hormone of 244 amino acids that contains a signal peptide, allowing its secretion outside the adipocyte. Adiponectin is post-translationally modified by hydroxylation and glycosylation. Due to these modifications, it can form high molecular weight trimers, hexamers, and isoforms. There is an inverse relationship between circulating levels of adiponectin and the amount of fat mass present in an organism or its BMI. In a longitudinal study in adults, it was suggested that low plasma adiponectin concentration is a factor of risk to present DM2. This argument was based on the fact that the subjects who developed DM2 had concentrations 22% lower than those who did not. However, in this study, a group of subjects with type II obesity had a physiological range of plasma adiponectin and did not develop metabolic comorbidities. In mouse models treated with adiponectin, this is a protective factor for IR to a greater degree in obese mice than in those with an average weight. A study on Mexican children demonstrated that adiponectin has a robust inverse association with age, BMI, and insulin concentrations. Besides, adiponectin improves insulin sensitivity, inhibits gluconeogenesis in the liver, inhibits the expression of adhesion molecules, blocks macrophage migration, prevents foam cell formation, reduces B cell production, decreases T cells' response, and favors the translocation of glucose transporters.

Leptin

The leptin gene (from the Greek leptos, meaning thin) was discovered in 1994 in the mouse. It is found on chromosome 7q31.3 in humans and has been called the LEP and OB genes. The OB gene encodes a non-glycosylated hormone-like protein of 167 amino acids, with a signal sequence of 21 amino acids removed

CHAPTER 15

Molecular Basis of Hepatitis B

Saima Pervaiz[1,*] and **Sara Masood Cheema**[2]

[1] *Department of Immunology, University of Health Sciences, Lahore, Pakistan*

[2] *Department of Pathology, Azra Naheed Medical College, Superior University, Lahore, Pakistan*

Abstract: Hepatitis B virus (HBV) is a serious global health problem affecting millions worldwide. Chronic HBV infection can lead to liver cirrhosis and hepatocellular carcinoma, making it a major cause of morbidity and mortality. The molecular basis of HBV infection and pathogenesis is complex and involves multiple interactions between the virus and the host immune system. HBV is a partially double-stranded DNA virus replicating through reverse RNA intermediate transcription. The virus has several proteins, including the envelope protein (HBsAg), core protein (HBcAg), and polymerase (HBp), that play critical roles in virus entry, replication, and assembly. The viral genome is organized into four overlapping open reading frames (ORFs), each encoding a different viral protein. During HBV infection, the virus initially binds to heparan sulfate proteoglycans on the cell surface, followed by binding to specific receptors, such as the sodium taurocholate co-transporting polypeptide (NTCP). The virus then enters the cell through endocytosis, where it is uncoated and releases the viral DNA into the nucleus. The viral DNA is then transcribed by the host RNA polymerase II, producing viral mRNAs translated into viral proteins. One key factor determining the outcome of HBV infection is the host's immune response. The innate immune response plays an important role in controlling the initial phase of HBV infection, while the adaptive immune response, particularly the CD8+ T cell response, is critical for the clearance of the virus. However, in some cases, the immune response cannot clear the virus, leading to chronic infection. Understanding the molecular basis of HBV infection and pathogenesis is critical for developing effective treatments and vaccines. Current treatments for chronic HBV infection include nucleoside/nucleotide analogs and interferon-based therapies, which can suppress viral replication and reduce liver damage. However, these treatments are not curative and can have significant side effects. Vaccination against HBV is highly effective in preventing infection and is recommended for all individuals at risk of HBV infection.

Keywords: Adaptive immune response, Core protein, CD8+ T cells, Cell surface receptors, Envelope protein, Hepatitis B virus, Innate immune response, Nucleoside/nucleotide analogs, Liver cirrhosis, RNA intermediate.

[*] **Corresponding author Saima Pervaiz:** Department of Immunology, University of Health Sciences, Lahore, Pakistan; E-mail: saimapz@hotmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

INTRODUCTION

The term hepatitis means “inflammation of the liver”, and the most common causes of this inflammation are infections by one or more of the viruses known as hepatitis A, B, C, D, and E viruses (HAV, HBV, HCV, HDV, HEV, respectively). These viruses can cause an acute illness with symptoms lasting several weeks, such as jaundice (yellowing of the skin and eyes), dark urine, extreme fatigue, nausea, vomiting, and abdominal pain. Of the viruses mentioned above, HBV and HCV can be asymptomatic, with no apparent liver damage, in carriers. They can also cause clinical manifestations, such as the onset of chronic hepatitis with progression to liver cirrhosis, which may culminate in hepatocellular carcinoma.

In relation to HBV, advances in molecular biology have contributed considerably to preventing, diagnosing, and treating hepatitis B. The genomic organization and mechanisms of transcription and replication of the virus are now well understood. For this purpose, diagnostic methods are available that make it possible to directly identify the presence of the genome of this virus in biological samples. In addition, a highly effective recombinant vaccine is also available (offers 90% protection) with minimal side effects. The production of recombinant antigens has also improved serological diagnosis for detecting viral antigens and/or antibodies formed in response to infection. Despite the efforts made by different research groups, the ideal treatment has yet to be available. The drugs approved to treat hepatitis B are interferon, lamivudine, adefovir, dipivoxil, entecavir, and tenofovir. Prospective targets and ongoing investigations include agents that block the NTCP receptor and the development of short-interfering RNAs that may disrupt the transcription and translation of HBV DNA. There are also possibilities of developing immunomodulatory vaccines by including additional Pre S antigens into routine Hepatitis B vaccines. Antiviral therapy has shown a clinical improvement in some patients and a decrease in the occurrence of hepatocellular carcinoma in countries where this condition is highly endemic. However, drug-resistant variants of the virus are commonly present, highlighting the need to extend prevention and diagnostic measures worldwide and continue the search for new and more effective treatment strategies to eradicate hepatitis B [1].

GLOBAL EPIDEMIOLOGY

The World Health Organization (WHO) estimates that approximately 2 billion people are infected with HBV, of which 350 million are chronically infected. HBV is the tenth leading cause of death worldwide, and about 500,000 to 1.2 million people die each year from complications of HBV infection. The distribution of HBV infection in the world is heterogeneous, so it is divided into three geographic categories based on the prevalence of chronic HBsAg carriers:

(a) areas with a high prevalence (>8%); (b) areas with intermediate prevalence (2 to 8%), and c) areas with a low prevalence (< 2%). Among the regions with high prevalence are some African, Southeast Asia, and China countries. Areas of intermediate prevalence include Eastern and south-eastern Europe, North Africa, the Middle East, Japan, India, and parts of South America. The United States and Western Europe are low-endemic areas. In Mexico, the studies, although not representative of the general population, suggest that it is a low-endemic country with some regions of intermediate prevalence.

Routes of Transmission

- Perinatal (Transmission takes place during birth. Transplacental transmission is very unlikely and so far there is no evidence of transmission due to breastfeeding).
- Unsafe injections and transfusions of blood products.
- Use of non-sterile syringes and needles.

In high-prevalence countries, most infections are acquired perinatally and during early childhood, when the immune system is immature. In these regions, hepatocellular cancer is among the three leading causes of death. In countries with moderate and low endemicity, hepatitis B is acquired in young adults through sexual activity and intravenous drug use. In addition, hepatitis B is the leading viral infection that occurs in healthcare workers through an accidental puncture or by working without sufficient preventive measures, especially in countries where vaccination is not widely used.

HEPATITIS B VIRUS

To understand the concepts of the natural history of the hepatitis B virus, its prevention, diagnosis, and treatment, first of all, it is essential to know the fundamental aspects of HBV, its genome organization, and the most important characteristics of its proteins. HBV is classified within the family Hepadnaviridae of the Orthohepadnavirus genus, being tropic for the liver, and is characterized as a double-stranded circular DNA virus replicating through RNA intermediary by reverse transcription. The HBV genome can integrate into host chromosomes [2].

In vivo, the liver is the main organ of virus replication; however, viral genomes have been found outside the liver, as in the bile duct epithelium, the pancreas, renal cells, lymphocytes, and monocytes. The mature viral particle comprises a nucleocapsid surrounded by a lipid bilayer into which the envelope proteins are incorporated. Inside the nucleocapsid is the viral genome containing all the organized genetic information.

CHAPTER 16

Molecular Basis of Hepatitis C

Muneeb Ur Rehman[1] and **Hafiz Zain Ul Abideen**[2,*]

[1] *Medical College of Wuhan, University of Science and Technology, Hubei, Wuhan, China*

[2] *College of Basic and Clinical Medicine, Dali University, Dali, Yunnan, China*

Abstract: Hepatitis C virus (HCV) is a significant cause of chronic liver disease worldwide. The molecular basis of HCV infection and replication has been extensively studied, leading to the identification of vital viral proteins and their interactions with host factors. The HCV genome encodes a single polyprotein cleaved by host and viral proteases into individual proteins, including the core, envelope glycoproteins (E1 and E2), p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B. These viral proteins play critical roles in virus assembly, entry, replication, and evasion of host immune responses. The HCV envelope glycoproteins E1 and E2 are responsible for virus attachment and entry into host cells through interactions with various host receptors, including CD81, scavenger receptor class B type I (SR-BI), and tight junction proteins. The viral protein NS3 has multiple functions, including protease and helicase activities, which are critical for viral RNA replication. NS5A is an essential component of the viral replication complex and regulates viral RNA replication, virion assembly, and modulation of host immune responses. NS5B is the RNA-dependent RNA polymerase responsible for viral RNA synthesis. The molecular mechanisms underlying HCV-induced pathogenesis and the development of chronic infection remain poorly understood. However, recent studies have shed light on the interactions between HCV and host factors, including the innate and adaptive immune responses and the roles of viral proteins in modulating these responses. These insights have led to new antiviral therapies, including direct-acting antivirals (DAAs) that target viral proteins in RNA replication.

Keywords: Glycoproteins, Hepatitis C virus (HCV), HCV-induced pathogenesis, Polyprotein, Virus assembly, Viral proteins, Viral RNA replication.

INTRODUCTION

The Hepatitis C virus (HCV) is a tiny RNA virus enveloped and prefers the liver, resulting in acute and chronic hepatitis in human hosts. Globally, the number of individuals impacted amounts to 200 million, as reported by the World Health

* **Corresponding author Hafiz Zain Ul Abideen:** Medical College of Wuhan, University of Science and Technology, Hubei, Wuhan, China; E-mail: muneebrx2@gmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

Organization (WHO), which further states that approximately 3% of the global population suffers from chronic infection caused by this particular virus. Most infected patients remain asymptomatic for the acute phase. They cannot eliminate the virus, which causes hundreds of thousands of patients to die yearly from chronic complications such as cirrhosis and hepatocellular carcinoma (HCC). To date, no vaccine is available to prevent infection, and the most suitable to combat it is expensive. It has adverse side effects, is unsuitable for all infected subjects, and is only effective in about 50% of patients treated. One of the limitations in advancing knowledge of hepatitis C is that there is a lack of models for small animal study since this virus only infects chimpanzees and human beings. However, the development of *in vitro* culture systems is causing considerable advancement in the knowledge related to the replication of viruses and the development of new antiviral therapies [1].

GLOBAL EPIDEMIOLOGY OF HEPATITIS C

There are about 200 million people in the world affected by HCV [2]. The prevalence varies by region: it is the highest in North Africa and the Middle East, mainly Egypt, where a prevalence of 14%, followed by the rest of Africa (3%), China and other Asian countries (2.1%), Eastern Europe and the United States (1.6%). In northern European countries, such as the United Kingdom, a prevalence of 1.1% is reported. In Mexico, although studies are not available on a representative population, and according to the existing reports, most blood donors' prevalence of hepatitis C is around 0.4%.

Transmission Routes

Contact with infected blood or products is the most prevalent way that people contract HCV. The spread of HCV has been significantly impacted by injectable therapy and medicines.

Intravenous Drug Use (IVD): In developed countries, such as the United States, IVD is the leading risk factor, even though a decrease in cases has been observed. The prevalence of hepatitis C in users who had been consuming IVD for a year decreased to 10% from 1997 to 1998, while from 1988 to 1991, the reported prevalence was 65%. In countries with high prevalence, most infections are acquired by not having safe blood transfusion measures, administering injections with non-disposable syringes, and using inefficient cutting material in medical and dental clinics.

Sexual Contact: Of cases with acute hepatitis reported in the United States, 15 to 20% of patients are engaged in high-risk sexual activity (more than two sexual partners during six months or having relationships with infected people).

However, prevalence studies conducted on monogamous couples where one is infected with HCV suggest that the spread of hepatitis C by sexual contact is rare.

Perinatal Transmission: Transmission to newborns, the number of infected mothers varies from 4.6 to 10%. The only element causing the transmission of hepatitis C in newborns is the presence of viral RNA in mothers during childbirth.

Transmission to Health Personnel: In developed countries, it is rare to observe patient-to-patient messages. When detected, this is related to improper healing practices, such as using multi-dose vials, improper cleaning, lack of equipment, and lack of aseptic techniques. The transmission of HCV from infected patients to health personnel is mainly due to accidents with contaminated needles.

Structural Features of HCV

HCV belongs to the Flaviviridae family and is the only member of the Hepaciviral genus. There are six genotypes (1 to 6) identified according to their nucleic acid sequence, with a divergence in its complete genome of 30 to 50%; there are also more than 70 sub-genotypes (a, b, c, d, *etc*.), identified by differences in the comparison of the genome from 20 to 25% among the genotypes. Quasispecies refers to the mixture of HCV genetic variants in the same individual. Of the identified genotypes, the most resistant to treatment is genotype 1. HCV is a single-stranded RNA virus, with sense positive, replicated by an RNA-dependent RNA polymerase. Its main organ of replication is the liver; however, evidence is available of extrahepatic cells for HCV, such as B lymphocytes from peripheral blood, epithelial cells in the intestine, and the central nervous system. The viral particles, analyzed in the electron microscope, are observed heterogeneously: have at least two sizes (50 and 100 nm in diameter) and contain double-walled and beak-shaped projections on their surface [3].

As was previously mentioned, the HCV genome consists of a 9.6 kb positive-sense single-stranded RNA. It has a single open reading frame (MLA) that codes for a polypeptide precursor of about 3000 amino acids in length. At least ten distinct proteins are produced from this polyprotein after it is cleaved by viral proteases and the host (Fig. **1**).

- A vital membrane protein is called p7; three structural proteins are called core, E1, and E2.
- The intracellular steps of the HCV life cycle are coordinated by six non-structural proteins: NS2, NS3, NS4A, NS4B, NS5A, and NS5B.

CHAPTER 17

Molecular Basis of the Human Immunodeficiency Virus

Maliha Sarfraz[1,*], **Sanaullah Sajid**[2] and **Hayat Ullah**[3]

[1] *Department of Zoology, Wildlife and Fisheries, University of Agriculture Faisalabad, Sub-Campus, Toba Tek Singh, Pakistan*

[2] *Institute of Microbiology, University of Agriculture, Faisalabad, Pakistan*

[3] *Department of Chemistry, University of Okara, Okara, Punjab, 56300, Pakistan*

Abstract: Human Immunodeficiency Virus (HIV) is the causative agent of Acquired Immunodeficiency Syndrome (AIDS), a deadly disease that affects the human immune system. HIV is a retrovirus that infects T-cells, macrophages, and dendritic cells in the immune system, leading to their destruction and, ultimately, the onset of AIDS. The molecular basis of HIV infection involves the interaction of the viral envelope protein gp120 with the host cell receptor CD4 and a co-receptor such as CCR5 or CXCR4. This binding triggers a conformational change in gp120 that exposes a fusion peptide, allowing the viral envelope to fuse with the host cell membrane and release its contents into the cytoplasm. Once inside the host cell, the viral genome is reverse-transcribed into DNA, which is then integrated into the host cell genome by the viral integrase enzyme. This allows the virus to replicate with the host cell and evade the immune system's surveillance. Despite advances in antiretroviral therapy, HIV continues to pose a significant global health threat, with over 38 million people living with HIV/AIDS worldwide. Understanding the molecular mechanisms of HIV infection is critical for developing effective treatments and vaccines to combat this deadly disease.

Keywords: AIDS, Antiretroviral therapy, CD4, Co-receptor, Dendritic cells, HIV, Macrophages, Retrovirus, Reverse transcription, T-cells, Viral envelope protein, Viral genome.

INTRODUCTION

Viruses are classified into more than 60 families, depending on the type of nucleic acid they contain, the information in their genome, and their type of genes. Among these families, at least 20 infect humans. The causative agent of acquired immunodeficiency syndrome (AIDS) is a retrovirus known as the human immu-

[*] **Corresponding author Maliha Sarfraz:** Department of Zoology, Wildlife and Fisheries, University of Agriculture Faisalabad, Sub-Campus, Toba Tek Singh, Pakistan; E-mail: maliha.sarfraz@uaf.edu.pk

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

nodeficiency virus (HIV), which belongs to the Retroviridae family and the Lentivirus genus. HIV was first identified in France in 1983 by Luc Montagnier, the virus with an RNA genome that induces chronic cell infection by converting its RNA to proviral DNA, which integrates into the cell's genome and induces progressive damage to the immune system of the host. Infection by this lentivirus is characterized by a slow progression, with long periods of clinical latency, followed by the gradual onset of signs and symptoms, such as fever, lymphadenopathy, opportunistic diseases, and oncological and chronic degenerative diseases.

One of the main characteristics of HIV is its genetic variability, which shows during its evolution the transmission of simian virus genomic fragments to the human that has given rise to various types, subtypes, variants, and multiple recombinant forms. Another factor influencing genetic variability is the virus genome's high mutation rate, which is provided by the reverse transcriptase enzyme. This enzyme generates 3×10^5 mutations per nucleotide per cycle and a production of 1010 viral particles per day. Considering these two variables, it is estimated that, on average, 3×10^9 can be generated mutations per day in the viral population of each person with HIV. Most of the mutations generated in the variants have little or no effect on viral function or replication capacity; still, others can be lethal for the virus. However, when these mutations are located at key genomic sites, such as genes for viral enzymes, they affect the virus's ability or may confer resistance to one or more antiretrovirals. Another factor contributing to genetic diversity is the recombination of the two RNA strands, which constitutes an intrinsic part of the normal cycle of viral replication and can mediate the repair of defective viral genomes, increase viral diversity or accelerate the spread of mutations beneficial among viral quasispecies. Genetic variability impacts the pandemic of HIV through diagnosis, pathogenesis, progression, and transmission of the disease, the quantifiability of viral load, clinical management, response to antiretroviral treatment, and vaccine design [1].

NATURAL HISTORY OF DISEASE

The various clinical states of HIV infection result from changes in the balance of immunological and virological status. Immune status is determined by the count of TCD4+ lymphocytes and is expressed as cells/mm^3. Virologic status is determined by the copy number of HIV RNA (viral load) and expressed as copies/ml. These are characteristics of the infection and the gradual destruction of the CD4+ T lymphocyte population with constant production of new viral particles, which generates immunodeficiency and takes the patient to the last stage of HIV infection or AIDS. HIV-infected patients go through three clinical stages [2].

1. Primary infection or acute phase: seroconversion.
2. Asymptomatic state or clinical latency.
3. Symptomatic state or AIDS.

Acute Phase

This period begins with the patient's infection and may have no clinical manifestation since it is only symptomatic in approximately 40-80% of patients two to three weeks after infection. The symptoms can vary, including fever, mononucleosis, maculopapular rash, sore throat, general malaise, and joint pain. Signs and symptoms disappear in two or three weeks; later, the patient is asymptomatic between six and ten years. The first episodes that occur after HIV crosses the mucosal barrier are known as the window period, during which the production of IgM and IgG antibodies is not sufficient for the diagnostic tests to detect the immune response against HIV. The duration is controversial, although historically, it has been suggested that it takes approximately three months. Due to the nonspecific nature of symptoms during this stage, the diagnosis of HIV disease is rarely recognized in medical practice. Between two and six weeks after infection, the viral load in the plasma reaches hundreds of thousands of copies/ml. Due to the massive increase in viral replication, the number of CD4+ T cells decreases. Consequently, the elevated viral load also contributes to the transmission of the disease. It has been estimated that up to 30% of infections by HIV are transmitted during the acute phase.

Four to 12 weeks after infection, specific IgG antibodies appear against the HIV, and the specific cellular immune response is activated, the leading cause of the decrease in viral load and the corresponding increase in the CD4+ T lymphocyte count in patients. A balance is maintained between the number of viral particles produced and eliminated and the number of lymphocytes T CD4+ destroyed and generated each day six to twelve months after infection.

Clinical Latency Phase

Despite the absence of early symptoms, the majority of patients are diagnosed at this stage. During this stage, high viral replication persists, and the number of viral particles produced reaches more than 1010 particles per day. The ability to counteract these particles to the regeneration of CD4+ T lymphocytes, however, is not sufficient to halt the progression of the illness. With the appropriate use of antiretroviral treatment, which inhibits the replication of the virus in different stages of the viral replication cycle, the reconstitution of the immune system is possible, which prolongs the appearance of the symptoms of AIDS and increases the quantity and quality of life of people living with HIV. The duration of this asymptomatic, also known as chronic, phase is variable and depends on a variety

CHAPTER 18

Molecular Mechanisms of Human Immunodeficiency Virus Resistance to Antiretroviral

Saima Pervaiz[1,*] and **Shehryar Awan**[2]

[1] *Department of Immunology, University of Health Sciences, Lahore, Pakistan*

[2] *Department of Pathology, Azra Naheed Dental College, Superior University, Lahore, Pakistan*

Abstract: Antiretroviral therapy (ART) has transformed the treatment of human immunodeficiency virus (HIV) infection, improving life expectancy and quality of life for millions worldwide. However, the emergence of drug-resistant HIV strains poses a significant challenge to the effectiveness of ART. The molecular mechanisms underlying HIV resistance to antiretroviral drugs involve multiple genetic changes in the viral genome that reduce drug susceptibility, often through alterations in the viral enzymes targeted by the drugs. The primary targets of ART are the viral reverse transcriptase (RT), protease (PR), and integrase (IN) enzymes, which are essential for HIV replication. Resistance to nucleoside reverse transcriptase inhibitors (NRTIs) results from mutations in the viral RT enzyme that reduce drug incorporation into the viral DNA chain. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) bind to a hydrophobic pocket near the RT active site, and resistance to these drugs arises from mutations that alter the binding pocket conformation. Protease inhibitors (PIs) bind to the viral PR enzyme, and resistance results from mutations that alter the enzyme's conformation, reducing drug binding affinity. Integrase strand transfer inhibitors (INSTIs) bind to the viral IN enzyme, and resistance arises from mutations that affect drug binding or alter the IN active site. The emergence of drug-resistant HIV strains can also result from poor adherence to ART, leading to the selection of pre-existing resistant viruses or the development of new resistance mutations. In addition, the genetic diversity of HIV and the high viral replication and mutation rate contribute to the rapid evolution and emergence of drug-resistant strains.

Keywords: Antiretroviral therapy, Drug resistance, Human immunodeficiency virus, Nucleoside reverse transcriptase inhibitors (NRTIs), Integrase, Non-nucleoside reverse transcriptase inhibitors (NNRTIs), Protease, Protease inhibitors, Viral genome.

[*] **Corresponding author Saima Pervaiz:** Department of Immunology, University of Health Sciences, Lahore, Pakistan; E-mail: shehryarawan73@gmail.com

Sanaullah Sajid, Sajjad ur Rahman, Shahid Mahmood, Shayan Bashir & Mudasser Habib (Eds.)

INTRODUCTION

In the last two decades, there have been notable advancements in the management of human immunodeficiency virus (HIV) infection, resulting in improved life expectancy and enhanced quality of life for patients. These improvements can be attributed primarily to the expanding repertoire of Antiretrovirals (ARVs) designed for HIV treatment, which encompasses diverse molecular approaches. Despite the notable efficacy of Highly Active Antiretroviral Therapy (HAART), a treatment regimen that involves the administration of a combination of three or more medications with distinct modes of action, it is important to acknowledge that the total suppression of viral replication or complete elimination of HIV from the reservoir within the host has not been achieved. It is only feasible to refer to an efficacious control of viral replication for prolonged periods.

Failure to adhere to the therapeutic regimen after the first year of initiating treatment is expected in 10 to 30% of individuals on initial therapy and between 40 and 60% of patients who have previously failed treatment. The leading causes of antiretroviral treatment failure are resistance to ARVs, limited antiviral potency, inadequate patient adherence to the therapeutic regimen, defects in drug absorption, drug interactions, insufficient drug activation (cellular resistance), and, finally, viral replication in lymphoid sanctuaries. ARV resistance is the leading cause of treatment failure. It is due to a series of nucleotide substitutions in the genes coding for glycoproteins 120 and 41, the reverse transcriptase (RT) enzymes, protease, and integrase, targets on which these drugs act [1].

ANTIRETROVIRAL TREATMENT

Since the discovery of HIV, therapeutic strategies have focused on the virus's chemical structure and life cycle, which led to the creation of the drugs available today. Thus, as the understanding of the pathophysiological mechanisms involved, the interaction of the viral envelope with specific receptors in the cell membrane, replication, and even the release of new infectious particles, new options for inhibiting them emerge. In the virus cycle, critical steps have been identified and targeted by ARVs currently used to inhibit replication, such as entry, reverse transcriptase, protease, and integrase. ARV treatment offers clear benefits to HIV-infected individuals when HAART is used, which usually involves the combination of two nucleoside analogues and a protease inhibitor, improves the quality and quantity of the patient's life, as well as decreases the frequency of some infections by opportunistic microorganisms.

At present, the goal of treatment is to maintain the viral load at undetectable levels and to increase the CD4+ lymphocyte count progressively for as long as possible. There are more than 20 different ARVs for the treatment of HIV patients.

THERAPEUTIC TARGETS [2]

Envelope Proteins

HIV attaches to the cell through the interaction of the virus envelope glycoprotein complex (gp160) with the CD4+ lymphocyte surface receptor and, depending on the course of the infection, with the co-receptors CCR5 and CXCR4.15. The glycoprotein gp160 is composed of two subunits, the glycoprotein gp120, which is the one that interacts with cellular receptors, and the transmembrane subunit gp41, which is the cause of virus fusion. After the binding of gp120 and the CD4+ cell receptor, a conformational change occurs in the gp120 that allows a closer interaction between the co-receptors CCR5 and CXCR4 on the surface of the CD4+ cell and gp120, causing a change in gp41, which results in the exposure of the hydrophobic region of the central part of the gp41 protein. Close to this region are the HR1 and HR2, composed of six helices forming a "hairpin" structure and, consequently, the fusion pore, through which the viral RNA enters the CD4+ lymphocyte.

Cell Receptor Inhibitors

CCR5 and CXCR4 inhibitors are the most promising within this class of drugs since blocking the receptor achieves a similar effect to that seen in people who possess a mutation in the CCR5 allele and generate a truncated, non-functional protein. Thus, when homozygotes do not express CCR5 and have relative protection against HIV infection, heterozygotes, on the other hand, can acquire the infection, but disease progression is slower. On the other hand, CXCR4 is found in 50% of patients with advanced disease, and its presence in lymphocytes is associated with rapid progression. Since CXCR4 is expressed with its ligand SDF-1α (stromal cell-derived factor-1α), when SDF-1α is co-inhibited with CXCR4, hematopoiesis, cardiac, and brain development functions are altered parallelly.

Among the inhibitors of CCR5, we can find new drugs such as Maraviroc, a non-competitive inhibitor of the co-receptor binding site, and Vicriviroc, which, like TAK-770, induces conformational changes in CCR5 that affect gp120 binding. Among the CXCR4 inhibitors, AMD070 blocks SDF1α, with the side effects described in the preceding paragraph.

Fusion Inhibitor

At present, there is only one inhibitor of this class, Enfuvirtide. This peptide of 36 amino acids prevents the termination of the HIV fusion sequence. Enfuvirtide is a synthetic analogue of the peptide fragment HR2, which binds to an HR1 moiety

SUBJECT INDEX

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

R

S

T

U

V

X

Y

Sanaullah Sajid

Dr. Sanaullah Sajid is an expert in microbiology with more than seven years of experience. He has a degree in veterinary medicine, and an M.Phil. and a Ph.D. in microbiology. He has played key roles in developing new vaccines and treatments, including patents for innovative work on detecting antibiotic-resistant bacteria and developing an immune complex vaccine. He has written important books like the "Training Manual of Bacteriology" and "Training Manual of Immunology," which are go-to resources for professionals. His chapters on new diabetes drugs and the global effects of the coronavirus, are evidence of his ability to tackle crucial health issues. His book, "Fundamentals of Cellular and Molecular Biology," is a guiding book for future scientists.